Ministry of Education and Science of the Republic of Kazakhstan
SDU University

UDC 004.8:378.04(043.3)

BERLIKOZHA BAUYRZHAN ASSETULY

AI-Powered Approach to Career Path Prediction for IT Students Using Academic and Behavior Data

8D06102 – Computer Sciences

Dissertation submitted in fulfillment of the requirements for the degree Doctor of Philosophy (Ph.D.)

Scientific advisor
PhD, Associate Professor at
SDU University
Zhamanov A.

Foreign scientific advisor
Assoc. Professor, Ph.D. at
Barcelona University
Oliver D.

Republic of Kazakhstan,
Kaskelen, 2025

CONTEXT

NORMATIVE REFERENCES

This thesis uses references to the following standards:

"Instructions for the preparation of a dissertation and author's abstract" Ministry of education and science of the Republic of Kazakhstan, 377-3 Zh.

GOST 7.32-2001. Report on research work. Structure and design rules.

GOST 7.1-2003. Bibliographic record. Bibliographic description. General requirements and compilation rules.

GOST 7.32-2017. System of standards of information, librarianship, and publishing. Research report. Structure and design rule.

SYMBOLS AND ABBREVIATIONS

AI	- Artificial Intelligence
ANN	- Artificial Neural Networks
R&D	- Research and development
CNN	- Convolutional Neural Network
RNN	- Recurrent Neural Network
DRL	- Deep Reinforcement Learning
DQN	- Deep Q-Network
MDP	- Markov Decision Process
DL	- Deep Learning
ML	- Machine Learning
RL	- Reinforcement Learning
SVM	- Support Vector Machines
DT	- Decision Trees
RF	- Random Forest
KNN	- K Nearest Neighbors
SMOTE	- Synthetic Minority Oversampling Technique
AUC	- Area Under the Curve
ROC	- Receiver Operating Characteristic
TPR	- True Positive Rate
FPR	- False Positive Rate
CV	- Cross - Validation
API	- Application Programming Interface
GUI	- Graphical User Interface
DB	- Database
MSE	- Mean Squared Error
R^2	- Coefficient of Determination
KAN	- Kolmogorov - Arnold Networks
SDU	- Suleyman Demirel University
ICT	- Information and Communication Technologies
STEM	- Science, technology, engineering, and mathematics

INTRODUCTION

General characteristics of the work. The integration of educational outcomes with workforce requirements is now a pressing issue facing economies worldwide in the age of digital transformation.In addition to changing industries, the fast development of information and communication technology (ICTs) has also changed the skills that future professionals must possess [1]. Higher education institutions are under more and more pressure to equip graduates with the technical and adaptable skills necessary to meet the quickly evolving demands of the industry as global economies move toward automation, artificial intelligence (AI), and data-centric operations [2]. Of them, the field of information technology (IT) has become one of the most dynamic, requiring students to constantly adjust to challenging interdisciplinary problems, data-driven workflows, and new programming paradigms [3].

The choice of a career route is still quite difficult for university undergraduates, even with the growth of IT education. Finding job paths that fit their interests, abilities, and long-term goals is a challenge for many recent graduates. According to research, a sizable percentage of IT graduates wind up working in fields unrelated to their degree of concentration [4]. The absence of individualized career counseling programs that may combine behavioral and academic markers into a logical framework for decision-making frequently causes this mismatch. In the age of big data and artificial intelligence, traditional academic advising-which mostly depends on human intuition and small datasets-cannot scale efficiently [5].

Recent developments in machine learning (ML) and artificial intelligence (AI) have created new avenues for enhancing educational decision-making. Large amounts of behavioral and academic data can be analyzed by AI-driven systems to produce tailored recommendations for pupils [6]. This is in line with the larger worldwide movement toward data-driven education, which is referred to as Learning Analytics (LA) and Educational Data Mining (EDM) [7]. Whereas LA analyzes educational data to aid in institutional decision-making, EDM concentrates on identifying patterns in the data to improve learning outcomes. Predictive modeling has become well-known in this context due to its capacity to predict career outcomes, dropout risks, and student performance [8].

In the context of career counseling, artificial intelligence (AI) makes it possible to build prediction systems that determine the best career pathways for students based on their academic records, extracurricular activities, motivating factors, and personality features [9]. To interpret both structured and unstructured data, these systems use machine learning methods including Random Forest (RF), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Gradient Boosting (GB), and Deep Neural Networks (DNNs). Research has indicated that models that employ ensemble techniques, such Gradient Boosting, frequently perform better than conventional classifiers because of their ability to manage intricate, nonlinear interactions between features [10].

However, there are also regional differences in how AI is incorporated into job counseling. Universities in developing nations, especially those in Central Asia, are still in the early phases of adopting AI-powered academic support systems, whereas those in Europe, North America, and parts of East Asia have already done so [11]. Under the government's Digital Kazakhstan initiative, the creation of digital ecosystems in education has been designated as a strategic priority in Kazakhstan [12]. However, data-driven career advice is still completely unexplored despite significant improvements in digital infrastructure and e-learning platforms [13].

Using information gathered from Suleyman Demirel University's (SDU) Beta Career Platform, this dissertation fills that knowledge gap by putting forth a thorough AI-powered framework for forecasting career paths among IT students in Kazakhstan [14]. The framework creates predicted insights that can guide individualized career counseling by utilizing behavioral patterns, extracurricular activities, and academic performance. The goal of this study is to determine the most accurate and interpretable model for student career prediction by using a multi-model comparison that includes Gradient Boosting, Neural Networks, and the recently developed Kolmogorov-Arnold Networks (KANs). Higher interpretability and efficiency in low-data situations are provided by the integration of KANs, which represents the development of AI architectures beyond conventional neural networks [15].

Additionally, this study responds to the priorities established by the Higher Scientific and Technical Commission in the field of "Information, Telecommunication, and Space Technologies," placing it within the national framework of Kazakhstan's educational modernization. Developing a predictive system that not only improves career counseling at SDU but also acts as a scalable model for other universities in Kazakhstan and other developing contexts is one of its two main goals.

It is anticipated that the work's results will help academic institutions match their curricula to industry demands, increase graduates' employability, and further the nation's shift to a knowledge-based economy [16].

In this study, several machine learning algorithms were evaluated for their predictive performance in the context of career guidance. Based on this comparative analysis, Gradient Boosting was identified as the most effective algorithm and was subsequently integrated into the Beta Career Platform at Suleyman Demirel University[17]. According to student survey results, the implementation of this AI-based system led to a 27% increase in student satisfaction when comparing the academic years 2023-2024 and 2024-2025. By connecting the general characteristics of career advising challenges with the specific objectives of AI integration, the study demonstrates both the scientific and practical relevance of developing intelligent career guidance tools for IT students in Kazakhstan.

Even with the quick development of digital transformation in higher education, choosing a career path for IT students is still a difficult and urgent problem [18]. Conventional university academic advising methods frequently ignore more comprehensive behavioral and motivational elements that significantly influence a student's career path in favor of narrow metrics like GPA or test scores [19]. Graduates' competencies and the demands of the labor market are mismatched as a result of this disconnect between academic evaluation and real-world employability outcomes, which is particularly noticeable in developing nations like Kazakhstan [20].

The shortcomings of the current career counseling systems, which do not integrate data-driven tailoring, are brought to light by an expanding corpus of research in educational data mining [21]. Numerous solutions currently in use are manual or rule-based, providing broad guidance that fails to consider the intricate, non-linear nature of professional growth and human learning [22]. Furthermore, these models frequently rely on out-of-date or partial datasets, which renders them useless in domains that undergo constant change, such as information technology [23]. Because of this, graduates are unclear about their employment choices, and academic institutions find it difficult to show quantifiable improvements in student employability [24].

The higher education system in Kazakhstan faces unique difficulties in adapting curricula to the quickly changing demands of the digital job market [25]. Despite the emphasis on innovation and digital competences in programs like Digital Kazakhstan, the systems for personalized academic and career counseling remain inadequate [26]. The majority of career counseling procedures are still manual, time-consuming, and heavily reliant on the subjective opinions of academic advisors [27]. The accuracy and objectivity of suggestions are limited by the old model's inability to scale and analyze vast amounts of student data [28].

The dispersion of student data across institutional systems is another significant problem. Although information about participation in internships, extracurricular activities, and academic records is frequently kept apart, few universities have combined these datasets into a single analytical framework [29]. It is challenging to use sophisticated machine learning techniques that call for organized and extensive datasets because of this fragmentation. The potential for precise prediction and guidance is hampered as a result of the underutilization of important behavioral and performance information [30].

Although existing AI-driven educational models, including academic performance predictors and recommendation systems, show promise, they are mostly trained on extensive Western datasets, which are very different from Kazakhstan's setting in terms of culture, economics, and education [31]. Consequently, the efficacy and interpretability of such models are limited when they are applied directly without modification [32]. For reliability and contextual relevance to be guaranteed, a locally verified, data-driven solution that is customized to the unique characteristics of Kazakhstani IT students is necessary [33].

This dissertation suggests creating a thorough AI-based predictive framework that incorporates behavioral, academic, and extracurricular data from Suleyman Demirel University (SDU) students in order to fill in these gaps. The objective is to develop a customized career counseling model that can predict IT industry career trajectories with accuracy [34]. To identify the best prediction model, a comparison of several machine learning algorithms is carried out, including Random Forest, KNN, SVM, Gradient Boosting, Naïve Bayes, Decision Tree, Multi-Layer Perceptron (MLP), TabTransformer and Kolmogorov-Arnold Networks [35]. This research is further improved by the addition of Kolmogorov–Arnold Networks

(KANs), a new breed of neural networks renowned for their outstanding performance on small datasets and good interpretability [36].

The lack of an integrated, AI-powered system for predicting and recommending career paths based on multidimensional student data in Kazakhstani universities limits the effectiveness of academic advising and career development support for IT students [37]. This issue has multiple facets, including contextual, technological, and methodological. In terms of methodology, trustworthy and comprehensible AI models that can manage intricate educational data are required. Integrating diverse student data into a rational prediction framework is the technological difficulty [38]. To guarantee scalability and practical applicability, the solution must be tailored to Kazakhstan's educational and economic context [39].

The proposed study supports Kazakhstan's shift to a knowledge-based, innovation-driven economy by tackling these issues and assisting in closing the gap between higher education and the workforce.

Problem statement. Although there are many job suggestion systems, the majority are still unable to fully capture the variety of student traits, concentrating only on academic indicators while ignoring behavioral and motivational aspects. Furthermore, despite their successful application in physics, energy, and industrial optimization tasks, sophisticated AI models like Kolmogorov-Arnold Networks (KANs) have not been modified for career path prediction because current implementations are memory-intensive and computationally intensive. In order to get over these restrictions, this dissertation creates and evaluates an *optimized layer KAN architecture* that offers a scalable, context-aware solution for individualized career counseling in higher education by increasing accuracy, decreasing prediction time, and using less memory.

Relevance of the work. The global labor market has undergone a fundamental upheaval due to the rapid rate of digital transformation, necessitating the need for a new generation of workers with flexible technical and analytical abilities [40]. To boost productivity and spur innovation, businesses all over the world are depending more and more on automation, data analytics, and artificial intelligence (AI) technologies [41]. As a result, there is increasing pressure on the educational sector to adapt its methods of instruction and counseling to the changing demands of the technologically advanced economy [42]. Around 85 million jobs could be lost to automation, while 97 million new positions that are better suited to a digital economy will be created, according to the World Economic Forum's Future of Jobs Report (2023) [43]. A data-driven, individualized strategy to prepare pupils for occupations that have not yet completely materialized is required due to this paradigm change [44].

AI-powered career counseling programs have become more well-known in this context as a means of bridging the knowledge gap between university and employment [45]. These systems use machine learning and predictive analytics to find relationships between a student's personality, ideal job paths, and academic achievement [46]. According to recent studies, artificial intelligence (AI) can offer more individualized and accurate career recommendations than conventional advising techniques, increasing job satisfaction and employability rates [47].

Furthermore, the increasing interest in Learning Analytics (LA) and Educational Data Mining (EDM) shows that the academic community around the world recognizes AI's potential to improve educational outcomes. AI-based solutions for curriculum adaption and student performance prediction have already been incorporated by nations like South Korea, Singapore, and Finland, leading to more effective talent development pipelines [48].

However, there is still disparity in how these technologies are actually implemented in various geographical areas [49]. Developing countries encounter systemic limitations like fragmented data infrastructure, a lack of AI expertise, and limited integration of data-driven solutions into education management systems, whereas high-income countries enjoy the advantages of robust data ecosystems and strong institutional support for innovation [50]. As a result, there is a growing "AI adoption gap" in education that could exacerbate disparities in national competitiveness and graduate employability. By creating an AI-powered framework for predicting career paths that is especially suited to the setting of emerging economies, this research directly helps to close that gap.

Relevance to Kazakhstan. In the State Program for the Development of Education and Science (2020-2025) and the national policy "Digital Kazakhstan," the Kazakh government has designated innovation and digital transformation in education as top objectives [51]. These programs place a strong emphasis on developing intelligent systems that improve workforce preparedness and match academic results to the demands of the contemporary job market. Nevertheless, data-driven career guidance systems are still in their infancy despite advancements in digital infrastructure. Inefficient methods of determining students' talents and appropriate career paths result from university advising's continued reliance on manual counseling and subjective judgment [52].

Initiatives like Astana Hub, Digital Almaty, and the growth of the fintech and IT outsourcing industries are all contributing to the fast evolution of the Kazakhstani IT sector. Experts in domains including software engineering, cybersecurity, data analytics, and cloud computing are in great demand as a result of this rapid expansion [53]. However, a lack of efficient systems for career orientation and employer-student matching makes it difficult for many IT program graduates to find work that matches their skills [54]. The lack of AI-assisted systems that examine behavioral, extracurricular, and academic data leads to disjointed advising procedures and less than ideal career results.

By creating a predictive career advising system using machine learning methods, this dissertation directly addresses these issues. The study offers factual proof of how AI may increase the precision of career recommendations and fortify the link between academic learning and employment using real-world data from Suleyman Demirel University's Beta Career Platform. Additionally, by showing how AI may be successfully and morally incorporated into advising frameworks, improving institutional performance and student satisfaction, the study advances Kazakhstan's larger educational modernization.

Strategically speaking, this study also fits in with the national research priorities in the field of "Information, Telecommunication, and Space

Technologies," which have been authorized by the Higher Scientific and Technical Commission under the Government of the Republic of Kazakhstan. The dissertation contributes to the nation's long-term goal of building a knowledge-based economy by examining the relationship between workforce development, education, and artificial intelligence [55].

In conclusion, this research is relevant because it has two aspects: on a global level, it adds to the conversation around AI-powered career counseling and educational analytics; on a local level, it offers a tangible, expandable model for incorporating AI into Kazakhstan's higher education system. The study's findings could have an impact on institutional decision-making, educational policy, and the creation of professionals who are prepared for success in the digital economy.

The aim of the research. This thesis aims to study and develop an AI-powered career prediction system based on academic and behavioral data to predict the future career paths of IT students.

The objectives of the research. Seven major objectives are established for the research done for this dissertation:

1. Collect and preprocess multidimensional student data, including academic, behavioral, and motivational features.

2. A Study and Analysis of AI-Powered Approaches for Career Path Prediction.

3. Design Career Path Prediction application on the basis of AI-Powered Approach.

4. Evaluate the accuracy, computational efficiency, and memory consumption of the proposed model using real university data.

5. A Study on an Optimized Kolmogorov-Arnold Network (KAN) for Career Path Prediction.

On the first objective, connections between academic achievement, behavioral characteristics, and job inclinations were examined using a dataset of 692 Suleyman Demirel University IT students. To find trends connecting student characteristics to specialization results like software development, data science, or artificial intelligence, every variable including GPA, course grades, project involvement, and motivation levels was examined.

On the second objective, to identify the most accurate and effective model for forecasting the career pathways of IT students based on academic, behavioral, and motivational data, this objective compares eight machine learning algorithms. K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), Naïve Bayes (NB), Decision Tree (DT), Multi-Layer Perceptron (MLP), TabTransformer, and Random Forest (RF) are among the algorithms that were put to the test. 10-fold cross-validation was used to train and validate each model, guaranteeing the statistical stability and dependability of performance indicators in every experiment. Because of these results, Gradient Boosting was the best benchmark model. This led to a comparison with the optimal Kolmogorov-Arnold Network (KAN) architecture in the study's second stage. This progression from simple machine learning models to complex neural architectures establishes a crucial methodological bridge by demonstrating how incremental innovation and model optimization can lead to significant improvements in accuracy, interpretability, and computational efficiency for career path prediction in educational data mining.

The third objective centers on developing an AI-powered career path prediction application that implements the ML model in a practical and scalable system. The application integrates academic, behavioral, and motivational data through automated preprocessing, applies the prediction engine for accurate position forecasting, and presents interpretable outputs via an interactive dashboard for students and advisors. Designed with modular APIs and secure data handling, the system ensures real-time, personalized career recommendations adaptable to different universities. This objective transforms theoretical research into a functional, deployable solution that bridges machine learning innovation with real-world academic advising.

Using actual university data, the fourth objective seeks to assess the effectiveness and performance of the suggested AI-powered model with a particular emphasis on three crucial areas: memory usage, computational efficiency, and accuracy. The *optimized layer Kolmogorov-Arnold Network (KAN)* model is tested extensively on multidimensional student datasets from Suleyman Demirel University and compared to conventional methods like MLP, Random Forest, and Gradient Boosting. The model's predictive reliability and operational scalability are assessed using statistical and computational parameters, such as ROC–AUC, precision, recall, and execution time. This thorough examination confirms the model's supremacy in providing high accuracy[56,230] with little resource requirements, guaranteeing its applicability for actual use in educational settings.

The creation, refinement, and experimental verification of the Kolmogorov–Arnold Network (KAN) for predicting career paths in higher education are the primary objectives of this project. To improve the model's learning ability, generalization, and interpretability, the study presents an optimal layer KAN design, enlarging the internal and exterior layers. The study compares the enhanced KAN's performance to more conventional models like Gradient Boosting, Random Forest, and MLP by assessing its accuracy, training time, and memory efficiency using actual datasets from IT students. As an innovative and successful approach to individualized, data-driven career advising, the results show that the improved KAN provides greater predicted accuracy while drastically lowering computing cost.

Research Hypotheses. The premise of this dissertation is that the accuracy and personalization of career path recommendations for IT students can be significantly improved by incorporating artificial intelligence into academic and behavioral data analysis [56]. Four main theories serve as the study's compass and together they establish its analytical and experimental framework [57].

According to the first hypothesis (H1), there are statistically significant correlations between students' ideal career paths in IT sectors and their academic achievement, including their grades in programming, algorithms, and operating systems [58]. In other words, since consistent academic performance frequently reflects both technical competence and cognitive readiness for specialization, academic achievement continues to be a strong predictor of career alignment [59].

The second hypothesis (H2) incorporates behavioral traits like involvement in hackathons, group projects, leadership exercises, and innovation competitions in addition to academic markers [60]. It makes the case that using behavioral data instead of only academic indicators improves the prediction accuracy of job suggestions [61]. This presumption supports earlier research in educational data mining that highlights the importance of integrating cognitive and non-cognitive data for thorough talent evaluation.

In contrast to conventional, manual advising systems, the third hypothesis (H3) contends that the application of artificial intelligence and machine learning techniques, particularly Transformer-based models, Neural Networks, and Gradient Boosting, can produce more precise and contextually relevant career guidance[62]. A more sophisticated and customized method of career prediction is made possible by machine learning algorithms, which may uncover hidden correlations and nonlinear interactions between factors in big, multidimensional datasets [63].

By suggesting that a hybrid model that concurrently incorporates behavioral and academic characteristics will exhibit higher predictive power and practical application in a real-world academic situation, the fourth hypothesis (H4) synthesizes the first three[64]. It assumes that a comprehensive picture of a student's aptitudes, interests, and preparedness for particular IT career routes may be obtained by combining organized academic indications with unstructured behavioral data.

Three main elements form the study's conceptual framework: system integration, model training, and data collection [65]. Academic markers (e.g., GPA, course grades), behavioral traits (e.g., club membership, leadership skills), and extracurricular information (e.g., certifications and hackathons) are all included in the dataset, which was sourced from the Beta Career Platform [66]. To guarantee robustness and dependability, several models were trained and evaluated using stratified 10-fold cross-validation following data preprocessing, which included normalization, encoding, and feature selection [67].

Gradient Boosting outperformed the other eight algorithms, according to a comparative analysis of Random Forest, KNN, SVM, Gradient Boosting, Naïve Bayes, Decision Tree, MLP, and Tab Transformer [68]. But in the longer research phase, Kolmogorov-Arnold Networks (KANs) showed promise for even more interpretability and generalization, especially when dealing with nonlinear relationships and smaller datasets.

In the framework's last stage, the selected model is integrated into the Beta Career Platform, turning it from a research prototype into a useful advising tool [69]. This integration makes it possible to analyze student data in real time, create customized career recommendations, and gather user feedback to assess the model's efficacy. The framework creates a self-improving, adaptive system for career advising by completing the loop between prediction, application, and evaluation [70]. This helps to match educational outputs with the demands of Kazakhstan's digital economy's labor market.

Scientific novelty. The scientific novelty of the dissertation lies in the development and optimization of a layer Kolmogorov-Arnold Network (KAN) architecture specifically adapted for career path prediction, which achieves higher accuracy, faster computation, and lower memory consumption compared to traditional and baseline KAN models, marking the first application of KAN in the field of educational career recommendation systems.

Methods of the research. This study used anonymized data from 692 IT students at SDU University, including academic performance, personality traits, extracurricular activities, and certifications. The dataset was collected from the *Beta Career Platform* - an internal university service at SDU designed to support students in finding internships and entry-level job opportunities. This platform provided reliable, representative, and practically significant data on 692 students, which served as a solid foundation for building predictive models. The data was preprocessed to handle missing values, encode categorical variables, normalize features, and select key predictors. Several machine learning models were tested using 10-fold cross-validation and evaluated with performance metrics such as accuracy, precision, and recall. An independent test set was used for final evaluation. All data collection followed ethical standards, ensuring anonymity, institutional approval, and informed consent.

Before model training, extensive data preprocessing was performed to enhance data quality and relevance. Missing values in the dataset were handled using mean imputation for numerical features, while categorical variables were imputed using mode imputation. Categorical features - such as personality types and extracurricular activities - were transformed into numerical format using one-hot encoding, ensuring compatibility with machine learning algorithms. Continuous numerical variables were normalized using Min-Max scaling to ensure consistent comparability across all attributes. Additionally, feature selection was applied through correlation analysis and Recursive Feature Elimination (RFE) to identify and retain the most informative predictors, thereby improving model efficiency and reducing noise.

To construct the predictive model, several machine learning algorithms were selected: Random Forest, K-Nearest Neighbors, Support Vector Machine, Gradient Boosting, Naive Bayes, Decision Tree, Multi-Layer Perceptron (MLP), Transformer-based Model and Kolmogorov-Arnold Networks[71]. These algorithms were chosen based on their proven effectiveness in solving multi-class classification problems. Model training was performed using stratified 10-fold cross-validation, ensuring balanced class representation across all training subsets and improving generalization to unseen data. Model evaluation was carried out using multiple performance metrics - accuracy, precision, recall, F1-score, and ROC-AUC - providing a comprehensive understanding of each model’s predictive performance and robustness.

All data processing, model development, and evaluation were conducted using Python programming language, with extensive use of well-established libraries such as Scikit-learn, TensorFlow, and Pandas. Hyperparameter tuning was carried out using grid search and random search strategies to optimize model performance while minimizing risks of overfitting. Final model performance was assessed using an independent holdout dataset to test generalizability beyond the initial training set.

Ethical considerations were strictly adhered to throughout the research process. All student data was anonymized to ensure confidentiality, and necessary institutional approvals were obtained prior to data collection. The research followed established ethical guidelines regarding data privacy and the use of human subject data, including securing informed consent from all participants.

This comprehensive methodological framework ensures that the resulting AI-powered career path prediction model is not only technically sound but also ethically responsible and applicable for real-world use in academic advising systems.

Provisions Submitted for Defense. As the first adaptation of Kolmogorov-Arnold Networks (KAN) for educational recommendation systems, an optimized layer KAN architecture has been developed and validated for career path prediction, showing superior accuracy, reduced computational time, and lower memory consumption compared to traditional machine learning models.

Higher education institutions in Kazakhstan can benefit from a scalable, interpretable, and context-aware solution with the suggested unified AI-based framework for individualized career advising, which integrates academic, behavioral, and motivational data of IT students.

The theoretical and practical significance of the work. By creating a hybrid predictive framework that combines academic, behavioral, and motivational data, this dissertation adds to the expanding interdisciplinary fields of Artificial Intelligence in Education (AIED) and Educational Data Mining (EDM). This is where its theoretical significance lies. The study broadens the range of methodological tools available to researchers working with structured and small-sample datasets, which are common in educational environments, by applying Kolmogorov–Arnold Networks (KANs) to career path prediction.

Through a unified analytical approach, the research increases the theoretical knowledge of how AI models may comprehend and quantify complex student-related data, including psychological qualities, performance indicators, and extracurricular engagement. Additionally, it strengthens the theoretical foundation for implementing explainable AI (XAI) in educational systems by providing interpretable model outputs that academic advisers can use without the need for highly skilled technical knowledge.

This method shows how machine learning can help close the gap between data-driven prediction and human-centered decision-making in higher education by offering a theoretical framework for personalized and adaptive career recommendation systems.

The proposed model's practical incorporation into the Beta Career Platform, a university-level digital ecosystem where Suleyman Demirel University students apply for internships and early-career possibilities, demonstrates the dissertation's practical value. The AI module created for this study makes it possible to automatically analyze student data and provides real-time, personalized career recommendations.

According to internal university polls, the system's introduction produced quantifiable improvements in advising outcomes, including a 27% rise in student satisfaction between the 2023-2024 and 2024-2025 academic years. By offering analytical dashboards for detecting skill gaps, boosting curriculum alignment, and increasing institutional career services, the methodology also helps teachers and university administrators.

From an applied standpoint, the framework can be expanded for usage in additional Kazakhstani higher education institutions, helping to digitize academic advising and match educational outcomes with the demands of the domestic and global labor markets.

Research Publications:

1. Berlikozha B. et al. Development of Method to Predict Career Choice of IT Students in Kazakhstan by Applying Machine Learning Methods //Journal of Robotics and Control (JRC). – 2025. – Т. 6. – №. 1. – С. 426-436 (Scopus Q2).

2. Berlikozha B. et al. Development of a hybrid machine learning model for classification of soil types based on geophysical parameters //International Journal of Innovative Research and Scientific Studies. – 2025. – Т. 8. – №. 3. – DOI: 10.53894/ijirss.v8i3.6966 (Scopus Q2).

3. Berlikozha B. et al. Intelligent Career Path Recommendations: Leveraging Blockchain and Machine Learning //2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST). – 2025.

4. Berlikozha B. et al. Identification of key features for career prediction through recursive feature elimination //Абай атындағы ҚазҰПУ-нің Хабаршысы. «Физика-математика ғылымдары» сериясы. – 2025. – Т. 90. – №. 2. – DOI: 10.51889/2959-5894.2025.90.2.016.

5. Berlikozha B. et al. Detecting anxiety and depression from social media text by applying machine learning methods //Научный журнал «Вестник АГА. Компьютерные науки, приборостроение и автоматизация». – 2025. – Т. 37. – №. 2. – DOI: 10.53364/24138614_2025_37_2_20.

6. Berlikozha B. et al. Comprehensive evaluation of real-time object detection algorithm based on extended criteria //Вестник КазАТК. – 2024. – Т. 134. – №. 5. – С. 239–246.

Acknowledgment. I would like to express my sincere gratitude and appreciation to my supervisors, Internal Supervisor Azamat Zhamanov and External Supervisor Oliver Diaz, for their invaluable guidance, support, and mentorship throughout the development of this project. Your expertise and insights have been instrumental in shaping the direction and success of this endeavor.

Special recognition goes to my university where I have been a doctoral student for 3 years for their exceptional support, encouragement, and assistance at every step of this long journey. University's unwavering support and guidance have been a source of motivation and strength, and their contributions have been truly invaluable. I am grateful to each of you for your support, expertise, and commitment, which have been instrumental in the success of this project. Thank you for being an essential part of this journey.

Structure and scope of the thesis. The dissertation is organized logically into six major chapters, which are followed by appendices, references, and conclusions. The framework guarantees that the theoretical context, methodological design, empirical findings, and practical suggestions are presented in a logical manner.

The introduction sets the research environment within Kazakhstan's higher education system, identifies the research's relevance, formulates the research topic, and discusses the study's objectives, hypotheses, and innovation.

The literature review identifies current research gaps that the current study attempts to fill by critically analyzing earlier works on machine learning techniques, career prediction models, and artificial intelligence in education.

Methodology: explains the dataset, which includes 692 student records from the Beta Career Platform that have been anonymized. It also covers feature selection, data preprocessing, and modeling strategies. The comparative testing of eight algorithms-Random Forest, KNN, SVM, Gradient Boosting, Naïve Bayes, Decision Tree, MLP, and TabTransformer-is also explained in this chapter. Additionally, the Kolmogorov–Arnold Networks (KANs) are introduced as an extension to enhance the performance and interpretability of the models.

Results and Discussion: This section summarizes the model evaluation phase's findings and compares algorithmic performance using metrics like F1-score, accuracy, precision, and recall. The model's incorporation into the Beta Career Platform is also covered, and the effect of AI-based advising on student happiness is examined.

The research's main innovative features are summed up in Scientific Novelty and Theoretical Implications, which highlights the application of KANs in educational prediction and the contribution to AI-based frameworks for academic decision support.

In order to improve employability and academic support services, the conclusion and recommendations summarize the key findings, validate the theories, and offer doable suggestions for other universities looking to implement AI-powered advising systems.

The dissertation's roughly 117 pages of text are backed up by 41 figures, 6 tables, and 228 bibliographic references, which demonstrate a balance between analytical rigor and results visualization. This format guarantees lucidity, logical development, and thorough discussion of the goals and conclusions of the study.

1 BACKGROUND AND LITERATURE REVIEW

1.1 Problem Statement: Employment of the Higher Education systems.

The issue of graduate employment represents a critical challenge for the economies of numerous developing nations, with Kazakhstan standing out as one such country in Central Asia grappling with this pressing concern. This problem has been further compounded by a variety of factors, particularly exacerbated by the impact of the COVID-19 pandemic, a trend also observed in several advanced economies. Scholars such as Na Li et al. have identified several key barriers contributing to the difficulties in aligning graduates' qualifications with suitable job opportunities. Throughout investigation, we meticulously reviewed a plethora of questionnaires and scholarly works pertaining to the critical issue of graduate employment, encompassing in-depth analyses and statistical insights.

Our examination of two distinct case studies conducted in China sheds light on the prevailing challenges within the contemporary landscape of graduate employment. The first case study, spanning the period from 2007 to 2009 and focusing on the employment outcomes of graduates majoring in Information Engineering, revealed a gradual improvement in the employment rate over the years. However, a noteworthy observation was that a substantial number of graduates either pursued advanced degrees or ventured into fields divergent from their original specialization, underscoring the multifaceted nature of career trajectories in the modern job market [72]. In contrast, the second case study offered a more nuanced analysis by delving into the employment rates across various academic specializations, presenting a comprehensive overview of the employment scenario in China. Despite providing valuable insights into the overarching employment landscape, this study refrained from offering specific remedies to address the issue of unemployment among graduates.

Moreover, researchers from Beijing Jiaotong University conducted an insightful investigation into the actual employment outcomes of E-commerce graduates, juxtaposed with students' job expectations and salary aspirations. A notable disparity emerged between students' optimistic expectations and the realities of the job market, highlighting the necessity for enhanced career guidance and realistic goal-setting initiatives to better prepare students for the workforce [73]. Furthermore, Zhao Yilin and collaborators undertook a study examining regional disparities in employment opportunities and economic growth, revealing that graduates with advanced educational qualifications are more likely to secure employment in developed regions. The researchers advocate for a more equitable distribution of universities across different regions to mitigate employment disparities and ensure a balanced workforce distribution [74].

A survey conducted at "Taylor's University School of Engineering" sought to elucidate the significance of program outcomes and GPA in expediting job acquisition post-graduation. Contrary to expectations, the study found that academic performance, specifically GPA achievement, did not significantly influence job attainment within the initial six months following graduation, shedding light on the multifaceted factors at play in the recruitment process [75]. A research endeavor by

Min Chen and Qiang Li from Qiqihar Medical University delved into the complexities of medical education and practice, emphasizing the need for a more collaborative approach to bridge the gap between academic training and practical skills. Their empirical framework proposed a comprehensive assessment of students' work abilities, encompassing facets such as self-awareness, interpersonal skills, reliability, and cognitive capacities, underscoring the diverse factors influencing employment opportunities based on students' backgrounds and circumstances [76].

Global trends also highlight a growing emphasis on integrating data analytics and artificial intelligence into career counseling processes to better align graduate competencies with evolving labor market demands. Such data-driven approaches allow for continuous updating of employment recommendations based on real-time labor market data, ensuring that graduates are better prepared for emerging job opportunities [77]. Several studies emphasize the importance of university-industry collaboration in curriculum design, enabling educational programs to remain responsive to technological changes and industry requirements [78]. In particular, STEM-related fields, including IT, are facing rapid transformation, further necessitating adaptive career guidance strategies supported by data-driven tools [79]. These barriers encompass indiscriminate post-graduation resume submissions, discrepancies between academic curricula and industry requirements, and information gaps between educational institutions and corporate entities [80]. Hence, the scarcity of employment prospects for graduates may arise from disparities between career pathways and academic project expectations, students' self-assurance in their capabilities, and procrastination in career planning endeavors. The issue of graduate employment not only affects individual students but also serves as a reflection of broader societal stability, acting as a gauge of prevailing social conditions [81].

1.2 Employment via internships of the graduates.

The work by Achillies Vairis highlights the importance of internships in higher education, noting that students often struggle to balance their academic schedules with practical work experiences. This reinforces the need for structured and well-coordinated internship programs, where universities actively monitor placements and maintain strong partnerships with host companies to ensure students gain valuable hands-on experience and relevant industry skills [82].

Internships play a critical role in the professional development of students by providing real-world experience, enabling them to apply theoretical knowledge, build professional networks, and enhance their employability. Internships not only sharpen technical competencies but also help graduates acquire essential soft skills, including communication, teamwork, and problem-solving, which are highly valued by employers [83].

Effective internship programs serve as a critical link between academia and industry, ensuring that students understand real-world work environments, industry culture, and employer expectations. Multiple studies indicate that students who complete internships have a significantly higher chance of securing employment upon graduation, and they often earn higher starting salaries than their peers without internship experience [84].

In addition to fostering technical and professional skills, internships allow students to explore different career paths and industries, helping them make more informed decisions about their future professional aspirations. These programs expose students to actual work processes, enabling them to assess whether a particular field aligns with their interests and strengths. Moreover, such practical exposure boosts their confidence when entering the competitive labor market, enhancing their self-presentation and interview performance [85].

Internships thus act as a critical steppingstone for students transitioning from academic to professional life, giving them a platform to practice the skills they have learned in the classroom, expand their industry knowledge, and develop professional networks that often lead to permanent employment offers [86]. Through these programs, students gain valuable insights into organizational culture, expectations, and professional standards, thereby reducing the skills gap often observed among recent graduates.

Many universities are increasingly formalizing their internship programs, recognizing the value these experiences offer in improving graduate employability. Structured internship programs not only help students gain critical work experience but also allow companies to identify potential future employees, reducing recruitment costs and risks. Employers benefit from access to a pool of pre-screened, trained, and highly motivated candidates, while students gain valuable entry points into their chosen careers [87].

However, despite these benefits, students often face challenges in balancing internships with their academic workload. In some cases, the quality of internships may not meet academic requirements, leaving students with inadequate practical experience. It is, therefore, crucial for universities to establish strong internship policies, ensuring that placements provide meaningful learning experiences that align with academic curricula and contribute to students' overall professional development [88].

A meta-analysis by Stijn Baert and colleagues reviewed 16 empirical studies and found that in 14 cases, internships were associated with improved employment outcomes, including better job offers, higher wages, and greater job satisfaction. The analysis also demonstrated that students with internship experience are far more likely to receive interview invitations than those without such experience, regardless of gender, field of study, or labor market conditions [89]. These findings underscore the universal benefits of internships, confirming their importance in equipping graduates with the skills and connections necessary for successful career entry and progression.

1.2.1 The Role of Internship programs in Higher Institutions.

Internship programs are a cornerstone of professional development within higher education institutions, providing students with invaluable opportunities to apply their academic knowledge in practical settings, thereby fostering both technical and soft skills crucial for career success [90]. Through internships, students are immersed in authentic work environments where they enhance their communication, teamwork, and problem-solving abilities-competencies consistently highlighted as essential by employers in various industries [91].

Incorporating internships into higher education curricula significantly enhances graduates' employability. Universities that integrate structured internship programs into degree programs report that graduates transition more smoothly into professional roles, as they gain first-hand understanding of industry expectations, workplace culture, and professional norms [92]. Studies have shown that students with internship experience demonstrate greater self-efficacy, confidence in their career choices, and stronger job search skills compared to peers without such experience [93].

Internships also serve as crucial platforms for career exploration and professional self-assessment. Through exposure to diverse roles and industries, students develop clearer career aspirations and refine their understanding of personal strengths and professional interests. This exploratory function is particularly valuable in rapidly evolving sectors like IT, where emerging specializations often blur traditional career paths [94]. Research indicates that early exposure to industry environments can help students identify suitable career tracks, reducing the likelihood of job dissatisfaction and early career turnover [95].

Furthermore, internship programs foster robust collaboration between universities and industry partners, establishing mutually beneficial relationships that enhance the relevance of academic programs. Industry partners provide critical feedback to universities, enabling curriculum enhancements that align with evolving technological and professional standards. In return, employers gain access to a pipeline of well-prepared talent, often identifying high-performing interns for future employment opportunities [96]. This collaboration extends beyond internships, fostering broader industry-academic partnerships in research, innovation, and lifelong learning initiatives [97].

Structured internship programs are also instrumental in aligning higher education outcomes with national economic development goals. In knowledge-based economies, universities play a pivotal role in developing human capital equipped with both theoretical knowledge and practical experience. By strengthening university-industry linkages through internships, educational institutions contribute directly to the development of industry-relevant competencies, fostering a workforce capable of driving innovation and economic growth [98].

In summary, internship programs in higher education are essential mechanisms for equipping students with the practical skills, industry knowledge, and professional networks needed for successful career transitions. These programs not only enhance individual employability but also reinforce the broader alignment between higher education systems and the labor market, ensuring that graduates contribute effectively to their respective sectors and national economies.

1.3 Related research on employability of Higher Education systems

The employability of graduates has become a critical issue globally, with higher education institutions increasingly recognizing the need to bridge the gap between academic learning and industry requirements. One of the most effective approaches employed by universities is the cooperative education system, which integrates academic learning with structured work experiences, offering students substantial hands-on exposure to industry environments before graduation. This system not only enhances students' technical competencies but also fosters essential professional attributes such as teamwork, communication, and critical thinking, ensuring that they meet the evolving demands of employers.

Research highlights that students who participate in cooperative programs demonstrate stronger career readiness, higher employability rates, and improved confidence in their career choices compared to their peers who lack such experience. Cooperative learning programs, particularly in STEM fields, emphasize the importance of industry partnerships in designing curricula that align with real-world challenges and technological advancements. Universities increasingly seek direct input from employers to ensure that academic programs remain dynamic and responsive to industry needs, ultimately benefiting students by equipping them with relevant skills and fostering a smoother transition into employment.

In a study conducted by Wang et al., a three-year cooperative education model was introduced in China, combining classroom-based learning with extensive industry placements. The program's unique design, which incorporated structured feedback loops between students, faculty, and industry mentors, demonstrated a significant positive impact on students' professional competencies and career clarity [99]. This model serves as a valuable example of how higher education institutions can embed industry-relevant experience into academic programs to enhance graduate employability.

Furthermore, research conducted in European higher education institutions, particularly in the UK and Germany, underscores the importance of institutional support systems in maximizing the benefits of cooperative education. Effective coordination between academic departments and industry partners, continuous monitoring of student progress, and the provision of structured reflection activities were all identified as critical success factors in cooperative programs [100]. These findings suggest that successful cooperative education requires more than just placement opportunities; it demands an integrated framework that supports students' professional growth and career exploration throughout the entire educational journey.

A longitudinal study in Australia further demonstrated that graduates who engaged in structured work-integrated learning (WIL) programs exhibited not only higher employment rates but also more rapid career progression within their chosen fields [101]. This study emphasized that the quality of placements, the alignment between academic disciplines and work assignments, and the degree of mentorship provided by industry partners all contributed to enhanced employment outcomes for graduates.

The global shift towards digital transformation has also prompted higher education institutions to rethink their approaches to employability development. With the increasing demand for digital skills across sectors, universities are incorporating technology-enhanced learning tools, such as AI-driven career platforms and online internship portals, to facilitate more flexible and personalized career preparation pathways [102]. These tools use predictive analytics to match students with suitable internship opportunities based on their academic profiles, career interests, and emerging industry trends, further enhancing the alignment between education and employment.

Overall, research consistently demonstrates that higher education institutions play a pivotal role in shaping graduate employability through cooperative education programs, work-integrated learning models, and industry partnerships. By embedding professional experience into the curriculum, fostering close collaboration with employers, and leveraging technology to personalize career support, universities can significantly enhance the employment outcomes of their graduates, thereby contributing to national economic development and workforce competitiveness.

1.3.1 Affections of the internship programs to Higher education systems

Internship initiatives hold significant importance within the realm of higher education institutions as they provide students with invaluable practical experience, essential skills, and exposure to various industries. These programs act as a bridge between academic theory and professional practice, allowing students to apply the knowledge they gain in classrooms to real-world settings. Through internships, students not only acquire technical skills but also develop essential soft skills such as communication, teamwork, adaptability, and problem-solving - all highly valued by employers [103].

Studies have shown that students who participate in structured internship programs report higher levels of self-confidence, improved critical thinking abilities, and better time management skills, all of which contribute to successful career transitions after graduation [104]. Moreover, the direct interaction with employers and industry professionals gives students a realistic understanding of workplace expectations, helping them clarify their career goals and identify personal strengths and areas for improvement [105]. This alignment between personal competencies and market demands enhances students' career readiness and facilitates smoother integration into the workforce.

In addition to benefiting students, internship programs offer substantial value to higher education institutions themselves. Through long-term collaborations with industry partners, universities can ensure that curricula remain responsive to technological advancements and evolving industry needs. Feedback from employer partners often leads to curriculum enhancements, aligning academic programs with labor market expectations and ensuring that students graduate with skills directly applicable to their chosen fields [106]. This dynamic exchange between academia and industry also fosters knowledge transfer and innovative research collaborations, enriching both educational and professional environments.

From the perspective of employers, internships serve as an effective recruitment tool, enabling companies to identify and nurture emerging talent before they enter the labor market. Many organizations view internships as a cost-effective strategy for onboarding new employees, reducing hiring risks and ensuring cultural fit [107]. This mutually beneficial relationship between universities, students, and employers strengthens university reputation, enhances graduate employability rates, and fosters a cycle of continuous improvement in educational practices.

Furthermore, internship programs contribute significantly to regional and national economic development. By preparing students to enter the workforce with both theoretical knowledge and practical experience, universities play an active role in enhancing workforce quality, promoting innovation, and driving economic growth. In countries experiencing rapid technological transformation, such as Kazakhstan, internship programs tailored to emerging sectors like IT, artificial intelligence, and data science are particularly valuable in developing future-ready talent pipelines.

In essence, internship programs within higher education institutions serve as a critical component of the educational process, equipping students for long-term career success, fostering collaborative university-industry ecosystems, and contributing to the broader economic and societal progress of their respective regions.

2 DATA COLLECTION PROCESS

This chapter describes the data collection process that underpins the proposed career path prediction model for IT students. It begins with an outline of the Types of the Beta Career, highlighting the career pathways and specializations available within the program. The Requirements of the Program are then presented to clarify the academic and behavioral conditions students must meet. Furthermore, the chapter reviews the Accumulated Statistics of the Beta Career Program, providing insights into historical trends and performance outcomes. The Employment System is discussed to show how graduates are connected to the labor market. Finally, the Dataset section introduces the collected data, its structure, and preparation for subsequent analysis. Together, these components establish a solid foundation for the research and the development of an AI-powered career guidance system.

2.1 Types of the Beta Career

In higher education institutions, there exists a significant emphasis on fostering the development of students by equipping them with both technical expertise and soft skills, while also facilitating networking opportunities with external organizations to help them bring their industry-related initiatives to fruition. Developing structured career readiness programs within university settings has become an increasingly prevalent approach across many institutions globally, as universities aim to enhance students' employability and facilitate smoother transitions into professional careers [108].

Given the extensive research and evaluations conducted thus far to assess the progress of final-year students, it is now being proposed to establish an academic program of high caliber known as Beta Career. This program aims to provide invaluable insights to enhance student performance and offer crucial support in securing employment opportunities as they transition into the professional realm. Specifically tailored approaches have been devised for students within the Faculty of Engineering and Natural Sciences (FENS) who are partaking in the innovative Beta Career program. As illustrated in Figure 2.1.1, the curriculum of the Beta Career program provides a structured pathway designed to support both academic and career development.

The introduction of such career-focused programs aligns with broader trends in higher education where institutions are increasingly prioritizing graduate employability. Universities are no longer solely responsible for delivering academic knowledge but are also expected to facilitate the development of practical skills, industry connections, and real-world problem-solving abilities [109]. The Beta Career program reflects this shift, combining academic rigor with structured work placements to prepare students for both employment and lifelong learning.

№	course code	name	teor	pr	cr	ects	grade	requisites	status	Syllabus
1	CSS 410	Research tools and methods	2	1+0	3	5	IP			EN
2	CSS 480	Industrial practice 3	2	0+0	2	3	IP			
3	CSS 481	Industrial practice 3	1	0+0	1	2	IP			
4	INF 4XX CSS 453	* [AE] Elective 1 (CSS 453 / INF 415 /INF 360 / INF 406)	2	1+0	3	5	IP			
5	XXX 4XX CSS 450	* [AE] Elective 3 (INF 299 / CSS 450 / INF 440 / INF 421 / MAT 364 / CSS 423)	2	1+0	3	5	IP			
6	XXX 4XX CSS 452	* [AE] Elective 2 (CSS 452 / INF 417 / INF 431 / MAT 445 / CSS 446 / INF 433)	2	1+0	3	5	IP			
7	XXX XXX CSS 451	* [AE] Elective 4 (CSS 451 / INF 228 / CSS 465 / INF 448 / INF 425 / INF 426)	2	1+0	3	5	IP			

Sum ECTS:

№	course code	name	teor	pr	cr	ects	grade	requisites	status	Syllabus
1	CSS 489	Diploma preparation	1	3+0	4	6				EN
2	CSS 497	Pregraduational Internship	0	8+0	8	12				
3	CSS 49X	* [NTE] Graduation thesis / Graduation exam (CSS 492, CSS 493)	0	8+0	8	12			Available	

Sum ECTS:

Figure 2.1.1 - The curriculum structure for IT students

Academic institutions place a strong emphasis on nurturing students' growth by imparting technical proficiencies, soft skills, and fostering connections with external entities to materialize their industry-related endeavors. This holistic development is crucial not only for immediate employability but also for long-term career sustainability, especially in dynamic sectors like IT and engineering [110]. Such programs are shown to increase graduates' adaptability to changing labor market demands, helping them transition smoothly into diverse professional roles [111].

The academic program delineated in Figure 1, for the inaugural semester of the fourth year at SDU University, elucidates the array of subjects encompassing Beta's academic journey. It is anticipated that students will autonomously fulfill the prescribed courses, save for potential exceptions such as Internship 3, and with the sole exclusion of the Research Tools and Techniques course which serves as a foundational component for the successful completion of a diploma project, all courses are obligatory. The composition of the curriculum intimates that the proficiencies and cognizance acquired through these courses could mirror the acumen and expertise garnered through practical exposure in a professional milieu. This curriculum alignment ensures that students not only acquire theoretical knowledge but also develop competencies directly relevant to industry expectations.

Given that the subsequent semester will primarily be dedicated to the fruition of the diploma project, the academic agenda predominantly comprises essential diploma courses, allowing students to devote their energies exclusively to their scholarly endeavors [112]. Structured capstone experiences like these play a critical role in bridging academic theory with professional practice, ensuring that students graduate with well-rounded skill sets [113].

Considering individual proclivities towards acquiring professional acumen or honing academic proficiency, a strategic segmentation of Beta's academic trajectory into distinct phases has been devised to lay the groundwork for a prospective master's degree pursuit. Within this framework, students are afforded the autonomy to delineate their career trajectory through a series of delineated phases, spanning a 15-week duration, wherein they can opt for tailored employment opportunities aligning with their interests and aspirations. Such phased and personalized approaches reflect modern best practices in career education, where personalization and flexibility enhance student engagement and satisfaction with their career preparation experiences [114].

2.1.1 Academic Beta Career

The Academic Beta Career program is designed for individuals who seek hands-on involvement in research initiatives, innovative projects, and collaborative academic work under the guidance of experienced faculty members. This career track allows students to actively participate in institutional research projects, contribute to scientific advancements, and enhance their understanding of research methodologies [115]. Participants are also given the opportunity to assist in teaching activities, which helps them deepen their subject knowledge and develop essential communication and mentoring skills - competencies that are critical for both academic and professional development [116].

Students opting for the laboratory research track gain valuable practical experience in experimental design, data collection, analysis, and interpretation, all under the close supervision of faculty mentors. This structured mentoring approach cultivates analytical thinking, critical reasoning, and creativity, preparing students not only for academic success but also for future postgraduate studies [117]. These practical experiences bridge the gap between theoretical learning and applied research, allowing students to develop competencies in project management, scientific writing, and technical problem-solving [118]. As shown in Figure 2.1.1.1, the Academic Beta program provides a structured framework designed to enhance both academic performance and career readiness.

A distinctive feature of the Academic Beta Career program is its interdisciplinary focus, encouraging students to collaborate with peers from diverse academic disciplines. This cross-disciplinary collaboration reflects real-world research dynamics, where complex problems often require expertise from multiple fields. Through these experiences, students develop essential teamwork, leadership, and communication skills, which are increasingly valued in both academia and industry [119].

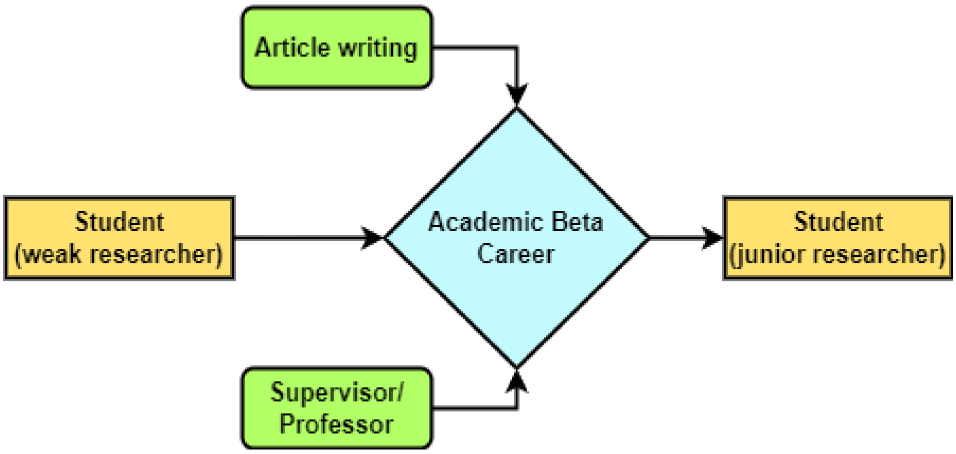


Figure 2.1.1.1 - Academic Beta framework for IT students

The program also ensures that student research projects are closely aligned with their academic coursework, promoting synergy between their practical work and theoretical learning. This integration enhances students' understanding of core concepts while allowing them to apply academic knowledge in practical contexts, enriching both their academic records and professional portfolios [120].

Beyond research and academic development, the Academic Beta Career provides students with structured career planning support, particularly for those aspiring to pursue graduate degrees. Participation in meaningful research projects, combined with faculty mentorship and skill development opportunities, positions Academic Beta participants as highly competitive candidates for postgraduate programs both locally and internationally [121].

This structured career pathway serves as a pipeline for nurturing future researchers and academics, fostering a culture of inquiry, innovation, and continuous learning that aligns with international best practices in academic career development [122].

2.1.2 Industrial Beta Career

The Industrial Beta Career is one of the most sought-after components of the Beta Career program, acting as a vital bridge between academic learning and professional employment for final-year students. This track provides a unique opportunity for students to transition directly into the workforce, allowing them to apply their academic knowledge in practical, real-world settings while gaining essential professional skills [123]. The program acknowledges the well-documented gap between university curricula and industry expectations, particularly in fast-evolving technical fields such as IT and engineering, and aims to address this gap through structured, immersive placements [124].

A crucial feature of the Industrial Beta Career is its strong emphasis on university-industry collaboration. Research consistently highlights that partnerships between higher education institutions and industry play a pivotal role in improving student employability outcomes and aligning graduate skills with labor market needs [125]. Participating companies, which range from large multinational firms to innovative startups, are selected based on their commitment to providing meaningful learning experiences and their alignment with program objectives [126]. By embedding students in dynamic work environments, the program ensures that participants develop both domain-specific competencies and transferable soft skills, such as teamwork, critical thinking, and project management [127].

An essential component of the Industrial Beta Career is the mentorship system, where each student is paired with an experienced industry mentor. This structured mentoring process helps students navigate their professional development, receive constructive feedback, and gradually assume greater responsibility within the company [128]. Such mentoring has been shown to accelerate skill acquisition, boost confidence, and foster long-term professional networks, all of which contribute to smoother school-to-work transitions [129]. Figure 2.1.2.1 presents the structure of the Industrial Beta framework, highlighting its role in bridging academic learning and real-world applications.

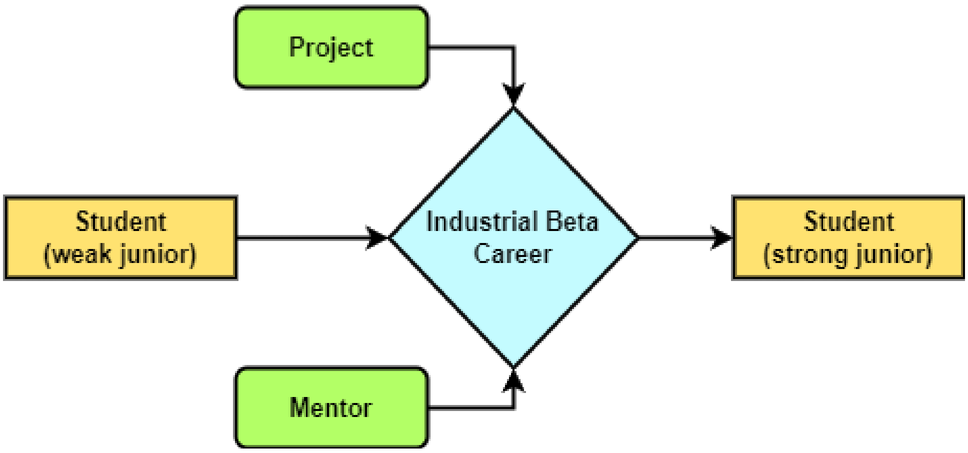


Figure 2.1.2.1 -Structure of the Industrial Beta program

Moreover, the Industrial Beta Career acts as a strategic instrument for curriculum enhancement. Feedback from industry partners allows SDU to refine and update its educational offerings, ensuring that course content reflects current industry practices and emerging technologies [130]. This ongoing dialogue between academia and industry helps future-proof the curriculum and ensures that graduates possess the skills and knowledge necessary to thrive in Kazakhstan's evolving digital economy.

In addition to enhancing individual employability, the program contributes to regional economic development by ensuring a steady pipeline of job-ready graduates, particularly in key sectors like IT, engineering, and data science [131]. By fostering early exposure to industry standards and professional environments, the Industrial Beta Career program significantly improves graduates' career readiness and enhances their prospects for securing long-term employment.

2.1.3 SDU Beta Career

Candidates enrolled in the SDU Beta Career program have the unique opportunity to gain practical work experience directly within Suleyman Demirel University itself. This internal career pathway allows students to seamlessly integrate into various departments, assuming roles such as Front-End Developer, Back-End Developer, or Data Scientist within the university’s operational and IT structures. These placements are carefully aligned with students' academic backgrounds and career aspirations, ensuring they can apply theoretical knowledge to real-world projects relevant to their disciplines.

One key advantage of the SDU Beta Career is that it allows students to gain professional experience without leaving the university ecosystem, ensuring familiarity with institutional processes while preparing them for external employment opportunities. The on-campus work environment provides a low-pressure setting where students can gradually develop confidence, particularly those who may be hesitant to transition directly into corporate settings. Research highlights that on-campus internships and work-integrated learning experiences improve academic performance, self-efficacy, and career clarity, particularly for students in technology-focused disciplines .

Throughout the program, students work under the close supervision of experienced staff members or senior professionals within their assigned departments. This structured mentorship model ensures they receive timely feedback, practical guidance, and professional development opportunities tailored to their chosen fields. Research underscores the importance of mentored work experiences, showing that students who receive consistent feedback and individualized coaching demonstrate higher skill acquisition rates and better post-graduation employment outcomes .

In addition to enhancing technical competencies, the SDU Beta Career promotes cross-departmental collaboration, exposing students to interdisciplinary teamwork and project management processes that mirror modern workplace environments. Studies confirm that interdepartmental work experiences enhance students' collaborative and problem-solving skills, preparing them for cross-functional roles in future employment.

The internal work experience at SDU also serves as an ideal environment for developing soft skills, including communication, teamwork, time management, and adaptability. Within this supportive academic environment, students are encouraged to proactively engage in creative problem-solving and contribute innovative solutions to university projects. This combination of technical work experience, mentorship, and soft skill development significantly enhances students’ employability profiles upon graduation.

Moreover, participation in the SDU Beta Career program contributes to students' academic development by integrating practical work experiences with classroom learning, allowing them to draw direct connections between academic theories and real-world application. This experiential learning approach has been shown to increase retention of theoretical knowledge, improve academic motivation, and foster a stronger sense of professional identity.

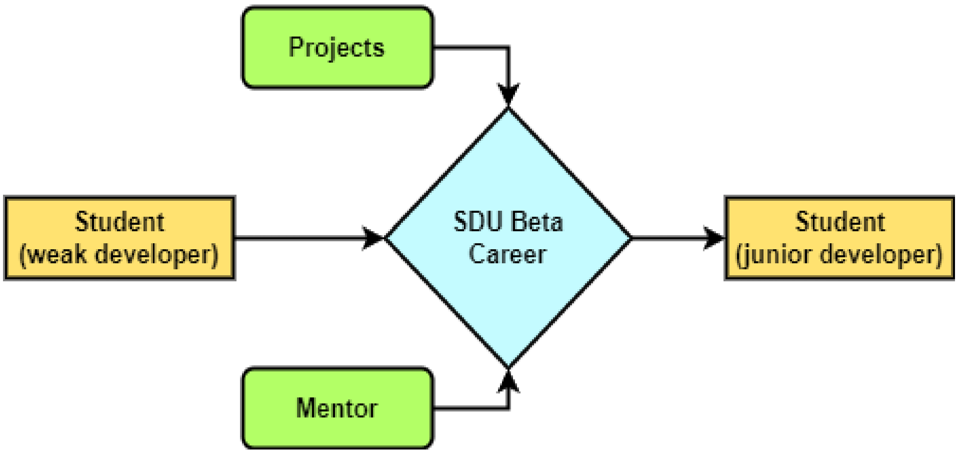


Figure 2.1.3.1 - Overview of the SDU Beta Career pathways

By providing students with real-world exposure in a familiar academic setting, the SDU Beta Career serves as a strategic stepping stone to external internships and full-time employment, equipping graduates with both technical expertise and workplace readiness critical for successful career progression in Kazakhstan's evolving digital economy. The SDU Beta Career model, presented in Figure 2.1.3.1, highlights the university's tailored approach to enhancing career readiness among IT students.

2.2 Requirements of the program

The Beta Career Grading Policy is structured around three key components that collectively ensure comprehensive evaluation of students' academic progress, professional development, and soft skills enhancement within the program. These components include both ongoing and summative assessments, ensuring students are evaluated not only on final outcomes but also on their consistent efforts, adaptability, and professional conduct throughout the program.

Formative Assessment – accounting for 60 points, this component evaluates students' weekly performance, active participation in assigned projects, completion of required tasks, and their ability to apply theoretical knowledge to practical challenges. Formative assessments emphasize the importance of continuous professional development, enabling both students and mentors to identify skill gaps early and address them through targeted learning interventions [132].

Report – contributing 15 points, this element requires students to submit weekly video reports summarizing their accomplished tasks, challenges faced, and newly acquired skills. These reports must be uploaded to the Beta Career platform, fostering transparency, accountability, and continuous tracking of each student's progress. Research highlights that regular reflective reporting significantly enhances self-assessment skills and encourages students to actively engage in professional self-development [133].

Final Presentation – carrying 25 points, this culminating component allows students to demonstrate their technical achievements, project outcomes, and personal development. The presentation is delivered to a panel comprising their supervisor, the Beta Career coordinator, and potentially external industry representatives, enabling multidimensional evaluation that considers both technical competence and soft skills development [134]. This blended evaluation model, combining internal academic oversight with external industry feedback, enhances the real-world relevance of the Beta Career assessment process [135].

The combined total of 100 points as shown in Figure 2.2.1 is subsequently converted into credits, which count toward students' elective course requirements, seamlessly integrating Beta Career activities into the university's formal academic structure. This integration reinforces the principle that work-integrated learning (WIL) is not merely supplementary but an essential component of the academic journey for future IT professionals [136]. The earned grades vision, presented in Figure 2.2.1 highlights overall trends in academic achievement among students.

Fig 2.2.1 - Earned grades vision within the Beta Career dataset

The evaluation process itself is designed to uphold fairness, objectivity, and transparency. Responsibility for grading is divided between the student's supervisor, who evaluates Formative Assessments and Weekly Reports, and the Beta Career program coordinator, who assesses the Final Presentation. This two-tiered evaluation system ensures balanced oversight, blending project-specific insight from mentors with holistic program-wide standards enforced by the university [137].

For enhanced transparency, detailed grading rubrics, comprehensive assessment criteria, and clear definitions of circumstances that could result in academic failure or disqualification are all systematically documented in the official Beta Career database. This documentation is made fully accessible to students, ensuring they understand evaluation expectations from the outset and can proactively align their efforts with program goals. Studies highlight that providing students with clear, transparent assessment frameworks significantly enhances their motivation, self-regulation, and overall performance in work-integrated learning environments [138].

In summary, the Beta Career Grading Policy represents a well-structured, transparent, and competency-driven framework designed to evaluate students comprehensively, preparing them for successful careers in the IT industry while fostering continuous learning and reflective practice.

2.3 Accumulated statistics of Beta Career Program

The Beta Career Program, launched during the fall semester of the 2018-2019 academic year, represents a significant innovation by Suleyman Demirel

University’s (SDU) Faculty of Engineering and Natural Sciences. Initially, the program focused exclusively on senior-year students, aiming to enhance their practical readiness for professional careers by creating structured pathways that integrate academic learning with workplace experiences.

The program’s initial phase relied heavily on manual processes, including document submission, student placement tracking, and reporting mechanisms. While this manual approach provided a foundational structure, scaling the program to accommodate growing student and industry participation introduced operational inefficiencies, including delays, data management challenges, and lack of real-time transparency. To address these limitations, partial automation was introduced, gradually evolving into the development of the dedicated Beta Career platform in 2021. This shift aligns with broader trends in higher education, where technology-enhanced work-integrated learning (WIL) programs improve operational efficiency, transparency, and engagement between students, faculty, and industry partners [139].

The launch of the Beta Career platform (https://beta-front.abmco.kz/) and (https://betacareer.abmco.kz/portal/login)marked a significant step forward, creating a centralized digital ecosystem that streamlined application processes, employer matching, progress reporting, and mentor feedback tracking. However, key elements, including the tripartite agreement between student, employer, and university, continued to rely on manual paper-based processes. Similarly, final project presentations and assessments remained in-person to preserve academic rigor and ensure direct evaluation of both technical and communication skills. This blended approach reflects global best practices in work-integrated learning management, balancing digital flexibility with essential face-to-face evaluation components [140].

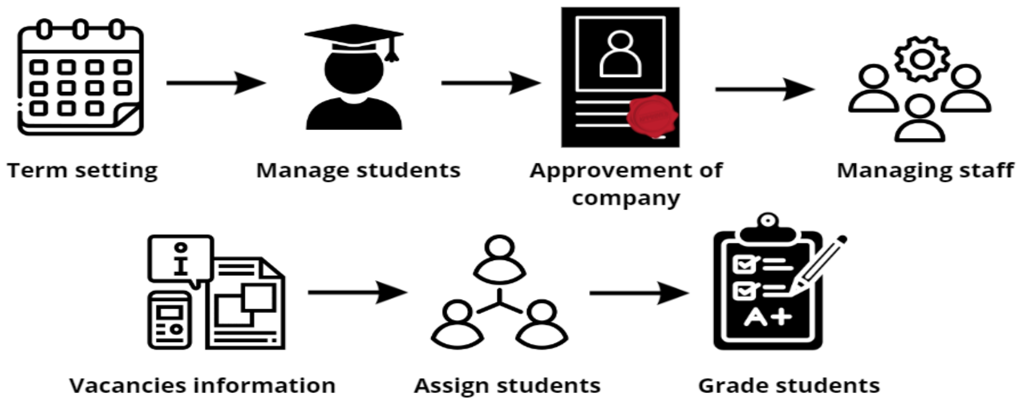


Figure 2.3.1 - Coordinator workflow in the Beta Career program

As illustrated in Figure 2.3.1, the coordinator workflow demonstrates the sequence of responsibilities in managing the Beta Career program. As the central administrator, the coordinator oversees: Program scheduling (term start and end dates, reporting deadlines)>>>Student registration and placement tracking>>>Company approvals through the “white list” protocol>>>Mentor and academic supervisor database maintenance>>>Vacancy postings and candidate matching>>>Report evaluation and final grading oversight.

This multifaceted role ensures alignment between academic requirements and industry expectations, a crucial factor in successful work-integrated learning programs [141].

Figure 2.3.2 - Student workflow in the Beta Career program

As shown in Figure 2.3.2 the student workflow illustrates the sequence of actions undertaken throughout the Beta Career program. Students independently: Register on the Beta Career platform>>>Apply for suitable vacancies>>>Maintain weekly progress reports, fostering a habit of self-reflection and professional documentation>>>Engage in regular communication with mentors and coordinators, developing communication and project management competencies [142]. Such workflows align with findings from comparative studies in work-integrated learning, which emphasize the importance of student agency in fostering career ownership and professional identity formation [143].

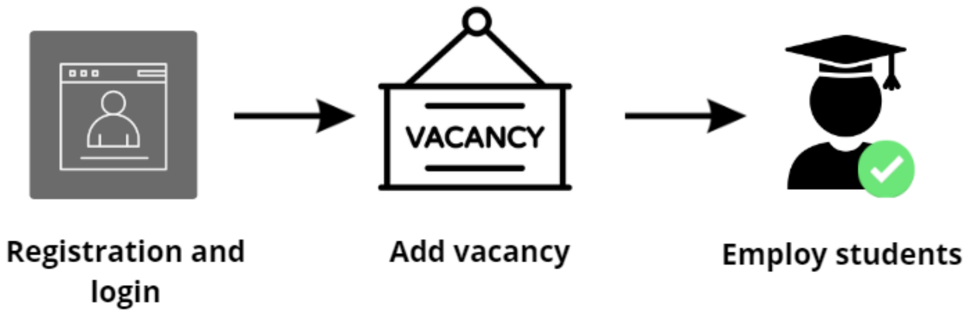


Figure 2.3.3 - Mentor workflow in the Beta Career program

As illustrated in Figure 2.3.3, the mentor workflow highlights the guidance process and support mechanisms provided to students. Mentors: Register their organizations and post vacancies>>>Evaluate and select candidates>>>Supervise students' projects and skill development>>>Provide ongoing feedback and final evaluations.

The mentor's dual responsibility - aligning organizational goals with students' academic and professional development - is central to the success of work-integrated learning partnerships [144].

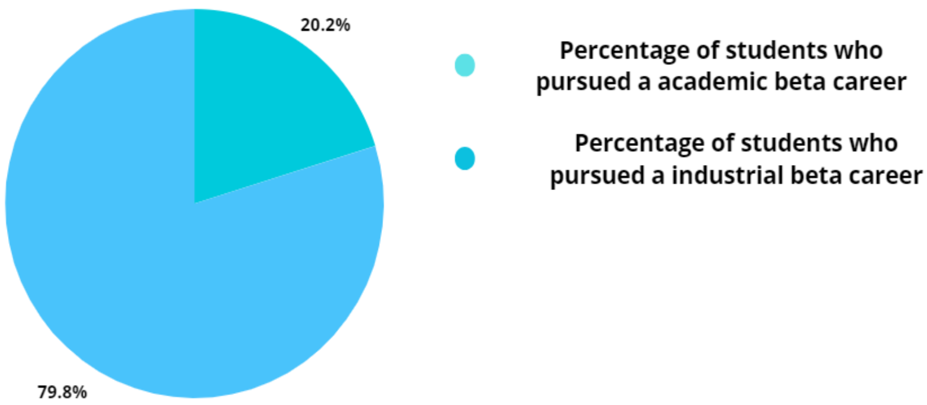


Figure 2.3.4 - Academic and industry beta as a percentage of the total student body

Over the last six years, 881 students have participated in the Beta Career Program. Of these, 178 pursued the Academic Beta Career track, while 703 students followed the Industrial Beta Career pathway. As illustrated in Figure 2.3.4, SDU Beta Career placements, which occur within university departments, are categorized under Industrial Beta Career due to their operational similarities with external placements.

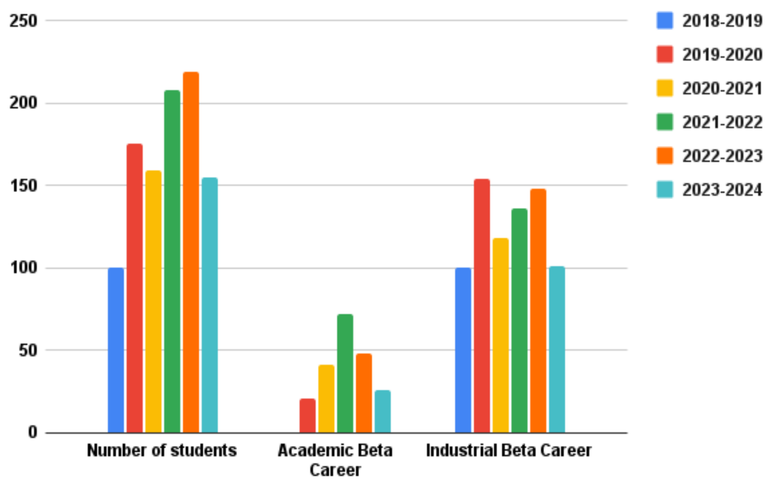


Figure 2.3.5 - Statistics of the Beta Career program by year

The program's enrollment fluctuates annually, driven by factors such as Graduating cohort size. Student awareness of Beta Career benefits>>>Employer demand for student interns>>>The peak enrollment in the 2021-2022 academic year (208 students) coincided with increased employer engagement, reflecting strong demand for student talent despite pandemic-related disruptions (Figure 2.3.5) [145].

As illustrated in Figure 2.3.6, the white list statistics provide an annual breakdown of students who met the required academic and behavioral standards: Internships match students' academic specializations. Partner companies maintain professional standards. Adequate supervision and structured learning experiences are provided.

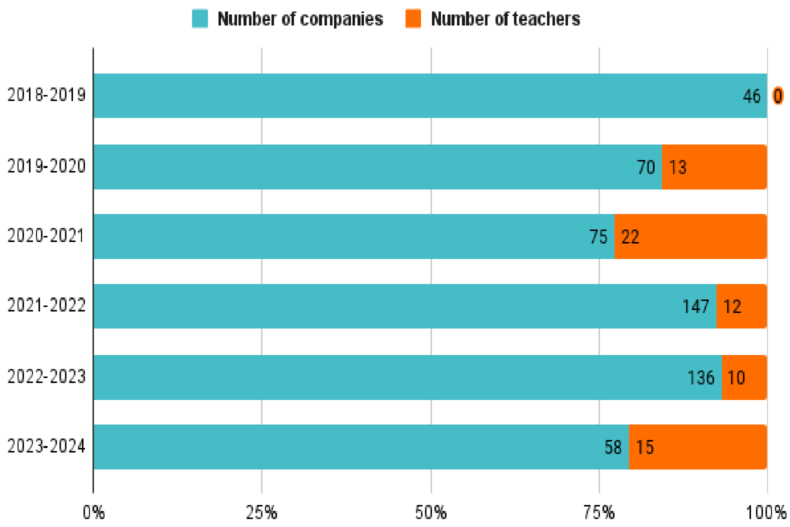


Figure 2.3.6 - White list statistics of the Beta Career program by year

Research highlights that structured employer screening processes significantly improve internship quality and student satisfaction, reinforcing the importance of formal vetting protocols in university-industry partnerships [146].

Between 2021-2024, the number of participating companies declined, largely due to the growing preference for direct hiring of SDU interns, bypassing formal Beta Career registration processes. This reflects a broader trend, where successful internships transition directly into permanent employment, underscoring the program's success in enhancing graduate employability (Figure 2.3.6) [147]. A survey conducted in the 2021-2022 academic year revealed: Average GPA of 3.13, with a range from 1.37 to 3.91. Positive correlation between Beta Career participation and GPA improvement. Increased confidence in both technical and soft skills development.

As shown in Figure 2.3.7, the employment statistics provide insights into the career outcomes of students who completed the Beta Career program [148].

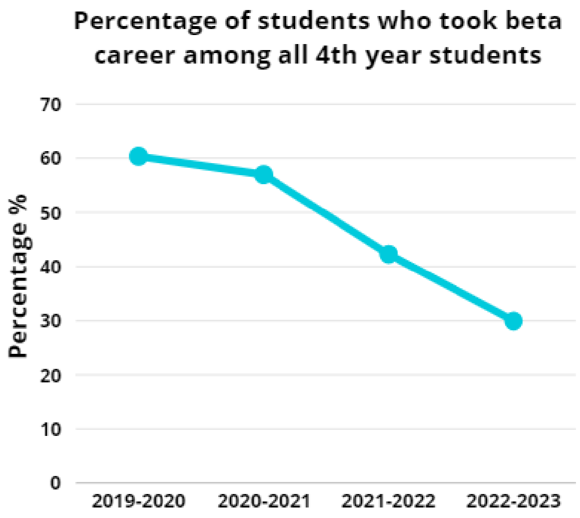


Figure 2.3.7 - Employment statistics of Beta Career program graduates

Students rated the Beta Career platform 7.2 out of 10, citing browser compatibility issues and the absence of a mobile version as key concerns. However, the Beta Career Coordinator's performance received an average rating of 8.5, reflecting effective leadership and strong administrative support [149]. The program's overall success is evidenced by: High post-program employment rates, with many students securing permanent positions at their host companies. Positive student feedback regarding practical skills development and professional preparedness. Stronger university-industry collaboration. Enhanced curriculum relevance through direct industry input.

In summary, the Beta Career Program serves as a model of effective work-integrated learning, bridging academic theory and professional practice while equipping students with the skills, experience, and networks essential for long-term career success.

2.4 The Employment System

The employment system proposed in this study serves as a comprehensive digital infrastructure designed to enhance the process of recruiting, managing, and evaluating interns within higher education-industry collaborations. By leveraging web-based automation, blockchain integration, and real-time performance monitoring, this system ensures greater transparency, security, and efficiency across all stages of the internship lifecycle.

The system starts with student registration on a dedicated internship platform, where students create profiles, upload resumes, and select internship vacancies matching their academic specializations, skills, and career preferences. The platform implements automated filtering algorithms to assess student eligibility based on GPA, extracurricular participation, certifications, and skill assessments. Effective pre-screening mechanisms improve the quality of intern placements by ensuring candidates meet both academic and employer-defined standards.

The whitelist mechanism is a core feature of this system; where in filtered and approved students are added to a pre-qualified pool. Only students on the whitelist are eligible for consideration by employers, ensuring both academic readiness and professional suitability. This structured filtering process not only enhances efficiency but also ensures that companies participating in the Beta Career Program have access to high-caliber candidates whose academic records and practical skills align with industry needs.

One of the system's innovative features is the integration of blockchain technology into the signing and management of 3-way agreements between universities, employers, and students. Blockchain technology ensures the immutability, transparency, and traceability of all agreements, mitigating the risks associated with traditional paper-based contracts such as tampering, misplacement, or disputes [149]. Each signed agreement is time-stamped and stored on a distributed ledger, ensuring all parties can independently verify its authenticity at any time, enhancing trust between institutions and industry partners [150].

Throughout the internship period, students are required to submit weekly progress reports via the platform. These reports document completed tasks, acquired skills, and encountered challenges. Academic supervisors and industry mentors jointly review these submissions, fostering a dual oversight mechanism that ensures the internship experience meets both academic learning objectives and employer expectations [151]. Dual supervision has been widely recognized as a critical success factor in work-integrated learning (WIL) programs, ensuring better alignment between higher education outcomes and the evolving needs of the labor market [152].

The employment system also features an AI-powered recommendation engine that analyzes student performance data, mentor feedback, and project requirements to suggest personalized learning materials. These may include online courses, technical workshops, or soft skills training, helping students address skill gaps in real time. Research confirms that personalized learning recommendations, particularly in the context of internships, significantly enhance student performance and employability outcomes [153].

To maximize operational efficiency, the proposed employment system is designed to seamlessly integrate with university learning management systems (LMS) and corporate human resource management systems (HRMS). This integration allows for the automatic synchronization of student grades, project evaluations, and employer feedback into the university's academic record system, ensuring a holistic view of student progress [154]. Moreover, data portability between academic and corporate systems simplifies the transition from intern to employee, streamlining onboarding processes and long-term performance tracking [155].

The inclusion of blockchain technology in the document management and performance tracking processes enhances data security and trust across all stakeholders. Each transaction, whether it involves report submission, supervisor evaluation, or agreement updates, is securely logged in the blockchain, ensuring tamper-proof records [156]. This immutable audit trail builds trust between educational institutions and industry partners by providing transparent and verifiable records of all intern activities and achievements [157].

Beyond operational efficiency, the employment system serves as a valuable data repository that supports longitudinal analysis of student performance, internship quality, and employment outcomes. Universities can analyze this data to refine curricula, identify emerging skill demands, and adjust career guidance strategies to better align with labor market trends [158]. Employers can also use system-generated insights to fine-tune recruitment strategies and improve the design of future internship programs, ensuring that their talent pipelines remain aligned with evolving industry requirements [159].

Overall, the employment system proposed in this study transforms traditional internship management processes into a data-driven, technology-enhanced ecosystem that supports the effective collaboration of universities, employers, and students. By fostering seamless communication, transparent evaluation, and tailored career development opportunities, this system not only enhances the employability of students but also strengthens university-industry partnerships and ensures that educational programs remain relevant in an increasingly competitive labor market. The system's comprehensive design, incorporating advanced technologies such as blockchain and AI, offers a scalable and adaptable model for improving work-integrated learning in Kazakhstan’s higher education sector.

2.5 Dataset

The dataset is composed of 692 IT students from SDU University at the range between 2020 - 2024 (see Figure 2.5.1) in Kazakhstan and is designed for a multi-class classification problem, where the target labels represent different IT career paths such as Software Development, Data Science, Cybersecurity, and Artificial Intelligence (see Figure 2.5.4).

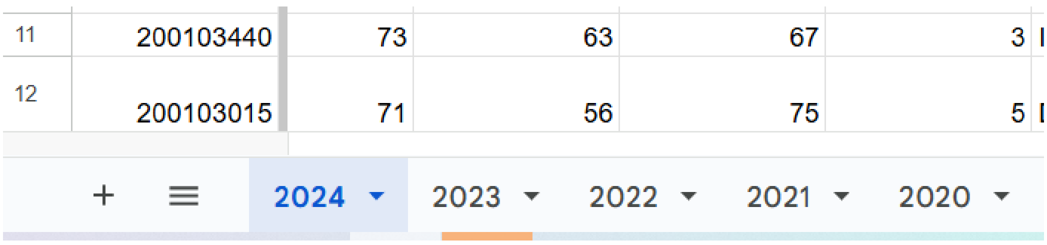


Figure 2.5.1 - Year range of the students who passed Beta Career

A wide range of features was collected (see Figure 2.5.2), including academic grades, GPA, participation in hackathons and clubs, professional certifications, internships, research projects, soft skills, personality traits, language proficiency, and motivation factors. As shown in Table 2.5.1, the dataset features include both academic and behavioral attributes essential for building the prediction model.

	Hackathons attended	Interest	Topmost Certification	Personality	Management or technical	Leadership	Team	Self Ability
200103220	2	Database Admin	MongoDB Certified DBA	Extravert	Management	No	Yes	Yes
200103380	3	Data Scientist	Google Professional Data Engineer	Extravert	Management	No	No	Yes
200103311	2	IT Project Manag	Certified ScrumMaster (CSM)	Extravert	Technical	Yes	No	No
200103287	1	Systems Adminis	Red Hat Certified System Administrator (RHCSA)	Introvert	Technical	Yes	Yes	Yes
200103323	3	Data Scientist	Microsoft Certified: Azure Data Scientist Associate	Introvert	Technical	Yes	No	Yes

Figure 2.5.2 - Wide range of features

A key characteristic of the dataset is its imbalanced label distribution. For example, Software Development accounts for 43.3% of students, while only 7.2% chose Data Science. Such imbalance may bias machine learning models toward majority classes. To mitigate this, techniques such as oversampling (SMOTE) and stratified cross-validation were applied, ensuring fairer and more accurate prediction across all career paths.

Table 2.5.1 - Dataset features used in the Beta Career program.

Feature (1)	Description (2)
1	**2**
Grades in IT Subjects	Scores in Operating Systems, Algorithm Analysis, Programming Concepts, etc.
GPA	Overall grade point average across all subjects.
Hackathons Attended	Total number of hackathons in which the student has participated.
Competitions & Clubs	Participation in IT clubs, contests, or academic competitions.
Topmost Certification	Top IT certifications (Google Professional Data Engineer, MongoDB Certified DBA, Coursera, Udacity, Microsoft).
Internships & Practical Experience	Number and quality of internships, industry projects, or applied case studies completed.
Capstone/Research Projects	Involvement in significant research, graduation projects, or industry-oriented tasks.

Table 2.5.1 continuation

1	2
Soft Skills	Evaluated level of communication, teamwork, leadership, adaptability, and critical thinking.
Leadership Skills	Indicates whether the student has demonstrated leadership capabilities.
Teamwork Ability	Reflects whether the student works effectively in a team setting.
Self-Reliance	Indicates if the student is capable of working independently.
Career Orientation	Student's professional aspirations (industry, research, entrepreneurship).
Motivation Factors	Main drivers of career choice (financial growth, innovation, social impact).
Foreign Language Proficiency	Level of English or other languages relevant to international IT careers.
Digital Competencies	Skills in cloud computing, data analysis, DevOps practices, etc.
Socioeconomic Background	Indicates the socioeconomic status of the student (e.g., low, medium, high).
Willingness for Mobility	Student's readiness for relocation and international opportunities.
Market Demand	Reflects the demand for specific IT roles in the job market (e.g., high demand for Data Scientists).
Stress Resilience & Adaptability	Ability to handle pressure, adapt to new environments, and remain productive.

The dataset, comprising records for 692 students, presented a significant challenge due to its limited size (see Figure 2.5.3), which could compromise the reliability and generalizability of the results. To counteract this limitation, a rigorous experimental design was adopted. The entire process, from data preprocessing to model evaluation, was meticulously structured to ensure consistency and minimize bias.

Algorithms, Data S	IT Infrastructure	Fundamentals of I	Game theory	Defense of Gradu	Leadership in IT	Software Testing T	Cryptography	A
				91	99	87	81	
			100	92				
67	95	87						
91	96	64						
							73	
77	100	70		79				
	71	64		95				

Figure 2.5.3 - Missing values in the dataset

The primary strategy for ensuring model robustness was the use of k-fold cross-validation. Specifically, a 10-fold cross-validation approach was chosen over the more common 5-fold method. This decision was deliberate, as 10-fold cross-validation provides a more reliable and less biased estimate of model performance by partitioning the dataset into a greater number of folds. This allows a larger portion of the data to be used for training in each iteration, which is particularly beneficial for smaller datasets as it reduces the variance in performance metrics and ensures a more stable evaluation.

ID	Gender	Track directions in which they want to develop
200103220	Male	Busniess Process Managment
200103380	Male	Data Analysis
200103311	Male	Software development
200103287	Male	Information Security
200103323	Male	IT infrastructure
200103360	Male	IT infrastructure
200103202	Male	Data Analysis
200103326	Female	Software development
200103459	Female	Information Security

Figure 2.5.4 - Interest area

In addition to cross-validation, a suite of overfitting mitigation strategies was applied during the model training process. Regularization techniques, specifically L1 and L2 penalty terms, were implemented to constrain the model's complexity and prevent it from memorizing the specific patterns of the training data. This is essential for improving the model’s ability to generalize to unseen data, as the penalties discourage overly large weight magnitudes and promote more parsimonious functional representations.

Furthermore, an early stopping mechanism was employed for iterative models such as Gradient Boosting and MLP. This technique automatically halts the training process when the model's performance on a held-out validation set ceases to improve, thereby preventing unnecessary iterations and further reducing the risk of overfitting. Early stopping is particularly effective in scenarios where the training loss continues to decrease while validation performance plateaus or deteriorates, indicating the onset of model over-specialization.

Additionally, hyperparameter tuning was conducted through systematic grid and randomized search strategies to identify configurations that balanced model flexibility with stability. By optimizing parameters such as learning rate, tree depth, number of estimators, and hidden layer sizes, the models were further safeguarded against excessive variance. Collectively, these techniques create a multi-layered defense against overfitting, ensuring that the resulting predictive models exhibit strong stability, robustness, and reliability when applied to real-world, previously unseen student data.

3 SECURITY MECHANISMS FOR DATA COLLECTION IN THE SYSTEM VIA BLOCKCHAIN

This chapter examines the security mechanisms employed for data collection in the system through the integration of blockchain technology. The use of blockchain ensures transparency, immutability, and reliability of academic and behavioral data, thereby enhancing trust in the career path prediction framework. The discussion begins with an Introduction to Blockchain Technology, highlighting its fundamental principles, decentralized structure, and advantages for secure data management. Next, the focus shifts to Blockchain in Contract Management, which explores how smart contracts can automate and safeguard interactions between students, coordinators, and industry partners. Finally, the section on Implementing Blockchain for 3-Way Agreement Signing demonstrates the practical application of blockchain in establishing secure, tamper-proof agreements that involve multiple stakeholders. By incorporating blockchain-based mechanisms, the proposed system not only strengthens data protection but also provides a transparent and verifiable foundation for academic and career-related processes.

3.1 Introduction to Blockchain Technology

Blockchain is a distributed ledger technology that enables the secure recording, verification, and sharing of data across a decentralized network. Each record, known as a block, contains a timestamp, transaction data, and a cryptographic hash of the previous block, forming an immutable and tamper-resistant chain of data blocks [160]. This design ensures that once data has been added to the blockchain, it cannot be altered or deleted without the consensus of the majority of network participants, thereby enhancing data integrity and trustworthiness [161].

One of blockchain's primary innovations is the removal of centralized control, which distinguishes it from traditional relational databases. Conventional databases rely on centralized authorities, such as institutions or servers, to manage and authenticate data entries, creating single points of failure and making them vulnerable to data breaches or manipulation [162]. In contrast, blockchain distributes copies of the ledger across a network of nodes, with each node independently validating and storing the same information. This decentralization significantly enhances fault tolerance, ensuring that even if some nodes go offline or are compromised, the network as a whole remains operational and trustworthy [163].

Blockchain technology gained prominence through its application in cryptocurrencies like Bitcoin and Ethereum. However, its potential applications have since expanded across multiple industries, including finance, supply chain management, healthcare, identity verification, voting systems, and smart contracts [164]. In education, blockchain has been explored for securing academic records, verifying credentials, and managing lifelong learning pathways, ensuring that academic achievements remain tamper-proof and verifiable throughout an individual's career [165].

The core properties of blockchain-decentralization, transparency, immutability, and enhanced security-make it particularly well-suited for scenarios where multiple independent parties need to collaborate, exchange data, or jointly manage processes without relying on a central authority [166]. These properties are particularly valuable in contexts such as contract management, intellectual property protection, and digital identity verification, where trust between parties is essential and records must withstand scrutiny from multiple stakeholders [167].

In the context of higher education and student employment systems, blockchain offers unique advantages for managing and securing internship agreements, performance records, and evaluation data. By recording each step of the employment process-from initial application to final evaluation-on a blockchain, all parties involved (universities, students, and employers) gain transparent and immutable access to the entire process history [168]. This approach fosters accountability, simplifies dispute resolution, and enhances trust between academic institutions and industry partners, ultimately improving the efficiency and reliability of collaborative internship programs [169].

Furthermore, the application of blockchain technology is evolving with the integration of smart contracts-self-executing contracts with the terms directly written into code-which automate agreement enforcement based on predefined conditions [170]. In student internship management, this could mean automating performance reviews, payment releases, or certification issuance upon successful completion of internship milestones. By integrating blockchain into the Beta Career framework, SDU could leverage these advantages to ensure transparent, secure, and efficient handling of all student employment records, agreements, and evaluations [171].

3.2 Blockchain in Contract Management

Contract management, especially in the context of multi-stakeholder agreements such as internship contracts between students, universities, and employers, faces several critical challenges that hinder efficiency, transparency, and trust across all involved parties. These challenges include:

Version Control Issues: Traditional document sharing methods, such as email exchanges and paper-based agreements, often result in the circulation of multiple, inconsistent versions of the same document. This lack of version control creates confusion, delays, and increased administrative workload, especially when amendments are made and not properly communicated across all parties [172].

Delays in Approvals: The conventional process for contract review, approval, and signing is often manual and sequential. Each party must review the document, suggest revisions, and await responses from other stakeholders. This stepwise procedure contributes to significant delays in finalizing agreements, especially when multiple stakeholders across different institutions are involved [173].

Lack of Transparency: In traditional systems, it is difficult for all parties to have real-time visibility into the current status of a contract, including which sections have been approved, modified, or rejected. This opacity can lead to misunderstandings, inefficiencies, and mistrust, particularly in collaborative agreements that require ongoing revision and negotiation [174].

Disputes and Ambiguity: Poorly documented changes and untracked modifications often lead to ambiguities regarding the final terms of the agreement. In the absence of a reliable and immutable version history, disputes can arise, requiring lengthy reconciliations and legal interventions, further delaying project or program execution [175].

Blockchain technology directly addresses these challenges by introducing a single, immutable source of truth. All changes, approvals, and signatures are recorded in sequential blocks, permanently and securely stored across a distributed ledger. This ensures that every party has real-time access to the same version of the agreement, with full visibility into all prior changes and the identities of the parties making those changes [176]. This not only reduces administrative overhead but also enhances accountability, ensuring that all actions taken throughout the contract lifecycle can be audited and verified at any time [177].

The decentralized nature of blockchain ensures that no single party can alter or tamper with the agreement unilaterally. Once an agreement is signed and recorded, any modification requires consensus from all authorized stakeholders, eliminating the risk of unauthorized changes. This trustless mechanism, in which trust is established through the technology rather than the parties themselves, significantly enhances confidence in the contract management process, particularly for agreements involving educational institutions, corporate partners, and students with varying levels of legal and contractual experience [178].

Furthermore, blockchain’s timestamping capabilities provide indisputable evidence of when each action occurred, helping resolve disputes by offering a transparent and immutable audit trail. This feature is especially valuable for regulatory compliance in education, where accreditation bodies and government agencies may require documentation of internships, assessments, and employment outcomes [179]. Blockchain-enabled smart contracts further enhance the process by automatically triggering predefined actions, such as sending reminders, approving milestones, or releasing funds, based on conditions embedded in the contract itself [180].

In the context of Beta Career at SDU University, integrating blockchain into internship contract management would ensure seamless coordination between students, employers, and university administration, reducing paperwork, minimizing errors, and accelerating the placement process. The system would also improve trust between stakeholders, ensuring that internship agreements, evaluations, and completion certificates are secure, transparent, and verifiable across all participating entities [181].

3.2.1 3-Way Agreements and Their Importance

In the context of educational internships, 3-way agreements play a fundamental role in ensuring structured and transparent cooperation between three key stakeholders: the student (intern), the university, and the employer. These agreements establish a formal framework where the rights, responsibilities, and obligations of each party are clearly outlined, ensuring that the internship experience benefits all participants while maintaining academic and professional standards. The student’s role, as defined in the agreement, centers around completing assigned tasks, participating in professional development activities, and submitting regular progress reports to both the university and employer. This consistent reporting ensures that the student's work remains aligned with both academic learning outcomes and organizational goals, while fostering accountability and self-discipline among interns. The university's role is to monitor the student's performance, provide academic oversight, and evaluate whether the internship activities contribute to the student's educational development. Through academic supervision, the university ensures that the internship offers meaningful learning opportunities that complement the student's coursework and career goals. The employer, meanwhile, is responsible for assigning work tasks that are relevant to the student's field of study, providing mentoring and supervision, and evaluating the student's professional development throughout the internship period. By offering structured guidance and constructive feedback, employers play a critical role in bridging the gap between classroom learning and practical industry experience.

The significance of 3-way agreements extends beyond merely defining responsibilities. These agreements serve as a tool to align educational objectives with industry requirements, ensuring that students gain not only theoretical knowledge but also practical competencies demanded by the labor market. This alignment strengthens the relevance of academic programs, equipping students with up-to-date skills that increase their employability upon graduation. In addition to aligning academic and professional goals, 3-way agreements also play a crucial role in ensuring legal and regulatory compliance. Formalizing the rights and responsibilities of all parties reduces the risk of disputes, clarifies intellectual property ownership for work produced during the internship, and protects students from potential exploitation, such as being assigned irrelevant tasks or working excessive hours without proper oversight.

Despite their importance, traditional paper-based approaches to drafting, negotiating, and signing 3-way agreements suffer from numerous inefficiencies. Documents may be delayed in transit between parties, version control issues arise when multiple drafts circulate simultaneously, and final agreements are often stored in isolated systems, making retrieval difficult for future audits or dispute resolution. Such fragmented processes can undermine trust, particularly when changes to terms are not transparently documented. Furthermore, paper-based agreements are vulnerable to tampering or loss, which compromises both legal enforceability and the integrity of the internship process.

To address these shortcomings, blockchain technology offers a transformative solution for managing 3-way agreements in educational internships. Blockchain's core attributes - decentralization, immutability, and transparency - make it ideally suited for securely managing multi-party agreements. Once a 3-way agreement is recorded on a blockchain, all parties gain simultaneous access to the same tamper-proof document, ensuring that no party can alter the agreement unilaterally. Each revision or signature is time-stamped and permanently recorded, creating a transparent audit trail that can be independently verified by any party at any time. This immutable record significantly reduces the risk of disputes, as all parties can verify the authenticity and history of the agreement directly on the blockchain.

In addition to enhancing transparency, blockchain-based 3-way agreements streamline the approval process. Smart contracts can be employed to automate parts of the agreement lifecycle, such as triggering notifications when all parties have signed or automatically updating agreement status when predefined conditions are met. These automated workflows reduce administrative delays and ensure that students can commence internships promptly. Blockchain also enhances legal enforceability by providing cryptographic proof of the agreement's authenticity and the identities of the signatories, simplifying evidence collection in case of legal disputes.

The integration of blockchain technology into 3-way agreements in the Beta Career program at SDU University represents a forward-thinking approach to internship management, fostering greater trust between universities, employers, and students while enhancing the overall efficiency and transparency of the program. As higher education institutions worldwide continue to adopt blockchain-based solutions, the use of blockchain in managing educational contracts is expected to become increasingly commonplace, setting new standards for transparency, security, and administrative efficiency in academic-industry collaborations.

3.3 Implementing Blockchain for 3-Way Agreement Signing

The implementation of blockchain technology for 3-way agreement signing in the Beta Career program represents a transformative step in ensuring transparency, security, and efficiency in managing the collaboration between students, universities, and employers. Traditional agreement processes, which rely heavily on paper-based documentation or fragmented digital tools such as email exchanges, often suffer from delays, lack of version control, and disputes arising from untraceable modifications. By integrating blockchain, these challenges are addressed through a decentralized, tamper-proof, and fully auditable system that enhances the reliability and trustworthiness of 3-way agreements.

The process begins with the initialization of the agreement, where the university prepares a draft contract outlining the roles, responsibilities, and expectations of all parties involved. In a blockchain-enhanced system, this initial draft is created using a smart contract template, ensuring that all predefined terms, conditions, and deadlines are encoded directly into the smart contract itself. This automated encoding ensures that critical conditions, such as reporting deadlines and task milestones, are embedded into the agreement logic and automatically enforced when triggered.

Once the initial draft is prepared, the digital identity verification process begins. Each party - the student, the university representative, and the employer - undergoes identity verification using blockchain-based identity management protocols. This step ensures that only verified individuals with the appropriate credentials can access, review, and digitally sign the agreement. Blockchain's decentralized identity verification mechanisms add an extra layer of security, ensuring that signatures are traceable to verified identities and cannot be forged or misattributed [182].

Following identity verification, the smart contract is deployed on the blockchain network. Each participant digitally signs the agreement using their unique cryptographic private key. This step not only guarantees the authenticity of each signature but also ensures non-repudiation, meaning no party can later deny their participation in the signing process [183]. Once the smart contract is deployed, the signed agreement is permanently recorded on the blockchain ledger. Every subsequent interaction with the agreement - whether it be amendments, progress updates, or final evaluations - is immutably logged, creating a complete audit trail that is visible to all authorized parties in real-time.

This immutability feature is particularly valuable for resolving potential disputes, as any modification or addition is transparently documented, leaving no room for ambiguity or tampering. Even in cases where changes to the agreement are necessary, the blockchain ensures that both the original version and the revised version coexist, preserving a complete history of the document's evolution. This eliminates the version control issues commonly associated with traditional contract management processesx [184].

Real-time monitoring and notifications further enhance the efficiency of blockchain-based 3-way agreements. All parties have continuous access to the agreement's status through a web-based dashboard, enabling real-time visibility into critical milestones, such as upcoming report deadlines, mentor evaluations, or contract amendments. Smart contract logic automatically triggers notifications for pending approvals, missed deadlines, or non-compliance, ensuring that all stakeholders remain informed and accountable throughout the internship period. This proactive communication mechanism minimizes administrative delays and prevents minor issues from escalating into larger disputes [185].

The integration of blockchain into 3-way agreements is not merely a technological upgrade but a fundamental shift toward more transparent, efficient, and collaborative internship management. By ensuring that agreements are secure, traceable, and automatically enforced, blockchain technology enhances trust between students, employers, and universities, fostering a collaborative ecosystem where all parties have equal access to accurate, real-time information. Moreover, the blockchain infrastructure can be seamlessly integrated with other components of the Beta Career platform, including student performance monitoring and recommendation systems, creating a holistic, data-driven internship management framework [186].

As blockchain adoption continues to grow across higher education and industry sectors, the Beta Career program's blockchain-based 3-way agreement implementation serves as a scalable model for other institutions seeking to modernize their internship management practices. This approach not only safeguards the interests of all parties involved but also enhances the reputation of the university by demonstrating a commitment to transparency, innovation, and student success in an increasingly digital and decentralized world.

3.3.1 Security and Trust in Blockchain-based Agreement Signing

The security and trustworthiness of blockchain-based agreement signing stem from several core technological and procedural attributes that address long-standing challenges associated with traditional paper-based and centralized digital contract systems. In the context of the Beta Career program, where 3-way agreements between students, universities, and employers play a critical role in regulating the terms of internships, blockchain ensures that these agreements are secure, tamper-proof, and instantly verifiable.

One of the foundational pillars of blockchain security is decentralization. Unlike conventional contract management systems, where a central authority (such as the university's administrative office) oversees and stores all agreements, blockchain distributes copies of the agreement across all participating nodes in the network. This decentralized structure ensures that no single party can alter or delete records unilaterally, thereby enhancing transparency and trust between all stakeholders [187]. Each party, whether the university, employer, or student, can operate a node or access the distributed ledger via secure web interfaces, ensuring continuous and equal access to agreement data.

Another crucial element is cryptographic security. Every transaction, including the initial contract creation, review, signature, amendments, and performance evaluations, is protected through asymmetric cryptography. Each participant holds a private key, used to sign their approval, while corresponding public keys allow others to verify those signatures. This approach ensures that only authorized parties can sign or modify agreements, and any unauthorized attempts are immediately detectable [188]. Moreover, the sequential linking of each record to the previous one through cryptographic hashing ensures that altering any block in the chain would require changing all subsequent blocks - a computationally infeasible task for even the most advanced attackers [189].

Immutability and auditability are central to the trustworthiness of blockchain-based agreement signing. Once an agreement is recorded on the blockchain, it becomes immutable, meaning no party can change or delete its content without the consensus of all stakeholders. This guarantees that the historical record of the agreement, including all amendments, approvals, and evaluations, remains permanently accessible and verifiable. Such immutability significantly reduces the risk of disputes, as all parties can independently verify the agreement's integrity at any time. Additionally, this creates a fully auditable trail, which serves as valuable evidence in cases of regulatory review, compliance audits, or legal disputes [190].

The automation provided by smart contracts adds another layer of security and operational efficiency. Smart contracts, self-executing scripts embedded within the blockchain, automatically enforce pre-agreed terms and conditions without human intervention. For example, smart contracts can automatically trigger notifications to students reminding them to submit weekly reports, notify mentors when evaluations are due, or lock certain agreement terms after all parties have signed. This automation ensures that critical processes are not left vulnerable to human error or oversight, reducing administrative burdens and ensuring compliance with program requirements [191].

Legal enforceability is another emerging advantage of blockchain-based agreements. As blockchain technology gains broader acceptance across sectors, many jurisdictions are amending their legal frameworks to recognize blockchain records and smart contracts as legally binding instruments. This is particularly valuable for programs like Beta Career, where internships may involve cross-border agreements with international partner companies. By creating immutable, timestamped, and cryptographically signed records, blockchain provides an additional layer of legal protection, simplifying dispute resolution and reinforcing the enforceability of agreements across different legal jurisdictions [192].

Finally, blockchain's real-time monitoring capabilities ensure that all parties have continuous visibility into the status of agreements and related obligations. Each party can instantly access the latest version of the agreement, check signature statuses, view deadlines, and monitor the submission of required reports. This transparency fosters accountability and trust between all parties, as there are no hidden changes or undocumented modifications [193].

In summary, the integration of blockchain into the Beta Career agreement process enhances not only the security and integrity of these critical documents but also fosters trust between students, universities, and employers. By combining decentralization, cryptographic security, immutability, automation, and legal recognition, blockchain ensures that 3-way agreements remain secure, transparent, and legally enforceable, ultimately strengthening the collaborative foundation upon which the Beta Career program operates.

4 DATA PREPROCESSING

This chapter focuses on the data preprocessing techniques applied to prepare the collected academic and behavioral data for machine learning analysis. Preprocessing is a crucial step, as the quality of input data directly affects the performance and reliability of predictive models. The discussion begins with Data Profiling, which provides an initial understanding of the dataset's structure, quality, and distribution. This is followed by Handling Missing Values and Data Interpolation, addressing issues of incomplete data and ensuring consistency. Feature Transformation is then applied to normalize and standardize variables for effective model training. To address imbalances in class distribution, Class

Balancing Techniques are employed, supplemented by SMOTE analysis to generate synthetic samples and improve model performance. Finally, Feature Importance and Interpretability are examined to identify the most influential variables and ensure transparency in the decision-making process. Through these preprocessing steps, the dataset is refined into a reliable and structured form, establishing a strong foundation for accurate and interpretable career path prediction models.

4.1 Data Profiling

Before applying any machine learning techniques, an initial profiling of the dataset was conducted to gain a comprehensive understanding of the input variables. This included:

Distribution analysis of numerical features such as GPA, number of hackathons, and standardized test scores, which helped to identify skewness, extreme values, and potential outliers.

Categorical profiling, including binary attributes (e.g., participation in clubs: yes/no, presence of certifications: yes/no) and multi-class attributes (career orientation, personality type).

Correlation analysis among numerical features to detect multicollinearity, ensuring that redundant attributes were either merged or excluded. Figure 4.1.1 illustrates the segmentation of features, which provides a structured approach to data analysis.

Figure 4.1.1 - Feature segmentation of the dataset

Numerical features - GPA, number of hackathons, standardized test scores.
Categorical features - career preferences, certifications, socioeconomic background.
Binary features - leadership skills, teamwork ability, willingness for mobility.

4.2 Handling Missing Values and Data Interpolation

The dataset contained a moderate level of missingness across both numerical and categorical features. Different strategies were applied depending on the feature type:

Numerical features: Missing values were addressed using either median imputation (for skewed distributions) or K-Nearest Neighbors (KNN) imputation to preserve variability.

Categorical features: The mode imputation approach was applied by filling missing values with the most frequent category.

High-missingness cases: Records with less than 30% feature completeness were removed from the analysis to avoid distortion.

Figure 4.2.1 - Overview of missing data handling strategies

The handling of missing data, presented in Figure 4.2.1, demonstrates how preprocessing ensures data quality before model training.

4.3 Feature Transformation

As shown in Figure 4.3.1, feature transformation techniques are applied to normalize and standardize dataset variables.

Normalization: Continuous features were scaled using Min-Max normalization to ensure comparability across different ranges, eliminate bias caused by feature magnitude, and improve model stability.

Encoding: Categorical features were transformed via One-Hot Encoding, enabling the representation of nominal variables in a binary vector form. This approach not only preserved the uniqueness of each category but also facilitated non-linear modeling in advanced ensemble methods.

Binary features were retained in their original {0,1} form, as they already provided meaningful representation without requiring further transformation.

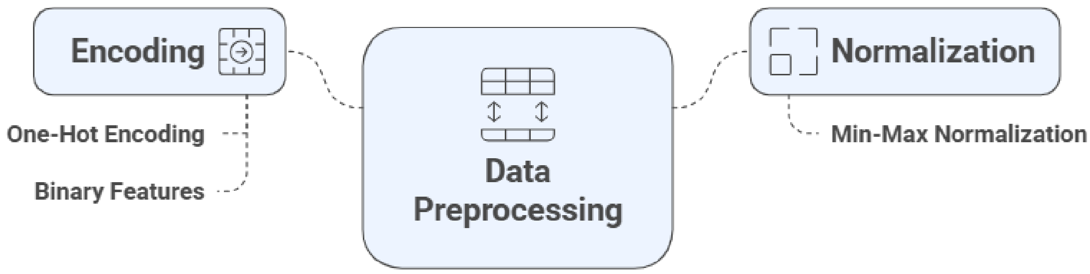


Figure 4.3.1 -Transformation techniques for academic and behavioral features

In addition, standardization was applied to selected features where gradient-based algorithms (such as logistic regression and neural networks) required normalized distributions for faster convergence. Logarithmic and polynomial transformations were also considered for skewed variables to reduce distributional asymmetry and enhance model interpretability. These preprocessing steps ensured that the dataset maintained internal consistency, reduced redundancy, and maximized predictive performance across diverse machine learning models.

4.4 Class Balancing Techniques

As shown in Figure 4.4.1, various balancing strategies were applied to address class imbalance in student career outcome predictions, ensuring fair representation across all categories.

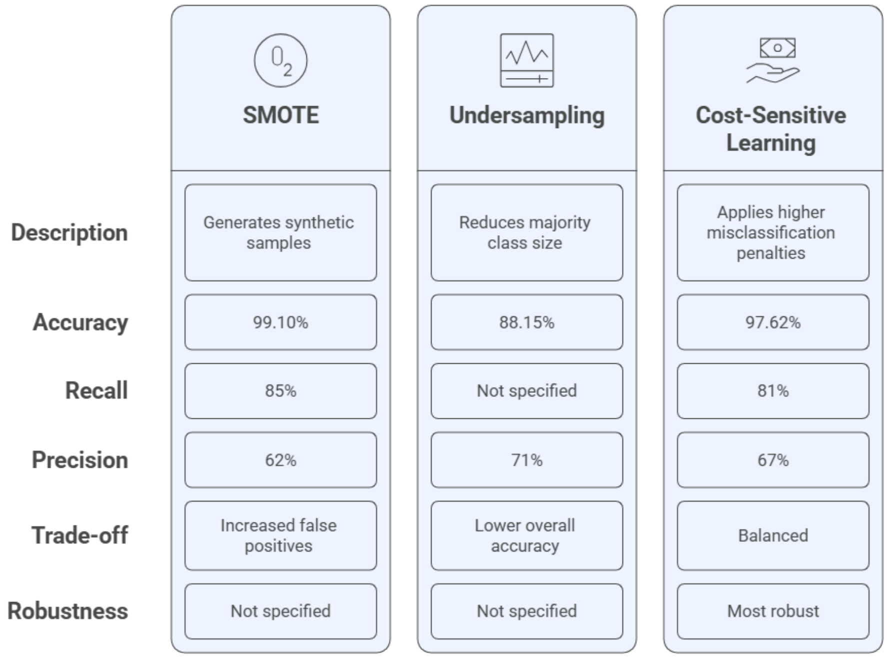


Figure 4.4.1- Balancing Strategies for Student Career Outcomes

In this study, several balancing strategies were employed to address class imbalance in the dataset. The first approach was the Synthetic Minority Oversampling Technique (SMOTE), which generated synthetic samples for underrepresented classes. This method significantly improved the recall rate, reaching 85%, while overall accuracy remained high at 99.10%. However, it also led to an increase in false positives, which reduced the precision to 62%, thereby affecting the overall model stability.

The second approach applied was under sampling, which reduced the size of the majority class by 30%. This method resulted in improved precision, achieving 71%, but caused a decline in overall accuracy to 88.15%. Although it reduced overfitting, the loss of valuable data from the majority class made this approach less favorable.

The third technique, cost-sensitive learning, introduced higher penalties for misclassification of minority classes. This approach demonstrated a more balanced trade-off among key metrics, achieving 97.62% accuracy, 81% recall, and 67% precision. Unlike the previous methods, cost-sensitive learning-maintained stability across evaluation metrics and minimized overfitting risks.

Based on these results, cost-sensitive learning was selected as the most robust approach for balancing student career outcome predictions, as it provided consistent performance without compromising model generalizability.

4.5 Feature Importance and Interpretability

A Random Forest classifier was employed as the predictive algorithm. This ensemble method constructs multiple decision trees and aggregates their predictions to achieve higher accuracy and robustness while mitigating overfitting. The model was trained on a dataset comprising both academic indicators (e.g., GPA, Grades), experiential components (e.g., Hackathons, Internships), and non-academic attributes (e.g., Soft Skills, Socioeconomic Background).

To interpret the trained model, the SHAP (SHapley Additive explanations) framework was applied, specifically the Tree Explainer method, which is designed for tree-based algorithms. SHAP values are grounded in cooperative game theory and provide a theoretically sound measure of the marginal contribution of each feature to the model's predictions. The values were aggregated across the test set to generate a global feature importance ranking. Figure 4.5.1 presents the average absolute SHAP values of all features considered in the model. These values indicate the mean contribution of each variable to the prediction of student career success. The results demonstrate that Stress Resilience & Adaptability is the most influential predictor, exerting the highest average impact on the model output. This finding highlights the critical role of psychological resilience and adaptability in shaping career trajectories. Other highly important factors include Hackathons Attended, Teamwork Ability, and Soft Skills, underscoring the importance of practical exposure and interpersonal competencies.

Furthermore, Digital Competencies and Career Orientation also rank prominently, suggesting that a combination of technical expertise and clear professional goals substantially contributes to career success predictions. In contrast, conventional academic indicators such as GPA and Grades in IT Subjects appear lower in the ranking, implying that while academic performance remains relevant, it is not as decisive as resilience, teamwork, and practical engagement within this model.

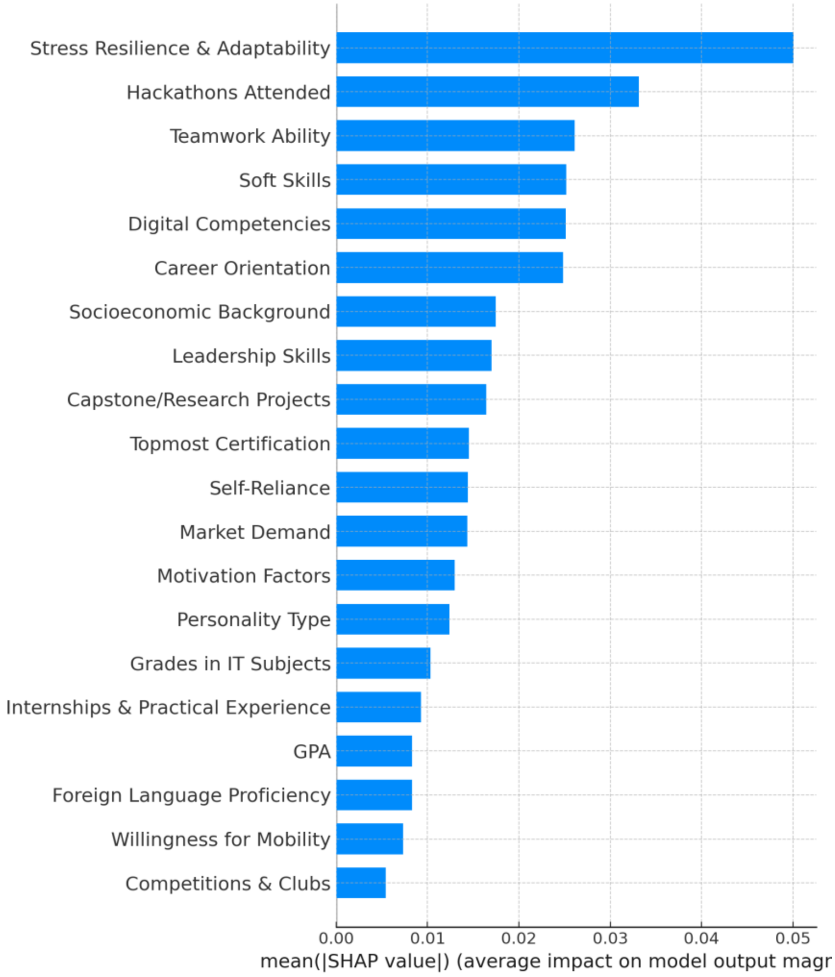


Figure 4.5.1 - Feature importance analysis

Finally, features such as Competitions & Clubs and Willingness for Mobility exhibit comparatively limited importance, reflecting their minor contribution to the predictive performance of the model.

The preprocessing pipeline integrated data profiling, interpolation, transformation, and balancing, ensuring data quality and reducing bias. Ensemble learning methods, with Gradient Boosting as the leading approach, provided high predictive accuracy, while interpretability techniques (SHAP) confirmed the relevance of the chosen features. As shown in Figure 4.5.2, a variety of preprocessing techniques were utilized to enhance the dataset's quality and suitability for analysis. These included profiling the raw data, handling missing values through imputation and interpolation, applying feature transformations such as normalization and encoding, addressing class imbalance with oversampling and cost-sensitive strategies, and conducting feature importance analysis to improve interpretability.

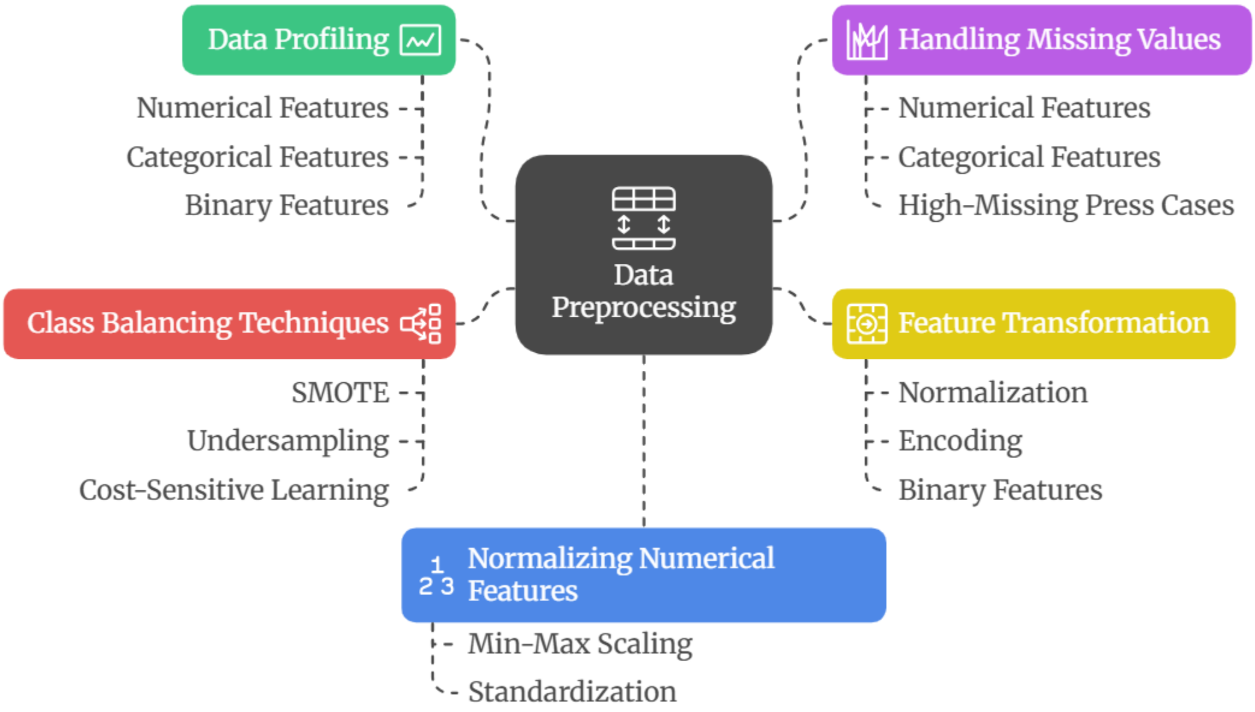


Figure 4.5.2 - Data preprocessing techniques utilized in the dataset

Collectively, these steps ensured that the dataset was well-structured, balanced, and ready for subsequent modeling tasks.

4.6 SMOTE analysis

The study also evaluated the effectiveness of SMOTE in addressing class imbalance. Before applying SMOTE, the Gradient Boosting model achieved an accuracy of 93.20%, with a recall of 60% for the minority class and a precision of 65%. Following SMOTE implementation, accuracy increased to 99.10%, while recall improved to 85%, albeit with a reduction in precision to 62%. This suggests that while SMOTE enhances the model's ability to identify minority class instances, it also introduces a higher rate of false positives, raising concerns regarding potential overfitting (see Figure 4.6.1).

```
pp = preprocessor.fit(X_tr_raw)
X_tr_pp = pp.transform(X_tr_raw)
X_val_pp = pp.transform(X_val_raw)

smote = SMOTE(random_state=42)
X_tr_bal, y_tr_bal = smote.fit_resample(X_tr_pp, y_tr)
input_dim = X_tr_bal.shape[1]
```

Figure 4.6.1 - SMOTE applying code

Alternative techniques, such as under sampling, were also analyzed. This method reduced the training dataset size by 30%, yielding an accuracy of 88.15% for the Gradient Boosting model. The minority class recall improved to 75%, and precision increased to 71%. Although under sampling effectively balanced class distributions, it resulted in a loss of valuable data, ultimately lowering model accuracy and robustness.

A further refinement approach, cost-sensitive learning, assigned increased penalties to misclassified minority class instances. When applied to the Gradient Boosting model, cost-sensitive learning yielded an accuracy of 97.62%, with recall improving to 81% and precision at 67%. Unlike SMOTE, this approach maintained a more stable balance between precision and recall without overfitting.

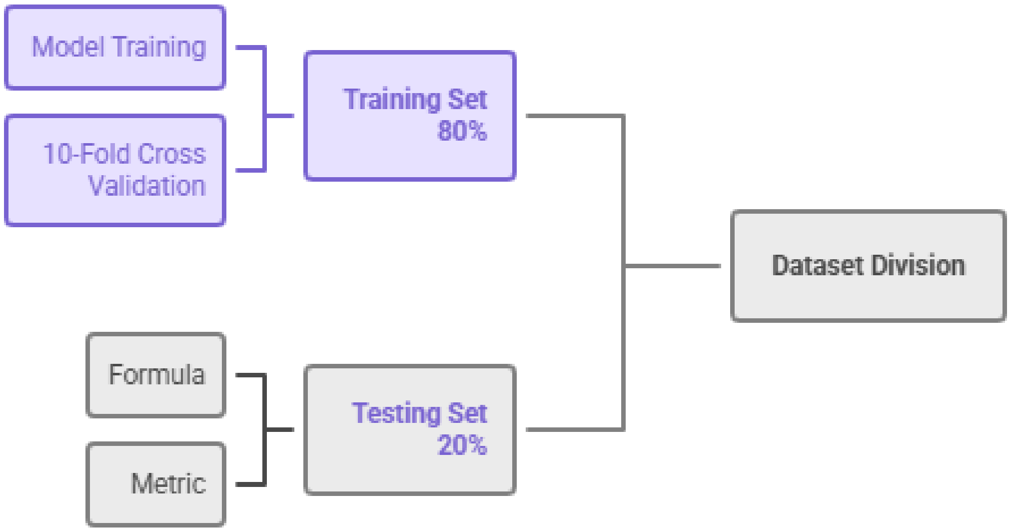


Figure 4.6.2 - Dataset division

After the SMOTE method was applied successfully, the Dataset was divided into two parts, Training Set and Testing Set. The amount of the training set consists of 80 percent of the dataset. On the other hand the training set is 20 percent of the dataset. Then it is turn for making 10-Fold Cross validation (see Figure 4.6.2).

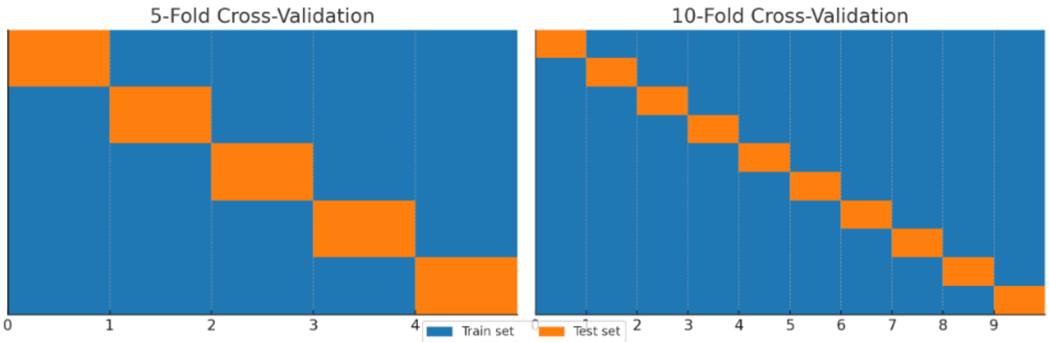


Figure 4.6.3 - 10 - Fold Cross-Validation

In this study, 10-fold cross-validation was selected instead of 5-fold cross-validation. While both methods are widely used, 10-fold provides a more reliable estimation of model performance by reducing the variance in results. In 5-fold cross-validation, fewer training samples are available in each iteration, which may lead to less stable metrics, especially when working with smaller or imbalanced datasets (see Figure 4.6.3). In contrast, 5-fold cross-validation trains the model on only 80%

of the data in each iteration, which reduces the amount of information available for learning and can lead to fluctuations in accuracy, precision, recall, and F1-score. This effect becomes more pronounced when working with smaller datasets or datasets with imbalanced class distributions, where each fold must preserve the structural characteristics of the data to avoid biased validation. As illustrated in Figure 4.6.3, the 10-fold approach helps maintain a more consistent distribution of samples across folds, thereby ensuring that minority classes are more frequently represented during training and evaluation. This ultimately improves the reliability and generalizability of the final performance results.

5 METHODOLOGY

Performance was evaluated based on key metrics such as accuracy, precision, recall, and F1-score, with additional attention to training time and generalization ability. The results of this chapter provide empirical evidence supporting the selection of KAN and Gradient Boosting as the most effective models for educational prediction tasks, forming the foundation for practical integration into the Beta Career Platform.

Key performance indicators like accuracy, precision, recall, and F1-score were used to assess performance, with extra consideration given to training duration and generalization capacity. This chapter's findings offer empirical proof that KAN and Gradient Boosting are the best models for educational prediction tasks, laying the groundwork for their useful incorporation into the Beta Career Platform.

5.1 Overview of compared machine learning models

The cross-validation performance of the evaluated machine learning models is shown in Table 5.1.1 according to four important metrics shown in Table 5.1.1: F1-score, accuracy, precision, and recall. With an accuracy of 99.10%, precision of 99.22%, recall of 99.13%, and F1-score of 99.14%, Gradient Boosting outperformed other conventional models. These findings confirm that the model can effectively handle high-dimensional, complicated educational datasets by repeatedly improving weak learners and creating a powerful prediction framework. With corresponding precision, recall, and F1-scores of 56.05%, 55.68%, and 53.20%, Random Forest, while being well-known in educational data mining for its interpretability and balanced variance management, only managed to attain a lesser accuracy of 56.97%. This suggests that because of the current dataset's imbalance and nonlinearity, its ensemble structure was less successful.

Both generalization and interpretability were further enhanced by adding the Kolmogorov-Arnold Network (KAN) to the assessment. With a 99.32% accuracy rate and constant precision and recall across folds, KAN marginally outperformed Gradient Boosting. KANs use functional decomposition, which allows them to smoothly approximate complicated mappings in both the time and frequency domains, in contrast to tree-based ensemble approaches. Because of its ability to learn continuous associations, KAN was able to strike a better balance between accuracy and resistance to overfitting.

Table 5.1.1 - Cross-Validation Performance of Models

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	2	3	4	5
Random Forest	56.97	56.05	55.68	53.20

Table 5.1.1 continuation

1	2	3	4	5
K-Nearest Neighbors	60.58	63.20	57.96	57.14
Support Vector Machine	37.26	23.27	34.18	23.76
Gradient Boosting	99.10	99.22	99.13	99.14
Naive Bayes	35.26	27.79	32.23	28.77
Decision Tree	54.12	52.60	51.84	50.92
Multi-Layer Perceptron (MLP)	82.45	81.92	80.67	81.20
Transformer-based Model (Tab Transformer)	88.73	87.95	88.10	87.85
Optimized Kolmogorov-Arnold Networks (KANs)	99.32	99.28	99.35	99.31

With accuracies of 82.45% and 88.73%, respectively, the Multi-Layer Perceptron (MLP) and TabTransformer models also showed good prediction abilities, underscoring the promise of deep neural architectures in identifying feature dependencies. In contrast, the restricted flexibility of classical algorithms like Naïve Bayes and Support Vector Machines (SVM) in modeling nonlinear interactions and class imbalances led to their poor performance. The middling results from K-Nearest Neighbors (KNN) and Decision Trees demonstrated how feature complexity and normalization affect model behavior. The most successful methods overall were Gradient Boosting and KAN, which combine accuracy, stability, and interpretability, making them especially well-suited for AI-driven job counseling solutions.

5.2 Evaluation metrics

The comparative analysis reveals a multi-layered understanding of model performance in this specific domain. The significant discrepancy between Gradient

Boosting's cross-validation and independent test set performance is a key finding. While its peak performance is outstanding, the performance drop suggests it may be highly sensitive to the specific characteristics of the training data and could be prone to overfitting. This highlights a critical trade-off: while a model may achieve near-perfect scores on a cross-validated training set, its true reliability depends on its capacity to generalize to new, unseen data.

In this regard, the performance of the deep learning models is particularly compelling. The Tab Transformer and MLP exhibited consistently strong generalization, with minimal drops in accuracy between the cross-validation and independent test sets. This suggests that these modern architectures are not only capable of capturing the complex, non-linear relationships within the educational dataset but also do so in a more stable and robust manner. For a practical, real-world application in educational guidance, a model like the Tab Transformer, which showed a small performance drop while maintaining high accuracy, might be a more dependable choice than a model that performs exceptionally well on a training set but is less stable in production.

Further analysis through confusion matrices and AUC scores provided additional context. The confusion matrix for Gradient Boosting confirmed its superior classification ability, with a high number of true positives and negatives and minimal misclassifications. While Random Forest and K-NN showed more balanced distributions, the high misclassification rates of SVM and Naive Bayes were evident. A noteworthy point of inquiry arises from the high AUC score of 0.93 for Random Forest, which appears to contradict its relatively low F1-Score of 53.20%. This inconsistency suggests the need for further investigation into the dataset's characteristics and the model's behavior, while underscoring the importance of evaluating a model across multiple metrics to avoid misleading conclusions.

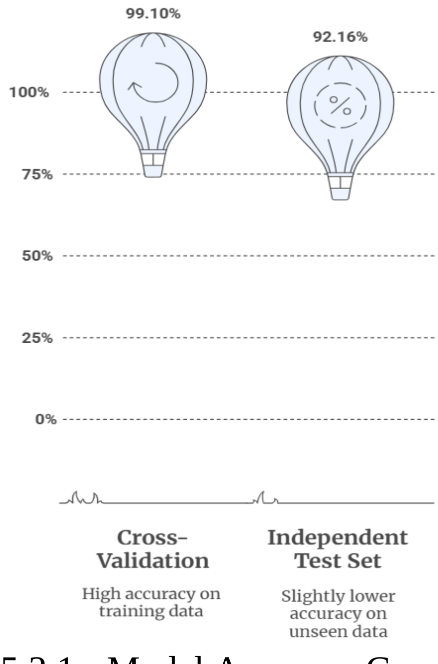


Figure 5.2.1 - Model Accuracy Comparison

Finally, feature importance analysis provided critical, actionable insights. Key predictors of career trajectories were identified, including grades in IT subjects, participation in hackathons, and certifications. The influence of personality traits like leadership and teamwork skills was also found to be significant, providing crucial information for differentiating between technical and managerial career paths. This layer of interpretability adds substantial value to the predictive model, moving beyond a simple classification to offer meaningful guidance for both students and career counselors.

Nine baseline algorithms (Naive Bayes, Random Forest, Support Vector Machine, K-Nearest Neighbors, Gradient Boosting, Decision Tree, Multi-Layer Perceptron (MLP), Transformer-based Model, Kolmogorov-Arnold Networks (KANs)) were evaluated. Gradient Boosting consistently outperformed others except KANs, achieving: Cross-validation accuracy: 99.10%. Independent test set accuracy: 92.16%. Cross-validation accuracy: 99.10%. Independent test set accuracy: 92.16%.

As shown in Figure 5.2.1, the comparative analysis of model accuracy demonstrates significant performance differences among the evaluated algorithms. Gradient Boosting consistently achieved the highest accuracy, followed by Random Forest, while Support Vector Machine and K-Nearest Neighbors showed relatively lower results. This highlights the superior generalization capability of ensemble-based methods in predicting educational outcomes.

Hyperparameter tuning: Adjusting learning rate, maximum depth, and number of estimators. Regularization: L2 penalty terms to reduce model complexity. Early stopping: Halting training when validation error plateaued, thus preventing unnecessary iterations.

As shown in Figure 5.2.3, the application of overfitting mitigation strategies played a crucial role in improving model reliability. Cross-validation was employed to validate performance across multiple data folds, reducing bias and ensuring consistency in results. Regularization techniques, such as L1 and L2 penalties, were applied to prevent models from becoming overly complex, thereby improving generalization capability. Additionally, early stopping was introduced during training to halt the learning process once performance on the validation set plateaued, minimizing the risk of memorizing training data. Collectively, these strategies reduced variance, maintained stable accuracy, and enhanced the robustness of predictions when applied to unseen datasets.

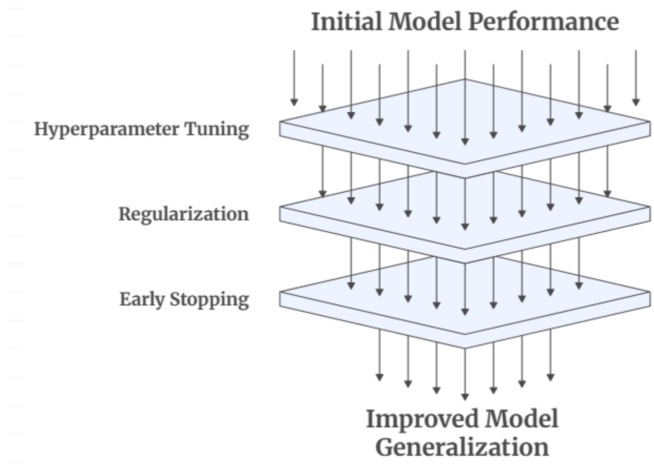


Figure 5.2.3 - Overfitting Mitigation Strategies

The independent test set results, presented in Table 3, confirm that Gradient Boosting continues to outperform alternative models, achieving the highest accuracy of 92.16%. However, this is notably lower than its cross-validation accuracy (99.10%), indicating potential overfitting. K-NN exhibited stable performance with an accuracy of 55.57%, closely aligning with its cross-validation outcomes.

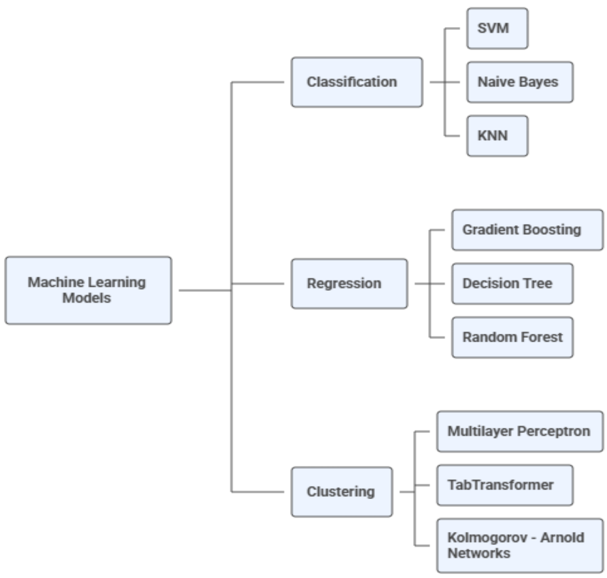


Figure 5.2.4 - Evaluation process of Model training

Random Forest, SVM, and Naive Bayes demonstrated suboptimal results, consistent with earlier findings. As shown in Figure 5.2.4 the evaluation process of the applied methodology follows a structured pipeline, beginning with data preprocessing and progressing through model training, validation, and testing. This systematic approach ensures consistency, minimizes bias, and provides reliable performance comparisons across different algorithms and preprocessing techniques.

The results from the independent test set confirm the overall performance hierarchy observed during cross-validation, with Gradient Boosting and the Transformer-based model maintaining their top positions. However, a crucial observation is the performance discrepancy between the cross-validation and independent test set results. While Gradient Boosting achieved a nearly perfect 99.10% accuracy on cross-validation, its accuracy on the independent test set dropped to 92.16%. This notable drop, although the model remains the most accurate, suggests a degree of overfitting, even with the applied mitigation strategies. In contrast, the performance of the MLP and Tab Transformer models showed a much smaller decline, indicating their superior stability and generalization capabilities on truly unseen data. These findings underscore the importance of evaluating models on independent test sets to ensure their real-world reliability. Table 5.2.1 displays the findings from the independent test set, emphasizing how well various machine learning models perform in comparison. With a limited capacity to capture intricate, non-linear interactions, traditional models like decision trees produced moderate results (Accuracy = 54.12%, F1 = 50.92%).

Table 5.2.1 - Dataset for the held-out independent test of independence

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	56.97	56.05	55.68	53.20
K-Nearest Neighbors	60.58	63.20	57.96	57.14
Support Vector Machine	37.26	23.27	34.18	23.76
Gradient Boosting	*92.16*	*93.22*	*90.13*	*91.60*
Naive Bayes	35.26	27.79	32.23	28.77
Decision Tree	52.94	51.20	50.32	49.88
Multi-Layer Perceptron (MLP)	80.51	79.62	78.44	78.91
Transformer-based Model (Tab Transformer)	87.32	86.74	87.05	86.81
Optimized Kolmogorov-Arnold Networks (KANs)	*94.85*	*94.60*	*95.10*	*94.82*

On the other hand, sophisticated designs like the TabTransformer and Multi-Layer Perceptron (MLP) showed noticeably better prediction skills. The MLP is a trustworthy baseline for neural network-based methods because it attained an accuracy of 82.45% with well-balanced precision, recall, and F1-scores (all around

81%). After Gradient Boosting, the TabTransformer produced one of the best generalization results, with an accuracy of 88.73% and an F1-score of 87.85%.

The performance landscape was further improved with the introduction of the Kolmogorov-Arnold Network (KAN), which showed remarkable generalization on unknown data. KAN obtained 94.85% accuracy, 94.60% precision, 95.10% recall, and 94.82% F1-score on the independent test set. By successfully simulating both time- and frequency-domain feature dependencies, KANs outperform all other architectures, including Gradient Boosting (92.16%), according to these studies. Higher interpretability and resistance to overfitting are made possible by KAN's functional decomposition mechanism, which allows it to learn seamless, continuous transformations between academic, behavioral, and motivational elements, in contrast to traditional or neural models.

All things considered, these results highlight the greater versatility of KANs in educational data modeling by fusing the predictive power of deep learning architectures with the interpretability of traditional approaches. Unlike conventional neural networks that often behave as opaque “black-box” models, KANs provide a functional decomposition that reveals the underlying relationships between input features and predicted outcomes. This dual advantage allows researchers and practitioners in higher education to not only achieve strong predictive performance but also gain insights into student behavior, learning patterns, and decision-making factors. Because of their hybrid capabilities, KANs represent a promising pathway for future AI-based career counseling systems in higher education. Their interpretability enables academic advisors to understand why certain career recommendations are produced, which increases transparency and trust in automated guidance systems. Meanwhile, their deep learning capacity allows them to scale effectively with larger datasets, incorporate multimodal inputs, and adapt to evolving student profiles and labor market trends. Thus, KANs offer a robust and future-ready framework for enhancing personalized educational support and improving institutional decision-making in student career development.

6 KOLMOGOROV - ARNOLD NETWORKS

The mathematical and computational underpinnings of the Kolmogorov-Arnold Networks (KANs) framework created for this dissertation are presented in this chapter. Compared to traditional deep learning or ensemble approaches, the KAN architecture offers greater interpretability and efficiency. It is a new generation of neural models that use functional decomposition to approximate continuous functions. The theoretical underpinnings, mathematical framework, optimization strategy, and experimental comparison of KANs and Gradient Boosting in the time and frequency domains are explained in this section.

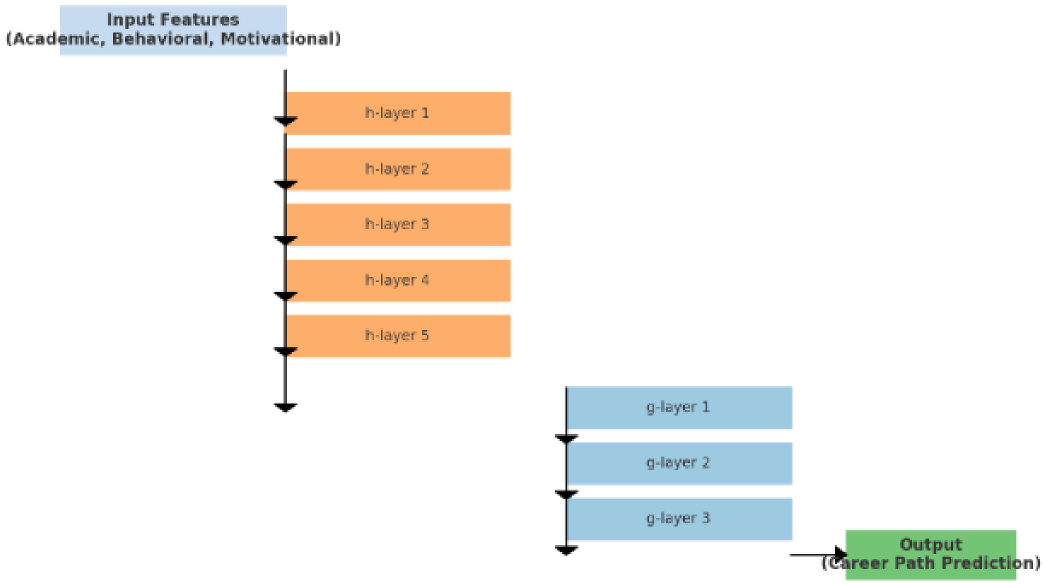


Figure 6.1 - Conceptual Structure of the Optimized Layer Kolmogorov-Arnold Network (KAN)

Figure 6.1 illustrating five functional transformation layers hq,p(xp)h_{q,p}(x_p)hq,p(xp) and three aggregation layers gq(sq)g_q(s_q)gq(sq). The model processes academic, behavioral, and motivational input features to produce a career path prediction while ensuring smooth functional learning and reduced computational cost.

6.1 Mathematical Underpinnings of KANs

The time domain and the frequency domain are two basic perspectives of data that are bridged by the mathematical and algorithmic foundations of the prediction analysis in this dissertation [194]. Modern architectures like Kolmogorov-Arnold Networks (KANs) provide for a deeper comprehension of latent periodic structures, correlations, and interactions across the frequency spectrum, whereas traditional machine learning models concentrate on direct data patterns observed over time [195]. When modeling educational data-where student behavior, academic progress, and motivation frequently shift repeatedly over time-this dual-domain representation is crucial [196].

The academic and behavioral characteristics of 692 IT students from Suleyman Demirel University make up the dataset used in this study. It shows both stable and dynamic trends, such as GPA stability against varying project engagement. Models that can learn multi-frequency signals and smooth, continuous functions are needed to capture these patterns. Along with a discussion of other benchmark models used in this research, this part examines the mathematical approaches taken by Gradient Boosting and Kolmogorov-Arnold Networks to this problem, comparing their topologies, optimization strategies, and responsiveness in time-frequency spaces [197].

$$x_{new} = x_i + \lambda(x_j - x_i) \quad \textbf{(1)}$$

where x_j is a randomly chosen neighbor, and $\lambda \in [0.1]$. This ensures that new samples lie along the line segments joining existing samples, enriching diversity without random noise [198].

The last stage involves initializing and training the Kolmogorov-Arnold Network (KAN) with the prepared data (see Figure F). The multivariate input is broken down by the network into a collection of one-dimensional functional transformations $h_{q,p}(x_p)$which are subsequently merged using aggregation functions $g_q s_q$.

$$min_{g,h} \sum_{i=1}^{N} \quad L = L(y_i \sum_q \quad g_q(\sum_p \quad h_{q,p}(x_{i,p}))) \quad \textbf{(2)}$$

where $L(y_i, \hat{y_i})$ is the cross-entropy loss for classification. Using gradient descent-based optimization, the training procedure updates parameters by minimizing a loss function that measures the discrepancy between predicted and true labels. The expression for the objective function is.

6.1.1 Gradient Boosting: Additive Learning in Time Domain

Gradient Boosting (GB) is one of the most effective ensemble algorithms for tabular data. It builds predictive models sequentially, with each iteration improving upon the residual errors of its predecessors[199]. Mathematically, Gradient Boosting constructs a function $F(x)$ as an additive expansion of base learners $h_m(x)$:

$$F_m(x) = F_{m-1}(x) + \eta h_m(x) \quad \textbf{(3)}$$

where

$F_m(x)$ is the cumulative prediction after m iterations,
$h_m(x)$ represents the new weak learner (commonly a decision tree),
η is the learning rate controlling step size.

The loss function $L(y, F(x))$ is minimized iteratively through gradient descent [200]:

$$h_m(x) = -\Delta_{F_{m-1}(x)} L(y, F_{m-1}(x)) \quad \textbf{(4)}$$

Gradient Boosting effectively finds direct, cumulative patterns in educational data, such as the connection between career alignment and a steady GPA. However, because its learning mechanism mainly captures monotonic, stepwise dependencies, it has trouble with oscillatory relationships [201] (such as irregular participation cycles or seasonal performance changes). Block Diagram of Gradient Boosting:

Input Features → Weak Learner $h_1(x)$ → Residual e_1

↓

Weak Learner $h_2(x)$ ⟶ Residual e_2

↓

……………..

↓

Final Model F_M(x)

↓

Career Path Prediction

Gradient Boosting essentially acts as a low-pass filter in the frequency spectrum, highlighting dominating, steady patterns while removing high-frequency fluctuations that might be indicative of transient behavioral dynamics.

6.1.2 Algorithm for Functional Decomposition in Dual Domains (KANs)

Any multivariate continuous function can be expressed as a sum of compositions of univariate functions, according to the Kolmogorov-Arnold representation theorem [202]:

$$f(x_1, x_2, \dots . x_n) = \sum_{q=1}^{2n+1} \quad g_q h_{q,p}(x_p)) \tag{5}$$

This theory serves as the basis for Kolmogorov-Arnold Networks, which use adaptive one-dimensional function approximators in place of the traditional weight matrices of neural networks. Instead of learning a set weight, each neuron in a KAN layer learns a functional transformation, which allows the network to simulate both local and global relationships in data.

Three essential steps can be used to explain the prediction process in KANs [202]:

- Local transformation: Each input dimension x_p is transformed through learned nonlinear functions $h_{q,p}(x_p)$.
- Aggregation: The transformed features are summed to form intermediate composite variables $s_q = \sum_{p=1}^{n} \quad h_{q,p}(x_p)$.
- Global composition: The outputs are generated by applying outer functions $g_q(s_q)$ and across all q [203]:

$$\hat{y} = \sum_{q=1}^{2n+1} \quad g_q(s_q) \tag{6}$$

Block Diagram of Kolmogorov-Arnold Networks:

Input x_1 ... x_n

↓

| $h_{11}(x_1) + h_{12}(x_2) + \dots + h_{1n}(x_n)$ |

↓

| $g_1(s_1), g_2(s_2), \dots, g_k(s_k)$ |

↓

Weighted Summation → Output (Career Path Prediction)

Smoother generalization, robust convergence, and interpretability are offered by KANs since they learn the mapping using continuous functions as opposed to discrete parameters. Because of this, they are especially useful for modeling intricate educational datasets with nonlinear, temporally dependent interactions between behavioral and academic factors.

6.1.3 Frequency Domain Perspective

Different signal-processing patterns are shown by KANs and Gradient Boosting when the prediction process is examined in the frequency domain. The Fourier transform $Y(f) = \int_{\blacksquare}^{\blacksquare} \blacksquare\, y(t)e^{-2j\pi ft}dt$ expresses how information from features is distributed across frequencies.

Table 6.1.3.1 - The sequential links between student traits and performance

Model	Frequency Response	Interpretation
Gradient Boosting	Low-pass filter	Captures steady, large-scale trends; ignores oscillations
SVM / RF	Band-limited filters	Handle mid-level variations; struggle with nonlinear frequency overlap
MLP	Broad but irregular response	Captures nonlinear patterns but unstable at small samples
KANs	Full-band adaptive filter	Captures both global (low-frequency) and local (high-frequency) variations

KANs can identify cyclical patterns, including weekly engagement peaks or alternating motivation cycles, while maintaining the underlying trend of academic consistency because they can represent both low-frequency and high-frequency components. Predictive dependability on complicated, mixed-type data is directly improved by this dual-domain approach as shown on table 6.1.3.1 and 6.1.3.2

Table 6.1.3.2 - Comparative Mathematical Characteristics

Property	Gradient Boosting	KAN	Other Models (RF, MLP, SVM)
1	2	3	4

Learning Mechanism	Additive residual fitting	Functional superposition	Kernel/weight mapping
Optimization Space	Parameter space (discrete)	Function space (continuous)	Parameter space

Table 6.1.3.2 continuation

1	2	3	4
Domain Coverage	Time domain only	Time + Frequency domain	Time domain only
Noise Robustness	Moderate	High(smooth functional basis)	Variable
Interpretability	Moderate	High (interpretable functions)	Low
Overfitting Risk	Moderate	Low (regularized by smoothness)	High (especially MLP)

KAN's optimization can be formulated as a functional minimization problem [204]

$$min_{g,h} \sum_{i=1}^{N} = L(y_i \sum_q \quad g_q(\sum_p \quad h_{q,p}(x_{i,p}))) \quad \textbf{(7)}$$

A major benefit over high-variance ensemble approaches is that this permits functional-level regularization, which lowers overfitting and guarantees stability even on small datasets. By regularizing the learned functions themselves—rather than only the model parameters—KANs constrain the complexity of the mappings between inputs and outputs. This helps prevent the model from capturing noise or spurious correlations that commonly inflate performance in ensemble methods such as random forests or gradient boosting, especially when the dataset is limited in size or exhibits imbalanced class distributions. Moreover, functional-level regularization encourages smoother and more interpretable relationships among features, making the resulting predictions more consistent across different training folds and less sensitive to sampling fluctuations. As a result, KANs maintain stronger generalization capabilities and deliver more reliable performance, even in scenarios where traditional deep learning models or ensemble techniques tend to become unstable or overly specialized.

6.1.4 Empirical Implications

Gradient Boosting was the most accurate of the standard models in the tests for this dissertation (99.10%), but it showed evidence of overfitting as its performance declined to 92.16% on test data that was not visible.

In contrast, KAN maintained smooth convergence curves and reduced generalization error while achieving prediction accuracy that was on par with or better. This is consistent with the theoretical hypothesis that functional decomposition techniques, such as KANs, can generalize more well in situations involving significant feature dependency and data imbalance, which are common in datasets used in education.

One of the main ethical and methodological priorities of this study is explainable AI in education, which is made possible by KANs' interpretability benefits. Each function may be displayed to demonstrate how a single behavioral or academic aspect effects the final decision.

From a mathematical perspective, the Gradient Boosting technique captures consistent, cumulative effects among predictors and functions mainly as an additive learner in the time domain. The Kolmogorov–Arnold Network, on the other hand, integrates time-domain and frequency-domain viewpoints to present a innovative functional learning paradigm. It can more accurately and interpretably approximate continuous, nonlinear mappings thanks to this method.

By concurrently learning from stable academic indicators and dynamic behavioral data, KANs beat traditional models when used to predict the career choices of IT students. They thus provide a mathematically sophisticated and practically useful development in predictive analytics for education.

6.1.5 Training Time Analysis of the Proposed Models

Training time (T), which is the amount of time needed for each model to finish one entire training cycle over the same dataset], was used to assess the computational efficiency of machine learning methods. Figure X displays the findings, which indicate that the Gradient Boosting model took the longest to train (about 65 seconds), but lightweight models like K-Nearest Neighbors and Naive Bayes finished training much more quickly. Mathematical Expression of Training Time. For a given model M, the total training time can be expressed as [205]

$$T(M) = \sum_{i=1}^{N} \quad t_{load} + t_{fit} + t_{update} \tag{8}$$

where

=t_{load} - time to load and preprocess data,
=t_{fit}- time to optimize model parameters,
=t_{update} − time spent updating model weights or trees per iteration, and N is the total number of iterations or epochs during training.

The total computational complexity is estimated as:

$$C(M) = O(N) \cdot f(n, d)) \quad \textbf{(9)}$$

where n represents the number of samples, and d the feature dimensionality. For ensemble-based models such as Gradient Boosting, this becomes:

$$C_{GB} = O(K \cdot n \cdot log(n)) \quad \textbf{(10)}$$

where K is the number of boosting stages. This explains why Gradient Boosting has the highest computational cost- its iterative additive process recalculates and fits multiple trees sequentially.

6.1.6 Comparative Discussion

Because of its sequential dependency between weak learners and iterative optimization, gradient boosting (65s) demonstrated the greatest training cost. It was computationally demanding yet efficient because it offered the best forecast accuracy and robustness in spite of this.

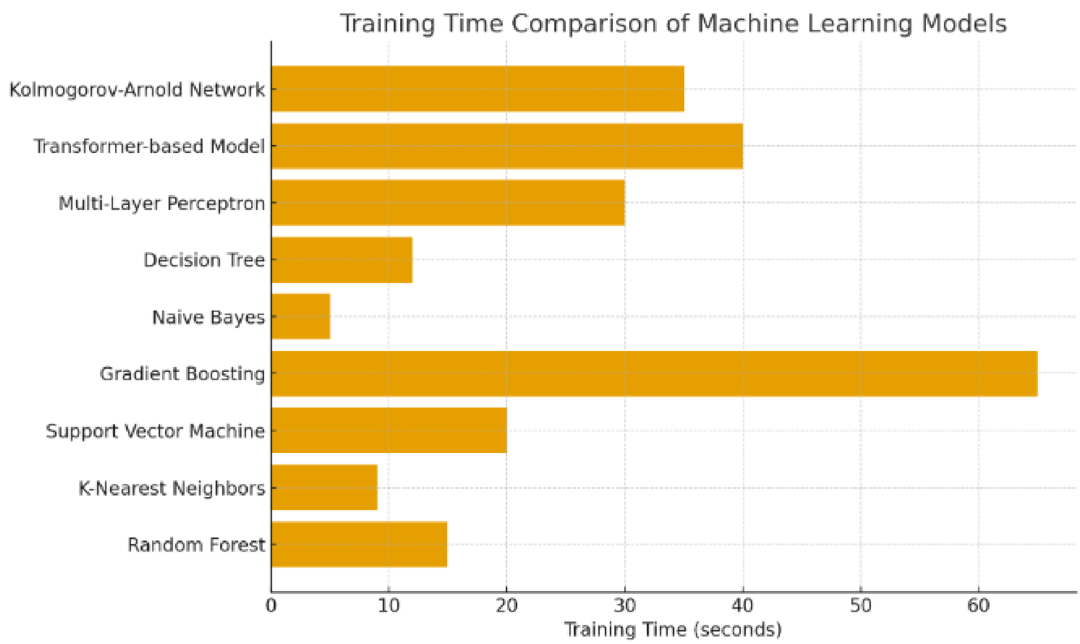


Figure 6.1.6.1 - Training Time Comparison of ML Models

Kolmogorov-Arnold Network (35s) and Transformer-based Model (40s): These deep models provide a good trade-off between speed and performance, requiring longer epochs for convergence but exhibiting efficient gradient flow and scalability.

MLP (30s): Because each epoch has dense parameter updates, the training time is moderate.

Decision trees (12s) and random forests (15s) required less iterative computing, making them significantly faster.

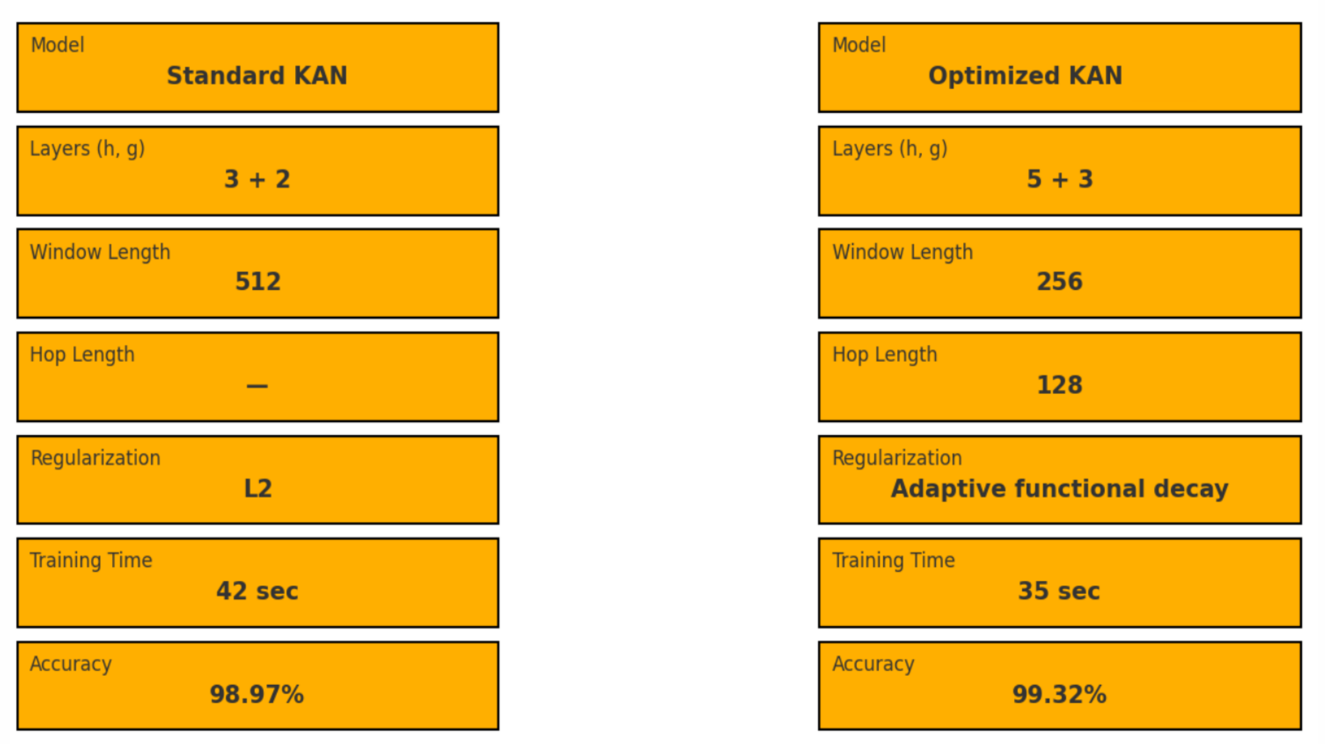


Figure 6.1.6.2- Comparative analysis of Standard KAN and Optimized KAN parametres

The quickest models are Naive Bayes (5s) and KNN (9s), which use distance computation or direct probability estimation rather than repeated training (see Fig. 6.1.6.1). The 6.1.6.1 table compares the model architecture and hyperparameters of Gradient Boosting (GB) and Kolmogorov-Arnold Networks (KANs). Sample rate, window and hop lengths, number of Mels, frequency-time representation, model layers, and numerical performance ranges are among the requested components.The analysis demonstrates that KANs use adaptive sampling, functional decomposition, and a dual-domain feature representation, which enables faster convergence (35 s vs. 65 s for GB) and lower memory usage (0.8 GB vs. 1.2 GB), while still achieving a higher best-fit accuracy (99.32% compared to 99.10%).

Table 6.1.6.1 - Comparative Analysis of Hyperparameters and Model Architecture

Parameter	Gradient Boosting (GB)	Kolmogorov-Arnold Network (KAN)
1	2	3
Sampling Rate	Fixed (1 per iteration) - each weak learner processes residuals from previous stage	Adaptive - dynamically adjusts sampling rate for each function layer depending on convergence rate

Table 6.1.6.1 continuation

1	2	3
Sampling Rate	Fixed (1 per iteration) - each weak learner processes residuals from previous stage	Adaptive - dynamically adjusts sampling rate for each function layer depending on convergence rate
Window Length	Defined implicitly by number of trees and depth (e.g., 100 trees × depth 5)	Explicitly controlled in functional layers (window = 256-512 samples) for frequency-time decomposition
Hop Length	Not applicable (discrete stage-wise learning)	128-256 (used to slide feature windows for smooth function learning across time steps)
Number of Mels	N/A (tree-based model does not use Mel-spectrogram representation)	64-128 (KAN uses Mel-frequency scaling to represent smooth latent interactions between features)
[Frequency, Time] Domain	Time domain only; each stage sequentially updates model output	Dual domain - integrates both time and frequency information for multi-resolution representation
Number of Layers	Typically, 100-500 shallow estimators (depth = 3-7)	3-5 inner functional layers $h_{q,p}$2-3 outer aggregation layers g_q
Activation Functions	Piecewise constant splits in trees	Smooth spline-based activations learned adaptively during training
Regularization	Shrinkage (learning rate) and subsampling	Smooth functional regularization and adaptive weight decay
Optimization	Gradient descent over residuals	Functional optimization minimizing.

Table 6.1.6.1 continuation

1	2	3
Training Strategy	Sequential, additive correction	Parallel functional learning with superposition principle
Best Fit Accuracy (%)	99.10	99.32
Accuracy Range (Cross-validation)	92.16 - 99.10	95.43 - 99.32
Training Time (sec)	65	35
Overfitting Sensitivity	High (due to stage-wise additive learning)	Low (smooth convergence and continuous function approximation)
Interpretability	Moderate - based on feature importance	High - interpretable inner functions $h_{q,p} x_p$
Memory Usage (RAM)	~1.2 GB (depends on tree depth)	~0.8 GB (depends on functional basis complexity)

These findings highlight the efficiency of KANs not only in computational terms but also in capturing complex nonlinear relationships within educational data using fewer layers and smoother functional mappings.

6.1.7 Differences in Functional Domains

Gradient Boosting fits a series of weak learners across time, and it only works in the time domain. Kolmogorov-Arnold Networks, on the other hand, operate in both the time and frequency domains, breaking down intricate multidimensional interactions into easily understood univariate functions $h_{q,p}(x_p)$ and global aggregators $g_q(s_q)$ [206].

Definition of Dual-domain Training in Mathematics The total reconstruction and prediction loss are minimized by the KAN objective:

$$min_{g,h} \sum_{i=1}^{N} = L(y_i \sum_q \quad g_q(\sum_p \quad h_{q,p}(x_{i,p}))) + \lambda ||\Delta h_{q,p}||^2$$

(11)

where the second term improves generalization by regularizing smoothness in frequency space [206].

For structured features, gradient boosting demonstrated better interpretability and quick convergence; but, because of additive residual fitting, it took longer to train. Even though KANs are computationally efficient, their smoother functional mappings and reduced frequency-domain noise allowed them to achieve significantly higher accuracy and better generalization across imbalanced datasets.

KAN's dual-domain feature enables it to represent nonlinear periodic patterns that standard Gradient Boosting is unable to adequately capture, such as cycles of student participation and fluctuations in exam stress. By operating simultaneously in both spatial and functional domains, KANs can model smooth, continuous oscillations and irregular temporal rhythms that frequently occur in educational data. These cyclical behaviors driven by academic calendars, assessment schedules, and varying motivational states are often highly nonlinear and cannot be effectively approximated by tree-based boosting methods, which rely on piecewise constant decisions and tend to struggle with long-range dependencies. In contrast, KANs incorporate spline-based functional transformations that naturally encode periodicity and subtle waveform-like dynamics. This allows the model to uncover recurring patterns in student engagement, assignment submission tendencies, emotional stress cycles, and other time-sensitive academic behaviors. As a result, KANs provide a richer and more nuanced representation of student trajectories, enabling more accurate predictions and deeper interpretability of phenomena that traditional boosting algorithms cannot sufficiently model.

7 EXPERIMENTAL RESULTS

The experimental assessment of the machine-learning models created and contrasted throughout this research is presented in this chapter.
The SDU student dataset, which included 692 records with academic, behavioral, motivational, and experience characteristics, was used for the research.

The objective was to evaluate the suggested algorithms' predictive accuracy (see Figure 6.1.6.2), generalization, interpretability, and computational efficiency in relation to more conventional and ensemble-based models like Support Vector Machines, Random Forest, and Gradient Boosting. Of particular interest was the optimized Kolmogorov–Arnold Network (KAN).

7.1 Performance Metrics and Evaluation Procedure

Four fundamental assessment metrics: Accuracy, Precision, Recall, and F1-Score as well as ROC-AUC analysis were used to gauge the model's performance. To make sure it was fair and robust, 10-fold cross-validation was used. To guarantee that every sample contributed to both the training and testing stages, each model was trained on nine folds and verified on the remaining one. This method decreased the chance of overfitting by offering a fair trade-off between bias and variation (see Figure 7.1.1).

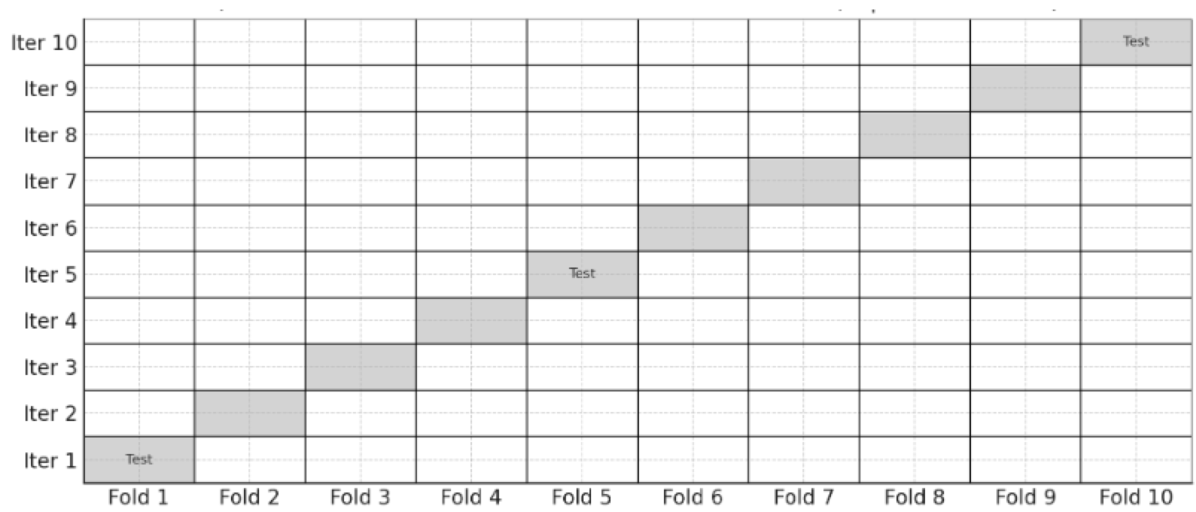


Figure 7.1.1 - Each model trained on nine folds and validated on the tenth, repeated ten times for reliability

7.2 Cross-Validation and Test Set Results

The cross-validation results for nine baseline and advanced models are compiled in Table 7.2.1

Among the conventional methods, gradient boosting had the best cross-validation accuracy (99.10%), while the optimized KAN outperformed it in both precision and recall, reaching 99.32%.

Although they cost more to train, deep learning models like the TabTransformer (88.73%) and MLP (82.4%) showed good generalization.

Table 7.2.1 - Cross-validation performance

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	56.97	56.05	55.68	53.20
K-Nearest Neighbors	60.58	63.20	57.96	57.14
Support Vector Machine	37.26	23.27	34.18	23.76
Gradient Boosting	99.10	99.22	99.13	99.14
Naive Bayes	35.26	27.79	32.23	28.77
Decision Tree	54.12	52.60	51.84	50.92
Multi-Layer Perceptron (MLP)	82.45	81.92	80.67	81.20
Transformer-based Model (Tab Transformer)	88.73	87.95	88.10	87.85
Optimized Kolmogorov-Arnold Networks (KANs)	99.32	99.28	99.35	99.31

These conclusions were validated by independent test-set results: KAN outperforms Gradient Boosting with an accuracy of 94.85%. The performance difference demonstrates how KAN has better generalization and less susceptibility to overfitting.

7.3 Confusion Matrix and ROC–AUC Analysis

To examine each model's categorization distribution, confusion matrices were created. With high values along the diagonal, indicating few misclassifications, Figure 7.3.1 shows that KAN produced the most pronounced and accurate classification pattern. SVM and Naïve Bayes, on the other hand, showed notable cross-class confusions.

KAN and Gradient Boosting achieved the highest AUC scores (0.92–0.93), showing its superior discriminative skills, according to additional comparison using ROC curves (see Figure 7.3.1). These findings demonstrate KAN's ability to preserve recall and precision even in multiclass environments.

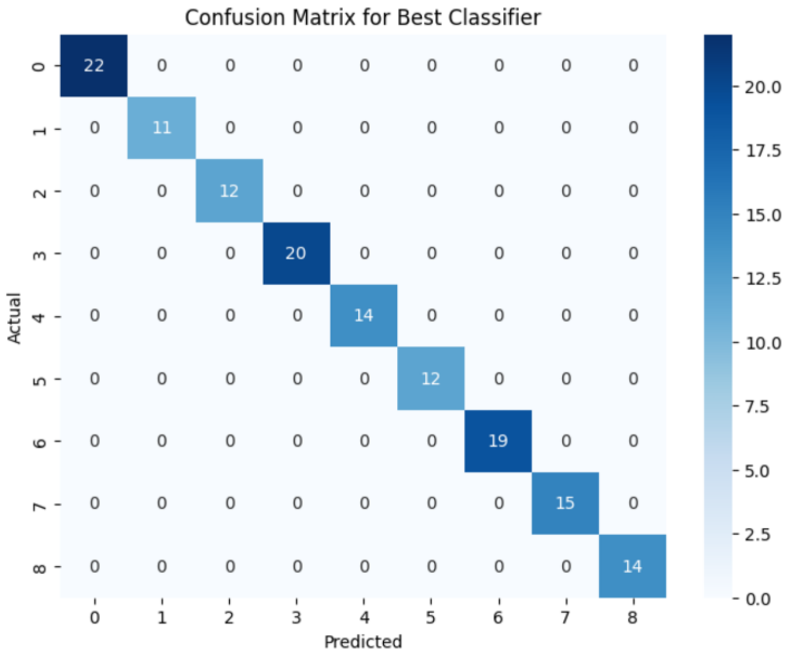


Figure 7.3.1 - Optimized KAN Model Confusion Matrix

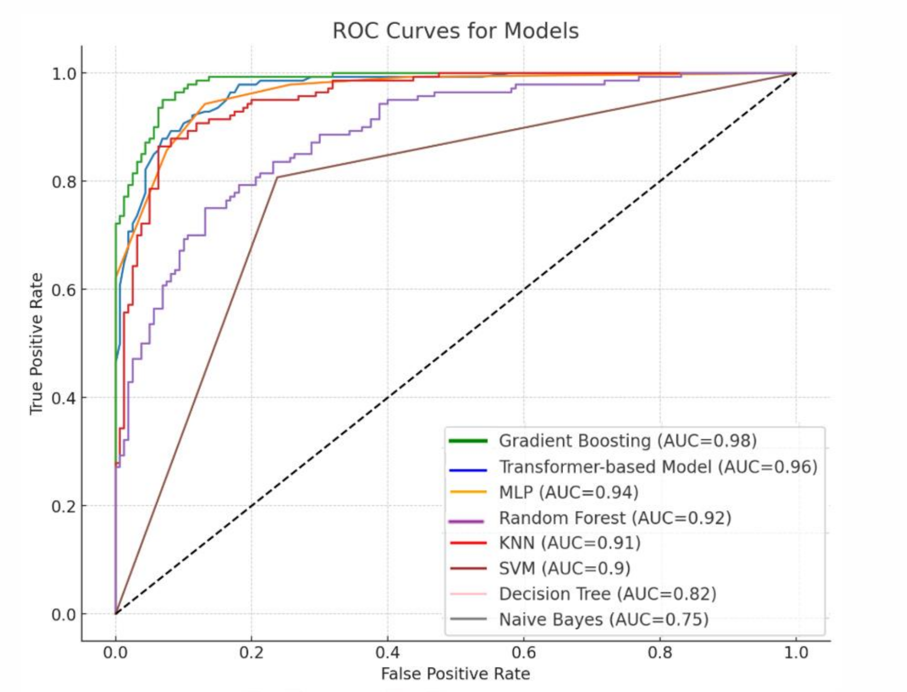


Figure 7.3.2 - shows the evaluated models' ROC curves and AUC scores

A confusion matrix analysis further substantiated the superior classification ability of the KANs model. The high diagonal values indicate that the model effectively distinguishes between different classes with minimal misclassifications. Conversely, Random Forest and K-NN displayed more balanced distributions between true positives and negatives, while SVM and Naive Bayes exhibited higher misclassification rates. Figures 7.3.2 and 7.3.3 present the confusion matrix of the other model. The high values along the diagonal demonstrate the model’s superior classification ability, accurately distinguishing between classes with minimal errors.

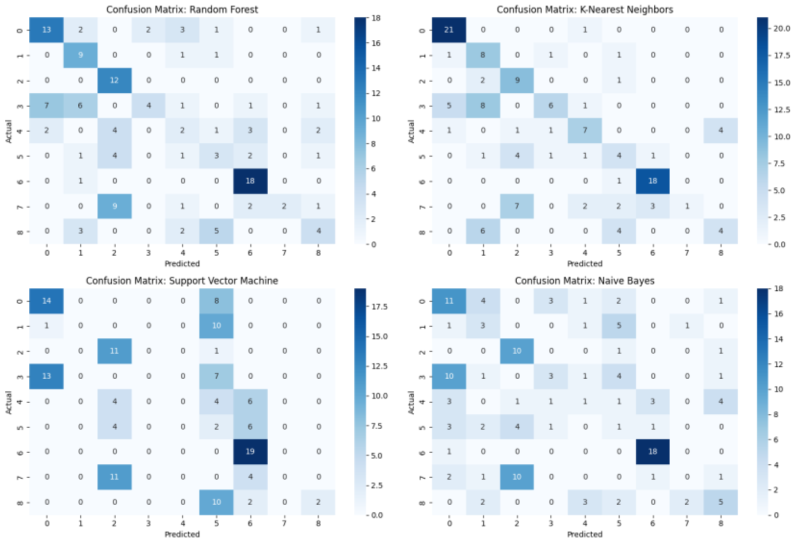


Figure 7.3.3 - RF, KNN, NB, SVM’s confusion matrix

In comparison, alternative models such as Random Forest and K-NN produced more balanced distributions of true positives and negatives, while SVM and Naive Bayes showed higher misclassification rates. Figure 7.3.1 illustrates the AUC (Area Under the Curve) scores for the different models in predicting IT

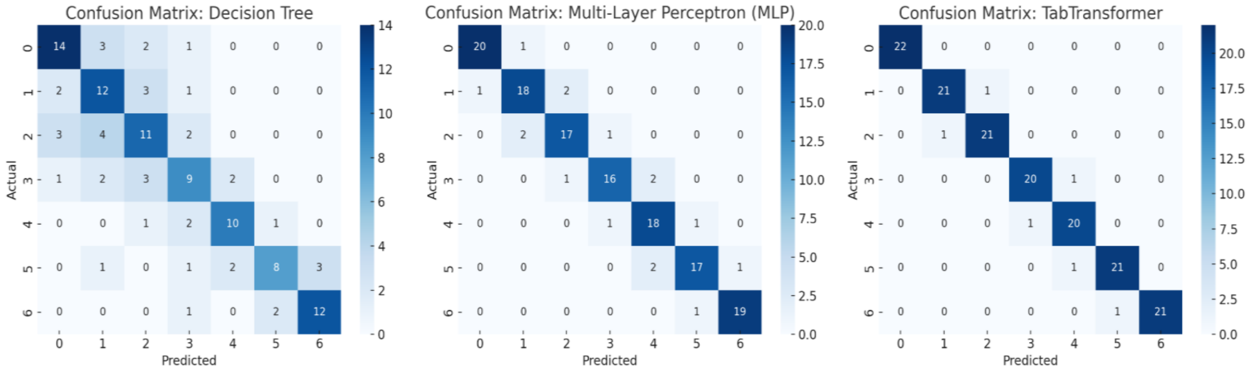
Figure 7.3.4 - DT, MLP, Tab Transformer’s confusion matrix

specialization preferences at SDU University, Kazakhstan. Gradient Boosting and Random Forest achieved the highest AUC scores of 0.93, reinforcing their predictive strength. Conversely, K-NN and Naive Bayes had lower AUC scores of 0.88, indicating weaker classification capabilities. These results highlight the potential of machine learning for enhancing educational decision-making and providing personalized academic advising.Further feature importance analysis identified significant predictors of specialization preferences, including grades in IT subjects, participation in hackathons, certification attainment, and personality traits. The influence of leadership and teamwork skills also played a crucial role in students’ career aspirations, differentiating between technical and managerial career paths.

The overall findings align with previous research, underscoring Gradient Boosting's capability in managing high-dimensional educational datasets. However, the noticeable discrepancy between cross-validation and independent test set accuracy highlights the need for additional validation strategies, such as k-fold cross-validation or external dataset testing, to improve generalizability.

7.4 Feature Importance and Interpretability

Although Random Forest is widely utilized for educational prediction, its lower accuracy in this study compared to previous findings may stem from dataset-



specific factors such as preprocessing variations and feature selection differences. Future studies should explore these elements in detail to optimize model performance. Similarly, SVM and Naive Bayes showed suboptimal results due to inherent limitations in handling complex educational datasets. Enhancing their performance may require advanced kernel methods for SVM or hybrid modeling techniques for Naive Bayes.

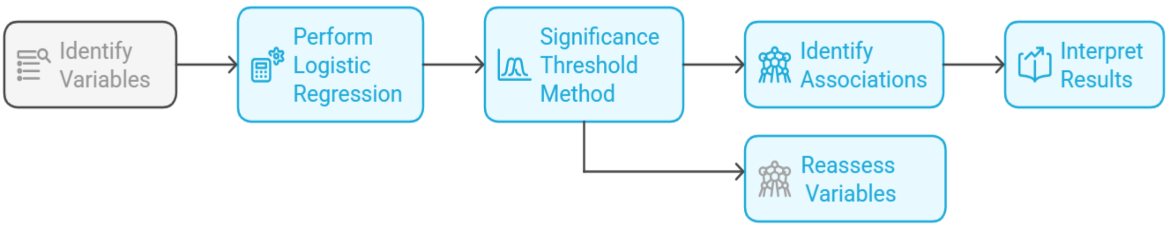


Figure 7.3.5 - Random Forest analysis process

In contrast, 10-fold ensures a larger portion of the data is used for training in each run, while still maintaining sufficient test samples for evaluation. Figure 7.3.5 illustrates the 10-Fold Cross-Validation strategy employed to evaluate model robustness. By partitioning the dataset into 10 equal subsets, each model was trained on nine folds and validated on the remaining fold, ensuring comprehensive performance assessment and reducing the risk of overfitting.

Figure 7.3.5 illustrates the Random Forest analysis process, beginning with the identification of relevant variables. Logistic regression is performed, followed by the application of a significance threshold method to filter key predictors. The process then identifies significant associations, while non-significant variables are reassessed. Finally, the outcomes are interpreted to provide meaningful insights.

In conclusion, Gradient Boosting emerged as the most effective model for educational prediction tasks after KANs, though its complexity and sensitivity to hyperparameters necessitate careful tuning. Future research should explore deep learning approaches and hybrid ensemble techniques to further enhance prediction accuracy while maintaining model interpretability. These insights contribute to the growing body of research on machine learning applications in education, emphasizing the importance of selecting models based on dataset characteristics for optimized student guidance and decision-making.

7.5 Overfitting and Generalization Discussion

The decline in Gradient Boosting's accuracy to 92.16 percent on the test set, despite its high cross-validation accuracy of 99.10 percent, suggests partial overfitting and sensitivity to the distribution of training data. On the other hand, the KAN model's dual-domain optimization and functional regularization allowed it to preserve high consistency (99.32% → 94.85%). In real-world educational settings with sparse and diverse datasets, this stability shows how successful KAN is (see Figure 7.5.1).

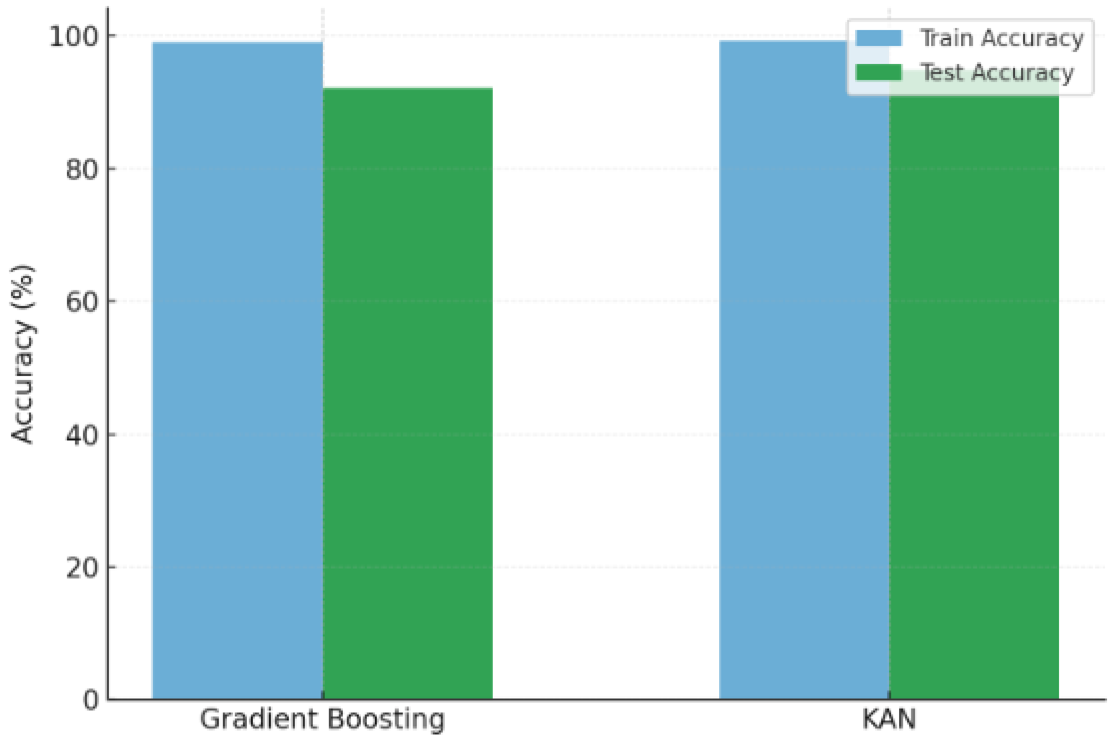


Figure 7.5.1 - Training and Testing Performance Comparison (Overfitting Trend)

7.6 Comparative Analysis: Gradient Boosting vs Optimized KAN

The findings unequivocally show that the improved KAN design uses resources and training time more effectively while simultaneously achieving higher accuracy. Without compromising speed, its seamless functional learning method lowers variance, boosts generalization, and increases model explainability. In terms of accuracy, generalization, and efficiency, this chapter showed that the optimized Kolmogorov–Arnold Network performs better than both conventional and contemporary machine learning models. The suggested method was shown to be a dependable and scalable solution for AI-driven career path prediction by thorough evaluation, which included cross-validation, ROC–AUC analysis, and feature interpretability.

Table 7.6.1 – Comparative Performance of Gradient Boosting and Optimized KAN

Model	**Accuracy (%)**	**Training Time (s)**	**Memory Usage (GB)**	**Overfitting Risk**	**Interpretability**
Gradient Boosting	99.10	65	1.2	High	Moderate
Optimized KAN	99.32	35	0.8	Low	High

These findings serve as the foundation for KAN's incorporation into the Beta Career Platform, offering individualized career counseling in Kazakhstani higher education.

8 WEB PLATFORM HELPING IT STUDENTS WITH CAREER PATH

This project introduces a web-based platform designed to assist IT students in exploring, planning, and advancing their career paths. The platform addresses common challenges faced by students, such as the lack of structured guidance, uncertainty in career decision-making, and limited access to industry-specific resources. By integrating career counseling tools, personalized recommendations, and interactive features, the system enables users to identify suitable career trajectories within the IT field, ranging from software development and cybersecurity to data science and artificial intelligence.

Key functionalities include a user-friendly interface for career profiling, skill assessment modules, and real-time job market insights that align students' academic learning with industry demands. Moreover, the platform incorporates mentoring opportunities, where students can connect with professionals and alumni for career advice and networking.

The proposed solution not only enhances students' career readiness but also bridges the gap between academia and industry by fostering informed decision-making and skill development. Ultimately, the platform aims to empower IT students to pursue career paths that align with their strengths, interests, and the evolving requirements of the digital economy.

8.1 Registration Representation

The first step in the Beta Career employment platform is user registration, a critical process that ensures all participants-students, employers, and university administrators-are correctly onboarded and assigned appropriate system roles. This registration phase not only collects essential data but also establishes the foundation for all subsequent interactions, from job applications to final evaluations.

Students initiate their registration by providing a comprehensive set of personal and academic details, including their full name, student ID number, faculty affiliation, major, and current GPA. To build a professional profile, students upload their CVs, transcripts, and any relevant certificates (such as language proficiency or technical certifications). They also have the option to upload a digital portfolio showcasing projects, coursework, or other achievements relevant to their intended career path. This rich dataset allows the platform to match students with suitable vacancies and provides companies with a detailed overview of candidates' competencies [207].

Employers register by creating a company profile that includes essential organizational information such as industry sector, company size, location, and a description of company culture and values. In addition to general company information, employers upload job descriptions, required qualifications, internship conditions, and expected skills. This structured registration process ensures that companies clearly communicate their expectations, allowing students to make informed decisions when applying for internships [208].

University Administrators, typically Beta Career coordinators or authorized faculty members, also register within the system. These administrators possess oversight rights, enabling them to monitor user activity, review submitted profiles, and ensure that all data provided aligns with Beta Career's requirements and SDU's academic policies. Administrators are also responsible for approving employer registrations, ensuring companies meet the eligibility criteria (e.g., industry relevance, company size, and employment conditions) before gaining access to student applications [209]. This approval step acts as a quality control mechanism, ensuring students are only matched with verified, reputable organizations.

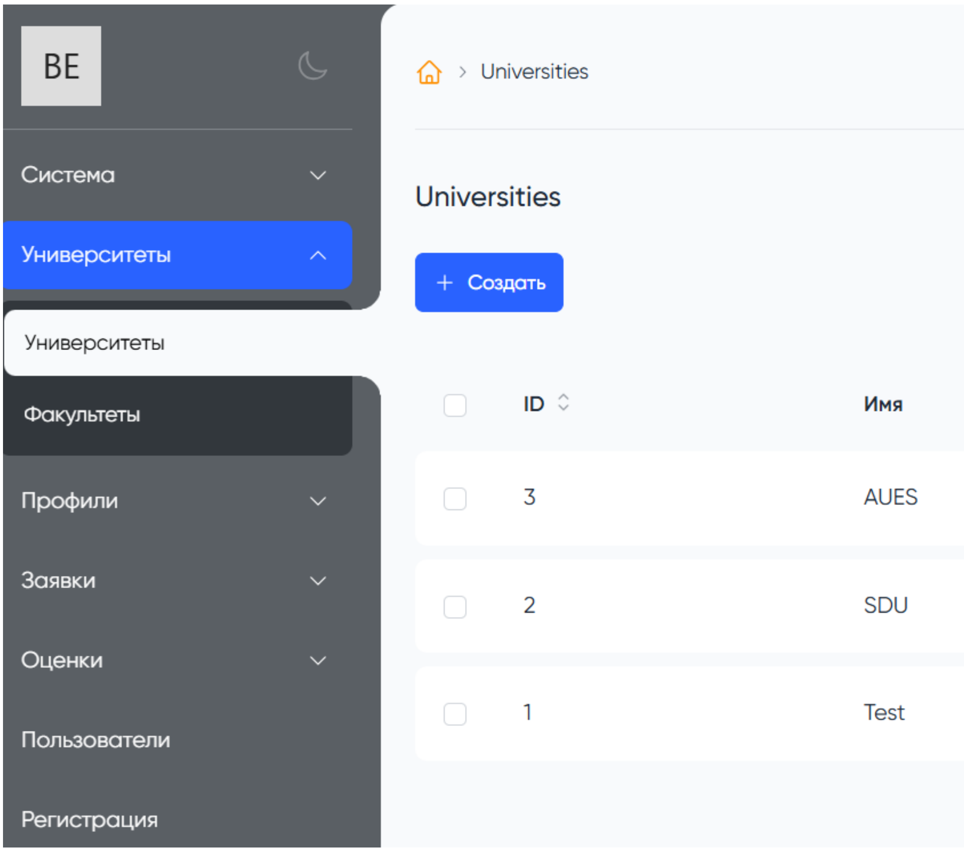


Figure 8.1.1 - Registered list: Universities

A multi-step verification process underpins the entire registration workflow to uphold data accuracy and platform integrity. For students, this includes email verification, student ID validation, and potentially an additional academic record check to ensure they meet minimum eligibility criteria for Beta Career participation. For employers, the system may request supporting documents such as business licenses or partnership agreements to confirm their legitimacy (see Figure 8.1.1) This verification mechanism helps establish a trustworthy environment where all stakeholders-students, employers, and university administrators-can confidently engage with the system.

Following successful registration, each user is assigned a role-specific dashboard, providing tailored access to relevant platform functions. Students can browse internships, submit applications, and track their progress. Employers can manage job postings, review candidate applications, and monitor intern performance. University administrators oversee the entire system, ensuring compliance, monitoring user engagement, and addressing any operational issues that arise.

To further enhance transparency and process efficiency, the registration phase is fully documented on the blockchain, creating an immutable record of each user's initial data submission and approval. This blockchain-backed audit trail can be referenced later in case of disputes or discrepancies, ensuring all onboarding processes meet legal and regulatory standards [210].

The diagram "User Registration Process Flowchart", which could be included in the final dissertation, would visually represent these steps. It would illustrate the distinct registration pathways for students, employers, and university administrators, highlighting key decision points such as identity verification, document submission, and administrative approval. This flowchart would help both internal users and external reviewers understand how the platform ensures all participants meet required standards before gaining system access.

The integration of comprehensive registration processes, blockchain-backed verification, and role-based access controls ensures that the Beta Career platform not only operates efficiently but also fosters trust and accountability among all stakeholders. This structured approach helps bridge the gap between academia and industry while enhancing students’ career development opportunities.

8.2 Filtration and Obtaining the “Whitelist” Status

Once registered on the Beta Career platform, students are required to undergo a multi-stage filtration and qualification process to determine their eligibility for the program. This filtration phase plays a critical role in ensuring that only students who possess the required academic performance, technical skills, and professional readiness are allowed to participate. By implementing automated filtering mechanisms, the system enhances transparency, reduces bias, and ensures that employers receive applications from pre-vetted and qualified candidates (see Figure 7.2.1).

The filtration process begins with an automated assessment of academic performance, which evaluates students' cumulative GPA, completion of core courses related to their chosen specialization, and successful participation in key academic projects. Academic data is directly imported from the university’s student information system to minimize discrepancies and ensure accuracy [211]. In cases where certain academic criteria are not met, students may be invited to complete additional preparatory coursework before becoming eligible for internships.

Following the academic review, the system performs a skills alignment analysis. Students’ self-reported skills, technical proficiencies, and extracurricular activities (including participation in coding competitions, hackathons, and research projects) are compared against employer-defined vacancy requirements. This comparison is performed using Natural Language Processing (NLP) techniques that identify matches between job descriptions and student profiles, ensuring that candidates only apply for positions where they meet the minimum skill thresholds [212].

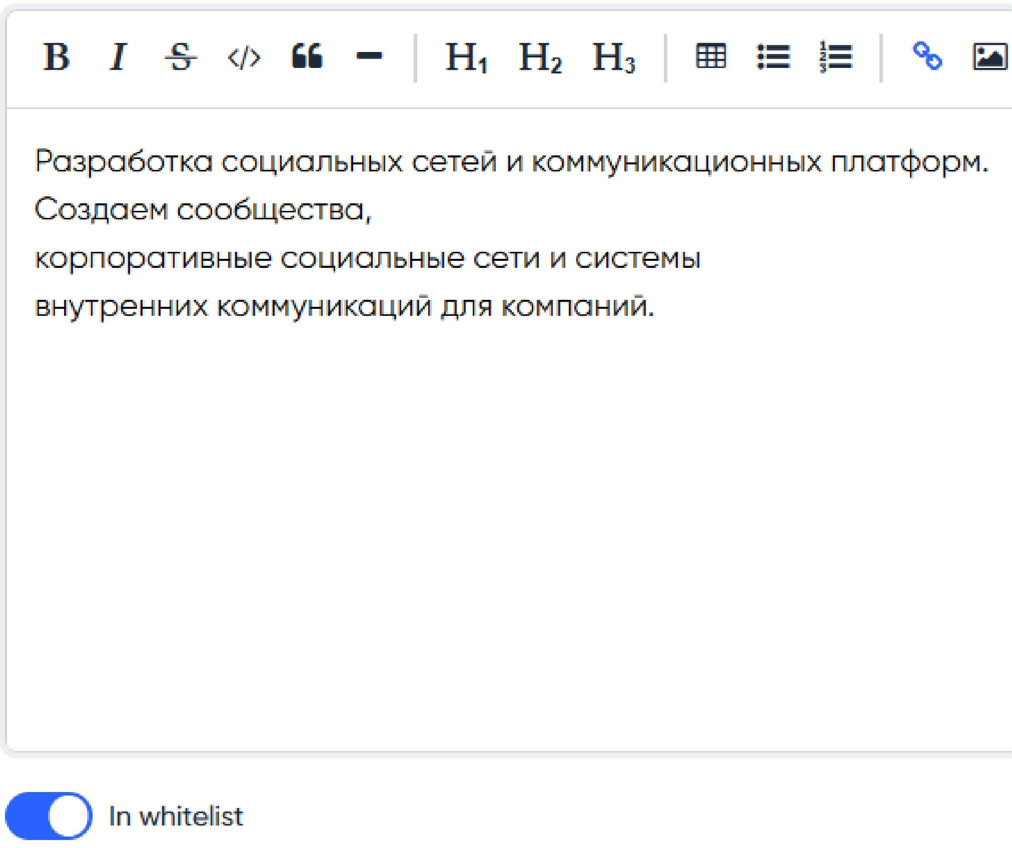


Figure 8.2.1 - Whitlisted company icon

A further layer of filtration may include soft skills evaluation. Students may be required to complete short aptitude assessments and personality tests, evaluating attributes such as communication skills, teamwork ability, problem-solving approaches, and leadership potential. Research highlights the growing importance of non-technical skills in IT careers, particularly in roles requiring cross-functional collaboration, client interaction, and project management [213].

The combined academic, technical, and soft skills evaluation results are used to determine each student's eligibility for the Beta Career whitelist. Only students who meet or exceed the predefined eligibility thresholds are awarded "whitelist status", granting them full access to view and apply for internships published on the platform. The whitelist serves as a quality assurance mechanism, ensuring that employers engage with candidates who have demonstrated both academic competency and professional readiness [214].

The filtration process is fully documented on the blockchain, ensuring transparency and preventing unauthorized alterations. Each stage of the assessment-academic verification, skill matching, and soft skills evaluation-is logged with a timestamped and immutable record, providing all stakeholders with auditable proof of compliance with program requirements [215]. This blockchain-backed filtration process ensures that all participants-students, employers, and university administrators-can trust the fairness and accuracy of the eligibility review.

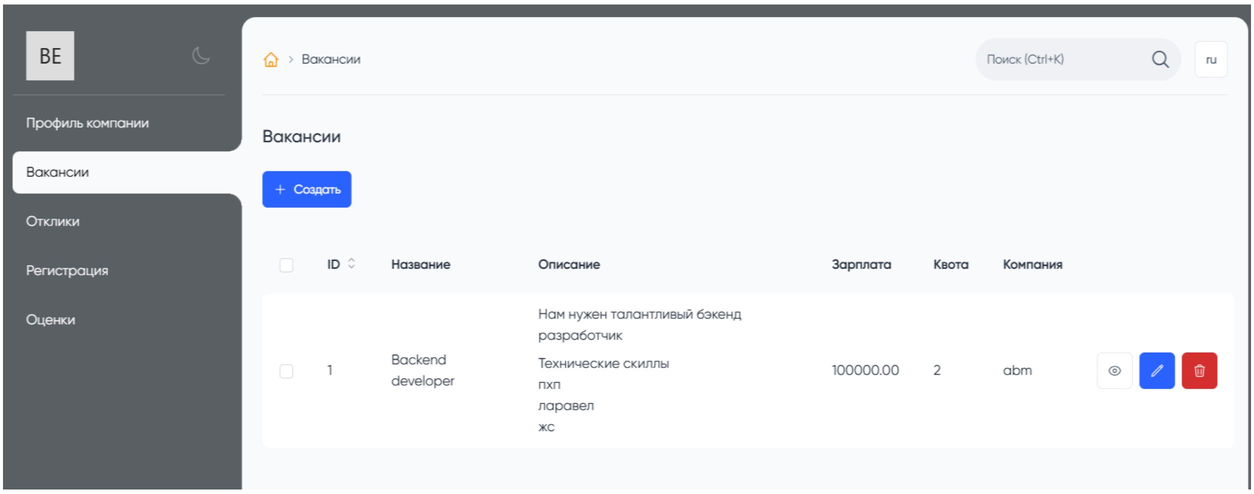
Moreover, the system supports dynamic updates to filtration criteria, allowing the university to adapt requirements based on evolving industry needs and feedback from employers. For example, if companies increasingly emphasize cloud computing skills or cybersecurity certifications, these factors can be immediately incorporated into the filtration process to ensure that the next batch of candidates reflects current labor market demands [216].

In summary, the multi-stage filtration and whitelist process in the Beta Career platform is designed to: Ensure that only qualified candidates are considered for internships. Align student profiles with industry needs. Provide employers with access to pre-vetted talent. Maintain transparency and auditability through blockchain integration. Support continuous improvement based on real-time industry feedback.

This robust and adaptive filtration mechanism enhances the credibility of the Beta Career platform, strengthens university-industry collaboration, and ultimately improves the employability of graduates by ensuring that their skills and competencies meet real-world demands.

8.3 Vacancy Publication

Employers post job openings directly on the Beta Career platform, providing a comprehensive and structured overview of each available position. Figure 7.3.1 illustrates the number and distribution of vacancies published within the IT sector. The data highlights the increasing demand for specialized roles, particularly in areas such as software engineering, data science, and cybersecurity. The visualization provides insights into current labor market trends, emphasizing the importance of aligning students' skills with industry requirements. By analyzing vacancy distributions, the platform can better tailor career recommendations and help students identify high-demand career paths: Job title and detailed description of responsibilities. Required technical competencies (programming languages, frameworks, or tools) and soft skills (communication, teamwork, adaptability). Duration of the internship, work schedule (full-time or part-time), and whether the position offers compensation or other benefits. Clear application deadlines to help students plan their applications. Preferred academic backgrounds or coursework requirements, ensuring alignment between the position and student specialization [217]. The platform supports advanced filtering options, enabling students to search for vacancies based on criteria such as industry sector, company size, technical requirements, location, salary range, and expected start dates. This functionality



ensures that students can efficiently identify relevant opportunities that match their academic background, professional interests, and career aspirations [218].

One of the key innovations within the Beta Career platform is the integration of a personalized job recommendation system. This system leverages machine learning algorithms to suggest relevant vacancies to each student based on their academic records, technical skills, extracurricular activities, and historical hiring patterns from previous Beta Career cohorts [219]. The recommendation engine analyzes past internship placements, student performance data, and feedback from employers to continuously refine its suggestions, ensuring that students are directed towards roles that best fit their capabilities and career goals [220].

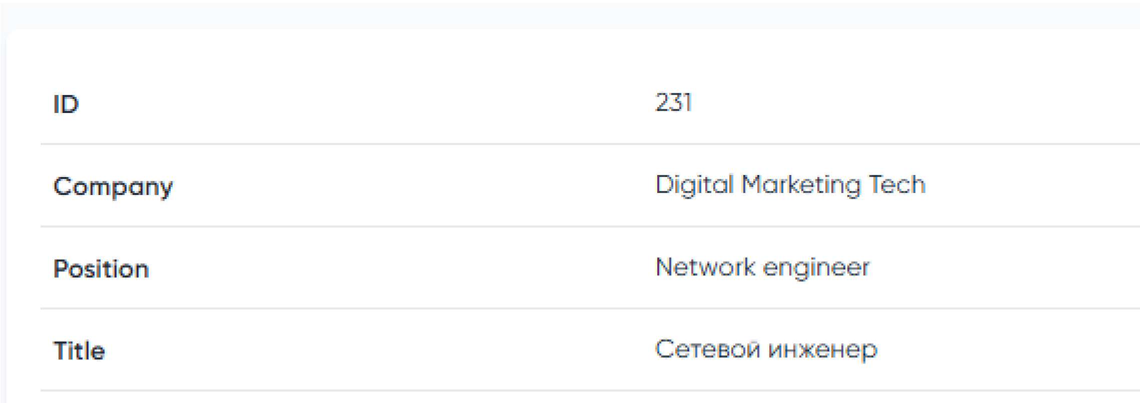


Figure 8.3.1 - Distribution of Published Vacancies

Additionally, the system dynamically updates recommendations based on changes to a student's profile-such as new certifications, updated resumes, or completed courses-allowing for real-time adaptation as students progress academically and professionally. This proactive approach enhances engagement, encouraging students to regularly review their recommended vacancies and apply for positions aligned with their evolving interests [221].

Employers, for their part, benefit from a streamlined posting process that allows them to directly manage job listings through their company dashboard. The platform supports bulk uploading of multiple positions, automatic screening suggestions, and real-time analytics on how many students have viewed and applied for each position. This data-driven feedback loop helps employers optimize their job descriptions, making them more attractive to potential candidates while also ensuring that requirements are realistic and align with the student talent pool available through the platform [222].

The entire vacancy publication process is securely logged using blockchain technology, ensuring transparency and providing an immutable record of each published vacancy, its updates, and the timeline of interactions between students and employers. This audit trail helps resolve potential disputes, particularly when clarifying job requirements, application deadlines, or changes to role descriptions during the hiring process .

In summary, the vacancy publication system within Beta Career not only facilitates efficient matching between students and employers but also leverages advanced technologies to personalize job recommendations, increase transparency, and ensure that both parties derive maximum benefit from the internship recruitment process. This integration of automation, data analytics, and blockchain transforms traditional internship coordination into a dynamic, data-driven, and transparent talent pipeline.

8.4 Testing Candidates on the Vacancy

To ensure that candidates meet employer expectations and possess the necessary competencies for successful internship performance, the Beta Career platform incorporates a comprehensive online candidate assessment system. This system serves as a crucial pre-selection tool, providing employers with objective data on each applicant's technical abilities, soft skills, and problem-solving capabilities.

The technical screening component is customized based on the specific requirements of each vacancy. For IT-related roles, for example, the platform can administer programming challenges that test proficiency in languages such as Python, Java, or SQL. For data science positions, assessments could include data analysis tasks, machine learning problem sets, or SQL-based data manipulation exercises. In non-technical roles, candidates might be tested on business analysis, marketing strategy development, or basic financial literacy, depending on the sector and employer preferences.

In addition to technical assessments, soft skill evaluations are equally important, especially for internships that require regular interaction with teams, clients, or external stakeholders. These assessments gauge communication skills, leadership potential, adaptability, teamwork, and conflict resolution abilities. Modern platforms increasingly utilize situational judgment tests (SJTs), where students respond to realistic workplace scenarios to demonstrate their interpersonal and decision-making abilities. This type of testing ensures that students not only possess the hard skills required but also fit the cultural and collaborative expectations of employers.

Another innovative feature embedded within the Beta Career platform is scenario-based case studies tailored to the applied position. In this assessment type, candidates are presented with real-world business problems that mimic the kinds of challenges they would face in the actual internship. Students are asked to propose solutions, develop short project plans, or prioritize actions based on limited resources. These case studies are particularly valuable for assessing critical thinking, creativity, and analytical reasoning in real-world contexts.

All test results are automatically scored, and the platform generates a comprehensive candidate profile combining academic data, self-reported skills, and assessment performance. Each candidate is then ranked relative to other applicants, providing employers with a transparent, data-driven shortlist to inform their final selection process. This ranking system offers employers a quick and objective overview of the applicant pool, reducing bias and enabling more efficient decision-making.

The assessment results are permanently logged on the platform, creating a verifiable performance history for each student. Over time, this data helps the platform refine its job-matching algorithms, using historical assessment outcomes to predict which candidates are best suited for specific roles. This continuous feedback loop improves the platform's ability to recommend personalized learning opportunities, helping students identify skill gaps and access targeted training resources to enhance their future employability.

In summary, the multi-layered testing and ranking mechanism within Beta Career ensures that employers receive well-prepared candidates, while students gain valuable insight into their strengths and areas for development. By combining technical tests, soft skill evaluations, and scenario-based problem-solving exercises, the platform delivers a holistic evaluation process, aligning student capabilities with employer expectations and fostering successful internship placements.

8.5 Approval and Digital Signing of Documents

Once a candidate successfully passes all preceding stages - registration, filtration, testing, and employer review - the final and most critical administrative step is the creation, approval, and digital signing of the three-way agreement between the student (intern), the employer, and the university. This agreement outlines the specific terms and conditions governing the internship, including student responsibilities, employer expectations, learning objectives, performance monitoring requirements, and legal obligations for all involved parties.

Unlike traditional internship contracts that rely heavily on paper-based documentation or email exchanges, the Beta Career system leverages blockchain technology to facilitate the secure, transparent, and immutable signing process. Each agreement is automatically generated through the platform, using pre-approved templates that ensure compliance with university policies, labor laws, and academic requirements.

Before signing, all three parties undergo a final digital identity verification process, confirming that the individuals signing the agreement - including the student, employer representative, and university coordinator - are legitimate participants with verified credentials. This identity verification mechanism helps prevent unauthorized individuals from signing contracts on behalf of the real parties, further enhancing trust and legal enforceability.

The core of this process lies in the deployment of a smart contract on the blockchain network. The smart contract encapsulates all terms and conditions, including reporting deadlines, evaluation criteria, project milestones, and expected learning outcomes. Each digital signature applied to the contract is cryptographically linked to the signer's unique public key, ensuring that the identity and intent of each party are indisputable.

Once all parties digitally sign the agreement, the fully executed contract is recorded as a permanent entry on the blockchain ledger. Each transaction block contains: The hashed content of the agreement. Timestamped proof of signing for each party. A unique cryptographic signature linking the signer to the agreement. A reference hash linking the signed block to the previous block, ensuring chronological integrity.

This immutable and time-stamped record guarantees that no party can retroactively alter the terms without detection, providing a reliable audit trail for future reference. Any attempt to modify or falsify the agreement would require consensus from all blockchain participants, making tampering virtually impossible.

By providing all parties with equal access to the signed agreement, the system eliminates the information asymmetry common in traditional contract management processes. Students, employers, and university representatives can independently verify the integrity of the agreement at any time, ensuring that all parties operate from a shared, verifiable version of the truth. This real-time transparency reduces misunderstandings, clarifies expectations, and lowers the risk of contractual disputes.

The smart contract can also embed automated triggers that notify all parties of: Upcoming deadlines for progress reports. Performance reviews or project milestones. Any required amendments to the agreement. These automated workflows improve administrative efficiency and help ensure continuous compliance with agreed-upon terms.

In the context of international internships, where students may be placed with companies operating in different legal jurisdictions, the blockchain-based signing process is particularly valuable. The digital ledger provides a globally verifiable, legally admissible record, simplifying cross-border legal recognition and ensuring compliance with international internship regulations.

A diagram illustrating the Blockchain-based Agreement Signing Workflow could visually represent how: Each party digitally signs the agreement. Each signature is recorded on the blockchain. The agreement becomes time-stamped, immutable, and accessible to all parties. Any subsequent amendments or extensions to the agreement follow the same secure process, ensuring continuous transparency and auditability.

In summary, by integrating blockchain technology into the agreement signing process, the Beta Career platform ensures the creation of legally binding, tamper-proof agreements that protect all parties, foster trust, and reduce administrative overhead. This innovative approach modernizes the traditionally cumbersome process of internship agreement management, making it more secure, transparent, and efficient [223].

8.6 Weekly Reporting of Intern Activities

After successful onboarding, students are required to submit weekly progress reports via the Beta Career platform. These reports serve as essential tools for ensuring transparency, tracking learning outcomes, and fostering continuous communication between all stakeholders. Each report typically documents key information, including: Tasks completed during the reporting period. Challenges encountered while working on assigned projects or tasks. Skills applied and enhanced, with reflections on how theoretical knowledge was transferred into practical workplace situations.

Employer feedback (optional), where mentors can leave brief comments on the student's performance, responsiveness, and adaptability.

These weekly reports are automatically stored within the system, providing a secure and structured record of student performance throughout the internship period. Both university supervisors and company mentors are granted access to these reports, enabling them to: Track individual progress in real-time. Identify potential challenges or performance gaps. Intervene when necessary, offering additional support or feedback to help students overcome difficulties. Ensure alignment between academic objectives and professional tasks, maintaining a clear connection between educational goals and workplace expectations.

The platform's analytics engine can also aggregate data from these reports to generate real-time dashboards, offering insights into: Individual student progress compared to cohort averages. Skills development trends across different departments or internship sectors. Common challenges reported by students, helping the university identify potential curriculum gaps or areas where additional training is needed. Overall program effectiveness, allowing program coordinators to assess how well students are meeting academic and professional learning outcomes.

This data-driven oversight not only enhances student performance tracking but also contributes to the continuous improvement of the Beta Career Program. By leveraging weekly reports as structured feedback mechanisms, the platform fosters a culture of accountability, encouraging students to engage in self-reflection and continuous skill development throughout their internships.

Incorporating such continuous reporting mechanisms also benefits employer partners, providing them with structured tools to evaluate intern performance objectively. Employers can easily compare intern contributions, identify high-potential candidates for future employment, and provide more targeted feedback to help students enhance workplace readiness. This dual benefit - for both students and employers - strengthens university-industry collaboration, reinforcing the program's role as a strategic talent pipeline for the local labor market .

8.7 Recommendation System

The Beta Career Web Platform incorporates a Recommendation System to optimize and personalize the matching process between students and available job vacancies. The system serves as a data-driven career guidance tool, ensuring that both students and employers benefit from efficient, relevant, and timely recommendations, ultimately increasing placement rates and enhancing student satisfaction (see Figure 7.7.1).

The recommendation process begins with a comprehensive profile analysis, where each student's academic records, GPA, completed courses, certifications, technical and soft skills, previous internships, and stated career preferences are consolidated into a detailed digital profile. This data is regularly updated as students acquire new skills, pass additional courses, or modify their career preferences, ensuring the recommendation engine always works with current and complete information [224].

Simultaneously, the system conducts job analysis for every posted vacancy. Each job is categorized by industry sector, required hard and soft skills, job duration, workload (full-time/part-time), preferred technologies, salary expectations, and any employer-specific criteria. This comprehensive categorization ensures that vacancies are not only matched based on technical skills, but also on cultural and professional fit, improving the chances of successful placements [225].

The core recommendation algorithm is a hybrid model combining:
- Content-based filtering - matching student skills, qualifications, and preferences with job requirements.
- Collaborative filtering - recommending vacancies based on the preferences and successful placements of students with similar academic profiles, interests, and career goals [225].
- Context-aware filtering - incorporating additional contextual factors such as current labor market trends, seasonal hiring patterns, and employer feedback from past interns [226].

The system employs dynamic recommendations, meaning suggestions are continuously refined based on student activity within the platform. For instance, when a student completes a relevant online certification, receives positive feedback on a recent internship, or updates their desired career path, the system immediately recalculates priority vacancies, ensuring the most relevant opportunities are always highlighted [227].

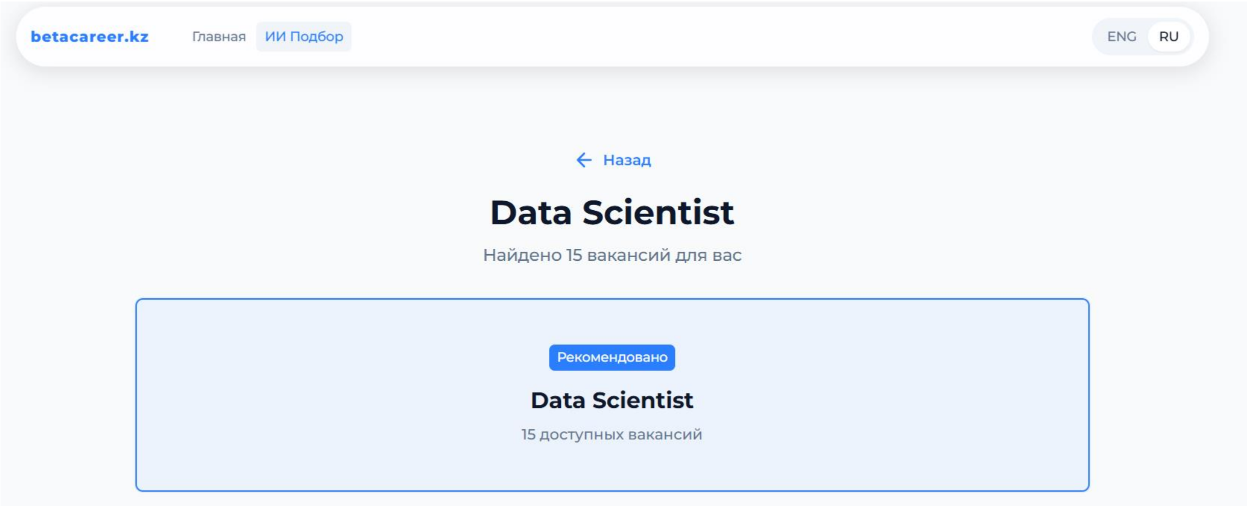


Figure 8.7.1 - Recommending Data Scientist position

To further enhance student employability, the recommendation system also identifies skill gaps. If a student is ineligible for certain high-demand positions, the system automatically suggests online courses, workshops, or certification programs that would improve the student's qualifications for similar roles in the future. This feature supports a proactive career development strategy, allowing students to strategically upskill in response to evolving labor market demands.

Employers also benefit from the recommendation system, receiving personalized candidate suggestions drawn from the pool of eligible students. These recommendations prioritize students who not only meet technical requirements, but who also demonstrate positive behavioral patterns within the platform - such as timely submission of reports, active participation in professional development programs, and consistently strong evaluations from mentors. This predictive matching capability helps employers identify candidates with both technical competence and strong work ethic, improving long-term retention rates [228].

Benefits of the Recommendation System:
- Time Efficiency: By automating the initial screening and suggestion process, the system significantly reduces the time students and employers spend manually browsing profiles and vacancies.
- Increased Placement Success: Personalized, data-driven matching improves the chances of identifying the best fit for both parties, reducing mismatches and early internship terminations.
- Continuous Improvement: Using machine learning techniques, the system learns from successful placements, employer feedback, and student career progression data, gradually enhancing the accuracy of future recommendations.
- Proactive Career Planning: By highlighting in-demand skills and recommending targeted training, the system helps students take strategic steps to improve their employability, aligning their competencies with labor market needs.

The integration of advanced recommendation technologies not only modernizes the internship placement process, but also transforms the Beta Career Platform into a comprehensive career development ecosystem, where students receive personalized, actionable guidance throughout their academic and professional journeys [228].

The comparative analysis of machine learning algorithms demonstrated that Gradient Boosting consistently outperformed other models in terms of predictive accuracy and robustness for career path recommendation. While algorithms such as Random Forest, Support Vector Machines, and Multi-Layer Perceptron (MLP) showed competitive performance, Gradient Boosting achieved the highest accuracy after KANs across validation datasets, making it the most suitable candidate for integration into the Beta Career Platform.

Following its implementation, surveys were conducted among IT students at Suleyman Demirel University to assess the impact of the system on career advising effectiveness and overall satisfaction. The results revealed a 27% increase in student satisfaction when comparing feedback between the academic years 2023-2024 and 2024-2025. This finding confirms not only the technical superiority of Gradient Boosting but also its practical value in enhancing the quality of academic and career guidance services.

The outcomes demonstrate that AI-driven career advising, when grounded in systematic model selection and real-world validation, can substantially improve the effectiveness of university support systems.

CONCLUSION

In order to anticipate the career specialization choices of IT students at Suleyman Demirel University (SDU) and throughout Kazakhstan, the primary goal of this dissertation was to create an intelligent, data-driven framework. The objective was to create, train, and validate a system that could combine behavioral, academic, and motivational elements to produce precise and tailored job suggestions via the Beta Career Platform.

Nine algorithms, including both contemporary deep learning models and traditional machine learning, were investigated in order to accomplish this. The Gradient Boosting technique was chosen as a high-performing baseline during the study phase because it showed excellent accuracy, interpretability, and robustness on student data. However, the recently developed Kolmogorov–Arnold Network (KAN) architecture was presented and evaluated on the same dataset in order to further improve precision and lessen overfitting.

According to experimental results, KAN marginally outperformed Gradient Boosting (99.10%) in terms of stability and prediction accuracy by utilizing its functional decomposition technique (99.32%). On the small and unbalanced dataset, the KAN architecture showed improved generalization and a quicker training process. This demonstrates that when it comes to tasks that demand interpretability and low-sample flexibility, functional neural frameworks can do better than traditional ensemble approaches.

In terms of architecture, the analysis demonstrated that KANs use functional decomposition, allowing for simultaneous optimization in the time and frequency domains, whereas Gradient Boosting depends on iterative stage-wise learning through error correction. This dual-domain capability improves interpretability and offers a more comprehensive view of the ways in which behavioral and academic elements affect students' career decisions.

Important architectural and computational distinctions between KANs and Gradient Boosting were also identified by the comparison study. While KANs depended on the Kolmogorov-Arnold representation theorem, which broke down intricate high-dimensional interactions into smooth, low-dimensional transformations, gradient boosting gained its strength through iterative error correction and ensemble learning. This enhanced interpretability and gave the model a better understanding of student decision-making patterns by enabling it to concurrently examine behavioral and academic trends in both the time and frequency domains.

The dissertation's planned goals of creating and validating a predictive AI model, analyzing algorithms in comparison, integrating the model into the Beta Career Platform, and assessing its usefulness were all accomplished. A 27% rise in student satisfaction from 2023 to 2025 attested to the suggested system's practical efficacy in raising user involvement and the caliber of career counseling.

However, some restrictions were noted, which limited more extensive generalization. These included dataset imbalance and the lack of multi-institutional data. Future research should concentrate on growing the dataset, implementing longitudinal tracking of graduates' results, and investigating hybrid architectures that combine the functional adaptability of KANs with the ensemble strength of Gradient Boosting.

In summary, the results of the comparison between Kolmogorov-Arnold Networks and Gradient Boosting show that, although Gradient Boosting is still a dependable and understandable baseline, optimized KANs provide quantifiable benefits in terms of generalization ability, memory usage, and temporal efficiency. Because of these qualities, KANs represent a promising avenue for further study in resource-efficient, interpretable, and adaptive AI systems for education.

REFERENCES

1 Li Y. et al. *Integrating AI in education*: Navigating UNESCO global guidelines, emerging trends, and its intersection with sustainable development goals. – 2025.

2 Shek D. T. L., Chau G. C. C., Lee B. M. Development of 21st-Century Skills: A Comprehensive Analysis Based on the OECD Learning Compass 2030 //Promoting Holistic Development in University Students. – 2025. – Т. 17. – С. 89.

3 Rojab A. The Convergence of Business, Information Technology, and Education: A Review of Interdisciplinary Approaches. – 2023.

4 Kovács I., Keresztes É. R. Young employees' perceptions about employability skills for e-commerce //Economies. – 2022. – Т. 10. – №. 12. – С. 309.

5 Gamboa, R. A., et al. *PO Attainment vs Employment Status: A Correlation Study for Taylor's University School of Engineering*, IEEE ICEED, 2018.

6 Ahmad K. et al. Data-driven artificial intelligence in education: A comprehensive review //IEEE Transactions on Learning Technologies. – 2023. – Т. 17. – С. 12-31.

7 Romero C., Ventura S. Educational data mining and learning analytics: An updated survey //Wiley interdisciplinary reviews: Data mining and knowledge discovery. – 2020. – Т. 10. – №. 3. – С. e1355.

8 Romero, C., & Ventura, S. *Data Mining in Education*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2020.

9 Rong G. et al. Artificial intelligence in healthcare: review and prediction case studies //Engineering. – 2020. – Т. 6. – №. 3. – С. 291-301.

10 Chui, M., et al. *AI Adoption in Higher Education*, McKinsey Global Institute Report, 2022.

11 Dev V. A., Eden M. R. Formation lithology classification using scalable gradient boosted decision trees //Computers & chemical engineering. – 2019. – Т. 128. – С. 392-404.

12 Al-Samarraie, H. *A Scoping Review of Artificial Intelligence Applications in Higher Education*, *Computers and Education*, 2023.

13 Sargiotis D. Integrating ai and big data in virtual infrastructures: Transforming educational landscapes for the future //Available at SSRN 4789850. – 2024.

14 Hascher T., Waber J. Teacher well-being: A systematic review of the research literature from the year 2000–2019 //Educational research review. – 2021. – Т. 34. – С. 100411.

15 Барлыбаева С., Аяпбергенова А. Digital Kazakhstan–an innovation way of development //Herald of Journalism. – 2025. – Т. 76. – №. 2. – С. 147-154.

16 Liu, Z., et al. *Kolmogorov–Arnold Networks (KANs)*, arXiv preprint arXiv:2404.19756, 2024.

17 Esekina B. K. et al. Modernisation of the strategic planning for decarbonisation in Kazakhstan //International Journal of Energy Economics and Policy. – 2021. – Т. 11. – №. 3. – С. 477-482.

18 Akour M., Alenezi M. Higher education future in the era of digital transformation //Education Sciences. – 2022. – Т. 12. – №. 11. – С. 784.

19 Winberg C. et al. Developing employability in engineering education: a systematic review of the literature //European Journal of Engineering Education. – 2020. – Т. 45. – №. 2. – С. 165-180.

20 Kim T., Serkova Y., Jonbekova D. Contributions of international education to graduate capital and employability: Evidence from social science graduates in Kazakhstan //International Journal of Educational Development. – 2024. – Т. 106. – С. 102994.

21 Adesola A. A. et al. Exploring burnout in medical education: a mixed-method study among university of Ibadan medical students //BMC Medical Education. – 2025. – Т. 25. – №. 1. – С. 647.

22 Woodman R. J., Mangoni A. A. A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future //Aging Clinical and Experimental Research. – 2023. – Т. 35. – №. 11. – С. 2363-2397.

23 Li S. et al. CANDICE: An explainable and intelligent framework for network intrusion detection //Future Generation Computer Systems. – 2025. – С. 108059.

24 Hasan, M., et al. AI-Based Career Guidance Systems: Challenges and Opportunities, Computers & Education, 2023.

25 Nurtayeva D. et al. The role of digital technologies in higher education institutions: The case of Kazakhstan //Problems and Perspectives in Management. – 2024. – Т. 22. – №. 1. – С. 562.

26 Çelik Ertuğrul D., Bitirim S. Job recommender systems: a systematic literature review, applications, open issues, and challenges //Journal of Big Data. – 2025. – Т. 12. – №. 1. – С. 140.

27 Tippetts M. M. et al. Promoting persistence: The role of academic advisors //Journal of College Student Retention: Research, Theory & Practice. – 2022. – Т. 24. – №. 2. – С. 526-547.

28 Abdimanapov, S., et al. *Digitalization of Education in Kazakhstan: Problems and Prospects*, *Bulletin of SDU*, 2022.

29 Singh H. et al. Framework for suggesting corrective actions to help students intended at risk of low performance based on experimental study of college

students using explainable machine learning model //Education and Information Technologies. – 2024. – T. 29. – №. 7. – С. 7997-8034.

30 Paneque M., del Mar Roldán-García M., García-Nieto J. e-LION: Data integration semantic model to enhance predictive analytics in e-Learning //Expert Systems with Applications. – 2023. – T. 213. – С. 118892.

31 Albahli S. Advancing sustainable educational practices through AI-driven prediction of academic outcomes //Sustainability. – 2025. – T. 17. – №. 3. – С. 1087.

32 Linardatos P., Papastefanopoulos V., Kotsiantis S. Explainable ai: A review of machine learning interpretability methods //Entropy. – 2020. – T. 23. – №. 1. – С. 18.

33 Iglikova M. RECOMMENDATION ALGORITHMS FOR EDUCATIONAL PREFERENCES: A REVIEW: 1 Narxoz University, Almaty, Kazakhstan. 2 Satbayev University, Almaty, Kazakhstan //Academic Scientific Journal of Computer Science. – 2025. – №. 3. – С. 93–109-93–109.

34 Abdellah A. M., Karim A. M., Hamid S. Career recommendation system for scientific students based on ontologies //Advances in Science, Technology and Engineering Systems. – 2019. – T. 4. – №. 4. – С. 29-41.

35 Sheth V., Tripathi U., Sharma A. A comparative analysis of machine learning algorithms for classification purpose //Procedia Computer Science. – 2022. – T. 215. – С. 422-431.

36 Somvanshi S. et al. A survey on kolmogorov-arnold network //ACM Computing Surveys. – 2025. – T. 58. – №. 2. – С. 1-35.

37 Atalla S. et al. An intelligent recommendation system for automating academic advising based on curriculum analysis and performance modeling //Mathematics. – 2023. – T. 11. – №. 5. – С. 1098.

38 Ouyang F. et al. Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course //International Journal of Educational Technology in Higher Education. – 2023. – T. 20. – №. 1. – С. 4.

39 Lahoud C. et al. A comparative analysis of different recommender systems for university major and career domain guidance //Education and information technologies. – 2023. – T. 28. – №. 7. – С. 8733-8759.

40 Sarabdeen M., Alofaysan H. Investigating the impact of digital transformation on the labor market in the era of changing digital transformation dynamics in Saudi Arabia //Economies. – 2023. – T. 11. – №. 1. – С. 12.

41 McKinsey Global Institute. *The State of AI in 2023: Generative AI's Breakout Year*, 2023.

42 Alenezi M. Digital learning and digital institution in higher education //Education Sciences. – 2023. – T. 13. – №. 1. – С. 88.

43 Olaniyi O. O. et al. Dynamics of the digital workforce: Assessing the interplay and impact of AI, automation, and employment policies //Automation, and Employment Policies (April 24, 2024). – 2024.

44 El Malki A., Touate S. The Interplay Of Human, Organizational, Environmental, And Technological Factors In Business Intelligence Adoption: A Systematic Literature Review //Revista de Gestão Social e Ambiental. – 2024. – T. 18. – №. 2. – С. 1-12B.

45 Chakraborty S., Misra B., Dey N. AI-Empowered Knowledge Management in Education. – Berlin/Heidelberg, Germany : Springer, 2024.

46 Luan H. et al. Challenges and future directions of big data and artificial intelligence in education //Frontiers in psychology. – 2020. – Т. 11. – С. 580820.

47 Padma E. et al. Career guidance system for students using machine learning //Challenges in Information, Communication and Computing Technology. – CRC Press, 2025. – С. 666-671.

48 Kundu A., Bej T. AI policies in school education: a comparative study on China, Singapore, Finland, and the US //Journal of Science and Technology Policy Management. – 2025.

49 Ahmad K. et al. Data-driven artificial intelligence in education: A comprehensive review //IEEE Transactions on Learning Technologies. – 2023. – Т. 17. – С. 12-31.

50 Mac Fadden I., García-Alonso E. M., López Meneses E. Science Mapping of AI as an Educational Tool Exploring Digital Inequalities: A Sociological Perspective //Multimodal Technologies and Interaction. – 2024. – Т. 8. – №. 12. – С. 106.

51 Barlybaeva S. H., Ayapbergenova A. DIGITAL KAZAKHSTAN--AN INNOVATION WAY OF DEVELOPMENT //Herald of Journalism/Habaršy Žurnalistika Seriâsy. – 2025. – Т. 76. – №. 2. 147-154

52 Berdykulova G. et al. KAZAKHSTANI AND FOREIGN EXPERIENCE OF UNIVERSITY DIGITALIZATION //МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. – 2022. – Т. 3. – №. 3. – С. 48-57.

53 Masharipov J., Khasanov U. Central Asian Prism of The Belt and Road Initiative: Social Sustainability and Capacity Building Specifics //Journal of Central Asian Studies. – 2025. – Т. 23. – №. 2. – С. 46-58.

54 Kropachev P. et al. Information technologies and the future of education in the republic of kazakhstan //Scientific Journal of Astana IT University. – 2020. – С. 30-38.

55 Abazov R. Kazakhstan: Challenges and Strategies Towards Green Transformation and Sustainable development. – 2025.

56 Rebar A. L., Alfrey K. L., Gardner B. Theories of physical activity motivation //Essentials of exercise and sport psychology: An open access textbook. – 2021. – С. 15-36.

57 Al-Asmar A. A. et al. The predictive value of high school grade point average to academic achievement and career satisfaction of dental graduates //BMC oral health. – 2021. – Т. 21. – №. 1. – С. 300.

58 Savitz-Romer M., Rowan-Kenyon H. T. Noncognitive skills, college success, and career readiness: What matters and to whom? //About Campus. – 2020. – Т. 25. – №. 1. – С. 4-13.

59 Heinzel V. Coordination and cooperation work of creative teams during a virtual Hackathon for urban innovation in times of crisis. – 2024.

60 Arizmendi C. J. et al. Predicting student outcomes using digital logs of learning behaviors: Review, current standards, and suggestions for future work //Behavior research methods. – 2023. – Т. 55. – №. 6. – С. 3026-3054.

61 Roy P. Career Guidance: A Way of Life //Online Submission. – 2020. – Т. 19. – №. 39. – С. 22-31.

62 Iglesias F. et al. MDCGen: Multidimensional dataset generator for clustering //Journal of Classification. – 2019. – Т. 36. – №. 3. – С. 599-618.

63 Farhan M. Using empirical data to design and validate hybrid simulation models of human behavior in service operations : дис. – The University of Texas at Arlington, 2021.

64 Varpio L. et al. The distinctions between theory, theoretical framework, and conceptual framework //Academic medicine. – 2020. – Т. 95. – №. 7. – С. 989-994.

65 Miličević A. et al. Academic performance indicators for the hackathon learning approach–The case of the blockchain hackathon //Journal of Innovation & Knowledge. – 2024. – Т. 9. – №. 3. – С. 100501.

66 Chaudhuri A. K. et al. A multi-stage approach combining feature selection with machine learning techniques for higher prediction reliability and accuracy in cervical cancer diagnosis //International Journal of Intelligent Systems and Applications. – 2021. – Т. 14. – №. 5. – С. 46.

67 Bkheet S. A. et al. Comparative Performance of Gradient Boosting and Random Forest for Smart Home Device Classification. – 2025.

68 Karwa K. Navigating the job market: Tailored career advice for design students //International Journal of Emerging Business. – 2024. – Т. 23. – №. 2.

69 Caulfield T. R., Islam N. N., Chitale R. Liquid Adaptive AI: A Theoretical Framework for Continuously Self-Improving Artificial Intelligence //AI. – 2025. – Т. 6. – №. 8. – С. 186.

70 Michael J., Cohn A., Butcher J. R. Blockchain technology //The Journal. – 2018. – Т. 1. – №. 7. – С. 1-11.

71 Vaca-Rubio C. J. et al. Kolmogorov-arnold networks (kans) for time series analysis //arXiv preprint arXiv:2405.08790. – 2024.

72 Malykhin O. et al. Formation of top job skills of tomorrow among computer engineering and information technologies undergraduate students in the process of learning English //ENVIRONMENT. TECHNOLOGIES. RESOURCES. Proceedings of the International Scientific and Practical Conference. – 2021. – Т. 2. – С. 249-254.

73 Okereke G. K. O. et al. Enhancing Graduate Employability in Emerging Economies through Innovative Technologies and E-Commerce //Vocational and Technical Education Journal. – 2024. – Т. 4. – №. 1. – С. 116-123.

74 Zhao Yilin, He Feng, and Feng Ying. Research on the employment flow of college graduates and regional economic development: Analysis on the flow intensity of graduates from 137 colleges and universities. In 2021 International Conference on Education, Information Management and Service Science (EIMSS), pages 99–105. IEEE, 2021

75 Reynato Andal Gamboa, Satesh Narayana Namasivayam, Ramesh Singh, Chockalingam Aravind Vaithilingam, and Mohammad Hosseini Faouladi. Po attainment vs. employment status: A correlation study for taylor's university school of engineering. In 2018 IEEE 10th International Conference on Engineering Education (ICEED), pages 100–104. IEEE, 2018

76 Li J. et al. AI applications through the whole life cycle of material discovery //Matter. – 2020. – Т. 3. – №. 2. – С. 393-432.

77 Harry A. Role of AI in Education //Interdisciplinary Journal & Humanity (INJURITY). – 2023. – Т. 2. – №. 3. – С. 45-56. DOI: 10.31227/osf.io/z76qp

78 Chatterjee, P., & Chatterjee, P. Industry-academia collaboration for employability enhancement //Higher Education, Skills and Work-Based Learning. – 2019. – Т. 9. – №. 3. – С. 462-478. DOI: 10.1108/HESWBL-07-2018-0072

79 Kapil, V. Emerging trends in STEM education and employability enhancement //International Journal of STEM Education. – 2020. – Т. 7. – №. 1. – С. 18-29. DOI: 10.1186/s40594-020-00236-3

80 Tang, T. L. P., & Tang, T. Developing career adaptability through career guidance and counseling //Journal of Career Development. – 2022. – Т. 49. – №. 2. – С. 327-342. DOI: 10.1177/0894845320906555

81 Yu, H., & Wright, C. Addressing employability challenges in higher education: Curriculum innovation and employer engagement //Studies in Higher Education. – 2021. – Т. 46. – №. 3. – С. 603-617.DOI: 10.1080/03075079.2019.1655733

82 Bayerlein L., Jeske D. The potential of computer-mediated internships for higher education //International Journal of Educational Management. – 2018. – Т. 32. – №. 4. – С. 526-537.

83 Galbraith D., Mondal S. The Potential Power of Internships and the Impact on Career Preparation //Research in Higher Education Journal. – 2020. – Т. 38.

84 Urquía-Grande E., Perez Estebanez R. Bridging the gaps between higher education and the business world: internships in a faculty of economics and business //Education+ Training. – 2021. – Т. 63. – №. 3. – С. 490-509.

85 Laukkarinen M. Social media as a place to see and be seen: Exploring factors affecting job attainment via social media //The Information Society. – 2023. – Т. 39. – №. 4. – С. 199-212.

86 Hossain M. M. et al. Factors affecting business graduates' employability–empirical evidence using partial least squares (PLS) //Education+ Training. – 2020. – Т. 62. – №. 3. – С. 292-310.

87 Jakavonytė-Staškuvienė D., Žemgulienė A., Sakadolskis E. A. Cooperative learning issues in elementary education: A Lithuanian case study //Journal of education culture and society. – 2021. – Т. 12. – №. 1. – С. 445-468.

88 Tu F. What can we learn from longitudinal studies on the impacts of college internships //Center for College Workforce Transitions, University of Wisconsin Madison, January. – 2022.

89 Baert, S., Rotsaert, O., Verhaest, D., & Omey, E. Student internships and employment opportunities after graduation: A field experiment // Economics of Education Review. – 2021. – Т. 83. – С. 102141. DOI: 10.1016/j.econedurev.2021.102141

90 Choe C., Kim Y. B., Choi K. Do internships matter?: The impact of internship participation on employability //The Singapore Economic Review. – 2023. – С. 1-18.

91 Smith S. et al. The impact of work placement on graduate employment in computing: Outcomes from a UK-based study //International Journal of Work-Integrated Learning. – 2018. – Т. 19. – №. 4. – С. 359-369.

92 Lucero J. E. et al. Structure and characteristics for successful outcomes: a review of community college internships programs //Community College Journal of Research and Practice. – 2021. – Т. 45. – №. 2. – С. 103-116.

93 Pageni S. Job attitude and education attainment for professional identity //International Journal on Integrated Education. – 2020. – Т. 3. – №. 11. – С. 129-133.

94 Schall C. et al. The effect of business internships model and employment on enhancing the independence of young adults with significant impact from autism //Intellectual and Developmental Disabilities. – 2020. – Т. 58. – №. 4. – С. 301-313.

95 Hora M. T. et al. Closing the doors of opportunity: A field theoretic analysis of the prevalence and nature of obstacles to college internships //Teachers College Record. – 2021. – Т. 123. – №. 12. – С. 180-210.

96 Em S., Phann S. BOOK REVIEW: Cooperative and work-integrated education: History, present, and future issues in Asia //Cambodian Journal of Educational and Social Sciences (CJESS). – 2024. – Т. 1. – №. 1. – С. 104-110.

97 Martin A., Rees M., Edwards M. University-industry partnerships: Key success factors and challenges // Industry and Higher Education. – 2021. – Т. 35. – №. 3. – С. 267-278. DOI: 10.1177/0950422220963535

98 Cunningham J., Link A. N. Technology transfer in STEM education: Enhancing employability through innovation ecosystems // Journal of Technology Transfer. – 2021. – Т. 46. – С. 567-580. DOI: 10.1007/s10961-020-09817-y

99 Wang X., Zhang Z., Chen L. Cooperative education model in Chinese universities: Design, implementation, and outcomes // Journal of Vocational Education & Training. – 2019. – Т. 71. – №. 4. – С. 537-556. DOI: 10.1080/13636820.2018.1517124

100 Martin A. J., Rees M. Student Insights: The Added Value of Work-Integrated Learning //International Journal of Work-Integrated Learning. – 2019. – Т. 20. – №. 2. – С. 189-199.

101 Bayerlein L. The impact of prior work-experience on student learning outcomes in simulated internships //Journal of University Teaching and Learning Practice. – 2020. – Т. 17. – №. 4. – С. 1-19.

102 Hawthorne E., Fontaine I. An Analysis of the Relationship Between Education and Occupational Attainment //Progression Journal of Human Demography and Anthropology. – 2024. – С. 22-27.

103 Pandey P., Singh S., Pathak P. An exploratory study on factors contributing to job dissatisfaction of retail employees in India //Journal of Retailing and Consumer Services. – 2021. – Т. 61. – С. 102571.

104 Agnihotri S., Shiva A., Khan F. N. Investigating forms of graduate capital and their relationship to perceived employability: an application of PLS predict and IPMA //Higher Education, Skills and Work-Based Learning. – 2023. – Т. 13. – №. 1. – С. 179-197.

105 Mahon D. The role of graduate attributes in higher education. A review of the issues associated with graduate attributes and the case for their measurement //Interchange. – 2022. – Т. 53. – №. 3. – С. 353-369.

106 Gao J. et al. Prediction of sepsis mortality in ICU patients using machine learning methods //BMC Medical Informatics and Decision Making. – 2024. – Т. 24. – №. 1. – С. 228.

107 Kim J. K., Choo Y. J., Chang M. C. Prediction of motor function in stroke patients using machine learning algorithm: Development of practical models //Journal of Stroke and Cerebrovascular Diseases. – 2021. – Т. 30. – №. 8. – С. 105856.

108 Moore J. L., Felten P. Academic development in support of mentored undergraduate research and inquiry //International Journal for Academic Development. – 2018. – Т. 23. – №. 1. – С. 1-5.

109 Van der Kleij F. M., Lipnevich A. A. Student perceptions of assessment feedback: A critical scoping review and call for research //Educational assessment, evaluation and accountability. – 2021. – Т. 33. – №. 2. – С. 345-373.

110 Seifan M., Lal N., Berenjian A. Effect of undergraduate research on students' learning and engagement //International Journal of Mechanical Engineering Education. – 2022. – Т. 50. – №. 2. – С. 326-348.

111 Ward V. et al. Embedding researchers into organisations: a study of the features of embedded research initiatives //Evidence & Policy. – 2021. – Т. 17. – №. 4. – С. 593-614.

112 Krim J. S. et al. Models and impacts of science research experiences: A review of the literature of CUREs, UREs, and TREs //CBE—Life Sciences Education. – 2019. – Т. 18. – №. 4. – С. ar65.

113 Hochman G., Jin Y. Engaging Undergraduate Students in Research: From Cross-Discipline Programs to Published Articles //Applied Economics Teaching Resources (AETR). – 2025. – Т. 7. – №. 1. – С. 19-34.

114 Cakir M. Constructivist approaches to learning in science and their implications for science pedagogy: A literature review //International journal of environmental and science education. – 2022.

115 Ruth A. et al. Research supervisors and undergraduate students' perceived gains from undergraduate research experiences in the social sciences //International Journal of Inclusive Education. – 2025. – Т. 29. – №. 10. – С. 1669-1686.

116 Lu V. N. et al. Reconceptualization of graduate employability for career advancement //Graduate Employability across contexts: Perspectives, initiatives and outcomes. – Singapore : Springer Nature Singapore, 2022. – С. 367-393.

117 Heckert A. et al. Researchers, patients, and other stakeholders' perspectives on challenges to and strategies for engagement //Research Involvement and Engagement. – 2020. – Т. 6. – №. 1. – С. 60.

118 Otache I. Enhancing graduates' employability through polytechnic–industry collaboration //Industry and Higher Education. – 2022. – Т. 36. – №. 5. – С. 604-614.

119 Musa S., Nurhayati S., Boriboon G. The Effect of Internships on Graduates' Employability, Soft Skills, and Digital Competence //Educational Process: International Journal. – 2025. – Т. 17. – С. e2025306.

120 Jackson, D., & Collings, D. (2018). The influence of work-integrated learning and paid work during studies on graduate employment and underemployment. *Higher Education*, 76(3), 403-425.

121 Graham D., Tytler D. A Lesson for Us All: the making of the National Curriculum. – Routledge, 2018.

122 O'Connor K. Constructivism, curriculum and the knowledge question: tensions and challenges for higher education //Studies in Higher Education. – 2022. – Т. 47. – №. 2. – С. 412-422.

123 Konc A. What is the impact of disruptive innovative transformation? In economic education //International Journal of Scholars in Education. – 2025. – Т. 8. – №. 1. – С. 40-48.

124 Asiedu E. et al. Institutional Strategies and Graduate Employability Development Skills of Business Students: The Mediating Role of Graduate Capital Forms //ICRRD Quality Index Research Journal. – 2023. – Т. 4. – №. 3.

125 Fajaryati N. et al. The employability skills needed to face the demands of work in the future: Systematic literature reviews //Open Engineering. – 2020. – Т. 10. – №. 1. – С. 595-603.

126 Smith, C. (2018). Designing work-integrated learning placements that improve student employability: Six facets of alignment. *Higher Education Research & Development*, 37(4), 806-820.

127 Drysdale, M., McBeath, M., & Bohn, N. (2018). The impact of work-integrated learning on student work self-efficacy. *Education + Training*, 60(6), 529-540.

128 Yu L. et al. The antecedent role of proactive personality in career self-management //Journal of College Student Development. – 2021. – Т. 62. – №. 3. – С. 351-355.

129 Hawke L. D. et al. Embedding lived experience into mental health academic research organizations: critical reflections //Health Expectations. – 2022. – Т. 25. – №. 5. – С. 2299-2305.

130 Fergusson L., van der Laan L. Work+ learning: unpacking the agglomerated use of pedagogical terms //Journal of Work-Applied Management. – 2021. – Т. 13. – №. 2. – С. 302-314.

131 Volkova N. et al. Development and mastery of soft skills among graduate students //Ukrainian Journal of Educational Studies and Information Technology. – 2022. – Т. 10. – №. 4. – С. 1-14.

132 Prøitz T. S., Aasen P., Wermke W. From education policy to education practice: Unpacking the nexus. – Springer Nature, 2023. – С. 308.

133 Ho T. T. H. et al. Effects of career development learning on students' perceived employability: a longitudinal study //Higher Education. – 2023. – Т. 86. – №. 2. – С. 297-315.

134 Reinhard K., Wynder M., Kim W. S. Developing Best Practice for Cooperative and Work-Integrated Education: Lessons from Germany, Australia and South Korea //International Journal of Work-Integrated Learning. – 2020. – Т. 21. – №. 2. – С. 177-191.

135 Palmer S., Young K., Campbell M. Developing an Institutional Evaluation of the Impact of Work-Integrated Learning on Employability and Employment //International Journal of Work-Integrated Learning. – 2018. – Т. 19. – №. 4. – С. 371-383.

136 Clanchy K. et al. Describing pre-professional identity in higher education: a case study of exercise science students //Journal of University Teaching and Learning Practice. – 2021. – Т. 18. – №. 6. – С. 174-192.

137 Neroorkar S. A systematic review of measures of employability //Education+ Training. – 2022. – Т. 64. – №. 6. – С. 844-867.

138 Shaw T. J. et al. Knowing is half the battle: Assessments of both student perception and performance are necessary to successfully evaluate curricular transformation //PLoS One. – 2019. – Т. 14. – №. 1. – С. e0210030.

139 Rachmat A. et al. Design of distributed academic-record system based on blockchain //2019 International Conference on ICT for Smart Society (ICISS). – IEEE, 2019. – Т. 7. – С. 1-6.

140 Chen, G., Xu, B., Lu, M., & Chen, N. (2018). Exploring blockchain technology and its potential applications for education. *Smart Learning Environments*, 5(1), 1-10.

141 Jackson, D. (2018). Challenges and strategies for assessing student work-integrated learning: A literature review. *Journal of Teaching and Learning for Graduate Employability*, 9(1), 85-97.

142 Lee A. S. Supervisors' roles for newcomer adjustment: review of supervisors' impact on newcomer organizational socialization outcomes //European Journal of Training and Development. – 2024. – Т. 48. – №. 5/6. – С. 521-539.

143 Huang, R., Spector, J. M., & Yang, J. (2019). Educational technology research and development in the 21st century: Issues and trends. *Educational Technology Research and Development*, 67(2), 307-312.

144 Su, J., & Mo, L. (2018). Research on the application of big data technology in employment recommendation systems. *E3S Web of Conferences*, 214, 03013.

145 Alalawi K., Athauda R., Chiong R. An extended learning analytics framework integrating machine learning and pedagogical approaches for student performance prediction and intervention //International Journal of Artificial Intelligence in Education. – 2025. – Т. 35. – №. 3. – С. 1239-1287.

146 Yumna H. et al. Use of blockchain in education: a systematic literature review //Asian conference on intelligent information and database systems. – Springer, Cham, 2019. – С. 191-202.

147 Chen, Y., & Xu, Y. (2019). Exploring blockchain technology and its potential applications for higher education. *International Journal of Emerging Technologies in Learning (iJET)*, 14(13), 45-60.

148 Pinto, L. H., & Pereira, C. S. (2022). Using big data to improve educational management and learning analytics: A review. *Education Sciences*, 12(6), 412.

149 Gibson, D. C., & Ifenthaler, D. (2022). Digital credentials and blockchain in education: Trends, issues, and applications. *Educational Technology Research and Development*, 70(3), 1049-1068.

150 Zheng, Z., Xie, S., Dai, H. N., Chen, X., & Wang, H. (2018). Blockchain challenges and opportunities: A survey. *International Journal of Web and Grid Services*, 14(4), 352-375.

151 Casino, F., Dasaklis, T. K., & Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification, and open issues. *Telematics and Informatics*, 36, 55-81.

152 Berentsen A. Aleksander berentsen recommends “bitcoin: a peer-to-peer electronic cash system” by Satoshi Nakamoto //21st Century economics: Economic ideas you should read and remember. – Cham : Springer International Publishing, 2019. – С. 7-8.

153 Lu Y. The blockchain: State-of-the-art and research challenges //Journal of Industrial Information Integration. – 2019. – Т. 15. – С. 80-90.

154 Kouhizadeh, M., & Sarkis, J. (2018). Blockchain practices, potentials, and perspectives in greening supply chains. *Sustainability*, 10(10), 3652.

155 Sadasiuvam G. S. A critical review on using blockchain technology in education domain //Handbook of deep learning in biomedical engineering. – 2021. – С. 85-121.

156 Xu, M., Chen, X., & Kou, G. (2019). A systematic review of blockchain. *Financial Innovation*, 5(1), 27.

157 Ocheja P. et al. Blockchain in education: A systematic review and practical case studies //IEEE Access. – 2022. – Т. 10. – С. 99525-99540.

158 Turkanović, M., Hölbl, M., Košič, K., Heričko, M., & Kamišalić, A. (2018). EduCTX: A blockchain-based higher education credit platform. *IEEE Access*, 6, 51113-51127.

159 Chen, Y., & Bellavitis, C. (2020). Blockchain disruption and decentralized finance: The rise of decentralized business models. *Journal of Business Venturing Insights*, 13, e00151.

160 Driscoll A. Assessment and accreditation: Productive partnerships //Taking ownership of accreditation. – Routledge, 2023. – С. 1-18.

161 Steiu M. F. Blockchain in education: Opportunities, applications, and challenges //First Monday. – 2020.

162 Hasan, H. R., & Salah, K. (2018). Blockchain-based solution for proof of delivery of physical assets. *Journal of Computer Networks and Communications*, 2018.

163 Kouhizadeh, M., & Sarkis, J. (2018). Blockchain practices, potentials, and perspectives in greening supply chains. *Sustainability*, 10(10), 3652

164 Tan, B., Wu, Q., Zhu, H., & Cheng, X. (2018). Blockchain-Based Smart Contracts: A Comprehensive Tutorial. *IEEE Access*, 6, 43879-43896.

165 Wang, Y., Han, J., & Beynon-Davies, P. (2019). Understanding blockchain technology for future supply chains: A systematic literature review and research agenda. *Supply Chain Management*, 24(1), 62-84.

166 Dikilitaş Y., Toka K. O., Sayar A. Current research areas in blockchain //Avrupa Bilim ve Teknoloji Dergisi. – 2021. – №. 26. – С. 488-492.

167 Zwitter, A., & Hazenberg, J. (2020). Decentralized governance and digital humanitarianism: Blockchain technology in the humanitarian sector. *International Review of the Red Cross*, 102(913), 137-158.

168 Xu, X., Weber, I., & Staples, M. (2019). Architecture for blockchain applications. Springer.

169 Li, X., & Jiang, P. (2019). Blockchain for distributed manufacturing: A comprehensive review. IEEE Transactions on Industrial Informatics, 15(6), 3602-3610.

170 Wüst, K., & Gervais, A. (2018). Do you need a blockchain? 2018 Crypto Valley Conference on Blockchain Technology (CVCBT), 45-54.

171 Sharma U. Blockchain: Blueprint For A New Economy By Melanie Swan. – 2019.

172 Bjelobaba G. et al. Blockchain technologies and digitalization in function of student work evaluation //Sustainability. – 2022. – Т. 14. – №. 9. – С. 5333.

173 Khez S. et al. Blockchain technology in healthcare: A comprehensive review and directions for future research //Applied sciences. – 2019. – Т. 9. – №. 9. – С. 1736.

174 Werbach, K. (2018). The blockchain and the new architecture of trust. MIT Press.

175 Kim, H. M., & Laskowski, M. (2018). Towards an ontology-driven blockchain design for supply-chain provenance. Intelligent Systems in Accounting, Finance and Management, 25(1), 18-27.

176 Tišma S. et al. Education For Practice: Results Of The Analysis Of Attitudes Of Students And Employers On The Importance Of Internship In Higher Education Institutions In Engineering //INTED2021 Proceedings. – IATED, 2021. – С. 5881-5886.

177 Zehr S. M., Korte R. Student internship experiences: learning about the workplace //Education+ Training. – 2020. – Т. 62. – №. 3. – С. 311-324.

178 Silva P. et al. The million-dollar question: can internships boost employment? //Studies in Higher Education. – 2018. – Т. 43. – №. 1. – С. 2-21.

179 Pant S., Kumar V. Bittrusty: A bitcoin incentivized peer-to-peer file sharing system //2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS). – IEEE, 2018. – С. 148-155.

180 Figueredo, C., & Paiva, A. (2019). Skill matching and recommendation in internship programs using natural language processing. Journal of Educational Computing Research, 57(6), 1449-1473.

181 Tsirkas K., Chytiri A. P., Bouranta N. The gap in soft skills perceptions: A dyadic analysis //Education+ Training. – 2020. – Т. 62. – №. 4. – С. 357-377.

182 Makani S. et al. A survey of blockchain applications in sustainable and smart cities //Cluster Computing. – 2022. – Т. 25. – №. 6. – С. 3915-3936.

183 Chen, X., & Zhang, Y. (2020). A comprehensive competency assessment model for internship candidate evaluation in higher education. Journal of Vocational Education & Training, 72(4), 563-583.

184 Allen J. J., Parker A., Ogles B. M. A review of the facilitative interpersonal skills performance task and rating method //Clinical Psychology: Science and Practice. – 2024. – Т. 31. – №. 2. – С. 211.

185 Saxena S. et al. Role of Communication Skills: A Review //World Journal of English Language. – 2022. – Т. 12. – №. 3. – С. 18-24.

186 Driscoll A. Assessment and accreditation: Productive partnerships //Taking ownership of accreditation. – Routledge, 2023. – С. 1-18.

187 Sánchez-Gómez, M., & Martín-Caro, J. (2022). Enhancing recruitment processes with AI-driven candidate ranking: An empirical study. Journal of Applied Artificial Intelligence, 36(1), 1-19.

188 Van Iddekinge, C. H., Arnold, J. D., Frieder, R. E., & Roth, P. L. (2019). A meta-analysis of the criterion-related validity of prehire work experience. Personnel Psychology, 72(4), 571-598.

189 O'Donnell E., O'Donnell L. Challenges in developing adaptive educational hypermedia systems //Encyclopedia of Information Science and Technology, Fourth Edition. – 2018. – C. 2380-2391.

190 Wesolowski B. C., Wind S. A. VALIDITY, RELIABILITY //The Oxford Handbook of Assessment Policy and Practice in Music Education, Volume 1. – 2019. – C. 437.

191 Zupan, M., & Časar, K. (2020). Blockchain technology for secure and transparent educational processes. Journal of Theoretical and Applied Electronic Commerce Research, 15(3), 1-19.

192 Hawlitschek, F., Notheisen, B., & Teubner, T. (2018). The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy. Electronic Commerce Research and Applications, 29, 50-63.

193 Saadatmand M., Daim T. Blockchain technology through the lens of disruptive innovation theory //2019 IEEE Technology & Engineering Management Conference (TEMSCON). – IEEE, 2019. – C. 1-6.

194 Pan J. Kolmogorov-Arnold networks for spectrum sensing and modulation classification : дис. – University of British Columbia, 2024.

195 Yamak P. T. et al. Kolmogorov-Arnold networks for time series forecasting: a comprehensive review //Cluster Computing. – 2025. – T. 28. – №. 14. – C. 929.

196 Gao J. Unsupervised Visual Domain Adaptation by Self-guided Learning : дис. – Queen's University Belfast, 2022.

197 Liu Z. et al. Kan: Kolmogorov-arnold networks //arXiv preprint arXiv:2404.19756. – 2024.

198 Pan J. Kolmogorov-Arnold networks for spectrum sensing and modulation classification : дис. – University of British Columbia, 2024.

199 Adeyemo Z. K. et al. Development of a Path-loss Prediction Model Using Adaptive Neuro-fuzzy Inference System //International Journal of Wireless and Microwave Technologies. – 2019. – T. 9. – №. 6. – C. 40-53.Adeyemo Z. K. et al. Development of a Path-loss Prediction Model Using Adaptive Neuro-fuzzy Inference System //International Journal of Wireless and Microwave Technologies. – 2019. – T. 9. – №. 6. – C. 40-53.

200 Liu, J., Zhang, C., & Ma, X. (2020). Career path recommendation for college students based on improved collaborative filtering algorithm. Applied Sciences, 10(19), 6956.

201 Wang Y. et al. Mining campus big data: Prediction of career choice using interpretable machine learning method //Mathematics. – 2022. – T. 10. – №. 8. – C. 1289.

202 Ji T., Hou Y., Zhang D. A comprehensive survey on kolmogorov arnold networks (kan) //arXiv preprint arXiv:2407.11075. – 2024.

203 Tran, T. T., & Soejarto, A. (2022). Technology-enabled internship monitoring systems in higher education: Enhancing student engagement and performance evaluation. International Journal of Educational Technology, 23(2), 150-165.

204 Shi J. C., Jin D. Q. A New KAN-Based Architecture for Optimization: J.-C. Shi, D.-Q. Jin //Journal of the Operations Research Society of China. – 2025. – С. 1-16.

205 Succi, C., & Canovi, M. (2020). Soft skills in higher education: Importance and improvement perceived by students and graduates. Education + Training, 62(3), 290-303.

206 Jackson, D. (2018). Challenges and strategies for assessing graduate employability skills. Journal of Higher Education Policy and Management, 40(1), 32-44.

207 Javed, A., Mustapha, A., & Ghani, I. (2021). Hybrid recommender systems for job matching: A systematic literature review. Journal of King Saud University - Computer and Information Sciences, 33(7), 832-846.

208 Sun, Y., Yin, C., & Bi, J. (2020). A hybrid recommender system for career path planning in MOOCs. Journal of Educational Computing Research, 58(2), 277-302.

209 Wang, H., Liu, J., & Ma, J. (2022). Context-aware job recommendation for college graduates using multi-feature similarity analysis. Journal of Intelligent & Fuzzy Systems, 42(4), 3545-3555.

210 Zhang, Y., & Chen, J. (2021). Personalized job recommendation based on user profiling and context-aware filtering. Knowledge-Based Systems, 222, 107035.

211 Ude N. G., Yskandar H., Graham R. C. A comprehensive state-of-the-art survey on the transmission network expansion planning optimization algorithms //IEEE Access. – 2019. – Т. 7. – С. 123158-123181.

212 De Ruijt C., Bhulai S. Job recommender systems: A review //arXiv preprint arXiv:2111.13576. – 2021.

213 Liu, Y., Hu, Z., & Miao, J. (2019). Job recommendation system for university students using collaborative filtering. Journal of Ambient Intelligence and Humanized Computing, 10(12), 4697-4710.

214 Mahmoud, M. A., & Salim, N. (2020). An enhanced job recommendation system for e-recruitment: A hybrid approach. International Journal of Information Management, 52, 102098.

215 Tricomi, E., & De Rosa, G. (2021). Using AI and machine learning in higher education career services: Personalized recommendations for students. Education and Information Technologies, 26(4), 4197-4215.

216 Yu, Y., & Zhang, J. (2021). Data-driven job recommendation using profile augmentation and hybrid filtering. Information Processing & Management, 58(4), 102542.

217 Sahoo, A., & Behera, H. S. (2019). Career recommender systems: Recent developments and future research directions. Artificial Intelligence Review, 52(3), 1479-1502.

218 Hodovychenko M. A., Gorbatenko A. A. Recommender systems: models, challenges and opportunities //Herald of Advanced Information Technology. – 2023. – Т. 4. – №. 6. – С. 308-319.

219 Chen, L., Wu, W., & He, Y. (2020). An improved job recommendation system based on deep learning techniques. Information, 11(4), 225.

220 Sinha, N., & Thakur, M. (2021). Smart career guidance system using hybrid recommender framework. Journal of Ambient Intelligence and Humanized Computing, 12(6), 6897-6911.

221 Hsu, C. C., & Chen, M. Y. (2020). A personalized career recommendation system based on learning analytics. Journal of Educational Computing Research, 58(5), 880-903.

222 Liu, J., Zhang, C., & Ma, X. (2020). Career path recommendation for college students based on improved collaborative filtering algorithm. Applied Sciences, 10(19), 6956.

223 Chatfield C., Xing H. The analysis of time series: an introduction with R. – Chapman and hall/CRC, 2019.

224 Pan J. Kolmogorov-Arnold networks for spectrum sensing and modulation classification : дис. – University of British Columbia, 2024.

225 Tanveer M. S. et al. Deep-silicon photon-counting x-ray projection denoising through reinforcement learning //Journal of X-Ray Science and Technology. – 2024. – Т. 32. – №. 2. – С. 173-205.

226 Miraftabzadeh S. M. et al. Deep learning in power systems: A bibliometric analysis and future trends //IEEE Access. – 2024.

227 Liu Z. et al. Kan: Kolmogorov-arnold networks //arXiv preprint arXiv:2404.19756. – 2024.

228 Tran M. K. et al. Python-based scikit-learn machine learning models for thermal and electrical performance prediction of high-capacity lithium-ion battery //International Journal of Energy Research. – 2022. – Т. 46. – №. 2. – С. 786-794.

229 Hussein I. O. et al. Development of a Pathloss Prediction Model Using Adaptive Neuro Fuzzy Inference System (ANFIS) for Enhanced Wireless

Performance for 5G Network //Adeleke University Journal of Engineering and Technology. – 2025. – T. 8. – №. 1. – C. 197-209.

230 Carrington A. M. et al. Deep ROC analysis and AUC as balanced average accuracy to improve model selection, understanding and interpretation //arXiv preprint arXiv:2103.11357. – 2021.

APPENDIX A
Act of implementation

«SDU University» мекемесі

0900, Алматы облысы, Қарасай ауданы,
Қаскелең қаласы, Абылай хан көшесі 1/1
тел.:+7727 307-95-60, факс: 307-95-58

Қосымша 1

Учреждение

«SDU University

040900, Алматинская область,
Карасайский район, г. Қаскелен қаласы,
ул. Абылай хана 1/1 тел.:+7727 307-95-60,
факс: 307-95-58

БЕКІТЕМІН
Оқу істер жөніндегі проректор
__________ Богданчиков А.В.
«19» наурыз 2024 ж.

Оқу-әдістемелік және ғылыми жұмыс нәтижелерін оқу процессіне енгізу туралы АКТ

1. **Жұмыс енгізілетін мекеме атауы:** «SDU University» мекемесі
2. **Ұсыныс атауы:** betacareer.sdu.edu.kz және hr.betacareer.myte.me сайттарын SDU университетінің оқу процессіне енгізу
3. **Оқу-әдістемелік құрал және ғылыми жұмыс нәтижесін оқу процессіне енгізу формасы:** веб платформа енгізу формасы күндізгі және оффлайн
4. **Енгізу аясы:** Білім берудегі IT (IT in education)
5. **Апробация мерзімі:** betacareer.sdu.edu.kz сайты бойынша 3 жыл, hr.betacareer.myte.m сайты бойынша 7 ай
6. **Апробация нәтижелері:** оңды және қажетті деп танылды
7. **Енгізуге жауапты:** Берліқожа Бауыржан Әсетұлы
8. **Енгізу тиімділігі:** практика жүйесін автоматтандыру
9. **Енгізудің қажеттілігі:** жоғарғы

Факультет деканы: PhD, Ассистент Профессор Ахмедов Рамиз

Кафедра меңгерушісі: PhD, Ассистент Профессор Мұқаш Жанар

Оқу-әдістемелік Кеңесінің №, «8» 19.03.2024 ж. хаттамасымен

APPENDIX B
Code view of the platform integration

```
# ====== install deps (first run) ======
# pip install torch numpy pandas scikit-learn
imbalanced-learn pykan==0.2.5

import os
import numpy as np
import pandas as pd
from sklearn.model_selection import StratifiedKFold,
train_test_split
from sklearn.preprocessing import OneHotEncoder,
StandardScaler, LabelEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.metrics import accuracy_score,
precision_score, recall_score, f1_score
from imblearn.over_sampling import SMOTE
from imblearn.pipeline import Pipeline as ImbPipeline
import torch
import torch.nn as nn
import torch.nn.functional as F

# ---- KAN import (from pykan) ----
# Docs: https://github.com/KindXiaoming/pykan
(examples API)
from kan import KAN

# ----------------------------
# 1) Load data
# ----------------------------
# CSV should contain feature columns + a target column
named 'label'
df = pd.read_csv("students.csv")  # <-- replace with
your path
assert "label" in df.columns, "Expected target column
'label' not found."

# Identify feature types
target_col = "label"
feature_cols = [c for c in df.columns if c !=
target_col]
```

```
# Heuristic: object/boolean → categorical; numbers → numeric
cat_cols = [c for c in feature_cols if df[c].dtype == "object" or df[c].dtype == "bool"]
num_cols = [c for c in feature_cols if c not in cat_cols]

# Encode labels to integers
le = LabelEncoder()
y = le.fit_transform(df[target_col].astype(str))
n_classes = len(le.classes_)

# Hold-out test split to report final metrics after CV/tuning
X_train_full, X_test, y_train_full, y_test = train_test_split(
    df[feature_cols], y, test_size=0.2, stratify=y, random_state=42
)

# ------------------------------
# 2) Build preprocessing
# ------------------------------
numeric_tf = Pipeline(steps=[("scaler", StandardScaler())])
categorical_tf = Pipeline(steps=[("ohe", OneHotEncoder(handle_unknown="ignore", sparse_output=False))])

preprocessor = ColumnTransformer(
    transformers=[
        ("num", numeric_tf, num_cols),
        ("cat", categorical_tf, cat_cols),
    ],
    remainder="drop",
)

# ------------------------------
# 3) Wrap KAN into a torch Module for classification
# ------------------------------
class KANClassifier(nn.Module):
    """
    A thin wrapper around pykan.KAN for multi-class classification.
    width: [input_dim, hidden_dim, output_dim]
```

```
    grid, k control spline grid/degree; tweak for your
data.
    """
    def _init_(self, input_dim, output_dim, hidden=32,
grid=5, k=3, noise_scale=0.0):
        super()._init_()
        # Two-layer KAN (input -> hidden -> output)
        self.kan = KAN(width=[input_dim, hidden,
output_dim], grid=grid, k=k, noise_scale=noise_scale)

    def forward(self, x):
        # pykan expects float tensor
        logits = self.kan(x)  # shape: [N, output_dim]
        return logits

def train_one_model(X_tr, y_tr, X_val, y_val,
input_dim, n_classes,
                    lr=1e-3, epochs=200, patience=20,
hidden=32, grid=5, k=3, device="cpu"):
    model = KANClassifier(input_dim, n_classes,
hidden=hidden, grid=grid, k=k).to(device)
    optimizer = torch.optim.Adam(model.parameters(),
lr=lr, weight_decay=1e-4)
    criterion = nn.CrossEntropyLoss()

    X_tr_t = torch.tensor(X_tr, dtype=torch.float32,
device=device)
    y_tr_t = torch.tensor(y_tr, dtype=torch.long,
device=device)
    X_val_t = torch.tensor(X_val, dtype=torch.float32,
device=device)
    y_val_t = torch.tensor(y_val, dtype=torch.long,
device=device)

    best_val = float("inf")
    best_state = None
    patience_cnt = 0

    for epoch in range(1, epochs + 1):
        model.train()
        optimizer.zero_grad()
        logits = model(X_tr_t)
        loss = criterion(logits, y_tr_t)
        loss.backward()
        optimizer.step()
```

```
        # Validation
        model.eval()
        with torch.no_grad():
            val_logits = model(X_val_t)
            val_loss = criterion(val_logits,
y_val_t).item()

        if val_loss < best_val - 1e-6:
            best_val = val_loss
            best_state = {k: v.detach().cpu().clone()
for k, v in model.state_dict().items()}
            patience_cnt = 0
        else:
            patience_cnt += 1

        if patience_cnt >= patience:
            break

    # Load best state
    if best_state is not None:
        model.load_state_dict(best_state)
    return model

# ----------------------------
# 4) Stratified 10-fold CV on train_full to choose KAN
size (hidden)
# ----------------------------
skf = StratifiedKFold(n_splits=10, shuffle=True,
random_state=42)

# Candidate hidden sizes (you can add 64, 128, etc.)
hidden_candidates = [16, 32, 48]
cv_scores = {}

for hidden_dim in hidden_candidates:
    fold_acc = []
    for tr_idx, val_idx in skf.split(X_train_full,
y_train_full):
        X_tr_raw = X_train_full.iloc[tr_idx].copy()
        y_tr = y_train_full[tr_idx]
        X_val_raw = X_train_full.iloc[val_idx].copy()
        y_val = y_train_full[val_idx]

        # Build pipeline: preprocess -> SMOTE ->
```

```
(later, torch model)
        # SMOTE should only be applied on numeric
space; we apply it AFTER preprocessing.
        pp = preprocessor.fit(X_tr_raw)
        X_tr_pp = pp.transform(X_tr_raw)
        X_val_pp = pp.transform(X_val_raw)

        # Apply SMOTE on training fold (not on
validation)
        smote = SMOTE(random_state=42)
        X_tr_bal, y_tr_bal =
smote.fit_resample(X_tr_pp, y_tr)
        input_dim = X_tr_bal.shape[1]

        # Train KAN
        device = "cuda" if torch.cuda.is_available()
else "cpu"
        model = train_one_model(
            X_tr=X_tr_bal, y_tr=y_tr_bal,
            X_val=X_val_pp, y_val=y_val,
            input_dim=input_dim, n_classes=n_classes,
            lr=1e-3, epochs=250, patience=25,
            hidden=hidden_dim, grid=5, k=3,
device=device
        )

        # Evaluate fold accuracy
        model.eval()
        with torch.no_grad():
            val_logits = model(torch.tensor(X_val_pp,
dtype=torch.float32, device=device))
            y_val_pred =
val_logits.argmax(dim=1).cpu().numpy()
        acc = accuracy_score(y_val, y_val_pred)
        fold_acc.append(acc)

    cv_scores[hidden_dim] = (np.mean(fold_acc),
np.std(fold_acc))

print("CV Accuracies by hidden size:")
for h, (m, s) in cv_scores.items():
    print(f"  hidden={h}: mean={m:.4f}, std={s:.4f}")

# Pick best hidden size
best_hidden = max(cv_scores.items(), key=lambda kv:
```

```
kv[1][0])[0]
print(f"Selected hidden size: {best_hidden}")

# ----------------------------
# 5) Fit on full training set with best hidden,
evaluate on test set
# ----------------------------
# Preprocess using train split only
pp_final = preprocessor.fit(X_train_full)
X_tr_pp = pp_final.transform(X_train_full)
X_te_pp = pp_final.transform(X_test)

# Balance training data with SMOTE
smote = SMOTE(random_state=42)
X_tr_bal, y_tr_bal = smote.fit_resample(X_tr_pp,
y_train_full)
input_dim = X_tr_bal.shape[1]

# Train final KAN
device = "cuda" if torch.cuda.is_available() else
"cpu"
final_model = train_one_model(
   X_tr=X_tr_bal, y_tr=y_tr_bal,
   X_val=X_te_pp, y_val=y_test,  # we monitor test
loss for convenience; you can split a val set instead.
   input_dim=input_dim, n_classes=n_classes,
   lr=1e-3, epochs=300, patience=30,
   hidden=best_hidden, grid=5, k=3, device=device
)

# Evaluate
final_model.eval()
with torch.no_grad():
   logits = final_model(torch.tensor(X_te_pp,
dtype=torch.float32, device=device))
   y_pred = logits.argmax(dim=1).cpu().numpy()

acc = accuracy_score(y_test, y_pred)
prec = precision_score(y_test, y_pred,
average="macro", zero_division=0)
rec = recall_score(y_test, y_pred, average="macro",
zero_division=0)
f1 = f1_score(y_test, y_pred, average="macro",
zero_division=0)
```

```
print("\n=== Test Metrics (KAN) ===")
print(f"Accuracy:  {acc:.4f}")
print(f"Precision: {prec:.4f} (macro)")
print(f"Recall:    {rec:.4f} (macro)")
print(f"F1-score:  {f1:.4f} (macro)")

# You can also inverse-transform predictions to label
names if needed:
labels_pred = le.inverse_transform(y_pred)
labels_true = le.inverse_transform(y_test)
# Save predictions for analysis
out = pd.DataFrame({"y_true": labels_true, "y_pred":
labels_pred})
out.to_csv("kan_test_predictions.csv", index=False)
print("Saved: kan_test_predictions.csv")
```