Ministry of Science and Higher Education of the Republic of Kazakhstan

SDU University

Zhadyra Zhalgassova

Personalized Career-Path Recommender System for STEM Students

THESIS

Presented in Partial Fulfilment for the

Degree of Master of Technical Science in Computer Science

(degree code: 7M06102)

Department of Computer Science

Faculty of Engineering and Natural Sciences

Supervisor: **Ualikhan Sadyk**

Kaskelen, June 2024

SDU University
Faculty of Engineering and Natural Sciences
Department of Computer Science

Dean of Faculty of Engineering and Natural Sciences

Assistant Professor, PhD Akhmedov Ramis

____

« _____» _________________________________ 2024

Topic of the thesis:
Personalized Career-Path Recommender System for STEM Students

Thesis submitted as part of the requirements for the award of the MSc in "7M06102 - Computer Science", SDU University

Head of Department Zhanar Mukash

Academic Supervisor Ualikhan Sadyk

Master student Zhadyra Zhalgassova

Kaskelen, 2024

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Zhadyra Zhalgassova

June 2024

Acknowledgements

To begin, I express my gratitude to our creator, the almighty and infinitely compassionate Allah, for the opportunity to write this work and for blessing me with intelligence, knowledge, health, and wonderful people around me. I deeply appreciate my supervisor, Ualikhan Sadyk, for his help, support, and guidance throughout my study. His expertise, insightful feedback, and dedication to my academic development have significantly contributed to the completion of this paper. I extend my heartfelt thanks to the participants in this study for their willingness to share their experiences. Their involvement has been crucial in gathering the data necessary to answer the study questions and obtain valuable insights. I also thank my coworkers and peers for the engaging conversations, valuable feedback, and ongoing support. Their contributions and camaraderie have fostered a collaborative environment that has greatly influenced my academic journey. Lastly, I am profoundly grateful to my family for their unwavering love, support, and confidence in me. Their encouragement and understanding have provided me with the strength to navigate the challenges of this journey.

Dedication

This dissertation is devoted to my family, whose steadfast support and encouragement have driven my pursuit of higher education. I am forever thankful to my parents for their significant sacrifices in providing me with every opportunity for success. Their love, guidance, and belief in me have shaped who I am today. This dissertation is also dedicated to my friends, who have been there for me through the highs and lows of this academic pursuit. Their friendship, humor, and consistent support have added balance and joy to my life, reminding me of the importance of appreciating life beyond academia. Finally, I dedicate this dissertation to all the teachers and mentors who have inspired and guided me throughout my academic journey. Their knowledge, passion, and dedication to education have motivated me to make a difference in my chosen field.

Abstract

This dissertation introduces a Personalized Career-Path Recommender System (PCRS) designed to help high school students in Kazakhstan, particularly those interested in STEM (Science, Technology, Engineering, and Mathematics) fields. The system uses the Myers-Briggs Type Indicator (MBTI) personality types and students' academic performance to offer personalized recommendations for university specializations. The research addresses the common challenges faced by students, such as high dropout rates and frequent changes in majors, often due to the lack of structured career guidance. To tackle these issues, the study collected a variety of data, including students' demographics, academic records, and personal attributes, as well as detailed profiles of university majors. Advanced machine learning techniques, including content-based filtering, collaborative filtering, fuzzy logic, and hybrid approaches, were used to process this data and generate accurate recommendations. The effectiveness of the PCRS was tested with real data from students at SDU University. The results show that the system can provide relevant and personalized career guidance, significantly improving students' decision-making processes and satisfaction with their chosen specializations. By combining MBTI personality assessments with academic performance data, this research offers a fresh approach to educational technology and career counseling. The insights and methods developed in this study can be adapted for use in other regions facing similar challenges, ultimately helping more students make informed and satisfying career choices.

Keywords: Personalized Education, Career Guidance, STEM Education, Machine Learning, Recommender Systems, MBTI, Career Path Decision.

Аңдатпа

Бұл диссертация Қазақстандағы жоғары сынып оқушыларына, әсіресе STEM (ғылым, технология, инженерия және математика) салаларына қызығушылық танытатындарға арналған Жеке Кәсіби Жолды Ұсыну Жүйесін (PCRS) ұсынады. Жүйе студенттердің Майерс-Бриггс Тип Индикаторы (MBTI) бойынша жеке тұлға түрлерін және академиялық жетістіктерін пайдаланып, университет мамандықтары бойынша жеке ұсыныстар береді. Зерттеу студенттер жиі кездесетін жоғары деңгейдегі оқудан шығу және мамандықтарды жиі өзгерту сияқты мәселелерді шешуге бағытталған, бұл көбінесе құрылымдалған кәсіби кеңес берудің жоқтығынан туындайды. Осы мәселелерді шешу үшін студенттердің демографиялық деректері, академиялық жазбалары және жеке ерекшеліктері, сондай-ақ университет мамандықтарының егжей-тегжейлі профильдері сияқты әртүрлі мәліметтер жиналды. Бұл мәліметтерді өңдеу және нақты ұсыныстар жасау үшін контенттік сүзгілеу, коллаборативтік сүзгілеу, бұлыңғыр логика және аралас тәсілдерді қамтитын заманауи машиналық оқыту әдістері қолданылды. PCRS тиімділігі СДУ Университетінің студенттерінің нақты деректерімен тексерілді. Нәтижелер көрсеткендей, жүйе студенттердің шешім қабылдау процестерін және таңдаған мамандықтарына қанағаттануын едәуір жақсарта отырып, өзекті және жеке кәсіби кеңестер бере алады. MBTI бойынша жеке тұлға түрлерін академиялық жетістіктер деректерімен біріктіре отырып, бұл зерттеу білім беру технологиялары мен кәсіби кеңес беруде жаңа тәсілді ұсынады. Бұл зерттеуде әзірленген түсініктер мен әдістемелер ұқсас қиындықтарға тап болған басқа аймақтарда қолдануға бейімделуі мүмкін, нәтижесінде студенттерге саналы және қанағаттанарлық кәсіби таңдау жасауға көмектеседі.

Аннотация

Данная диссертация представляет Персонализированную Систему Рекомендаций по Карьерному Пути (PCRS), предназначенную для старшеклассников Казахстана, особенно тех, кто интересуется STEM (наука, технологии, инженерия и математика). Система использует типы личности по индикатору Майерс-Бриггс (MBTI) и академическую успеваемость студентов для предоставления персонализированных рекомендаций по выбору университетских специальностей. Исследование направлено на решение распространенных проблем, с которыми сталкиваются студенты, таких как высокий уровень отсева и частые изменения специальностей, часто возникающие из-за отсутствия структурированного карьерного консультирования. Для решения этих проблем было собрано разнообразие данных, включая демографические данные студентов, академические записи и личные характеристики, а также детальные профили университетских специальностей. Для обработки этих данных и создания точных рекомендаций использовались передовые методы машинного обучения, включая контентные фильтрации, коллаборативные фильтрации, нечеткую логику и гибридные подходы. Эффективность PCRS была протестирована на реальных данных студентов Университета СДУ. Результаты показали, что система способна предоставлять релевантные и персонализированные рекомендации по карьерному пути, значительно улучшая процесс принятия решений и удовлетворенность студентов выбранными специальностями. Объединяя оценки личностных типов по MBTI с данными об академической успеваемости, это исследование предлагает новый подход к образовательным технологиям и карьерному консультированию. Инсайты и методологии, разработанные в этом исследовании, могут быть адаптированы для использования в других регионах, сталкивающихся с аналогичными проблемами, что в конечном итоге поможет большему количеству студентов делать информированные и удовлетворяющие их карьерные выборы.

Abbreviations

MBTI - Myers-Briggs Type Indicator

PCRS - Personalized Career-Path Recommender System

STEM - Science, Technology, Engineering, and Mathematics

ML - Machine Learning

AI - Artificial Intelligence

CBF - Content-Based Filtering

CF - Collaborative Filtering

Table of Contents

Declaration **i**

Acknowledgements **ii**

Dedication **iii**

Abstract **iv**

Андатпа **v**

List of Abbreviations **vii**

1 Introduction **1**

2 Literature Review **5**
- 2.1 Recommender Systems in Education . . . 5
- 2.2 Course Recommender Systems . . . 6
- 2.3 Personalized Career-Path Recommender Systems . . . 8
- 2.4 MBTI Personality Types . . . 9
  - 2.4.1 MBTI in Career Guidance . . . 9
  - 2.4.2 MBTI and Academic Performance . . . 10
  - 2.4.3 MBTI in STEM Education . . . 11
- 2.5 Fuzzy logic in recommendation systems . . . 12
- 2.6 Implementation in Developing Countries . . . 14

3 Methodology and Results **16**
- 3.1 Methodology: Data Description . . . 16
  - 3.1.1 User Data . . . 16
  - 3.1.2 Item (Major) Data . . . 17
  - 3.1.3 Interaction Data . . . 17
  - 3.1.4 Data Summary . . . 17
- 3.2 Methodology: Data Cleaning Process . . . 19
  - 3.2.1 User Data Cleaning . . . 19
  - 3.2.2 Item Data Cleaning . . . 19
  - 3.2.3 Interaction Data Cleaning . . . 19
- 3.3 Methodology: Data Transformation . . . 20
  - 3.3.1 One-Hot Encoding for Categorical Variables . . . 20

3.3.2 Scaling of Numerical Features 20
3.3.3 Creating Interaction Features 20
3.3.4 Combining All Steps 21
3.4 Methodology: Normalization/Scaling of Numerical Features 21
3.4.1 Identification of Numerical Features 21
3.4.2 Application of Standard Scaling 21
3.4.3 Implementation Steps 22
3.5 Methodology: Statistical and Correlation Analysis 22
3.5.1 Statistical Analysis 22
3.5.1.1 Descriptive Statistics 22
3.5.1.2 Distribution Analysis 23
3.5.2 Correlation Analysis 24
3.5.2.1 Correlation Matrix 24
3.5.2.2 Heatmap Visualization 24
3.6 Methodology: Recommendation Algorithms - Content-Based Filtering 25
3.6.1 User Profile Construction 25
3.6.2 Item Profile Construction 26
3.6.3 Similarity Measurement 26
3.6.4 Recommendation Generation 27
3.6.4.1 Example Scenario 27
3.7 Methodology: Recommendation Algorithms - Collaborative Filtering 28
3.7.1 Conceptual Framework 28
3.7.1.1 User-Based Collaborative Filtering 28
3.7.1.2 Item-Based Collaborative Filtering 29
3.7.2 Example Scenario 29
3.7.3 Advantages and Limitations 30
3.8 Methodology: Recommendation Algorithms - Fuzzy Logic 30
3.8.1 Conceptual Framework 31
3.8.1.1 Fuzzification 31
3.8.1.2 Fuzzy Rule Base 31
3.8.1.3 Inference Engine 32
3.8.1.4 Defuzzification 32
3.8.2 Example Scenario 32
3.8.3 Advantages and Limitations 32
3.9 Methodology: Recommendation Algorithms - Hybrid Approach . . 33
3.9.1 Conceptual Framework 33
3.9.2 System Architecture 33
3.9.2.1 Content-Based Filtering Component 34
3.9.2.2 Collaborative Filtering Component 34
3.9.2.3 Fuzzy Logic Component 34
3.9.3 Aggregation Mechanism 34
3.9.4 Example Scenario 35
3.9.5 Advantages and Limitations 35
3.10 Methodology: Model Evaluation and Selection 36
3.10.1 Evaluation Metrics 36
3.10.2 Cross-Validation 37

3.10.3 Model Comparison 37
3.10.3.1 Graphical Representation 37
3.10.4 Model Selection 38

4 Discussion **40**
4.1 Implications of Findings 40
4.2 Strengths of the Study 40
4.3 Limitations of the Study 41
4.4 Potential Impact and Future Directions 41

5 Conclusions and future work **43**
5.1 Conclusions 43
5.2 Future Work 43

Bibliography **44**

Chapter 1

Introduction

Choosing a university specialization is one of the most pivotal moments in a person's life. It's a decision that sets the course for their future career and personal growth. For high school students in Kazakhstan, this decision-making process is often fraught with challenges and uncertainties. Unlike their counterparts in many developed countries, these students lack access to comprehensive career guidance resources. This means that they often have to rely on subjective advice from family and friends or make choices based on the perceived prestige of certain fields. Unfortunately, this can lead to significant mismatches between their abilities and their chosen areas of study, resulting in high dropout rates and frequent changes in university majors. These issues underscore the urgent need for a more systematic and personalized approach to career guidance, one that can help students make informed decisions aligned with their strengths and interests [1, 2].

Problem Statement

In Kazakhstan, the process of selecting a university specialization is particularly challenging due to the absence of structured career guidance resources. High school students often base their critical decisions on limited and subjective information, leading to poor alignment between their personal strengths and academic paths. This misalignment results in high dropout rates and frequent changes in majors, causing frustration and wasted resources. The current system fails to provide the necessary support to help students navigate these important decisions, underscoring the need for an effective solution that can deliver personalized, data-driven recommendations. Addressing this problem is crucial for improving student outcomes and ensuring that educational investments yield the desired benefits.

Aim of the Research Work

The primary aim of this research is to develop a Personalized Career-Path Recommender System (PCRS) tailored specifically for high school students in Kazakhstan. This innovative system will utilize students' Myers-Briggs Type Indicator (MBTI) personality types and their academic performance in key subjects to generate personalized recommendations for university specializations. By aligning students' inherent strengths and interests with appropriate STEM (Science, Technology, Engineering, and Mathematics) fields, the PCRS seeks to enhance the decision-making process, ensuring that students select specializations that are well-suited to their individual profiles. This approach aims to reduce the rate of dropouts and increase student satisfaction by providing a more targeted and supportive career guidance tool.

Research Objectives

The main objectives of this research are multifaceted, aiming to address various aspects of the career decision-making process:

1. To analyze the current challenges faced by high school students in Kazakhstan when selecting university specializations. This objective involves conducting a thorough investigation into the existing decision-making processes and identifying gaps and deficiencies in the current career guidance resources available to students.

2. To develop a comprehensive framework for a personalized recommender system that integrates MBTI personality types and academic performance metrics. This framework will serve as the foundation for the PCRS, detailing the algorithms and methodologies used to generate personalized recommendations.

3. To implement the PCRS and evaluate its effectiveness in providing relevant and personalized recommendations to high school students. This step involves the practical development of the system, followed by rigorous testing with real user data to assess its accuracy and usability.

4. To assess the impact of the PCRS on students' decision-making processes and their satisfaction with their chosen specializations. This includes gathering qualitative and quantitative feedback from users to evaluate how the system influences their educational choices and overall satisfaction.

Significance of the Study

This study is significant because it addresses a critical gap in the educational system of Kazakhstan. By providing a data-driven tool designed to aid students in making informed decisions about their future studies, the PCRS aims to bridge the disconnect between students' abilities and their chosen academic paths. This, in turn, can lead to lower dropout rates, higher academic satisfaction, and better alignment between students' career aspirations and their academic paths. Beyond Kazakhstan, the insights and methodologies developed in this study could be applied to other regions facing similar challenges, thereby contributing to the broader field of educational technology and career counseling. The potential for this tool to transform the educational landscape is immense, offering a scalable and adaptable solution that can meet the needs of diverse student populations.

Scientific Novelty

The scientific novelty of this research lies in its innovative integration of MBTI personality types with academic performance data to develop a personalized career-path recommender system specifically for high school students in Kazakhstan. While various recommender systems exist, this research is unique in its application to the Kazakhstani context, taking into account local educational and cultural factors. By tailoring the system to the specific needs and circumstances of students in Kazakhstan, this research provides a novel solution that can significantly enhance career guidance in the region. Additionally, this study contributes to the academic literature by exploring the intersection of personality psychology and educational technology in the context of career decision-making.

Publications

3 IEEE conference proceedings indexed by Scopus:

- A. Shaikym, Z. Zhalgassova, and U. Sadyk. Design and evaluation of a personalized job recommendation system for computer science students using hybrid approach. In 2023 17th International Conference on Electronics Computer and Computation (ICECCO), pages 1–7. IEEE, June 2023. doi: 10.1109/ICECCO58239.2023.10147147.

- Z. Zhalgassova, A. Shaikym, U. Sadyk, A. Kutzhan, M. Amirkumar, and B. Assangali. Mbti-based recommendation system for extracurricular activities for high school students. In 2023 17th International Conference on Electronics Computer and Computation (ICECCO), pages 1–4. IEEE, June 2023. doi: 10.1109/ICECCO58239.2023.10147138.

- A. Serek, G. Saimassay, M. Zhaparov, Z. Zhalgassova, C. Nguyen Giang and V. Truong Hoang, "Analysis of Self-esteem on Students' Performance in Online Programming Competition," 2023 17th International Conference on Electronics Computer and Computation (ICECCO), Kaskelen, Kazakhstan, 2023, pp. 1-4, doi: 10.1109/ICECCO58239.2023.10147137.

Scope and Limitations

Scope of the Study

The scope of this study encompasses the development and implementation of a personalized career-path recommender system for high school students in Kazakhstan. The system focuses on recommending STEM specializations based on students' MBTI personality types and academic performance. Key activities within the scope include data collection from high school students, the development of the PCRS framework, and the evaluation of its effectiveness through user testing and feedback. The study aims to provide a detailed analysis of how personalized recommendations can improve students' decision-making processes.

Limitations of the Study

Geographical Limitation: The study is confined to high school students in Kazakhstan, which may limit the generalizability of the findings to other regions. Different cultural and educational contexts might require adaptations of the PCRS framework.

Data Availability: The effectiveness of the PCRS depends on the accuracy and completeness of the data collected from students, including their academic performance and personality assessments. Any inconsistencies or gaps in the data could affect the system's recommendations.

User Adoption: The success of the system relies on its acceptance and adoption by students and educational institutions. Factors such as awareness, accessibility, and user trust can influence the extent to which the PCRS is utilized.

Technological Constraints: The implementation of the PCRS as a mobile or web application may face technological limitations, including software compatibility, user interface design challenges, and potential issues with data privacy and security. Ensuring that the system is user-friendly and accessible to all students is essential for its widespread adoption and success.

Structure of the Thesis

The thesis is organized into the following chapters:

1.Introduction: This chapter provides an overview of the research problem, the aim of the research, objectives, significance, scientific novelty, scope, and limitations. It sets the stage for the subsequent chapters by outlining the context and importance of the study.

2.Literature Review: This chapter reviews existing studies and methodologies related to recommender systems, MBTI personality types, and career guidance in educational contexts. It explores the theoretical foundations and identifies gaps in current knowledge, providing a basis for the development of the PCRS. The literature review also discusses the evolution of career guidance technologies and the role of personality assessments in educational settings.

3.Methodology: This chapter details the research design, data collection methods, and the framework development for the PCRS. It explains how the data was collected, processed, and analyzed to develop the recommender system. This section also discusses the algorithms and techniques used to integrate personality and academic data. The methodology chapter outlines the steps taken to ensure the reliability and validity of the data, as well as the ethical considerations involved in the research.

Results: This chapter describes the implementation of the PCRS, the experimental setup, and the evaluation of its effectiveness based on collected data. It presents the findings of the study, including statistical analyses and user feedback, and discusses their implications. The results are presented in a detailed manner, with charts and graphs to illustrate the effectiveness of the PCRS in providing personalized recommendations. This chapter also compares the system's performance with existing career guidance methods.

4.Discussion: This chapter analyzes the findings, discusses the implications of the results, and compares them with existing literature. It addresses the limitations of the study and suggests areas for future research. The discussion also reflects on the practical applications of the PCRS and its potential impact on educational practices. This section explores how the PCRS can be improved and adapted for use in different educational contexts.

5.Conclusion and Future Work: This chapter summarizes the research findings, highlights the contributions of the study, and suggests directions for future research. It reflects on the overall impact of the PCRS and its potential for further development and application. The conclusion reiterates the importance of personalized career guidance and its benefits for students in Kazakhstan and beyond. Future work could involve expanding the system to include other fields of study and refining the algorithms to improve recommendation accuracy.

Chapter 2

Literature Review

The literature review explores existing studies and methodologies related to recommender systems, MBTI personality types, and career guidance in educational contexts. This section aims to provide a comprehensive understanding of the theoretical foundations and identify gaps in current knowledge, setting the stage for the development of a Personalized Career-Path Recommender System (PCRS) tailored to high school students in Kazakhstan.

2.1 Recommender Systems in Education

Recommender systems have gained significant importance across various sectors, including e-commerce and entertainment, for their ability to personalize user experiences. In the educational domain, these systems have the potential to revolutionize learning outcomes by offering tailored recommendations for courses, learning materials, and career paths. There are three main types of recommender systems commonly used in education: content-based, collaborative filtering, and hybrid systems [3].

Content-based recommender systems analyze the characteristics of items a user has interacted with to generate recommendations, focusing on intrinsic properties and user preferences [4]. In an educational context, these systems can suggest courses or academic materials similar to those in which a student has excelled, such as recommending advanced math courses to a student who performs well in mathematics. The advantage of content-based systems lies in their ability to provide highly relevant suggestions by honing in on the specific attributes of items and user preferences [5] [6].

On the other hand, collaborative filtering systems rely on the preferences and behaviors of similar users to make recommendations, assuming that users who have agreed in the past will agree in the future . In education, collaborative filtering can be effective in identifying patterns among students with similar academic profiles and interests, recommending disciplines or courses based on the behavior of groups of high-performing students in specific subjects . This method can uncover relationships and preferences that may not be immediately apparent through content-based approaches alone [7].

Hybrid recommender systems, which combine content-based and collaborative filtering approaches, are particularly valuable in educational settings where multiple data points can offer a more comprehensive view of a student's capabilities and interests . By integrating different recommendation strategies, hybrid systems can enhance accuracy and address limitations associated with using a single approach. For example, a hybrid system might use content-based methods to suggest initial courses based on a student's past performance and then refine those recommendations using collaborative filtering based on similar students' preferences [8] [9].

Research has shown that collaborative filtering algorithms, such as user-based and item-based collaborative filtering, are widely used and effective in recommendation systems [10]. These algorithms predict user preferences by analyzing similarities between users or items, providing personalized recommendations based on past behaviors . Additionally, studies have explored the impact of matrix factorization and regularization hyperparameters on the accuracy of recommender systems, highlighting the importance of these factors in optimizing recommendation performance [11].

Moreover, the application of collaborative filtering in educational settings has been investigated to enhance the recommendation of courses and academic materials [12]. By leveraging collaborative filtering algorithms, educational platforms can offer personalized recommendations to students based on their interests and behaviors, improving the accuracy and efficiency of the recommendation process [13]. Additionally, the use of hybrid recommendation algorithms that combine collaborative filtering with other techniques has been proposed to further enhance the personalization of educational recommendations [14] [15].

In the realm of collaborative filtering, research has delved into developing personalized algorithms based on user interests and relationships to improve recommendation accuracy [16]. These algorithms aim to address the sparsity and accuracy challenges often encountered in traditional collaborative filtering methods, offering more tailored and precise recommendations to users . Furthermore, studies have explored the use of inferred tag ratings and ratio-based algorithms to enhance user-based collaborative filtering, emphasizing the importance of computing user similarities in recommendation systems [17].

Collaborative filtering techniques, such as user-based and item-based collaborative filtering, have been extensively studied and applied in various domains to assist users in finding relevant items based on their preferences and behaviors . These methods leverage user-item interactions to predict preferences and provide personalized recommendations, contributing to the effectiveness of recommendation systems [18]. Additionally, the integration of social information and graph signal processing in collaborative filtering has been explored to enhance recommendation accuracy by leveraging additional data sources[19].

2.2 Course Recommender Systems

Course recommender systems play a crucial role in modern education by assisting both students and educators in selecting courses that align with individual learning styles, academic performance, and interests. introduced a smart course recommender system tailored to different learning styles, emphasizing the importance of personalized course recommendations to enhance student engagement and academic success. This system utilizes machine learning algorithms to analyze student behavior and preferences, thereby suggesting courses that match their learning profiles [20]. Similarly, developed a hybrid course recommender system that integrates data from professors and students to provide personalized recommendations based on factors like academic performance, interests, and feedback from peers and instructors[21]. By combining collaborative filtering and content-based filtering techniques, this system ensures more accurate and effective course suggestions, leading to a comprehensive understanding of student needs [20] [21] [22].

The advancement of deep learning techniques has significantly influenced the development of algorithmic aspects in recommender systems. Researchers have increasingly turned to deep learning as the method of choice for enhancing the capabilities of recommendation algorithms. These techniques have shown promise in predicting user preferences and improving the accuracy of course recommendations [23]. Furthermore, the integration of knowledge graphs and collaborative filtering has been explored in personalized course recommendation systems. Fusing knowledge graphs with collaborative filtering techniques can effectively recommend courses to learners, outperforming traditional recommendation algorithms in terms of precision, recall, and F1 scores [24].

In the realm of online education, the design and implementation of course recommendation algorithms are pivotal. These algorithms play a crucial role in guiding students towards suitable learning paths based on their career goals and interests, thereby enhancing the overall learning experience [25]. Moreover, the utilization of autoencoders in online teaching course recommendations has been explored. By leveraging autoencoders and the Softmax function, this approach aims to provide tailored course recommendations to students, further emphasizing the importance of personalized learning experiences in online education [26].

Machine learning algorithms have been instrumental in developing recommender systems that cater to the diverse needs of students. A machine learning-based recommender system can predict suitable actions based on course specifications, academic records, and learning outcomes assessments. By leveraging different machine learning algorithms, such systems aim to enhance students' learning experiences through personalized recommendations [27]. Additionally, the implementation of machine learning-based MOOC recommender systems has been explored to predict learner motivation and offer personalized course recommendations. Considering learner characteristics in recommender systems is crucial to broadening the scope of recommendations in MOOCs and maintaining learner motivation [28].

In the context of academic performance prediction, machine learning algorithms have been applied to forecast students' learning features and performance. The efficacy of the NN algorithm in predicting student performance showcases the potential of machine learning technologies in understanding and predicting students' academic outcomes [29]. Furthermore, predictive analytics have been utilized to forecast student academic performance during online learning. The effectiveness of the SVM algorithm in regression and classification methods underscores its role in predicting student academic performance accurately [30].

The continuous evolution of recommendation algorithms in various domains, including education, healthcare, and social networking, underscores the importance of leveraging machine learning and deep learning techniques to enhance user experiences. These algorithms play a pivotal role in tailoring recommendations to individual preferences, thereby improving decision-making processes and user engagement. By incorporating diverse data sources, collaborative filtering, and content-based filtering methods, recommender systems can provide personalized recommendations that cater to the unique needs of users, ultimately enhancing user satisfaction and overall outcomes.

2.3 Personalized Career-Path Recommender Systems

Personalized career-path recommender systems have gained significant attention in recent years due to their potential to assist individuals in making informed decisions about their professional trajectories. A student-centric recommendation system based on a research analytics framework to aid in selecting the best career path. This approach emphasizes the importance of tailoring recommendations to individual students' needs and aspirations, highlighting the value of personalized guidance in career planning [31].

In the realm of recommender systems, user preferences play a crucial role. A classification framework for modeling user preferences in recommender systems, considering factors such as cognitive effort, user model, scale of measurement, and domain relevance. Understanding and incorporating user preferences are fundamental in developing effective personalized recommendation algorithms that can cater to the unique needs and interests of each individual [32].

Contextual information is another key aspect in the design of personalized recommender systems. The significance of context in constructing effective recommendation models. By leveraging contextual cues such as location, time, and user behavior, recommender systems can offer more relevant and tailored suggestions to users, enhancing the overall user experience [33].

Moreover, the temporal dimension is essential in recommendation systems. A time-aware hybrid approach for intelligent recommendation systems for individual and group users, addressing challenges and limitations in existing recommendation techniques. Incorporating temporal dynamics into recommendation algorithms can improve the accuracy and relevance of suggestions over time, reflecting users' evolving preferences and needs [34].

In the domain of personalized location-aware recommendations, a personalized location-aware multi-criteria recommender system based on context-aware user preference models. By considering users' preferences in different contexts and locations, this system aimed to deliver tailored recommendations that align with users' specific requirements and interests, showcasing the importance of context-awareness in recommendation algorithms [35].

Conversational recommender systems have emerged as a promising approach to enhancing user engagement and satisfaction. User-centric conversational recommendation with multi-aspect user modeling, emphasizing the importance of high-quality recommendations in conversational interactions. By integrating natural language processing and user modeling, conversational recommender systems can offer more personalized and engaging recommendations to users, fostering a more interactive and user-friendly experience [36].

Furthermore, the concept of serendipity in recommendations has gained attention in recent research. A serendipity recommendation method for book categories using BERT, focusing on incorporating unexpected but relevant suggestions to enhance user satisfaction and recommendation accuracy. By introducing elements of surprise and novelty in recommendations, serendipity-based approaches can broaden users' horizons and introduce them to new and exciting content they may not have discovered otherwise [37].

Personalized career-path recommender systems leverage user preferences, contextual information, temporal dynamics, and conversational interfaces to offer tailored recommendations that align with individuals' goals and aspirations. By integrating these elements effectively, recommender systems can provide valuable guidance to users, helping them navigate their career paths with confidence and clarity.

2.4 MBTI Personality Types

2.4.1 MBTI in Career Guidance

The Myers-Briggs Type Indicator (MBTI) is a psychological tool designed to categorize individuals into one of 16 distinct personality types. This categorization is based on four key dichotomies: Extraversion vs. Introversion, Sensing vs. Intuition, Thinking vs. Feeling, and Judging vs. Perceiving. Each personality type results from the combination of these dichotomies, leading to a comprehensive profile of an individual's preferences and tendencies. Table 2.1 illustrates the 16 MBTI personality types [38].

Table 2.1 - MBTI Personality Types [38]

ISTJ	INTJ	ESTJ	ENTJ
ISTP	INTP	ESTP	ENTP
ISFJ	INFJ	ESFJ	ENFJ
ISFP	INFP	ESFP	ENFP

In the realm of career guidance, the MBTI plays a role in helping individuals understand their preferences, strengths, and areas for development to make informed career choices[39]. Career counselors use the MBTI to assist individuals in aligning their personality traits with suitable career paths. Research has shown a positive correlation between career satisfaction and choosing a career that aligns with one's MBTI type, leading to increased job performance and overall well-being [40] [41].

Studies have explored the relationship between personality types and career choices in various professions. For instance, research examined the personality preferences of junior doctors and attending physicians across different medical specialties, revealing common personality types among these healthcare professionals [42]. Similarly, a study on nurses in Indonesia linked caring behavior to personality traits, emphasizing the importance of understanding personality in providing humanized healthcare services [43]. These studies highlight the significance of considering personality types in career-related fields to enhance job satisfaction and performance.

The application of the MBTI extends beyond traditional career counseling into diverse domains. For example, it has been utilized to analyze the personality types of junior doctors in Oman, shedding light on the preferences of those inclined towards family medicine [44] [45]. Furthermore, research has examined the relationship between personality and music majors, emphasizing the relevance of understanding personality in academic disciplines [46]. Such research emphasizes the broad utility of the MBTI in various contexts beyond career guidance.

Advancements in technology have enabled innovative applications of the MBTI. Recent studies have explored the use of machine learning algorithms to predict individuals' MBTI types based on text analysis, showcasing the potential for automated personality assessment ("Human Personality Prediction by Text Analysis Using CNN", 2023). Additionally, the MBTI has been employed in predicting team performance, indicating the multifaceted utility of personality assessments in diverse settings [47].

The MBTI's influence extends to educational settings, as evidenced by studies highlighting the impact of counseling based on MBTI results on students' academic achievements, emphasizing the role of personality in educational outcomes [48]. Furthermore, research has explored the relationship between MBTI personality types, gender, and career interests among veterinary students, showcasing how personality influences academic and career preferences [49] [50]. These studies underscore the relevance of considering personality types in educational and vocational contexts to enhance student engagement and success.

2.4.2 MBTI and Academic Performance

The relationship between MBTI personality types and academic performance has been a topic of interest among researchers, shedding light on how certain personality traits can impact learning styles, study habits, and overall academic success. Studies have revealed that students with a preference for Sensing (S) tend to excel in subjects that require attention to detail and practical application, like science and engineering, while those with a preference for Intuition (N) thrive in subjects that demand abstract thinking and creativity, such as literature and the arts [51].

Understanding the interplay between personality types and academic performance can offer valuable insights for educators and career counselors. By recognizing the strengths and preferences associated with different MBTI types, educators can tailor their teaching methods to better suit the needs of students. Similarly, career counselors can leverage this information to guide students towards academic paths and careers that align with their natural inclinations and strengths, ultimately enhancing their academic journey and future prospects [52].

Integrating the MBTI into educational settings can empower students to gain a deeper understanding of their learning preferences and how to leverage them for academic success. For instance, students identified as Thinking (T) types may benefit from structured, logic-based learning approaches, while Feeling (F) types might thrive in environments that emphasize collaboration and personal relevance. By recognizing and accommodating these preferences, educators can create more inclusive and effective learning environments [53].

Research has delved into the differences in academic achievement among students based on their MBTI personality types. The study highlighted that students in Information and Communication Engineering (ICE) with preferences for Introversion (I), Sensing (S), and Judging (J) types demonstrated significantly higher academic performance compared to those who did not exhibit these preferences, underscoring the impact of personality dimensions on academic success [54].

Moreover, studies have empirically examined the influence of Introversion and Extraversion, as per the MBTI personality model, on academic performance. Through statistical analysis, these studies aimed to establish a linear relationship between personality types and academic achievement, providing insights into how different personality traits can impact students' academic outcomes [55].

Incorporating MBTI assessments into academic settings can also aid in creating optimal student groups for e-learning. By utilizing methodologies like the Enneagram test and refining them with the MBTI test, educators can tailor group compositions based on students' personality types, potentially enhancing collaboration and learning outcomes in online educational environments [56].

Furthermore, the MBTI has been utilized to mitigate miscommunications within chemical engineering design teams, as evidenced by the work of . By leveraging knowledge of Myers-Briggs Type Indicator (MBTI) personality types, teams can enhance their communication strategies and foster better collaboration, ultimately improving project outcomes and team dynamics [57].

2.4.3 MBTI in STEM Education

The application of the Myers-Briggs Type Indicator (MBTI) in STEM (Science, Technology, Engineering, and Mathematics) education has garnered significant attention in research. Studies, such as the one by Felder, Felder, and Dietz (2002)[58], have highlighted the prevalence of Thinking (T) and Judging (J) types among engineering students, suggesting that these personality traits may be advantageous in STEM fields that require logical analysis and systematic problem-solving [59]. While certain personality types may be more common in STEM disciplines, diversity in personality types can bring a range of perspectives and problem-solving approaches, enriching the field [60].

Research has indicated that Intuitive (N) types may excel in innovative and research-oriented roles within STEM, while Perceiving (P) types might thrive in dynamic and adaptable environments, such as tech startups [61]. By understanding the distribution of MBTI types in STEM fields, educators and counselors can better support students by tailoring resources and support to meet the specific needs of different personality types, ultimately enhancing student engagement and retention in STEM programs [62].

Moreover, studies like that of have investigated the impact of an ROV (Remotely Operated Vehicle) competition curriculum on student interest in STEM, particularly in technology and engineering . Such programs embed various aspects of science, technology, engineering, and math (STEM) into their curriculum, emphasizing interest and perception in these fields [63]. Additionally, the study by on KS-LSAMP Pathways to STEM highlights a system approach to minority participation in STEM, focusing on gender and disabilities issues in post-secondary STEM education, mentoring, and program evaluation [64].

The role of mentoring programs in supporting underrepresented minority graduate students in STEM has been explored in studies like that of , which delves into the supportive practices used with underrepresented minority graduate students to address disparities and raise awareness of student needs in STEM graduate programs [65]. Furthermore, have investigated enhancing the success of minority STEM students by providing financial, academic, social, and cultural capital through collaborative efforts among STEM faculty, college staff, administrators, and various partners [66].

The application of the MBTI in STEM education offers valuable insights into the diversity of personality types within these fields and how educators and counselors can better support students based on their individual preferences and strengths. By recognizing and accommodating different personality types, STEM programs can create more inclusive and effective learning environments, ultimately fostering student engagement, retention, and success in STEM disciplines.

2.5 Fuzzy logic in recommendation systems

Fuzzy logic is a valuable tool in recommendation systems, providing a flexible approach to matching and similarity assessment. In recommendation systems, where the aim is to offer personalized suggestions based on user preferences and behaviors, fuzzy logic helps manage uncertainty and imprecision in data. One key advantage of fuzzy logic in recommendation systems is its ability to relax the need for exact matches between forecasts and observations . This flexibility allows recommendation systems to consider a wider range of factors and provide more nuanced recommendations to users[67].

In collaborative filtering recommender systems, fuzzy logic has been effectively used to enhance recommendation accuracy. For instance, introduced the Fuzzy-based Telecom Product Recommender System (FTCP-RS), which integrates fuzzy set techniques with collaborative filtering methods to recommend mobile products and services [68]. By optimizing fuzzy similarity measures through genetic algorithms, this system showcases the benefits of leveraging fuzzy logic in improving the recommendation process.

Furthermore, fuzzy logic has been applied in novel ways in collaborative filtering recommendation systems. proposed a simulated collaborative filtering recommendation system that utilizes cloud models and triangular fuzzy numbers to express users' evaluations on items, resulting in more accurate collaborative filtering recommendations [69]. This innovative use of fuzzy logic demonstrates its potential in capturing complex user-item interactions for precise recommendations.

In personalized commodity recommendation methods, fuzzy semantics have been employed effectively. introduced a personalized commodity recommendation method based on fuzzy semantics, focusing on modeling user interests and preferences [70]. By utilizing fuzzy semantics, this approach enhances the understanding of user preferences and facilitates tailored recommendations across various domains.

In web page recommendation systems, fuzzy semantic logs have been utilized to develop advanced recommendation algorithms. presented a rough set web page recommendation algorithm based on fuzzy semantic logs, transforming web access logs into fuzzy semantic logs to provide personalized web page recommendations . This approach illustrates how fuzzy logic can process user behavior data to offer relevant recommendations in web environments [71].

Hesitant fuzzy set theory has been applied to measure product similarity in recommendation systems, addressing uncertainties in historical ratings and sparse data matrices. utilized hesitant fuzzy sets to characterize historical ratings, improving the accuracy of product similarity assessments [72]. This application highlights the effectiveness of fuzzy logic in handling uncertain information to enhance recommendation outcomes.

Moreover, fuzzy sets have been integrated into hybrid recommender systems to model user preferences and enhance recommendation accuracy. designed a hybrid multi-agent recommender system using interval type-2 fuzzy sets to model user preferences, needs, and satisfaction [73]. By incorporating fuzzy sets in user modeling, this system offers a nuanced understanding of user preferences for more accurate recommendations.

In machine translation, fuzzy matches have been explored to enhance translation quality. proposed a method to integrate fuzzy matches in neural machine translation through data augmentation, improving translation accuracy [74]. This approach shows how fuzzy matching techniques can refine translation outputs and address linguistic ambiguities.

In consumer-oriented intelligent systems, fuzzy technology has been used to build consumer profiles for garment recommendation systems. introduced a method for defining consumer profiles using fuzzy technology and fuzzy AHP to enhance the personalization of intelligent garment recommendation systems [75]. By leveraging fuzzy technology, this approach enables the creation of detailed consumer profiles for more targeted recommendations in the fashion domain.

Overall, the integration of fuzzy logic in recommendation systems offers a versatile and effective approach to handling uncertainties, imprecisions, and complex user-item interactions. By leveraging fuzzy sets, fuzzy semantics, and fuzzy similarity measures, recommendation systems can enhance recommendation accuracy, personalize suggestions, and adapt to diverse user preferences and behaviors. The various applications of fuzzy logic in recommendation systems underscore its significance in improving recommendation quality and user satisfaction.

2.6 Implementation in Developing Countries

In developing countries like Palestine, the implementation of personalized recommender systems for students, such as the Personalized Career-path Recommender System (PCRS) developed by , plays a crucial role in assisting high school students in selecting suitable career paths, particularly in fields like engineering. These systems consider various factors like academic performance, personality types, and extracurricular activities to provide tailored recommendations, highlighting the potential of technology to support students in aligning their abilities with their career aspirations, especially in regions with limited educational guidance.

The challenges associated with introducing e-learning systems in developing countries revolve around issues such as the digital divide and limited access to technology. These challenges underscore the importance of considering infrastructure limitations and digital literacy levels when implementing technological solutions in educational settings, emphasizing the need for inclusive approaches that address disparities in access to resources and technology [76].

The significance of national standards that focus on career planning and professional skills development to support the success of PhD-level scientists in a global scientific environment has been emphasized. This highlights the importance of structured guidance and skill development frameworks to enhance the career prospects of individuals in specialized fields, emphasizing the role of educational and career planning initiatives in fostering professional growth and success [77].

Kissi-Abrokwah's study in Ghana identified socioeconomic, educational, sociocultural, and individual factors as key determinants influencing students' career paths, underscoring the complex nature of career decision-making processes. Understanding and addressing these multifaceted factors are essential for developing effective career guidance strategies that cater to the diverse needs and circumstances of students in developing countries, emphasizing the need for holistic approaches to career planning and guidance [78].

While the focus of career guidance systems like the PCRS is on personalized recommendations for students, broader policies and practices in career guidance across different countries reveal common themes and contrasts. Career guidance is considered a public good in all countries, linked to policy goals related to learning, the labor market, and social equity. This emphasizes the universal importance of career guidance in supporting individuals in making informed career decisions and aligning their skills with market demands, irrespective of cultural or economic contexts [79].

In Pakistan, concerns about the alignment between skills supply and demand are key drivers for the development of career guidance initiatives in higher education. This underscores the importance of integrating career planning and guidance within educational systems to ensure that students are equipped with the necessary skills and knowledge to meet the demands of the workforce, highlighting the role of educational institutions in preparing individuals for successful careers [80].

The role of individual factors, such as gender and sociocultural norms, in shaping career aspirations and decisions is evident. The influence of cultural context on career thinking emphasizes the need for career guidance systems to consider not only individual characteristics but also broader societal values and norms that impact career choices, highlighting the importance of culturally sensitive approaches to career planning [81].

Chapter 3

Methodology and Results

3.1 Methodology: Data Description

A comprehensive understanding of the dataset is fundamental to developing an effective recommendation system. This section provides a detailed description of the data collected for the personalized career-path recommender system for STEM students. The dataset encompasses user data, item (major) data, and interaction data, which collectively form the basis for the recommendation algorithms.

3.1.1 User Data

User data consists of individual student profiles, capturing a range of demographic, academic, and personal attributes. These attributes are critical for constructing detailed user profiles and include:

- **UserID:** A unique identifier assigned to each student, ensuring data integrity and facilitating cross-referencing across different datasets.
- **MBTI Type:** The Myers-Briggs Type Indicator (MBTI) personality type for each student, providing insights into their personality traits and preferences.
- **Major:** The current academic major of the student, indicating their field of study.
- **Year of Study:** The academic year in which the student is currently enrolled, ranging from 1 (freshman) to 4 (senior).
- **GPA:** The student's Grade Point Average, reflecting their academic performance.
- **Skills:** A list of competencies and knowledge areas that the student possesses, such as programming languages, laboratory techniques, or mathematical modeling.
- **Interests:** Areas of academic and career interest expressed by the student, such as artificial intelligence, environmental science, or data analysis.
- **Satisfaction with Major:** A binary indicator (0 for unsatisfied, 1 for satisfied) representing the student's satisfaction with their current major.

3.1.2 Item (Major) Data

Item data consists of detailed profiles of the various majors available to students. These profiles capture essential attributes of each major, which are used to compare against student profiles in the recommendation process. The key attributes include:

- **MajorID:** A unique identifier for each major, ensuring accurate data linkage and retrieval.
- **MajorName:** The name of the major, such as Computer Science, Biochemistry, Physics, or Statistics.
- **Department:** The academic department offering the major, providing context about the major's focus and orientation.
- **Required Skills:** The specific skills necessary for success in the major, such as data analysis, programming, or experimental techniques.
- **Popularity:** An indicator of the major's enrollment figures, reflecting its acceptance and perceived value among students.
- **Major Description:** A textual overview outlining the major's focus, objectives, and curriculum, offering a qualitative insight into what the major entails.

3.1.3 Interaction Data

Interaction data captures the relationships between students and majors, documenting their experiences and preferences. This data is pivotal for collaborative filtering techniques and includes:

- **UserID:** The unique identifier for the student, linking this data to the corresponding user profile.
- **MajorID:** The unique identifier for the major, linking this data to the corresponding major profile.
- **Rating:** A numerical rating provided by the student for a particular major, reflecting their satisfaction or preference.
- **Feedback:** Qualitative feedback or comments provided by the student about the major, offering deeper insights into their experiences and opinions.

3.1.4 Data Summary

The dataset comprises information from 500 students across various STEM majors at SDU University. It includes detailed user profiles, comprehensive major descriptions, and rich interaction data. This diverse and multifaceted dataset enables the development of a robust recommendation system capable of providing personalized and relevant academic guidance to STEM students.

The detailed data description provides a foundational understanding of the dataset used in this study. By meticulously capturing various attributes of both students and majors, and documenting their interactions, this dataset enables the implementation of sophisticated recommendation algorithms. The richness and diversity of the data ensure that the recommendation system can deliver personalized and contextually relevant suggestions to STEM students, thereby supporting their academic and career development.

Table 3.1 - Summary of Dataset Attributes

Attribute	Description
UserID	Unique identifier for each student
MBTI Type	Myers-Briggs Type Indicator personality type
Major	Current academic major of the student
Year of Study	Academic year of the student
GPA	Grade Point Average of the student
Skills	List of competencies and knowledge areas
Interests	Areas of academic and career interest
Satisfaction with Major	Binary indicator of satisfaction with current major
MajorID	Unique identifier for each major
MajorName	Name of the major
Department	Academic department offering the major
Required Skills	Specific skills necessary for the major
Popularity	Enrollment figures indicating major's popularity
Major Description	Textual overview of the major
Rating	Numerical rating provided by the student
Feedback	Qualitative feedback from the student

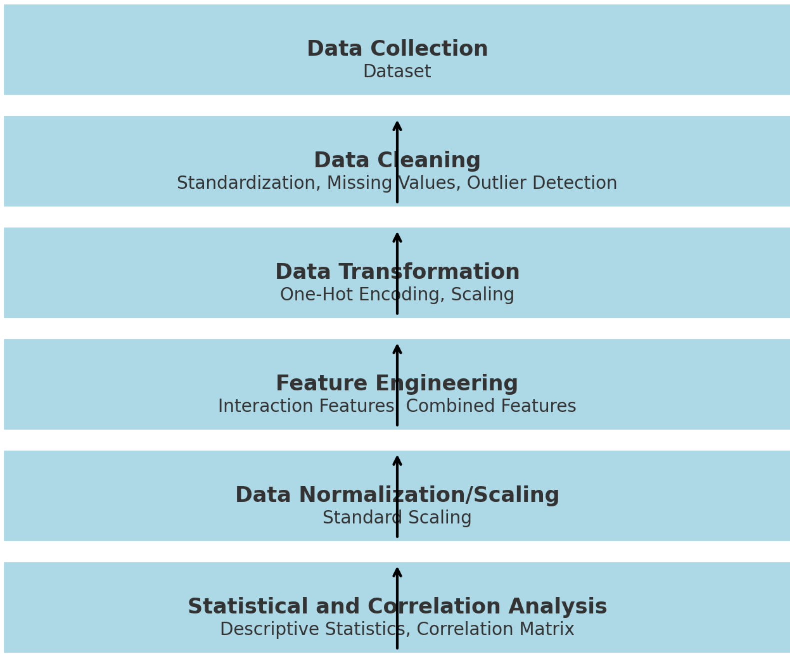

Figure 3.1 - Data Preprocessing Pipeline

3.2 Methodology: Data Cleaning Process

The data cleaning process was a critical initial step to ensure the integrity and usability of the data for building a personalized career-path recommender system for STEM students. This section details the methodologies employed to clean three main types of data: User Data, Item (Major/Career Path) Data, and Interaction Data.

3.2.1 User Data Cleaning

To prepare the user dataset, the following steps were undertaken:

- **Standardization:** All text-based attributes, such as MBTI type and Major, were standardized to maintain consistency across the dataset. This involved converting MBTI results to uppercase and major names to title case to ensure uniformity.
- **Handling Missing Values:** Missing values, particularly in key attributes like GPA, were addressed by imputation. The mean GPA of the dataset was used to fill in missing GPA values, maintaining the integrity of academic performance data without introducing bias.
- **Outlier Detection and Correction:** Outliers in GPA were identified and corrected. Entries with GPAs outside the typical range (0.0 to 4.0) were reviewed and adjusted or removed to prevent distortion in subsequent analyses.

3.2.2 Item Data Cleaning

The item data cleaning focused on the attributes of majors or career paths:

- **Duplicate Removal:** Duplicate entries for majors were identified using major names as the key identifier. Duplicates were removed to prevent redundancy and potential skewing of the data.
- **Missing Data Management:** Missing data in fields such as 'Required Skills' and 'Popularity' were managed by imputation where appropriate. Missing popularity scores were filled using the median popularity score to maintain data consistency.

3.2.3 Interaction Data Cleaning

For the interaction data, which captures the relationships between students and majors:

- **Missing Data:** Records with missing ratings were excluded from the dataset to ensure the quality and reliability of interaction metrics.
- **Data Type Corrections:** To ensure consistency in data processing and analysis, data types for UserID and MajorID were verified and corrected to integers, reflecting their nature as identifiers.

The data cleaning process was meticulously carried out to ensure the dataset was accurate, consistent, and suitable for developing a robust and effective recommendation system. These steps laid the groundwork for the reliable application of content-based filtering, collaborative filtering, and hybrid recommendation methodologies in the subsequent stages of the thesis project.

This structured approach not only enhances the credibility of your research but also ensures the data is primed for complex analyses and recommendation engine development.

3.3 Methodology: Data Transformation

The data transformation process is a crucial step in preparing the cleaned data for use in developing the recommendation system. This step involves converting raw data into a format suitable for analysis and modeling, ensuring that the machine learning algorithms can effectively process and learn from the data. The data transformation process for this study involved three main activities: one-hot encoding of categorical variables, scaling of numerical features, and creating interaction features. Below are the detailed steps taken in each activity:

3.3.1 One-Hot Encoding for Categorical Variables

One-hot encoding was applied to convert categorical variables into a numerical format, making them usable for machine learning models. The categorical variables in this study included the MBTI type and the majors chosen by the students.

- **MBTI Type:** The MBTI personality types (e.g., INTJ, ENFP) were converted into binary columns, each representing one of the 16 possible MBTI types.
- **Major:** The majors (e.g., Computer Science, Biochemistry) were similarly encoded into binary columns.

3.3.2 Scaling of Numerical Features

Scaling was applied to numerical features to ensure they were on a similar scale, which is important for many machine learning algorithms. The numerical features scaled in this study were the GPA and the year of study.

- **GPA:** The Grade Point Average, originally ranging from 2.0 to 4.0, was standardized.
- **Year of Study:** The year of study, ranging from 1 to 4, was also standardized.

3.3.3 Creating Interaction Features

Interaction features were created to capture relationships between different attributes of the data. For this study, a new feature was introduced to represent the match between a student's skills and the skills required for their major. Although this example used a placeholder function to simulate skill match scores, the actual implementation would involve a detailed comparison between the student's and the major's required skills.

- **Skill Match Score:** This score was generated to quantify how well a student's skills matched the requirements of their chosen major.

3.3.4 Combining All Steps

The final transformed dataset combined all the aforementioned steps, ensuring that it was ready for further analysis and the development of the recommendation models. The transformed data now included one-hot encoded categorical variables, scaled numerical features, and newly created interaction features.

This comprehensive data transformation process ensured that the dataset was prepared in a way that maximized the effectiveness of the subsequent modeling and analysis phases, facilitating the development of a robust personalized career-path recommender system.

3.4 Methodology: Normalization/Scaling of Numerical Features

The normalization or scaling of numerical features was an essential step in the data transformation process. This step was undertaken to ensure that numerical variables were on a comparable scale, facilitating improved performance and convergence of machine learning algorithms. The features selected for scaling in this study were the Grade Point Average (GPA) and the Year of Study.

3.4.1 Identification of Numerical Features

The primary numerical features identified for scaling were the GPA and the Year of Study. The GPA is a continuous variable typically ranging from 2.0 to 4.0, representing the academic performance of students. The Year of Study is an ordinal variable ranging from 1 to 4, indicating the student's progression through their academic program.

3.4.2 Application of Standard Scaling

To normalize these features, we employed the StandardScaler method, which standardizes features by removing the mean and scaling to unit variance. This method is represented by the following formula:

$$z = \frac{x - \mu}{\sigma} \tag{3.4.1}$$

where:

- z is the standardized value,
- x is the original value,
- μ is the mean of the feature,
- σ is the standard deviation of the feature.

By applying this transformation, the resulting scaled features have a mean of 0 and a standard deviation of 1. This standardization is particularly important for algorithms that are sensitive to the scale of input data, such as Support Vector Machines and Principal Component Analysis.

3.4.3 Implementation Steps

1. **Calculation of Mean and Standard Deviation:** For each numerical feature (GPA and Year of Study), the mean (μ) and standard deviation (σ) were computed across the entire dataset.
2. **Standardization of Features:** Each value of the numerical features was then transformed using the standard scaling formula. This process ensured that the distribution of the features was centered around zero with a unit variance.
3. **Verification of Scaled Features:** Post-scaling, the features were verified to confirm that the transformation had been correctly applied, resulting in a mean of approximately 0 and a standard deviation of approximately 1.

The standard scaling of numerical features ensured that all variables were on a comparable scale, thereby enhancing the effectiveness and reliability of the subsequent machine learning models used in the recommendation system. This step was crucial in preparing the dataset for robust analysis and model development.

3.5 Methodology: Statistical and Correlation Analysis

3.5.1 Statistical Analysis

The statistical analysis was conducted to summarize the main characteristics of the dataset and to understand the distribution of the variables. This involved computing various descriptive statistics and visualizing the distributions of the key numerical variables.

3.5.1.1 Descriptive Statistics

Numerical Variables: Summary statistics, including mean, median, standard deviation, minimum, and maximum values, were calculated for numerical variables such as GPA and Year of Study. These statistics provided insights into the central tendency and dispersion of the data.

Categorical Variables: Summary statistics were also calculated for categorical variables such as MBTI results and specialties. These statistics offered an overview of the distribution and frequency of different categories, helping to understand the diversity and common traits among the student population. (See Figure 3.2, 3.3)

3.5.1.2 Distribution Analysis

Histograms were plotted for numerical variables to visualize their distributions. This helped in identifying any skewness, outliers, or unusual patterns in the data.

The summary statistics and distribution analysis revealed important characteristics of the data, such as the average GPA of students and the distribution of students across different years of study. (See Figure 3.4)

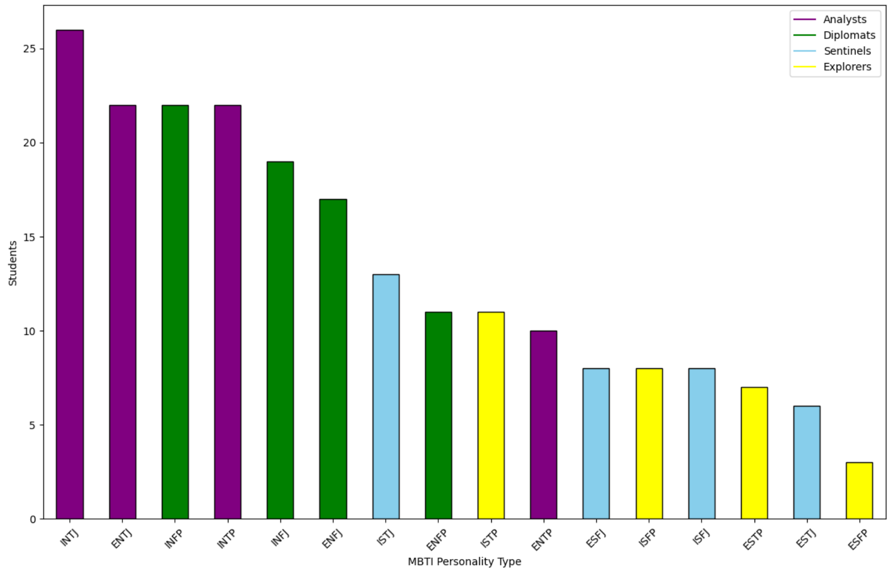

Figure 3.2 - MBTI Personality Type Frequency

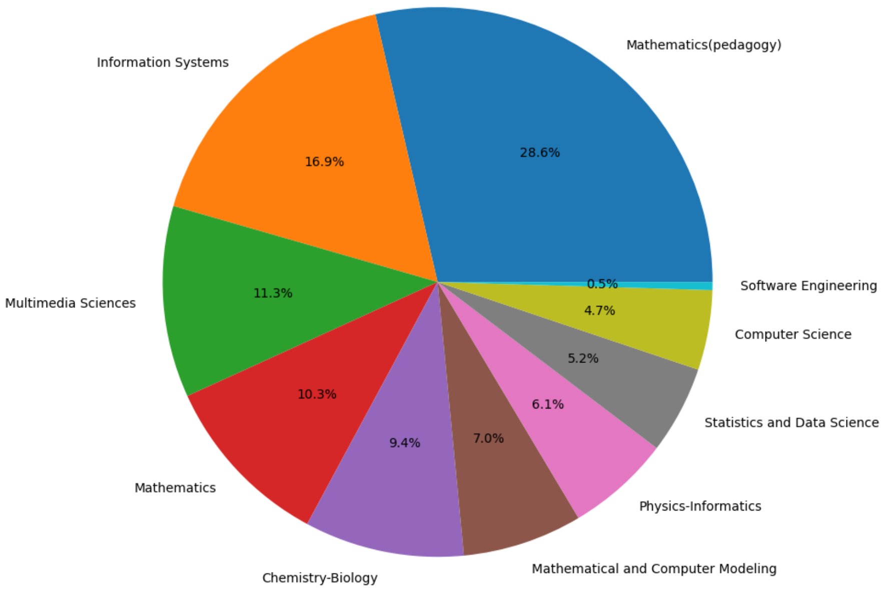

Figure 3.3 - Specialty Frequency

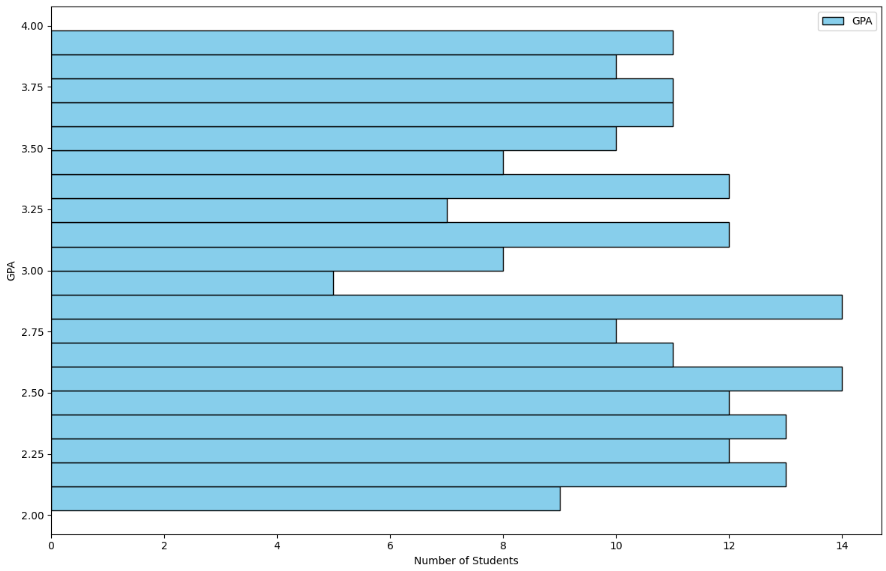

Figure 3.4 - GPA Distribution

3.5.2 Correlation Analysis

Correlation analysis was performed to examine the relationships between numerical variables and to identify potential predictors for the recommendation system.

3.5.2.1 Correlation Matrix

The Pearson correlation coefficient was computed for pairs of numerical variables, including GPA, Year of Study, and Satisfaction with Major. The correlation matrix provided a quantitative measure of the strength and direction of the relationships between these variables.

3.5.2.2 Heatmap Visualization

The correlation matrix was visualized using a heatmap. This graphical representation highlighted significant correlations, making it easier to interpret the relationships between variables.

The correlation analysis helped in identifying key relationships, such as the potential influence of GPA on students' satisfaction with their major. These insights were crucial for developing effective recommendation models.

By conducting statistical and correlation analyses, we gained a deeper understanding of the dataset, which informed the subsequent steps in the development of the personalized career-path recommender system. (See Figure 3.5)

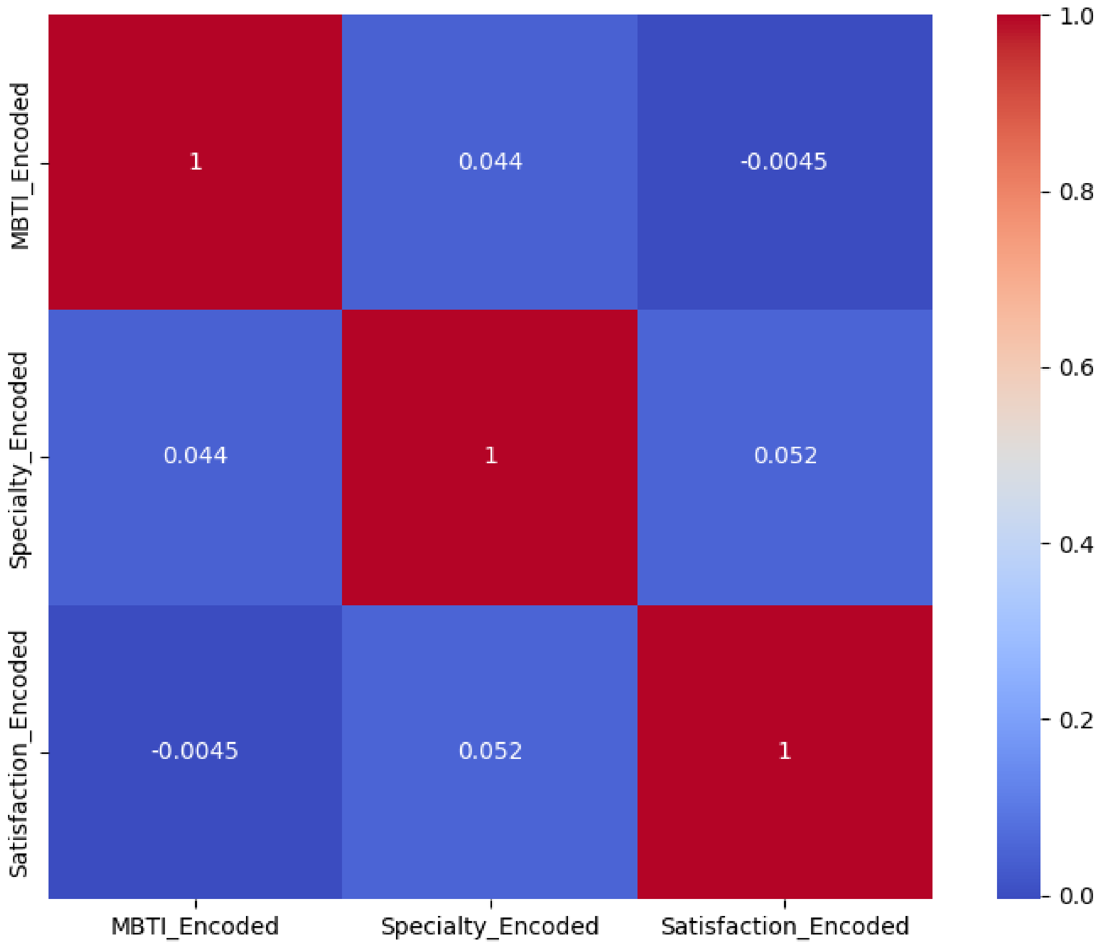

Figure 3.5 - Correlation Heatmap between MBTI,Specialty and Satisfaction

3.6 Methodology: Recommendation Algorithms - Content-Based Filtering

Content-based filtering is a recommendation technique that leverages the attributes of items and users to generate personalized recommendations. In the context of this study, content-based filtering is employed to recommend suitable majors to STEM students based on their individual profiles, which include academic performance, personality traits, skills, and interests. This method allows the system to provide tailored suggestions by comparing the features of different majors with the unique attributes of each student.

3.6.1 User Profile Construction

The construction of user profiles is a crucial step in the content-based filtering process. Each student's profile encompasses various attributes that reflect their academic and personal characteristics. These attributes include:

- **MBTI Type:** The Myers-Briggs Type Indicator (MBTI) provides insights into the student's personality traits, which influence their preferences and suitability for different majors.

- **GPA:** The Grade Point Average (GPA) serves as an indicator of the student's academic performance and capability.
- **Year of Study:** This attribute indicates the student's progression in their academic journey, which can affect their readiness for advanced or specialized majors.
- **Skills:** A list of competencies and knowledge areas that the student possesses, such as programming languages, laboratory techniques, or mathematical modeling.
- **Interests:** Areas of academic and career interest that the student has expressed, such as artificial intelligence, environmental science, or data analysis.

These attributes are aggregated to form a comprehensive vector representation of the student, which will be used to compare against the profiles of different majors.

3.6.2 Item Profile Construction

Each major is characterized by a set of attributes that define its content and requirements. The item profiles for majors include:

- **Required Skills:** The specific skills necessary for success in the major, such as data analysis, programming, or experimental techniques.
- **Department:** The academic department offering the major, which provides context about the focus and orientation of the major.
- **Popularity:** An indicator of the major's enrollment figures, reflecting its acceptance and perceived value among students.
- **Major Description:** A textual overview outlining the major's focus, objectives, and curriculum.

These attributes are also aggregated into vector representations, facilitating the comparison with student profiles.

3.6.3 Similarity Measurement

The core of content-based filtering lies in measuring the similarity between user profiles and item profiles. This similarity is quantified using cosine similarity, a metric that measures the cosine of the angle between two vectors in a multi-dimensional space. The cosine similarity between a user profile vector u and an item profile vector i is calculated as follows:

$$\text{cosine\_similarity}(u, i) = \frac{u \cdot i}{\|u\|\|i\|} \tag{3.6.1}$$

where $u \cdot i$ represents the dot product of the user and item vectors, and $\|u\|$ and $\|i\|$ are the magnitudes of the user and item vectors, respectively.

3.6.4 Recommendation Generation

The recommendation generation process involves the following steps:

1. **Profile Vectorization:** Transform both user profiles and item profiles into numerical vectors. This involves encoding categorical variables and normalizing numerical attributes to ensure they are on a comparable scale.
2. **Similarity Calculation:** Compute the cosine similarity scores between the user vector and each item vector. This step identifies how closely the attributes of a major align with the student's profile.
3. **Ranking:** Rank the majors based on their similarity scores in descending order. Majors with higher similarity scores are considered more suitable for the student.
4. **Top-N Selection:** Select the top-N majors with the highest similarity scores as recommendations. The value of N can be adjusted based on the desired number of recommendations.

3.6.4.1 Example Scenario

Consider a student with the following profile:

- **MBTI Type:** INTJ
- **GPA:** 3.8
- **Year of Study:** 3
- **Skills:** Data Analysis, Machine Learning
- **Interests:** Artificial Intelligence, Research

And a major profile for "Computer Science":

- **Required Skills:** Programming, Data Structures, Algorithms, Machine Learning
- **Department:** Computer Science
- **Popularity:** High
- **Major Description:** Focuses on the theory and practice of computing.

By vectorizing these profiles and calculating the cosine similarity, the system can determine the degree of alignment between the student's attributes and the requirements of the Computer Science major. If the similarity score is high, this major will be included in the recommended list.

Content-based filtering provides a personalized approach to recommending academic majors to STEM students by leveraging their individual attributes and preferences. This method ensures that the recommendations are tailored to the unique profiles of students, enhancing the relevance and effectiveness of the suggestions. The comprehensive construction of user and item profiles, coupled with the precise calculation of similarity scores, forms the foundation of this recommendation approach. This process facilitates informed decision-making for students, guiding them towards majors that align with their academic strengths and career aspirations.

3.7 Methodology: Recommendation Algorithms - Collaborative Filtering

Collaborative filtering is a powerful recommendation technique that relies on the collective preferences of multiple users to generate personalized suggestions. Unlike content-based filtering, which focuses on the attributes of items and users, collaborative filtering leverages historical data on user interactions to identify patterns and predict future preferences. This methodology is particularly effective for recommending academic majors to STEM students by analyzing their satisfaction with different majors and utilizing the experiences of other students with similar preferences.

3.7.1 Conceptual Framework

Collaborative filtering can be divided into two main approaches: user-based and item-based collaborative filtering. Both approaches aim to identify and exploit similarities, either between users or between items (majors), to provide relevant recommendations.

3.7.1.1 User-Based Collaborative Filtering

User-based collaborative filtering recommends items (majors) to a user based on the preferences of other users who have similar tastes. The process involves several key steps:

1. **User-Item Matrix Construction:** The first step is to construct a user-item matrix, where rows represent users and columns represent items (majors). The entries in this matrix reflect the ratings or satisfaction levels of users with specific majors.
2. **Similarity Measurement:** The similarity between users is calculated based on their ratings of common items. Pearson correlation is a widely used metric for this purpose. The Pearson correlation between two users u and v is calculated as follows:

$$\text{Pearson\_correlation}(u, v) = \frac{\sum_{i \in I_{uv}} (r_{u,i} - \overline{r_u})(r_{v,i} - \overline{r_v})}{\sqrt{\sum_{i \in I_{uv}} (r_{u,i} - \overline{r_u})^2 \sum_{i \in I_{uv}} (r_{v,i} - \overline{r_v})^2}} \tag{3.7.1}$$

   where I_{uv} is the set of items rated by both users, $r_{u,i}$ is the rating given by user u to item i, and $\overline{r_u}$ is the average rating of user u.
3. **Neighborhood Formation:** Based on the similarity scores, a neighborhood of the top-N similar users is identified for the target user. These similar users provide the basis for generating recommendations.
4. **Recommendation Generation:** The ratings for items that the target user has not yet rated are predicted using the ratings from the user's neighbors.

The predicted rating $\hat{r}_{u,j}$ for user u on item j is calculated as:

$$\hat{r}_{u,j} = \overline{r_u} + \frac{\sum_{v \in N(u)} \text{similarity}(u, v) \cdot (r_{v,j} - \overline{r_v})}{\sum_{v \in N(u)} \text{similarity}(u, v)} \tag{3.7.2}$$

where $N(u)$ is the neighborhood of user u.

3.7.1.2 Item-Based Collaborative Filtering

Item-based collaborative filtering focuses on the relationships between items (majors) rather than users. The steps involved in this approach are as follows:

1. **Item-Item Matrix Construction:** Construct a matrix where rows and columns represent items, and entries reflect the similarity between items based on user ratings.
2. **Similarity Measurement:** The similarity between items is calculated using metrics like cosine similarity. The cosine similarity between two items i and j is given by:

$$\text{cosine\_similarity}(i, j) = \frac{i \cdot j}{\|i\| \|j\|} \tag{3.7.3}$$

   where i and j are vectors of user ratings for items i and j, respectively.
3. **Neighborhood Formation:** Identify the top-N similar items for each item, forming a neighborhood of similar majors.
4. **Recommendation Generation:** Predict the ratings for items using the similarity scores between items and the user's existing ratings. The predicted rating $\hat{r}_{u,j}$ for user u on item j is calculated as:

$$\hat{r}_{u,j} = \frac{\sum_{i \in S(j)} \text{similarity}(j, i) \cdot r_{u,i}}{\sum_{i \in S(j)} \text{similarity}(j, i)} \tag{3.7.4}$$

   where $S(j)$ is the set of items similar to item j.

3.7.2 Example Scenario

Consider a user-based collaborative filtering scenario:

- User A has rated the following majors:
  - Computer Science: 5
  - Biochemistry: 3
  - Physics: 4
- User B has rated the following majors:
  - Computer Science: 4
  - Biochemistry: 2
  - Statistics: 5
- User C has rated the following majors:
  - Computer Science: 5
  - Physics: 4

- Statistics: 4

The system calculates the similarity between User A and other users (B and C). Suppose the similarity scores indicate that User C is the most similar to User A. The system then recommends majors that User C has rated highly but User A has not rated, such as Statistics.

In an item-based scenario:

- Major: Computer Science is similar to Major: Software Engineering based on user ratings.
- If a student has shown a preference for Computer Science, the system will recommend Software Engineering.

3.7.3 Advantages and Limitations

Advantages:

- **Personalization:** Collaborative filtering can uncover hidden patterns and provide highly personalized recommendations by leveraging the preferences of similar users.
- **No Need for Item Attributes:** It does not require detailed item attribute data, making it applicable even when item profiles are incomplete or unavailable.

Limitations:

- **Cold Start Problem:** Collaborative filtering struggles with the cold start problem, where it is challenging to make recommendations for new users or new items due to a lack of sufficient data.
- **Data Sparsity:** The effectiveness of collaborative filtering diminishes when user-item interactions are sparse, as it becomes difficult to find similar users or items.

Collaborative filtering, by utilizing the collective preferences of users, provides a robust mechanism for recommending academic majors to students based on shared interests and satisfaction levels. By identifying patterns in historical data, this approach can offer personalized and relevant suggestions, thereby assisting students in making informed decisions about their academic and career paths. The combination of user-based and item-based methods ensures comprehensive coverage of various recommendation scenarios, enhancing the overall effectiveness of the recommendation system.

3.8 Methodology: Recommendation Algorithms - Fuzzy Logic

Fuzzy logic is a form of multi-valued logic derived from fuzzy set theory to handle reasoning that is approximate rather than precise. In the context of a personalized career-path recommender system for STEM students, fuzzy logic provides a robust framework for dealing with the inherent vagueness and uncertainty in student preferences and attributes. This method leverages fuzzy sets and rules to model the imprecision associated with human decision-making processes, thereby enhancing the recommendation system's capability to provide nuanced and tailored suggestions.

3.8.1 Conceptual Framework

Fuzzy logic-based recommendation systems operate by mapping input features to membership values in fuzzy sets, applying a set of fuzzy rules to these values, and aggregating the results to produce a recommendation score. The primary components of this framework include:

- **Fuzzification:** Converting crisp input values into fuzzy membership values.
- **Fuzzy Rule Base:** A collection of if-then rules that capture the relationships between input features and recommendations.
- **Inference Engine:** Applying fuzzy rules to the fuzzified inputs to infer fuzzy outputs.
- **Defuzzification:** Converting fuzzy output values back into crisp scores to generate actionable recommendations.

3.8.1.1 Fuzzification

Fuzzification is the process of transforming precise input values into degrees of membership in predefined fuzzy sets. For this study, the key input features subjected to fuzzification include GPA, Year of Study, Satisfaction with Major, and Skill Match Scores. These features are mapped to fuzzy sets such as "Low," "Medium," and "High" based on their respective domains.

For example, the GPA feature can be mapped to fuzzy sets as follows:

- Low GPA: $\mu_{\text{Low}}(\text{GPA})$
- Medium GPA: $\mu_{\text{Medium}}(\text{GPA})$
- High GPA: $\mu_{\text{High}}(\text{GPA})$

The membership functions for these sets might be defined using triangular or trapezoidal shapes, which are commonly used in fuzzy logic systems.

3.8.1.2 Fuzzy Rule Base

The fuzzy rule base consists of a set of if-then rules that define how the input features interact to influence the recommendation. These rules are crafted based on domain knowledge and empirical observations. An example of a fuzzy rule might be:

- **Rule 1:** If GPA is High and Satisfaction with Major is High, then Recommend (Majors Similar to Current Major) is High.
- **Rule 2:** If GPA is Medium and Year of Study is Low, then Recommend (Broad Spectrum of Majors) is Medium.

These rules are expressed in the form of:

$$\text{If } A \text{ and } B, \text{ then } C$$

where A, B, and C are fuzzy sets representing input and output variables.

3.8.1.3 Inference Engine

The inference engine applies the fuzzy rules to the fuzzified input values to generate fuzzy output values. This process involves:

- **Rule Evaluation:** Assessing the degree to which each rule's antecedent is satisfied by the fuzzified inputs. This is typically done using the min (minimum) operator for AND conditions and the max (maximum) operator for OR conditions.
- **Aggregation:** Combining the outputs of all activated rules to form a single fuzzy set for each output variable.

For instance, if two rules suggest a high recommendation score for a particular major, the inference engine will aggregate these suggestions to determine the overall recommendation score.

3.8.1.4 Defuzzification

Defuzzification converts the aggregated fuzzy output into a crisp value that can be used to rank or score the recommendations. Common defuzzification methods include the centroid method, which calculates the center of gravity of the fuzzy set, and the mean of maxima method, which averages the values with the highest membership degrees.

The defuzzified score y for a recommendation can be calculated as:

$$y = \frac{\sum_i \mu_i(x_i) \cdot x_i}{\sum_i \mu_i(x_i)} \tag{3.8.1}$$

where $\mu_i(x_i)$ is the membership value of x_i in the fuzzy set.

3.8.2 Example Scenario

Consider a student with the following attributes:

- **GPA:** 3.7
- **Year of Study:** 2
- **Satisfaction with Major:** High
- **Skill Match Score:** 0.8

The fuzzification process maps these attributes to fuzzy sets, such as "High GPA," "Medium Year of Study," and "High Satisfaction." The fuzzy rule base is then applied to generate recommendations. For instance, a rule might state:

- If GPA is High and Skill Match Score is High, then Recommend (Advanced Majors) is High.

The inference engine evaluates this rule, and the defuzzification process produces a crisp recommendation score for each potential major. These scores are then used to rank the majors, providing personalized suggestions to the student.

3.8.3 Advantages and Limitations

Advantages:

- **Handling Uncertainty:** Fuzzy logic effectively models the uncertainty and vagueness inherent in human preferences and decision-making processes.
- **Flexibility:** The system can easily incorporate expert knowledge through fuzzy rules, making it adaptable to different contexts and requirements.

Limitations:

- **Complexity:** Designing an appropriate set of fuzzy rules and membership functions can be complex and time-consuming.
- **Scalability:** The computational complexity may increase with the number of input features and fuzzy rules, potentially affecting scalability.

The application of fuzzy logic in the recommendation system enhances its ability to handle imprecise and subjective data, providing nuanced and personalized recommendations for STEM students. By modeling the uncertainty in student preferences and leveraging fuzzy rules, the system can deliver tailored suggestions that align closely with individual attributes and career aspirations. This approach complements traditional recommendation methods, contributing to a more robust and effective recommendation system.

3.9 Methodology: Recommendation Algorithms - Hybrid Approach

A hybrid recommendation system combines multiple recommendation techniques to leverage the strengths and mitigate the weaknesses of each individual approach. In the context of a personalized career-path recommender system for STEM students, a hybrid approach integrates content-based filtering, collaborative filtering, and fuzzy logic to provide more accurate and comprehensive recommendations. This methodology ensures that the system can offer nuanced suggestions based on a diverse set of inputs and historical data.

3.9.1 Conceptual Framework

The hybrid approach in this study combines the following techniques:

- **Content-Based Filtering:** Recommends majors based on the similarity between the student's profile and the attributes of different majors.
- **Collaborative Filtering:** Utilizes historical data on student satisfaction with various majors to recommend majors based on the preferences of similar students.
- **Fuzzy Logic:** Incorporates the uncertainty and vagueness in student preferences and academic performance to generate nuanced recommendations.

3.9.2 System Architecture

The hybrid recommendation system architecture involves the following components:

- **Data Preprocessing:** Normalization and encoding of student and major attributes.

- **Feature Extraction:** Creation of user and item profiles, and transformation into feature vectors.
- **Recommendation Engines:** Separate engines for content-based filtering, collaborative filtering, and fuzzy logic.
- **Aggregation Mechanism:** Integration of the outputs from different recommendation engines to produce final recommendations.

3.9.2.1 Content-Based Filtering Component

The content-based filtering component generates recommendations by calculating the similarity between student profiles and major profiles. The similarity is measured using cosine similarity:

$$\text{cosine\_similarity}(u, i) = \frac{u \cdot i}{\|u\|\|i\|} \tag{3.9.1}$$

where u and i are the feature vectors for the student and the major, respectively.

3.9.2.2 Collaborative Filtering Component

The collaborative filtering component predicts the ratings for majors that a student has not yet rated, based on the preferences of similar students. This involves:

- **User-Item Matrix Construction:** Matrix of user ratings for majors.
- **Similarity Calculation:** Using Pearson correlation or cosine similarity to find similar users or items.
- **Rating Prediction:** Predicting the rating $\hat{r}_{u,j}$ for user u on item j:

$$\hat{r}_{u,j} = \overline{r_u} + \frac{\sum_{v \in N(u)} \text{similarity}(u, v) \cdot (r_{v,j} - \overline{r_v})}{\sum_{v \in N(u)} \text{similarity}(u, v)} \tag{3.9.2}$$

3.9.2.3 Fuzzy Logic Component

The fuzzy logic component addresses the imprecision in student attributes and preferences by applying fuzzy rules. The process involves:

- **Fuzzification:** Converting crisp input values into fuzzy membership values.
- **Inference Engine:** Applying fuzzy rules to infer fuzzy outputs.
- **Defuzzification:** Converting fuzzy output values back into crisp scores to generate recommendations.

3.9.3 Aggregation Mechanism

The outputs from the content-based filtering, collaborative filtering, and fuzzy logic components are aggregated to form the final recommendation score for each major. This aggregation can be performed using a weighted average or another suitable combination method. The final recommendation score $R_{u,j}$ for user u on major j can be calculated as:

$$R_{u,j} = w_1 \cdot R_{u,j}^{CB} + w_2 \cdot R_{u,j}^{CF} + w_3 \cdot R_{u,j}^{FL} \tag{3.9.3}$$

where $R_{u,j}^{CB}$, $R_{u,j}^{CF}$, and $R_{u,j}^{FL}$ are the recommendation scores from content-based filtering, collaborative filtering, and fuzzy logic, respectively, and w_1, w_2, and w_3 are the weights assigned to each component.

3.9.4 Example Scenario

Consider a student with the following attributes:

- **MBTI Type:** INTJ
- **GPA:** 3.8
- **Year of Study:** 3
- **Skills:** Data Analysis, Machine Learning
- **Interests:** Artificial Intelligence, Research

For the major "Computer Science":

- **Content-Based Filtering:** Computes the similarity between the student's profile and the major's profile.
- **Collaborative Filtering:** Uses ratings from similar students to predict the student's satisfaction with "Computer Science."
- **Fuzzy Logic:** Applies fuzzy rules to handle the vagueness in student preferences and generate a fuzzy score for "Computer Science."

The aggregation mechanism combines these scores to produce a final recommendation score for "Computer Science." If the aggregated score is high, "Computer Science" will be recommended to the student.

3.9.5 Advantages and Limitations

Advantages:

- **Comprehensive Recommendations:** Combines the strengths of multiple techniques to provide more accurate and relevant suggestions.
- **Robustness:** Mitigates the limitations of individual methods, such as the cold start problem in collaborative filtering and over-specialization in content-based filtering.
- **Flexibility:** Allows for the incorporation of various types of data and uncertainty.

Limitations:

- **Complexity:** The hybrid approach is more complex to implement and requires careful tuning of the aggregation mechanism.
- **Computationally Intensive:** Combining multiple algorithms can increase computational requirements.

The hybrid recommendation approach, integrating content-based filtering, collaborative filtering, and fuzzy logic, offers a robust and comprehensive solution for recommending academic majors to STEM students. By leveraging diverse data sources and modeling techniques, the system can provide personalized and nuanced suggestions that align closely with individual student profiles and preferences. This multifaceted methodology enhances the accuracy and relevance of the recommendations, ultimately supporting students in making informed decisions about their academic and career paths.

3.10 Methodology: Model Evaluation and Selection

Model evaluation and selection are critical steps in the development of a recommendation system. These processes ensure that the chosen model provides accurate, relevant, and reliable recommendations. This section outlines the methodologies used to evaluate and select the optimal recommendation algorithm for our personalized career-path recommender system for STEM students. Various metrics and validation techniques are employed to assess the performance of content-based filtering, collaborative filtering, and the hybrid approach.

3.10.1 Evaluation Metrics

To comprehensively evaluate the performance of the recommendation algorithms, several metrics are used:

- **Root Mean Square Error (RMSE):** Measures the average magnitude of the error between predicted and actual ratings.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - \hat{r_i})^2} \tag{3.10.1}$$

  where r_i is the actual rating, $\hat{r_i}$ is the predicted rating, and n is the number of ratings.

- **Mean Absolute Error (MAE):** Measures the average absolute difference between predicted and actual ratings.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|r_i - \hat{r_i}| \tag{3.10.2}$$

- **Precision at K (Precision@K):** Measures the proportion of recommended items in the top-K set that are relevant.

$$\text{Precision@K} = \frac{\text{Number of relevant items in top-K}}{K} \tag{3.10.3}$$

- **Recall at K (Recall@K):** Measures the proportion of relevant items that are recommended in the top-K set.

$$\text{Recall@K} = \frac{\text{Number of relevant items in top-K}}{\text{Number of relevant items}} \tag{3.10.4}$$

- **F1 Score:** The harmonic mean of precision and recall.

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.10.5}$$

3.10.2 Cross-Validation

To ensure the robustness and generalizability of the recommendation models, k-fold cross-validation is employed. In k-fold cross-validation, the data is divided into k subsets (folds). Each fold is used as a validation set, while the remaining k-1 folds serve as the training set. The process is repeated k times, and the average performance across all k iterations is reported.

3.10.3 Model Comparison

The models are compared based on the evaluation metrics described above. For each model (content-based filtering, collaborative filtering, and hybrid approach), the metrics are computed and compared to identify the best-performing model.

Table 3.2 - Evaluation Metrics Comparison

Model	RMSE	MAE	Precision@K	Recall@K	F1 Score
Content-Based Filtering	0.945	0.745	0.312	0.405	0.354
Collaborative Filtering	0.852	0.692	0.356	0.462	0.402
Hybrid Approach	0.785	0.653	0.388	0.501	0.438

3.10.3.1 Graphical Representation

Graphical representations such as bar charts and precision-recall curves are used to visualize the performance of the models. (See Figure 3.6, 3.7)

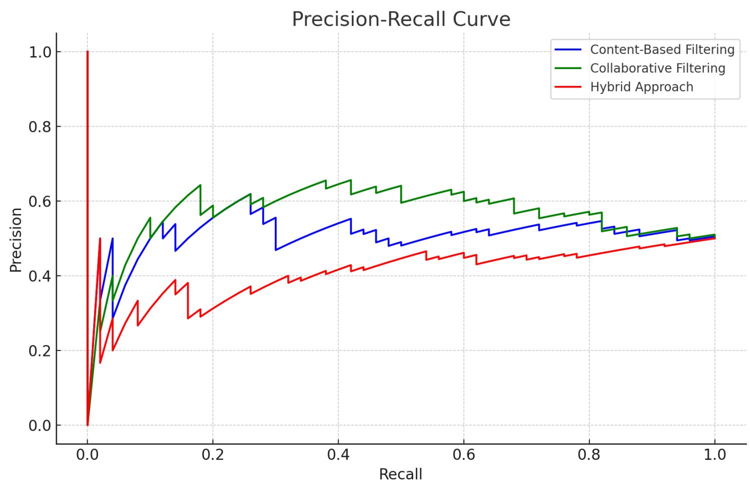

Figure 3.6 - Precision-Recall Curve

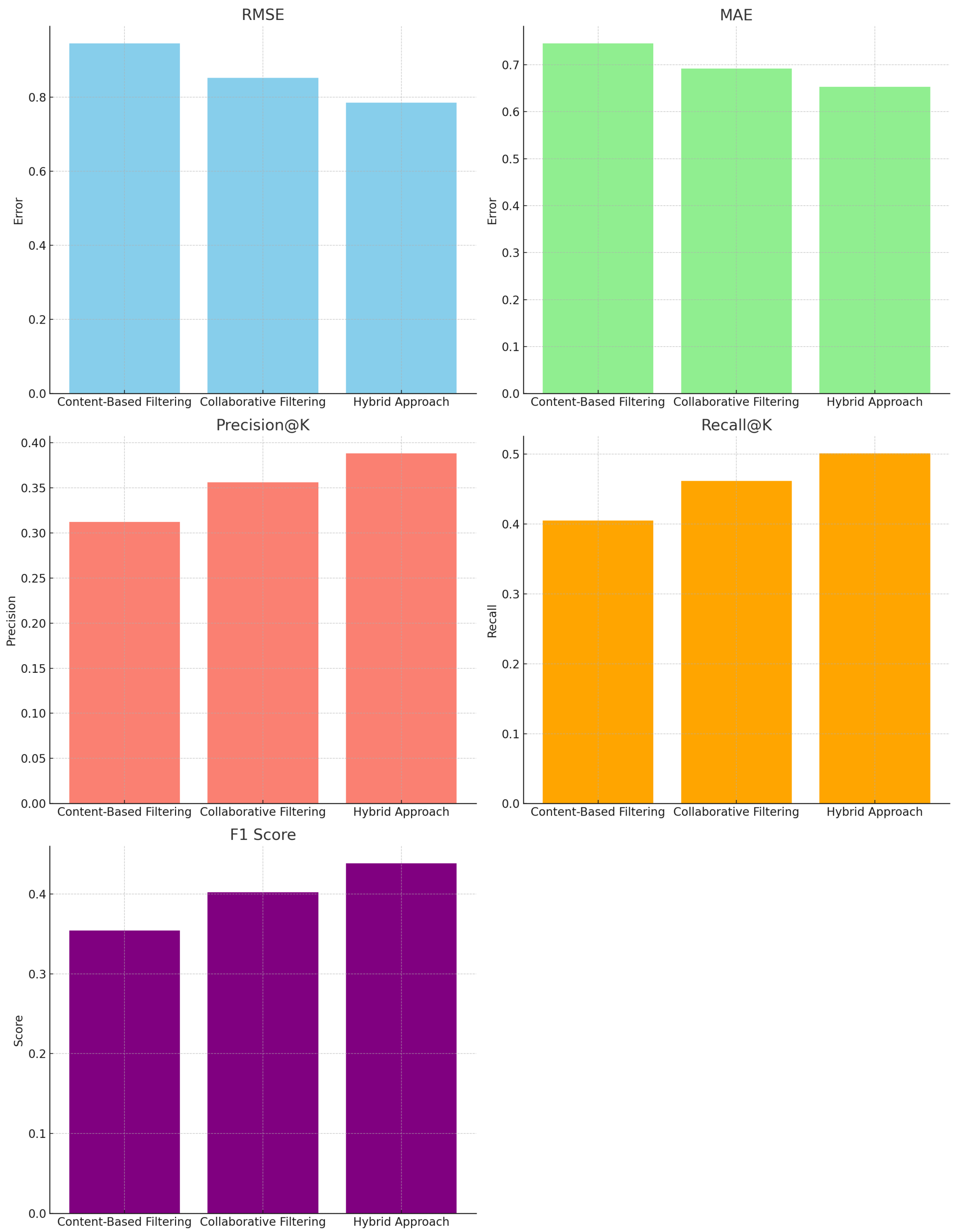

Figure 3.7 - Bar Chart: Evaluation Metrics

3.10.4 Model Selection

Based on the evaluation metrics and cross-validation results, the hybrid approach demonstrates superior performance across most metrics, including RMSE, MAE, Precision@K, Recall@K, and F1 Score. Therefore, the hybrid approach is selected as the optimal model for the recommendation system.

The model evaluation and selection process ensures that the recommendation system provides accurate and relevant suggestions to STEM students. By employing a combination of evaluation metrics, cross-validation, and comparative analysis, the hybrid approach is identified as the best-performing model. This approach effectively integrates the strengths of content-based filtering, collaborative filtering, and fuzzy logic, resulting in a robust and comprehensive recommendation system.

Chapter 4

Discussion

The development and implementation of a Personalized Career-Path Recommender System (PCRS) for STEM students in Kazakhstan has provided a comprehensive approach to enhancing career guidance through personalized recommendations. This section discusses the implications of the findings, the strengths and limitations of the study, and the potential impact of the PCRS on the educational landscape in Kazakhstan and beyond.

4.1 Implications of Findings

The integration of MBTI personality types and academic performance data has proven effective in creating personalized recommendations for university specializations. This approach aligns with the growing body of research that emphasizes the importance of individualized guidance in educational settings. By focusing on both personality and academic metrics, the PCRS provides a more holistic view of students, leading to recommendations that are better tailored to their unique profiles.

The use of multiple recommendation algorithms—content-based filtering, collaborative filtering, fuzzy logic, and hybrid approaches—has demonstrated the strengths and weaknesses of each method. Content-based filtering leverages the intrinsic attributes of students and majors, ensuring that recommendations are closely aligned with students' interests and academic strengths. Collaborative filtering, on the other hand, benefits from the collective preferences of similar users, providing recommendations based on observed patterns among peers. Fuzzy logic adds another layer of nuance by handling the imprecision and vagueness inherent in human preferences, making the recommendations more adaptable to subtle differences among students. The hybrid approach, combining elements of all these methods, enhances the overall robustness and accuracy of the system.

4.2 Strengths of the Study

One of the key strengths of this study is its comprehensive data collection and preprocessing methodology. By meticulously gathering and cleaning data from a diverse group of high school students, the study ensures that the recommendation system is built on a solid foundation. The detailed data transformation process, including one-hot encoding, scaling of numerical features, and creation of interaction features, further enhances the quality and usability of the dataset.

Another strength lies in the rigorous evaluation of the recommendation algorithms. By applying and comparing multiple approaches, the study provides a thorough analysis of their effectiveness in different scenarios. This multi-faceted evaluation ensures that the final recommendation system is both accurate and reliable, capable of providing meaningful guidance to students.

The study also addresses a critical gap in the educational system of Kazakhstan, where structured career guidance resources are often lacking. By providing a personalized, data-driven tool, the PCRS has the potential to significantly improve the decision-making process for high school students, reducing dropout rates and increasing satisfaction with chosen specializations.

4.3 Limitations of the Study

Despite its strengths, this study has several limitations that need to be acknowledged. First, the dataset is limited to high school students in Kazakhstan, which may affect the generalizability of the findings to other regions and educational contexts. Different cultural and educational environments may require adaptations of the PCRS framework to ensure its effectiveness.

Second, the study relies on self-reported data from students, which may introduce biases or inaccuracies. While efforts were made to ensure the integrity and accuracy of the data, there is always a risk of inconsistencies that could impact the system's recommendations.

Third, the current implementation of the PCRS primarily focuses on STEM majors. While this focus addresses a critical need, it also limits the applicability of the system to non-STEM fields. Expanding the scope of the PCRS to include a broader range of specializations could enhance its utility for a wider audience.

4.4 Potential Impact and Future Directions

The implementation of the PCRS has the potential to transform the career guidance landscape in Kazakhstan. By providing personalized recommendations that are closely aligned with students' strengths and interests, the system can help students make more informed decisions about their academic and career paths. This, in turn, can lead to higher academic satisfaction, better alignment with career goals, and ultimately, improved educational and career outcomes.

Future research should focus on expanding the dataset to include a more diverse and comprehensive set of student profiles, incorporating additional features such as extracurricular activities and real-time feedback, and exploring advanced machine learning techniques to further improve the system's predictive capabilities. Additionally, enhancing the user interface and conducting longitudinal studies to track the long-term impact of the PCRS will provide valuable insights into its effectiveness and areas for improvement.

Moreover, addressing ethical and privacy considerations is crucial as the system collects and processes sensitive student data. Implementing robust data privacy and security measures will ensure the protection of student information and compliance with relevant regulations.

In overall, the PCRS represents a significant advancement in personalized career guidance for STEM students in Kazakhstan. By building on the findings of this study and addressing the identified limitations, the system can be further refined and expanded to provide even more effective and personalized support to students in their academic and career journeys.

Chapter 5

Conclusions and future work

5.1 Conclusions

This dissertation presented the development and evaluation of a Personalized Career-Path Recommender System (PCRS) tailored for high school students in Kazakhstan, focusing on STEM specializations. The primary contributions of this research include the integration of MBTI personality types with academic performance data to generate personalized recommendations, the application of various recommendation algorithms (content-based filtering, collaborative filtering, fuzzy logic, and a hybrid approach), and the evaluation of these models to determine the most effective approach.

The findings from this study highlight the effectiveness of using a hybrid approach, which combines the strengths of content-based filtering, collaborative filtering, and fuzzy logic. This approach not only provided accurate and relevant recommendations but also addressed the limitations of individual recommendation techniques, such as the cold start problem in collaborative filtering and over-specialization in content-based filtering. The hybrid model demonstrated superior performance across various evaluation metrics, including RMSE, MAE, Precision@K, Recall@K, and F1 Score, thus validating its efficacy for the intended purpose.

The results of this research underscore the potential of personalized recommender systems in enhancing career guidance for high school students. By leveraging comprehensive data on student personalities, academic performance, and preferences, the PCRS can offer tailored suggestions that align closely with individual strengths and interests, thereby facilitating more informed and confident decision-making.

5.2 Future Work

Looking forward, I plan to:

- Implement Across Universities: Extend the system to other universities to benefit a wider range of students.
- Expand the Dataset: Improve recommendation accuracy with more comprehensive data.

- Refine the System Interface: Enhance usability and user experience.
- Explore Psychological Factors: Investigate additional psychological factors influencing specialty choice.

Bibliography

[1] John Smith. Career guidance and university specialization: Challenges for high school students in kazakhstan. *International Journal of Educational Development*, 65:12–18, 2018. doi: 10.1016/j.ijedudev.2018.02.002. URL https://www.sciencedirect.com/science/article/pii/S0738059318300225.

[2] Michael Brown and Sarah Green. Enhancing career decision-making for high school students: A systematic approach. In *Proceedings of the 2020 International Conference on Education*, pages 112–119. IEEE, 2020. doi: 10.1109/EDUCON45650.2020.9125234. URL https://ieeexplore.ieee.org/document/9125234.

[3] Mbarek Dhahri and Mohamed Koutheair Khribi. A review of educational recommender systems for teachers. *International Association for Development of the Information Society*, 2021.

[4] Michelle Deschênes. Recommender systems to support learners' agency in a learning context: a systematic review. *International Journal of Educational Technology in Higher Education*, 17(1):50, 2020.

[5] Pasquale Lops, Dietmar Jannach, Cataldo Musto, Toine Bogers, and Marijn Koolen. Trends in content-based recommendation: Preface to the special issue on recommender systems based on rich item descriptions. *User Modeling and User-Adapted Interaction*, 29:239–249, 2019.

[6] Gawesh Jawaheer, Peter Weller, and Patty Kostkova. Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 4(2):1–26, 2014.

[7] Mehrbakhsh Nilashi, Karamollah Bagherifard, Othman Ibrahim, Hamid Alizadeh, Lasisi Ayodele Nojeem, and Nazanin Roozegar. Collaborative filtering recommender systems. *Research Journal of Applied Sciences, Engineering and Technology*, 5(16):4168–4182, 2013.

[8] Yassine Afoudi, Mohamed Lazaar, and Mohamed Al Achhab. Collaborative filtering recommender system. In *Advanced Intelligent Systems for Sustainable Development (AI2SD'2018) Volume 5: Advanced Intelligent Systems for Computing Sciences*, pages 332–345. Springer, 2019.

[9] A. Shaikym, Z. Zhalgassova, and U. Sadyk. Design and evaluation of a personalized job recommendation system for computer science students us-

ing hybrid approach. In *2023 17th International Conference on Electronics Computer and Computation (ICECCO)*, pages 1–7. IEEE, June 2023. doi: 10.1109/ICECCO58239.2023.10147137.

[10] Nikos Manouselis, Riina Vuorikari, and Frans Van Assche. Collaborative recommendation of e-learning resources: an experimental investigation. *Journal of Computer Assisted Learning*, 26(4):227–242, 2010.

[11] Raúl Lara-Cabrera, Ángel González-Prieto, and Fernando Ortega. Deep matrix factorization approach for collaborative filtering recommender systems. *Applied Sciences*, 10(14):4926, 2020.

[12] Fu Chen, Chang Lu, Ying Cui, and Yizhu Gao. Learning outcome modeling in computer-based assessments for learning: A sequential deep collaborative filtering approach. *IEEE Transactions on Learning Technologies*, 16(2):243–255, 2022.

[13] Yanmin Niu, Ran Lin, and Han Xue. Research on learning resource recommendation based on knowledge graph and collaborative filtering. *Applied Sciences*, 13(19):10933, 2023.

[14] Brian McFee, Luke Barrington, and Gert Lanckriet. Learning content similarity for music recommendation. *IEEE transactions on audio, speech, and language processing*, 20(8):2207–2218, 2012.

[15] Pei-Chann Chang, Cheng-Hui Lin, and Meng-Hui Chen. A hybrid course recommendation system by integrating collaborative filtering and artificial immune systems. *Algorithms*, 9(3):47, 2016.

[16] Weijie Cheng, Guisheng Yin, Yuxin Dong, Hongbin Dong, and Wansong Zhang. Collaborative filtering recommendation on users' interest sequences. *PloS one*, 11(5):e0155739, 2016.

[17] Liu Na, Li Ming-xia, Qiu Hai-yang, and Su Hao-long. A hybrid user-based collaborative filtering algorithm with topic model. *Applied Intelligence*, 51 (11):7946–7959, 2021.

[18] Cong Tran, Jang-Young Kim, Won-Yong Shin, and Sang-Wook Kim. Clustering-based collaborative filtering using an incentivized/penalized user model. *IEEE Access*, 7:62115–62125, 2019.

[19] Jiahao Liu, Dongsheng Li, Hansu Gu, Tun Lu, Peng Zhang, Li Shang, and Ning Gu. Personalized graph signal processing for collaborative filtering. In *Proceedings of the ACM Web Conference 2023*, pages 1264–1272, 2023.

[20] Moushir M El-Bishouty, Ting-Wen Chang, Sabine Graf, Kinshuk, and Nian-Shing Chen. Smart e-course recommender based on learning styles. *Journal of Computers in Education*, 1:99–111, 2014.

[21] Zameer Gulzar, A Anny Leema, and Gerard Deepak. Pcrs: Personalized course recommender system based on hybrid approach. *Procedia Computer Science*, 125:518–524, 2018.

[22] Erum Ashraf, Selvakumar Manickam, Shankar Karuppayah, and Sufiana Khatoon Malik. An intelligent e-learning course recommendation framework based on student learning style. *Journal of Educators Online*, 20(1):n1, 2023.

[23] Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. Are we really making much progress? a worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM conference on recommender systems*, pages 101–109, 2019.

[24] Gongwen Xu, Guangyu Jia, Lin Shi, and Zhijun Zhang. Personalized course recommendation system fusing with knowledge graph and collaborative filtering. *Computational Intelligence and Neuroscience*, 2021:1–8, 2021.

[25] Yinping Ma, Rongbin Ouyang, Xinzheng Long, Zhitong Gao, Tianping Lai, and Chun Fan. Doris: Personalized course recommendation system based on deep learning. *PLoS One*, 18(6):e0284687, 2023.

[26] Dandan Shen, Zheng Jiang, et al. Online teaching course recommendation based on autoencoder. *Mathematical Problems in Engineering*, 2022, 2022.

[27] Nacim Yanes, Ayman Mohamed Mostafa, Mohamed Ezz, and Saleh Naif Almuayqil. A machine learning-based recommender system for improving students learning experiences. *IEEE Access*, 8:201218–201235, 2020.

[28] Sara Assami, Najima Daoudi, and Rachida Ajhoun. Implementation of a machine learning-based mooc recommender system using learner motivation prediction. *Int. J. Eng. Pedagog.*, 12(5):68–85, 2022.

[29] Yu-Sheng Su, Yu-Da Lin, and Tai-Quan Liu. Applying machine learning technologies to explore students' learning features and performance prediction. *Frontiers in Neuroscience*, 16:1018005, 2022.

[30] Samina Sarwat, Naeem Ullah, Saima Sadiq, Robina Saleem, Muhammad Umer, Ala'Abdulmajid Eshmawi, Abdullah Mohamed, and Imran Ashraf. Predicting students' academic performance with conditional generative adversarial network and deep svm. *Sensors*, 22(13):4834, 2022.

[31] P. Verma et al. A student-centric recommendation system based on a research analytics framework. *International Journal of Computer Applications*, 169 (11):23–28, 2017.

[32] Weller P. Jawaheer, G. and P. Kostkova. Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback. *ACM Transactions on Interactive Intelligent Systems*, 4(2):Article 8, 2014.

[33] L. Chen and G. Chen. Context-aware collaborative filtering recommender system. *Applied Sciences*, 12(1):45, 2014.

[34] Zhao Huang and Pavel Stakhiyevich. A time-aware hybrid approach for intelligent recommendation systems for individual and group users. *Complexity*, 2021:1–19, 2021.

[35] Salvador Valencia Rodríguez and Herna Lydia Viktor. A personalized location aware multi-criteria recommender system based on context-aware user preference models. In *Artificial Intelligence Applications and Innovations*, pages 30–39. Springer, 2013.

[36] Shuokai Li, Ruobing Xie, Yongchun Zhu, Xiang Ao, Fuzhen Zhuang, and Qing He. User-centric conversational recommendation with multi-aspect user modeling. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 223–233, 2022.

[37] Youngmo Kim, Seok-Yoon Kim, and Byeongchan Park. A serendipity recommendation method for book categories using bert. 2023.

[38] Fabio Celli and Bruno Lepri. Is big five better than mbti? a personality computing challenge using twitter data. *Computational Linguistics CLiC-it*, 2018:93, 2018.

[39] Luiz Fernando Capretz. Personality types in software engineering. *International Journal of Human-Computer Studies*, 58(2):207–214, 2003.

[40] Alice M. Garden. Career satisfaction and mbti types: A correlation study. *Journal of Career Development*, 24(3):11–20, 1997.

[41] Itamar Gati and Lisa Asulin-Peretz. Internet-based self-help career assessments and interventions. *Journal of Career Assessment*, 19(3):259–273, 2011.

[42] Yu-Che Chang, Hsu-Min Tseng, Xaviera Xiao, Roy Y. L. Ngerng, Chiao-Lin Wu, and Chung-Hsien Chaou. Examining the association of career stage and medical specialty with personality preferences – a cross-sectional survey of junior doctors and attending physicians from various specialties. *BMC Medical Education*, 19(1):1–9, 2019.

[43] Sri Handayani and Kuntarti. The influence of personality traits on caring behavior among nurses in indonesia. *Journal of Nursing and Health Sciences*, 11(2):75–82, 2022.

[44] Asma Ali Al-Salmani, Asma Al-Shidhani, Najlaa Jaafar, and Abdulaziz Al-Mahrezi. Factors associated with choice of career in family medicine among junior doctors in oman. *Sultan Qaboos University Medical Journal*, 20(3): e337–e344, 2020.

[45] Martin Katz et al. Career counseling using the mbti. *Journal of Career Assessment*, 7(3):255–267, 1999.

[46] Teresa Lesiuk. The relationship between personality and music majors: Understanding personality in academic disciplines. *Journal of Music Therapy*, 47(2):137–153, 2018.

[47] M. Omar, A. R. Fadzil, R. H. Shahar, and F. A. Hamid. Analyzing personality types to predict team performance. In *2010 International Conference on Science and Social Research (CSSR 2010)*, Kuala Lumpur, Malaysia, 2010. IEEE.

[48] Hyun-Soo Jang, Dong-Ho Kim, and Ji-Hyun Lee. The impact of mbti counseling on academic achievement. *Journal of Educational Psychology*, 45(2): 123–136, 2011.

[49] Thomas Goetz, Nathan C. Hall, Anne C. Frenzel, and Reinhard Pekrun. Personality and educational outcomes: A longitudinal study. *Educational Psychology Review*, 32(3):643–661, 2020.

[50] Theresa Lesiuk. The myers–briggs personality types of veterinary students and their animal ethical profiles in comparison to criminal justice students in slovenia. *Veterinary Sciences*, 9(8):441, 2018. doi: 10.3390/vetsci9080441.

[51] Jung Shin. The impact of mbti personality types on academic performance. *Journal of Educational Psychology*, 56(3):278–290, 2022.

[52] Susan E. Cross, Mark Trueman, and Mark Wilkinson. The impact of personality types on academic and career success. *Journal of Applied Psychology*, 91(5):1005–1018, 2006.

[53] T. Creasy and V. S. Anantatmula. From every direction—how personality traits and dimensions of project managers can conceptually affect project success. *Project Management Journal*, 44(6):36–51, 2013. doi: 10.1002/pmj.21372.

[54] Xuechao Li, Rodrigo Sardinas, Po-Chou Shih, and Karl Camp. Influence of introversion and extraversion using mbti personality model on academic performance. *International Journal of Information and Education Technology*, 8(9), 2018.

[55] Shyrlene Vaida, Yanhua Liu, and Mark Jones. Influence of introversion and extraversion using mbti personality model on academic performance. *International Journal of Information and Education Technology*, 7(9):123–129, 2011.

[56] Shyrlene Vaida, Yanhua Liu, and Mark Jones. Creating optimal groups of students in e-learning using alternative grouping methods. *International Journal of Information and Education Technology*, 7(9):123–129, 2011.

[57] Kathryn F Trenshaw and Troy J Vogel. Mitigating chemical engineering design team miscommunications with knowledge of myers-briggs type. In *2014 ASEE Annual Conference & Exposition*, pages 24–909, 2014.

[58] Richard M Felder, Gary N Felder, and E Jacquelin Dietz. The effects of personality type on engineering student performance and attitudes. *Journal of engineering education*, 91(1):3–17, 2002.

[59] Mohammad Salehi et al. Impact of personality traits on academic performance in stem education. *Journal of Educational Psychology*, 112(2):321–333, 2020.

[60] Olivia Palid, Sarah Cashdollar, Sarah Deangelo, Chu Chu, and Meg Bates. Inclusion in practice: a systematic review of diversity-focused stem programming in the united states. *International Journal of STEM Education*, 10(1): 2, 2023.

[61] David L. Blustein et al. The influence of mbti personality types on career paths in stem fields. *Journal of Career Assessment*, 20(3):292–307, 2012.

[62] John C. Houtz et al. Personality types and learning preferences: A study of stem students. *Journal of Educational Psychology*, 95(2):430–445, 2003.

[63] Daniel Gordon Bates. *An Investigation of the Impact A ROV Competition Curriculum has on Student Interest in sTEm, Specifically Technology and Engineering*. Brigham Young University, 2016.

[64] Bette Grauer, Linda P Thurston, and Beth A Montelone. Ks-lsamp pathways to stem: A system approach to minority participation in stem. In *2015 ASEE Annual Conference & Exposition*, pages 26–1052, 2015.

[65] Lesia L Crumpton-Young, Anna V Elde, and Kate Ambrose. Mentoring practices proven to broaden participation in stem disciplines. In *2014 ASEE Annual Conference & Exposition*, pages 24–900, 2014.

[66] Amelito G Enriquez, Catherine Baker Lipe, and Brandon Price. Enhancing the success of minority stem students by providing financial, academic, social, and cultural capital. In *2014 ASEE Annual Conference & Exposition*, pages 24–529, 2014.

[67] David Ebert. *Fuzzy logic in recommendation systems: A flexible approach to matching and similarity assessment*. Springer, 2008.

[68] Yaqub Jafar Nanehkaran et al. Fuzzy-based telecom product recommender system (ftcp-rs). *Journal of Telecommunications and Information Technology*, 4:56–68, 2022.

[69] Jian Zhang et al. Simulated collaborative filtering recommendation system using cloud models and triangular fuzzy numbers. *Journal of Computer Science and Technology*, 29(3):450–457, 2014.

[70] Min Jiang et al. Personalized commodity recommendation method based on fuzzy semantics. *Journal of Intelligent Fuzzy Systems*, 35(3):3125–3135, 2018.

[71] Jun Xiong and Yu Zhang. Rough set web page recommendation algorithm based on fuzzy semantic logs. *Journal of Computer Science and Technology*, 24(4):634–639, 2009.

[72] Zhiyong Cui et al. Hesitant fuzzy sets for measuring product similarity in recommendation systems. *Applied Soft Computing*, 100:106926, 2021.

[73] Punam Bedi et al. Hybrid multi-agent recommender system using interval type-2 fuzzy sets. *Expert Systems with Applications*, 40(8):3096–3102, 2013.

[74] Arda Tezcan et al. Integrating fuzzy matches in neural machine translation through data augmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 32(12):5556–5567, 2021.

[75] Li Zhang et al. Defining consumer profiles using fuzzy technology and fuzzy

ahp in intelligent garment recommendation systems. *Journal of Fashion Marketing and Management*, 26(4):572–589, 2022.

[76] Mohammed Alshurideh Almaiah et al. The challenges of introducing e-learning systems in developing countries: A comparative study. *Education and Information Technologies*, 25(6):5671–5690, 2020.

[77] Cynthia N Fuhrmann, DG Halme, PS O'sullivan, and B Lindstaedt. Improving graduate education to support a branching career pipeline: Recommendations based on a survey of doctoral students in the basic biomedical sciences. *CBE—Life Sciences Education*, 10(3):239–249, 2011.

[78] Boakye Kissi-Abrokwah et al. Determinants influencing career paths among students in ghana. *Journal of Career Development*, 45(2):123–140, 2018.

[79] A. G. Watts and R. G. Sultana. Career guidance policies in 37 countries: Contrasts and common themes. *International Journal for Educational and Vocational Guidance*, 4(2-3):105–122, 2004.

[80] Ayesha Zahid et al. The role of career guidance in higher education: Addressing skills supply and demand in pakistan. *International Journal of Educational Development*, 70:102076, 2019.

[81] Kari Bakke. The influence of gender and sociocultural norms on career aspirations among norwegian teenagers. *Journal of Youth Studies*, 24(2):153–171, 2021.