Ministry of Education and Science of the Republic of Kazakhstan
Suleyman Demirel University

Zhumat Abdimaulenov

Estimating Student's Motivation Level from Accumulated Data in The Context of Learning

A thesis submitted for the degree of
Master in Computer Science
(degree code: 7M06102)

Kaskelen, 2022

Ministry of Education and Science of the Republic of Kazakhstan
Suleyman Demirel University
Faculty of Engineering and Natural Sciences

Estimating Student's Motivation Level from Accumulated Data in The Context of Learning

A thesis submitted for the degree of
Master in Computer Science
(degree code: 7M06102)

Author: **Zhumat Abdimaulenov**

Supervisor: **PhD Birzhan Moldagaliyev**

Dean of the faculty:
Associate Professor, PhD Azamat Zhamanov

Kaskelen, 2022

Suleyman Demirel University

Faculty of Engineering and Natural Sciences

Department of Computer Science

Dean of Faculty

Associate Professor, PhD Zhamanov A.

«20» 06 2022

Topic of the thesis:

Estimating Student's Motivation Level from Accumulated Data in the Context of Learning

Thesis submitted as part of the requirements for the award of the MSc in "7M06102 - Computer Science", SDU, 2020-2022

Head of Department Assoc. Prof. Cemil Turan

Academic Supervisor PhD Birzhan Moldagaliyev

Master's student Zhuemat Abdimaulenov

Kaskelen, 2022

Abstract

The motivation of students to perform academically is closely linked to their academic performance. Teachers should identify students with low academic motivation as early as possible, just as they should identify those with an extremely high level of motivation. The purpose of this article is to identify the influence of academic motivation on behavior in learning management systems (LMS) courses as a means of developing a classification model to predict students' academic motivation. The research was conducted by students studying on the online platform eduway.kz. The classifiers were created using machine learning methods (ANN, decision trees and SVM). To determine if there was a significant difference between the models, a t-test was conducted. While all classifier models were successful in predicting motivation based on students' behavior in a learning management system course, the neural network model had the highest success rate.

Аңдатпа

Оқушылардың оқуға деген ынтасы олардың оқу үлгерімімен тығыз байланысты. Мұғалімдердің мотивация деңгейі жоғары оқушыларды анықтағаны сияқты, академиялық мотивациясы төмен оқушыларды да мүмкіндігінше ерте анықтауы маңызды. Бұл жұмыстың мақсаты студенттердің академиялық мотивациясын болжау үшін классификация моделін әзірлеу құралы ретінде оқытуды басқару жүйелері (LMS) курстарында оқушылардың академиялық мотивациясы мен олардың оқу барысындағы іс-әрекеттерімен байланысты анықтау болып табылады. Зерттеуге eduway.kz онлайн платформасында оқитын оқушылар қатысты. Машиналық оқыту классификаторы ретінде нейрондық желілерді, шешім ағаштарын және тірек векторлық машиналарды пайдаланылды. Барлық классификатор модельдері оқытуды басқару жүйесі курсындағы іс-әрекеттеріне негізделген студенттің академиялық мотивациясын анықтауда сәтті орындалды, әсіресе нейрондық желі моделі ең жоғары көрсеткішке ие болды..

Аннотация

Академическая мотивация студентов тесно связана с их успеваемостью. Учителя должны как можно раньше выявлять учащихся с низкой академической мотивацией, так же как учащихся с высоким уровнем мотивации.Целью данной статьи является выявление связей между академической мотивацией студентов и их поведением на курсах систем управления обучением (LMS) в качестве средства разработкп моделп классификации для прогнозирования академической мотивации студентов. В исследование участвовали ученики, обучающиеся на онлайн-платформе eduway.kz. Классификаторы машинного обучения были созданы с использованием нейронных сетей, деревьев решенпй и машин опорных векторов. В то время как все модели-классификаторы успешно определяли академическую мотивацию учащихся на основе их поведения в курсе системы управления обучением, модель нейронной сети имела самый высокий уровень успеха.

Contents

Chapter 1

Introduction

1.1 Motivation

Motivation for learning activity is a complex, multidimensional structure that includes not only motives, but also goals, strategies for responding to failures, perseverance, cognitive components and mechanisms. Motives are understood as both conscious and unconscious reasons for the activity carried out by the subject, behind which are his needs and values. The motives of activity differ not only by the criterion of strength (intensity), but also by the criterion of content (quality) and place in the hierarchy. The correlation between the content of motives and the nature of the activity performed determines the features of their influence on the effectiveness of the activity and the psychological well-being of its subject. The heterogeneity of educational motivation and, accordingly, the need to assess the quality of educational motivation was pointed out by L.I. Bozhovich, who identified two types of motives for learning activities - generated by the learning activity itself, related to the content and process of learning (cognitive) and broad social motives generated by the system of relations that exist between the student and the surrounding society. Classification L.I. Bozhovich during the second half of the 20th century was constantly refined in the works of A.K. Markova, M.V. Matyukhina, P.M. Yakobson, A.I. Savenkov. To date, there is no generally accepted idea of specific subtypes of cognitive and social learning motivation, there are no unambiguous grounds for their selection and the possibility of their reliable measurement. The value of the classification of motives proposed by L.I. Bozhovich, consists in distinguishing two fundamentally different types of educational motives - cognitive, or internal, inherent in the activity itself; and social, or external to it. The latter were subdivided into two types - broad social and narrow personal motives, which have different age dynamics and different influences on the effectiveness of the educational process.

In foreign psychology, internal motivation has been intensively studied within the framework of various theoretical paradigms as an independent entity (A. Gottfried, M. Csikszentmihyi, etc.), as well as in opposition to external motivation

(S. Harter). In the latter case, internal and external educational motives were simply opposed, considered as two poles of a dichotomy: on one lay the individual's interest in the activity itself, and on the other, the interest in the activity was caused by external reasons, while the motives of the first type were considered as conducive to the implementation of the activity, and the second - as interfering with it. With internal motivation, the activity performed in itself is of interest and value to the individual, it gives pleasure. In the case of external motivation, the activity performed is a means of achieving other results, external in relation to its content, to which the individual aspires. Studies show that intrinsic activity motivation is positively associated with creative problem solving, conceptual thinking, and cognitive flexibility, and is a predictor of more effective learning strategies, persistence, productive coping strategies, as well as academic achievement and psychological well-being. With internal motivation, the performance of an activity is accompanied by feelings of flow, joyful enthusiasm for the process of activity itself, enthusiasm, pleasure.

Overcoming the initial opposition of intrinsic and extrinsic motivation and identifying qualitatively different types of extrinsic motivation is proposed within the framework of the theory of self-determination. This theory argues that the external motivation of activity is a heterogeneous formation in which four types of regulation can be distinguished - external, introjected, identified and integrated, characterized by varying degrees of frustration of the need for autonomy (its maximum frustration occurs with external regulation). At the same time, the authors of the theory understand the need for autonomy as the desire to be an active subject, a source of one's own activity, to control one's behavior. As shown by studies conducted within the framework of the theory of self-determination on the material of different types of activity, in contrast to external identified, integrated and internal motivation (the so-called autonomous forms of regulation), external introjected and external motivation (controlled forms of regulation) are associated with less productive reactions to difficulties, less achievement in activities, relatively low psychological well-being, and a tendency to drop out of school prematurely. Ideas about the content and structure of achievement motivation, as well as specifically learning activities, were developed in the works of T.O. Gordeeva. Internal and external motivation of educational activity consist of motives that are relevant to the content of educational activity or external to it, respectively, aimed at satisfying or overcoming the frustration of other needs that are not directly related to the activity being performed. Internal educational motivation is a relatively homogeneous formation and is set by motives, which are based on the desire to satisfy human needs in knowledge, achievement and self-development. External motives are set by various kinds of external needs of the individual in relation to the educational process, the most characteristic of which are the needs for respect, autonomy and acceptance.

From the data obtained to date in the field of motivation of educational activity, it is clear that the allocation of types of external motivation according to the parameter of autonomy is certainly useful for understanding the mechanisms and predicting the well-being and success of the activity of the subject of the educational process. However, the selected types do not exhaust the whole variety of educational motives. Thus, one of the most characteristic types of external educational motives that regulate the performance of educational activities is the desire to achieve respect and recognition of significant others, as well as self-esteem due to the achievement of high results in activities.

Existing methods for assessing learning motivation:

1. The SRQ-A scale is designed to assess the motivation of primary and secondary school students, it does not include an indicator of amotivation.
2. The SIMS technique (the Situational Motivation Scale) is intended for situational assessment of intrinsic and extrinsic motivation, however, it contains the author's understanding of identified and external regulation, and there is no scale of introjected regulation. Interestingly, a number of authors, on the contrary, offer methods in which two subtypes of the latter are distinguished.
3. The AMS methodology (The Academic Motivation Scale) has French and English versions in versions for schoolchildren and students. A specific feature of AMS is the identification of three types of intrinsic motivation - the motivation for learning, achieving and experiencing the flow. However, the validity of the last subscale turned out to be low, and this led to the fact that in recent years only one scale of intrinsic motivation (cognition motivation) has been used. In addition, it was shown in different samples that the reliability of the scale of identified regulation in AMS was not high enough. The main problem lies in the fact that this technique is based on an understanding of introjected and external forms of external regulation, which is essentially different from that accepted in the framework of the theory of self-determination.

Contradictory operationalization of various types of extrinsic motivation identified in the theory of self-determination leads to different results of empirical studies obtained using these methods and generally does not contribute to the development of the theory and a better understanding of the phenomenon of extrinsic motivation.

The theoretical basis of many studies of motivation is the theory of self-determination by E. Deci R. Ryan [9]. The main types of motivation according to this theory are internal, external motivation and amotivation, which reflects the desire of the subject to leave the activity. The fundamental difference between the understanding of motivation from the standpoint of the theory of self-determination is that the so-called motivational continuum is considered, and not the opposition of external and internal motivation. The motivational continuum extends from amotivation to extrinsic motivation and its various forms and then to intrinsic motivation. Different types of motivation reflect the degree of subjective determination of behavior or self-determination. Later, the theory was developed thanks to R. Vallerandet [21], who expanded the understanding of intrinsic motivation. As a result, the characteristics of motivation were divided into general and particular factors of motivation. Internal, or introjected, motivation is understood as the desire for pleasure and pleasant emotions and includes cognitive motivation, motivation for achievement and self-development. External, or external, motivation describes situations where behavior is performed in order to achieve a goal and receive rewards from others, and includes motivation based on duty and self-esteem, as well as motivation based on the requirements of society, the group to avoid possible problems. Kornienko D.S., Derish F.V. Academic Motivation and Properties of the Dark Triad of Personality to withdraw from activities and is not divided into smaller factors.Based on these provisions, the Academic Motivation Scale was developed, which includes seven characteristics of motivation.When testing the validity of the methodology, the structure of academic motivation was reproduced [12]. only academic, but also sports motivation.In particular, on the basis of the existing scale, a methodology for studying sports motives was developed, which is quite common in the study of sports activities [13], we can state that the introjected motivation of polo is strongly associated with neuroticism, obligation motivation is associated with extraversion and conscience, and intrinsic motivation is generally associated with conscience [34]. A comparison of students with a predominance of intrinsic and extrinsic motivation according to the Big Five traits showed that intrinsically motivated students are friendly and open to experience, while extrinsically motivated students are emotionally unstable. are more conscious and open to experience, while the externally motivated are more likely to be extroverted and emotionally unstable. Students in whom amotivation predominates demonstrate less benevolence and consciousness. In domestic psychology, the features of internal and external academic motivation in the context of the theory of self-determination are considered in detail by T.O. Gordeeva [46]. there is no educational success. The study of the connections and relationships of various types of motivation with personal, social and cognitive characteristics will, according to T.O. Gordeeva [46], expand the understanding of the essence of motives.

Academic motivation refers to a person's behaviors with regard to academic performance [44]. The self-determination theory [9] serves as a theoretical framework in this study to understand academic motivation from the viewpoint of intrinsic and extrinsic motivations. Academic motivation has been well recognized to influence student academic performance. In higher education spaces, academic motivation is well known to affect student performance, but is generally overlooked. However, recognizing student academic motivation may very well assist educators to provide high-quality learning experiences and improve student performance. The issue of predicting student academic motivation has not been addressed by many researchers, despite considerable research dealing with academic motivation. Researchers typically employed traditional statistical methods. The literature also fails to adequately address student motivation in an online environment.

And this scientific work will answer the question: how to measure academic motivation; which machine learning models are more suitable for predicting the level of academic motivation; how suitable the data of the educational process are for predicting motivation.

1.2 Aims and Objectives

This work has two goals. Using methods of machine learning and student's behaviour data in the learning management systems as variables, we are exploring the possibility of creating a predictive model for student academic motivation based on student academic motivation. Secondly, we are seeking to analyze the efficacy of the machine learning models obtained in predicting student academic motivation in order to discover the best model. This study's objectives are to provide educators with an insight into the data from learning management system associated with academical motivation in order to use this data and to improve teaching and learning. Furthermore, this study analyzes how student behavior in education relates to detecting motivation of students. Through the focus on student behavior within learning courses and exploitation of ML methods, the paper fills a gap in the literature by developing a model that can accurately predict student academic motivation.

1.3 Thesis Outline

The first chapter is the Introduction chapter. In the introduction, it was written about the general understanding and the main characteristics of academic motivation. Types and approaches to measuring academic motivation. Also in this chapter my goals were written, why we decided to write on this topic. The second chapter is a Literature review. In this chapter, it was written about the works, one way or another related to my topic and written by scientists. And the third chapter was devoted to Machine learning methods. This chapter has written about different machine learning methods and how they work, as well as their advantages and disadvantages. The fourth chapter is about Learning Management System (LMS). This chapter has talked about the different types of LMS and directly about the LMS used in this work. The fifth chapter is devoted to Sample and research methodology. The sixth chapter is about the Results. And in the chapter Conclusion we complete our conclusion and further ways of development.

Chapter 2

Literature review

The information contained in LMS log files has been exploited for various research purposes by researchers. Many of the studies examine student performance (e.g. [38, 37]), but some studies also aim to increase student retention [27], measure a sense of community within classes [4] and extract quality characteristics of learning management systems (LMS) courses [45]. Since this review examines the use of classification methods in determining motivation of students within an online learning environment, it presents some results from several research works used log files and various methods to determine a student's academic motivation. Despite numerous studies examining motivation's relationship with academic performance (for example [1, 2], [11], [30], [35], [41]), predicting student motivation with LMS logs has been uncommon. Cocea and Weibelzahl [5] analyzed log files in order to determine whether they could predict motivation for a given student using five factors (ID, correct answers percentage in quizes, page views time, page visit number, spending time on quizes, and motivation). The researchers used their own methodology for assessing motivational levels, and emphasized that this was one of the main limitations of their study. Other approaches to assessing motivation levels were proposed by these authors in subsequent work ([7], [6]). Based on a review of 48 student log files, activities and timestamps were used to find out whether students were engaged, neutral or disengaged [7]. Data mining techniques analysed included Bayesian networks using K2 method, regression, classification, classification using J48 classifier, reduced-error pruning trees and decision trees with J48 classifier, badging using reduced-error pruning trees, and classification using regression trees and J48 classifier. They then factored in the amount of false answers on tests, the amount of true answers on tests, exams number, spending time on each exam, the amount of visited pages, the time average on each of them as important predictors of disengagement [7]. The size of the sample and the way students' motivation levels were determined in studies had limitations. Additionally, other researchers focused primarily on variables related to time. Among the limitations of Hershkovitz and Nachmias' study, they highlighted the lack of validation of the motivation classification based on engagement, energization and source of motivation. The following variables from the log files were analyzed using the clustering method: time spent on the platform, activity in sessions, duration of inactivity between sessions, pace of word work, exam results, and activity in game work. Tongsiri and Kularbphettong [25] also determined that student motivation is influenced by class activities. Using log files, they identified 19 variables largely related to the amount of time spent on certain activities and the grades received for doing those activities, then they used a decision tree to develop affiliation rules which helped to extract variables showing the influence on behavior. Researchers found data mining to be a useful tool to predict student outcomes. We used the Academic Motivation Scale (AMS) developed by Vallerand and co-researchers [21] to predict student motivation in this work. There are a number of studies in which the AMS is used to measure student motivation for academic success (e.g. [16], [24], [23], [28], [33]), but so far, there is no previous research that uses it to predict student motivation with log files. Additionally, this study differs from other studies in the type of input variables it uses, and while previous investigations have not addressed interdependencies among variables in the dataset, this research uses interdependencies to assure the reliability and effectiveness of the suggested model.

Chapter 3

Machine learning methods

3.1 Artificial neural networks (ANN)

The study and use of artificial neural networks, in principle, began quite a long time ago - at the beginning of the 20th century, but they really gained wide popularity a little later. This is due, first of all, to the fact that advanced (for that time) computing devices began to appear, the power of which was large enough to work with artificial neural networks. In fact, at the moment you can easily simulate a neural network of medium complexity on any personal computer. In neural networks, neurons are connected in a way that is specific to each individual. Here is an example of a neuron in Figure 3.1. The neuron is an element that

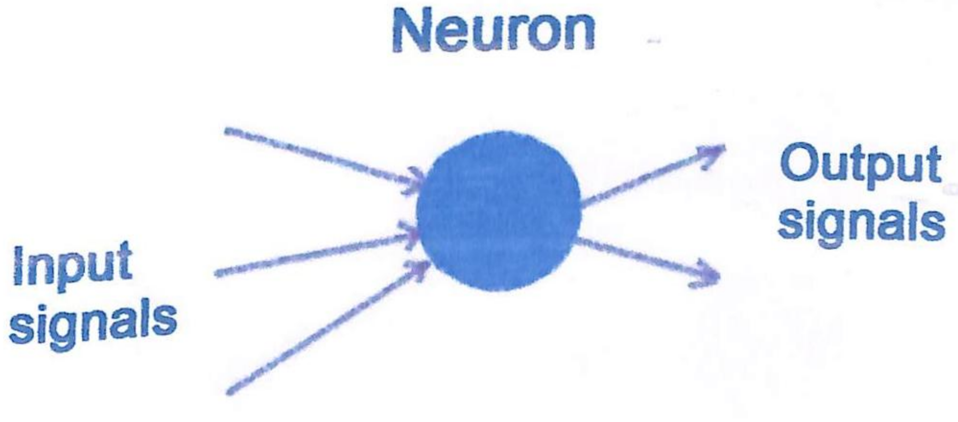

Figure 3.1: Representation of a neuron

produces a signal based on input signals according to a defined rule. This means that one neuron takes the following actions:

- Reception of signals from previous elements of the network
- Input signal combination

- Calculation of the output
- Sending the output signal to the next network element

The connections between neurons can vary depending on the network structure. In spite of this, the essential function of the neural network remains unchanged. Based on the totality of signals entering the network input, an output signal (or several output signals) is formed at the output. In this sense, you can consider a neural network as a black box with inputs and outputs. A great number of neurons are embedded within the black box. We have listed the main stages of the network, now let's dwell on each of them separately.
Since several input signals can come to each neuron, when modeling a neural network, it is necessary to set a certain rule for combining all these signals. And quite often the rule of summation of the weighted values of connections is used. What does weighted mean? Now let's figure it out ... Each connection in a network of neurons can be fully characterized using three factors:

- the first is the element from which the link originates
- the second is the element to which the link is directed
- the third is the weight of the connection.

The link weight determines whether the signal transmitted over this link will be boosted or attenuated. To explain simply, "on the fingers", then let's look at example in Figure 3.2. The output signal of neuron 1 is 5. The weight of the

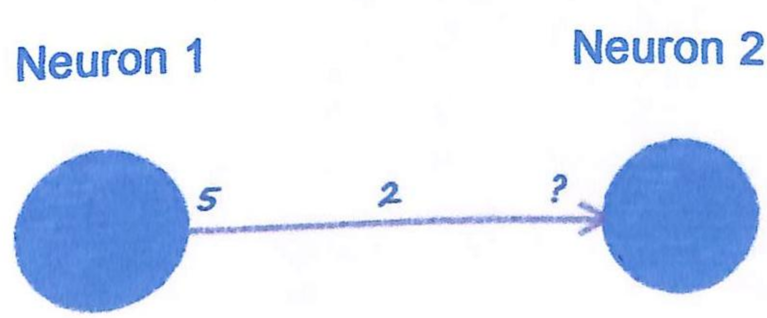

Figure 3.2: Link weight between neurons

connection between neurons is 2. Thus, to determine the input signal of neuron 2 coming from neuron 1, it is necessary to multiply the value of this signal by the weight of the connection (5 * 2). And if there are many signals, then they are all summed up. As a result, we get the following at the input of the neuron:

$$net_j = \sum_{i=1}^{N} x_i * w_{ij}$$

N is the number of elements that transmit their output signals to the input signal j. And w_{ij} the weight of the connection connecting neuron i with neuron j. By summing all the weighted inputs, we get the combined input of the network element. Most often, the structure of connections between neurons is represented as a matrix W, which is called a weight matrix. The matrix element w_{ij} in the formula, determines the weight of the link going from element i to element j. In order to understand how weight matrices are composed, let's look at a simple neural network in Figure 3.3.

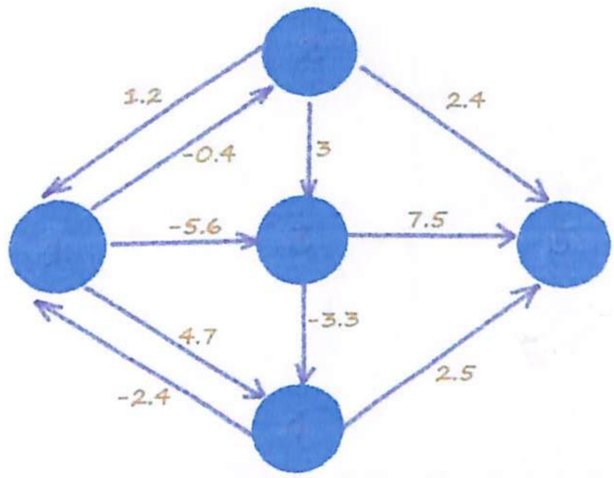

Figure 3.3: Example of neural network

The weight matrix of such a neural network will look like this:

$$W = \begin{bmatrix} 0 & -0.4 & -5.6 & 4.7 & 0 \\ 1.2 & 0 & 3 & 0 & 2.4 \\ 0 & 0 & 0 & -3.3 & 7.5 \\ -2.4 & 0 & 0 & 0 & 2.5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

For example, from the second element to the third there is a connection, the weight of which is 3. We look at the matrix, the second row, the third column - the number 3, everything is correct.

Consider the output signals. For each element of the network, there is a certain rule, according to which, from the value of the combined input of the element, its output value is calculated. This rule is called the activation function. And the output value itself is called the activity of the neuron. Absolutely any mathematical functions can act as activation functions, here are some of the most commonly used ones as an example:

- threshold function - if the value of the combined input is below a certain value (threshold), then the activity is equal to zero, if it is higher - to one.

- logistic function.

Let's consider one more small example, which is very often used in the literature to explain the essence of the operation of neural networks. 9 The purpose of the example is to calculate the XOR ratio using a neural network. That is, we will give different variants of signals to the input, and at the output we should get the result of the XOR operation for the input values (Figure 3.4). Elements 1 and 2

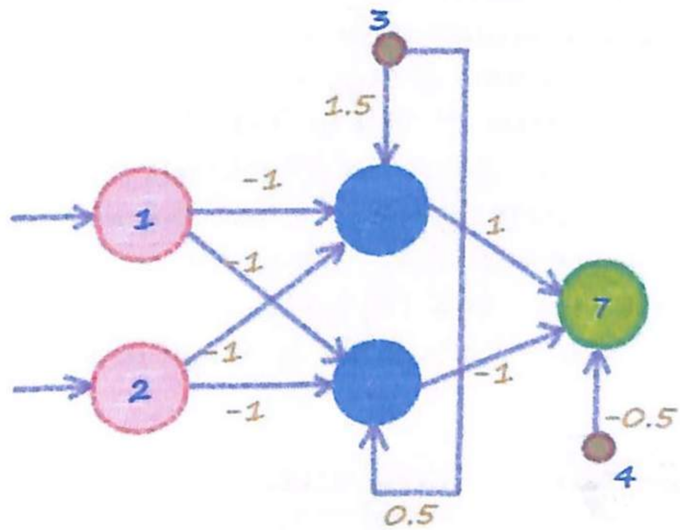

Figure 3.4: Example of XOR operation in neural network

are inputs, and element 7 is output. Neurons 5 and 6 are called hidden because they are not connected to the external environment. Thus, we got three layers - input, hidden and output. Elements 3 and 4 are called displacement elements. Their output (activity) is always 1. To calculate the combined input in this network, we will use the weighted link sum rule, and the threshold function will act as the activity function. If the combined input of the element is less than 0, then the activity is 0, if the input is greater than 0, then the activity is 1. Let's input one to the input of neuron 1, and zero to the input of neuron 2. In this case, the output should be 1 (0 XOR 1 = 1). Calculate the output value manually to demonstrate network performance.
Combined input of element 5:net5 = 1 * (-1) + 0 * (-1) + 1 * 1.5 = 0.5.
Element 5 activity: 1 (0.5 > 0).
Combined input of element 6: net6 = 1 * (-1) + 0 * (-1) + 1 * 0.5 = - 0.5.
Item 6 activity: 0.
Combined input of element 7: net7 = 1 * (1) + 0 * (-1) + 1 * (-0.5) = 0.5.
The activity of element 7, and at the same time the output value of the network, is equal to 1. This is what was required to be proved.
It's possible to try to use all possible values as input signals (0 and 0, 1 and 0, 0 and 1, 1 and 1), at the output we will always see the value corresponding to the truth table of the XOR operation. In this case, all the values of the weight

coefficients were known to us in advance, but the main feature of neural networks is that they themselves can correct the weight values of all connections in the process of network training.

A **fully connected neural network** has neurons transmitting their outputs to one another, including themselves. Every input signal is transmitted to each neuron. Several clock cycles in the network can produce the output signals of the network which can be all or some of the output signals of individual neurons.

Multilayer Neural networks (layers) are formed by condensing neurons into layers. Within each layer, there are a set of neurons sharing the same input signals. There is no requirement for any layer to have a specific number of neurons, and that number does not depend on that in other layers. Layers numbering from left to right compose the network in general. A network's inputs come from external signals, which are fed into its input layer (usually numbered zero) and its outputs are generated by its last layer. A multilayer neural network also includes a hidden layer besides the input and output layers. Connections from outputs of neurons of some layer q to inputs of neurons of the next layer (q+1) are called serial.

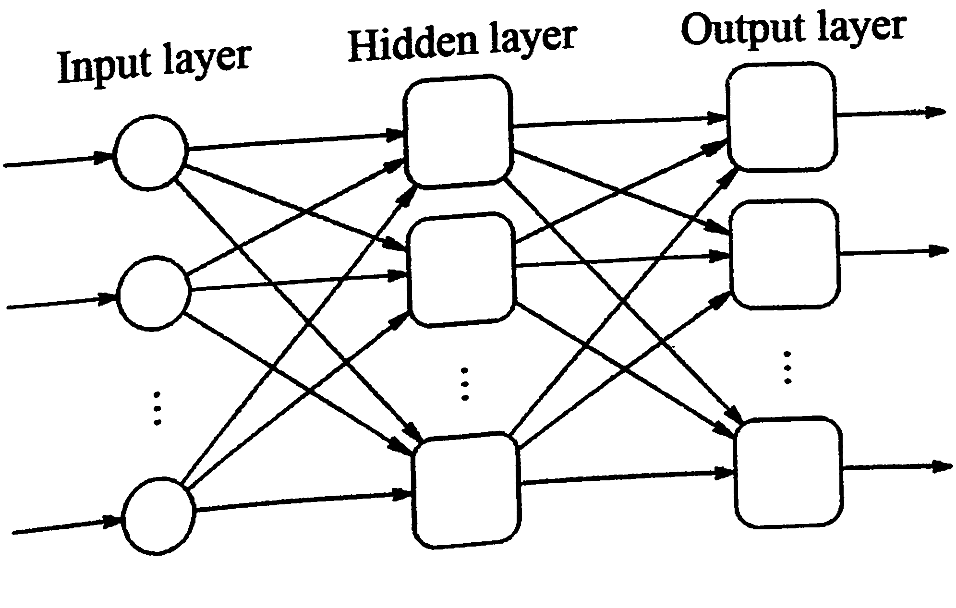

Figure 3.5: Multilayer Neural network

As ANN methods, the use of multilayer perceptron (MLP) and radial basic functions (RBF) played an important role. The components of an MLP network are the input, hidden, and output layers. MLP neural networks have only one hidden layer, but can have more or less hidden layers depending on the problem being solved. A neuron is not interconnected with neurons in any outer layers, but adjacent layer's neuron is connected with both outer layers through weighed connections. Haykin described that the basic characteristics are high connectivity and a differentiable activation function [18]. MLP networks are commonly trained using back-propagation algorithms. Training an MLP network entails creating an output adapted to a particular input and current weight, and then adjusting the weights according to the distinction between the output and target [29]. Hsieh describes the flow of signals in a MLP network describes the flow of signals in a MLP network [20]. Assuming that the x are inputs and the hidden layer is carried out by h neurons, an activation functions f and g, a weight parameter matrix w, an offset parameter b, and y as the output, then the following equations can be used to describe the flow of signals:

$$h_j = f\left(\sum_j w_{ji}x_i + b_j\right) \tag{1}$$

$$y_k = g\left(\sum_j \widetilde{w}_{kj}h_j + \widetilde{b}_k\right) \tag{2}$$

In this study, MLP neural network models were trained using backpropagation algorithms and conjugate gradient algorithms. Training involved various activation functions, including exponential, hyperbolic tangents, logistic functions, entropy, and identity functions. There were 3 to 20 hidden units, and 200 training cycles were fixed. The RBF neural network can also be referred to as a universal approximator, as it is one of the most popular alternatives to MLP [10]. It, like MLP, consists of three layers. RBF neural networks obtain input data from the input layer, but perform nonlinear transformations in the hidden layer. Used radial basis function for calculation of each unit [18]:

$$\varphi_j\,(x) = \varphi(\|x - x_j\|),\ \text{j=1, 2,..., N} \tag{3}$$

and x_j is the center of function. The input layer nodes do not have weighted connections with the hidden layer nodes [13]. Since the Gaussian function is often used as a radial basis function, it was chosen in this study. Accordingly, each unit within a hidden layer is described as:

$$\varphi_j\,(x) = exp\left(-\frac{1}{2\sigma_j^2}\|x - x_j\|^2\right),\ \text{j=1, .., N} \tag{4}$$

An x_j is the center of the jth Gaussian function with width j. Radial functional centers and weights for learning in RBF neural networks were also determined, as indicated in [10]. Both unsupervised and supervised learning were used as teaching methods [20]. A supervised algorithm minimizes the mean squared error (MSE) between output data and target data through supervised learning , whereas unsupervised learning determines the widths and centers of the radial-basis functions. RBF neural networks were configured with a minimum number of hidden units of 3 and a maximum number of 23 hidden units. Error functions were calculated using the cross entropy or sum of squares.

3.2 Classification tree

As a data structure, regression and classification trees are collectively known as Decision Trees (DT). They allow you to identify patterns in data so that you can leverage them. Decision trees are organized as a hierarchical structure consisting of decision nodes to evaluate the values of certain variables to predict the resulting value. The use of classification trees results in a symbolic designation of the class, and as a result of the use of regression trees, continuous values are returned. For training decision trees, sample data must be provided, so it is necessary to create or collect such data in advance. On the one hand, the data can be prepared by an expert, and on the other hand, the accumulation of a collection of facts related to the problem under consideration can be provided. Based solely on the interpretation (or classification) of such data, many applications, including AI models, can be created. In practice, each problem considered in this case can be represented using a set of attributes, for which the decision tree predicts an unknown attribute (solution).

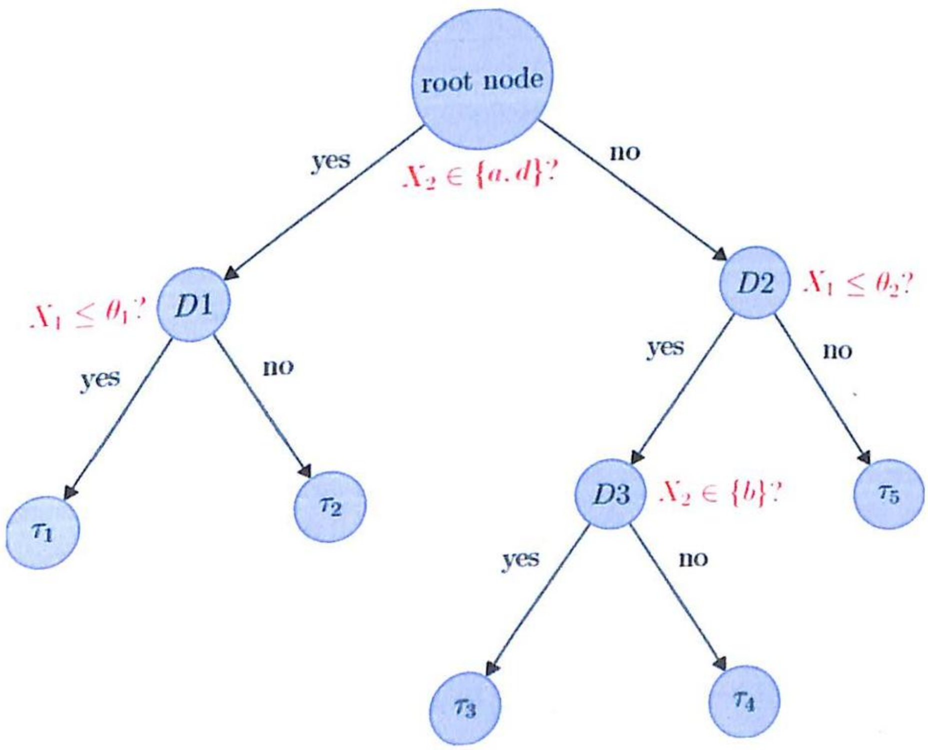

Figure 3.6: Decision tree

Decision trees can be used as a decision-making component during the implementation of game characters. A decision tree can provide the best course of action in any given situation. Each situation is represented by a set of attributes, so an analysis of the decision tree can provide a recommendation. In addition, regression trees can be used to assess the strengths of any entity (or to predict the outcome, either positive or negative). Before looking at what decision trees achieve, or even how they are constructed, it is necessary to understand the basic concepts behind the decision tree data structure.

Any decision tree outputs a predicted value resulting from the evaluation of some input attributes. Decision trees are classified into two different types: classification trees and regression trees. This difference does not depend on the types of input data, since trees of both types can take either continuous or symbolic values. The determining factor on which the tree type depends is the output value. A decision tree with continuous outputs is called a regression tree, while classification trees output specific values instead. The summary information that defines the differences between the types of decision trees is given in Table 3.1.

Table 3.1: Differences between the types of decision tree

Tree type	Prediction	Data type
Classification Tree	Discrete	Symbols
Regression Tree	Continuous	Real numbers

The first subsection of this chapter defines the concept of data sampling and shows its significance. Data samples are intended to be processed by decision trees and often also serve as input/output format definitions. The following sections look at ways to represent the structure of decision trees and reveal the secrets behind making decision trees so simple and fast. Finally, the most important parts of decision trees, such as decision nodes and leaf nodes, are explored in more detail.

Of course, without the use of data, not only pattern recognition, but also machine learning in general, becomes impossible. Datasets are often presented as separate samples. Such selections are sometimes called events, instances, patterns, and, of course, many other names are used to refer to them. In essence, any data sample is a set of attributes, which are also called prediction variables. Each attribute can be a continuous value (i.e. a floating point number) or a character (i.e. a set of unordered discrete values). It is also possible to apply additional attributes with special meaning, known as response variables (or dependent variables). Such an attribute can be expressed with a symbol representing discrete categories (in classification trees) or a continuous value (in regression trees). Response variables can serve as decision criteria for each sample in both classification and regression problems.

Any decision tree is essentially a tree graph, in the literal sense of this concept, formulated in computer science. This data structure consists of nodes connected to each other by edges (Figure 3.7). In this case, it is not allowed for the edges to form a cycle, since otherwise the tree turns into a graph that is different from a tree-like one (and when using such a graph for decision-making, difficulties arise).

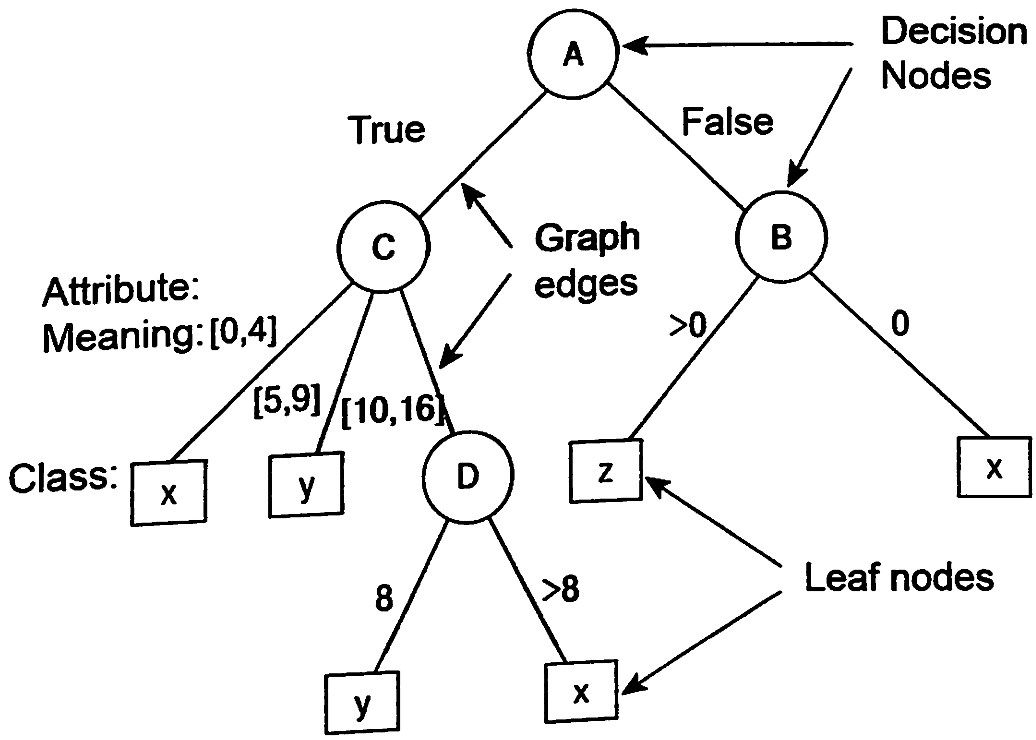

Figure 3.7: A simple decision tree with the root node shown at the top and each decision node represented by a circle (leaf nodes are indicated by rectangles)

There is one special node in the tree, known as the root. Essentially, this node is the base of the tree, since you can go from the root to any other node in the tree. Another special kind of node is the node at the end of any chain of consecutive edges, the leaf nodes.

The above description of the tree data structure is very broad. Any person with reasonable experience with programming languages should be familiar with this information. But if we are talking about the use of decision trees in artificial intelligence, then each of the above concepts takes on a special meaning.

Each level in the tree can be considered as one of the solutions; the decision node provides a test of the condition, and each edge denotes one of the possible options. More formally, it can be noted that the decision nodes contain selection criteria, and the edges express the mutually exclusive results of checking compliance with these criteria.

Essentially, each time the condition is checked, the data samples are sorted such that each data element is determined to correspond to only one edge. If all samples are considered as one common set of data, then the decision criteria break this set into non-overlapping subsets, as shown in Fig. 3.8 As a result of combining such checks into a certain hierarchy, the process of splitting all data into smaller and smaller parts is actually organized until a leaf node is reached. Each leaf node corresponds to a small but exclusive (non-repeating) part of the original set.

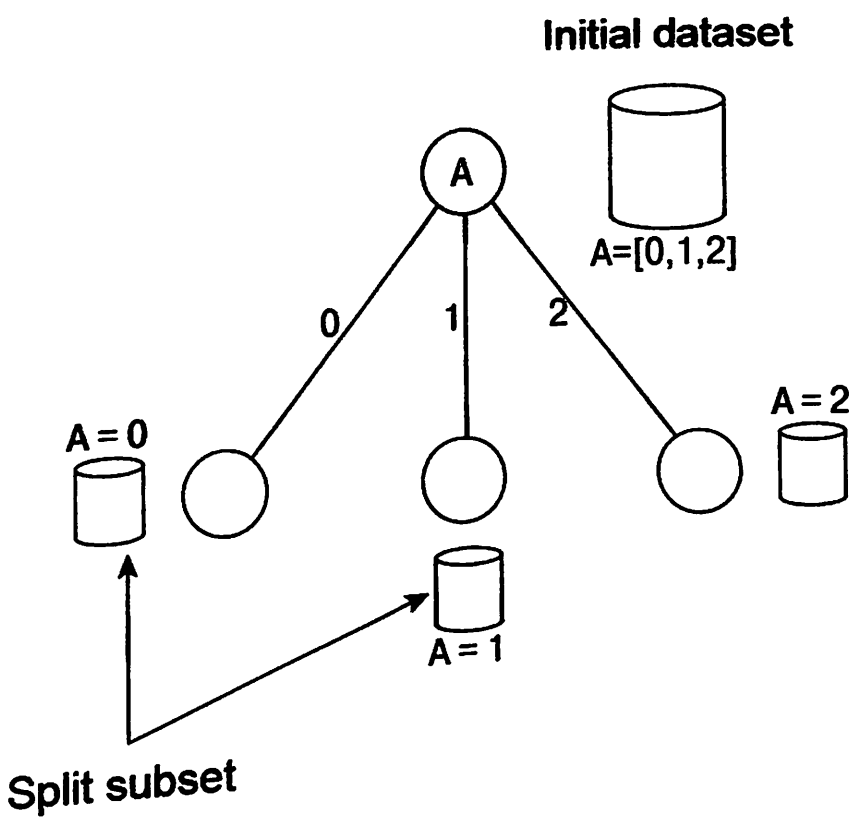

Figure 3.8: Splitting a data sample into mutually exclusive subsets using a decision node

There is no doubt that the number of possible ways to represent solutions must be quite large, so the choice of which way to use depends on the type of attribute being tested, as well as on the operation used in the condition test. Because attributes can be expressed as symbols or continuous values, the tests themselves can be organized as boolean conditions or continuous relationships. And the number of possible test results determines how many edges should come from the considered decision node. The most commonly used condition checks are listed below.

- **Checking a boolean value.** This test determines whether the application of a particular operator results in a true or false value. Obviously, the possible results of the test are true or false.
- **Sign definition.** When performing such a check, the sign of the expression is determined. The result can be either positive or negative. The specified check can be considered as a special case of checking a boolean value.
- **Class membership check.** When performing such a check, it is determined to which class the given symbol belongs. The result of the check is

the designation of one of the possible classes (the number of which can be arbitrary).

- **Validation of belonging to the range of values.** When conducting such a check, it should be clarified to which range of values this value belongs. Each of the possible outcomes indicates to which range, into which the entire range of the variable is divided, the given value belongs. Such a check can be thought of as a class membership check.

Typically, each decision node performs a single condition test covering only one attribute (as an example, you can specify the B==true test operation). This approach often becomes quite acceptable, since the decision tree allows you to create hierarchical combinations of checks to form more complex decision structures. However, some more complex varieties of decision trees allow for condition tests that consider more attributes (for example, the conditions A==1 and B<0 are checked simultaneously), which makes it possible to increase the accuracy of the tests, at the expense of complication. But in this case, the number of possible combinations increases, so the number of outgoing edges increases, as shown in Table 3.2.

Table 3.2: Four examples of simple condition checks and decision operations using more attributes

Condition check	Possible results
A (boolean value)	true
	false
B (sign)	B > 0
	B ≤ 0
C (belonging to the range)	C in [0..4]
	C in [5..9]
	C in [10..14]
A, B	A == true and B > 0
	A == false and B > 0
	A == true and B ≤ 0
	A == false and B ≤ 0

Generally speaking, an increase in the number of prediction variables used in a condition test at a decision node results in an increase in the number of outgoing edges. But it is possible to apply an arbitrary number of attributes and map them to a single boolean solution. This may require expressions such as (A and B) or C to be evaluated in the decision node. Similar expressions can also be used to convert continuous values to a single number. But the downside of using such complex expressions is that it makes the task of learning the decision tree more difficult.

The classification tree is a structure used to group objects into predefined classes based on their inherent characteristics [36]. There are a few main advantages listed in [20] of this method of classification. The classification tree presents an easy way to understand the relationship between the predictand and the predictors when they are used together. It should be noted that this method is suitable for reducing the number of predictors in the dataset [20]. It is made up of root nodes, test or internal, and terminals or leaf nodes. Identifying the difference between these nodes can be thoroughly understood by comparing incoming and outgoing edges. There is no incoming edge on the root node, whereas others have a single incoming edge [36]. Leaf nodes are distinguished from root and internal nodes. Leaf nodes differ in that they do not have outgoing edges, but one target class is assigned to each leaf node [36]. Nevertheless, the internal nodes include splits in which the predictor variable value is tested [32]. Decision trees begin their classification at the root of the tree and proceed to one leaf nodes where the decision is made [29]. As part of this investigation, adaptions of the CART (Classification and Regression Tree) algorithm and the QUEST (Quick Unbiased Efficient Statistical Trees) algorithm. Using the best leaf node and a predictor variable, were obtained univariate splits based on the discriminant. Therefore, it will help to determine the significance of the relationship between the predictor variable and class membership [22]. During the split process, the choice of the smallest p-value of the predictor variable is important [22]. The Gini measure of node impurity is the popular goodness measure used in CART (splitting criteria). According to [31], Gini is defined as follows when $l \in [1, k], nisonenode, pl(n)$ - the fraction of points at n that belong to l:

$$Gini\ (n) = \sum_{l=1}^{k} p_l(n)(1 - p_l(n)) \tag{5}$$

Additionally, Chi-square and G-square measurements of goodness were applied in addition to this measure. Split-selection tools based on discriminant-based univariate split and CART style searches for splits, with equal misclassification costs and estimated prior probabilities, were used. Direct stopping according to FACT (Fast Algorithm Classification Trees) was employed as a stop rule on 0.05 set as the fraction of objects. To select the split variables, the p-value was set at 0.05 and was performed 10-fold cross validation.

3.3 Support vector machines

There is a linear model for classification and regression problems known as the Support Vector Machine or SVM [42]. There are many practical problems involving linear or non-linear models that can be solved using this technique. In machine learning, support vector machines (SVM) are very powerful and flexible models that can be used to perform linear or nonlinear classifications, regressions, and outlier detections. Anyone interested in machine learning should have it in their toolkit, as it is one of the most popular models. SVM methods are particularly well suited for classifying complex but small to medium datasets.
A visual explanation of the SVM method is the best way to explain its fundamental idea. The iris dataset is shown in Figure 3.9. There is a clearly visible separation between the two classes (they are linearly separable). As can be seen on the left, the decision boundaries of the three linear classifiers are presented. In fact, a model that has dotted lines representing its decision boundaries cannot separate classes properly. The other two models perform great on this training set, but their decision boundaries are so close to the samples that these models probably won't perform as well on new samples. On the other hand, the solid line on the graph on the right indicates where the SVM classifier is expected to make a decision; the line not only indicates separation of two classes, but also indicates the distance from the nearest training sample. In a SVM classifier, the width of the band between two classes is set as wide as possible (represented by parallel dotted lines). It is sometimes referred to as margin-based classification.

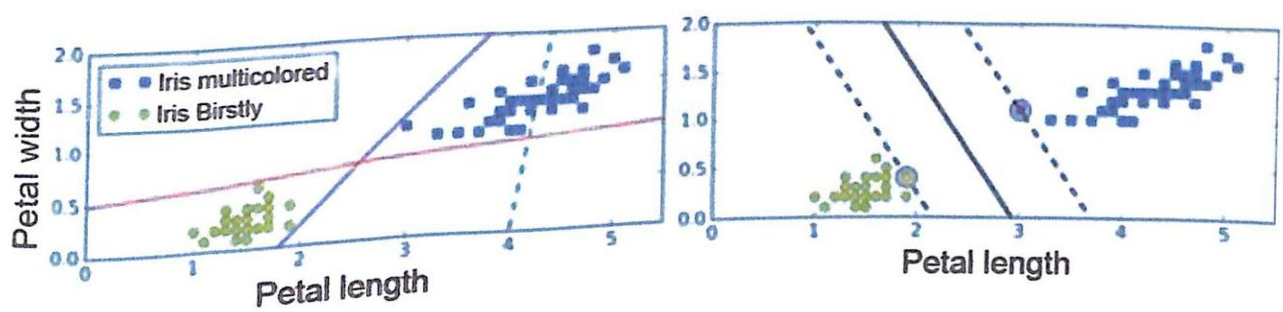

Figure 3.9: Wide gap classification

Note that adding additional "out-of-stripe" training samples will not affect the decision boundary at all: it is completely determined by the samples located at the edges of the strip (or "leans" on them). Such samples are called support vectors (support vector); in fig. 3.9 they are circled.
According to Figure 3.10, SVM methods exhibit sensitivity to feature sizes: the graph on the left has a vertical scale much greater than the horizontal, so the widest bar is near to the horizontal. After feature scaling (for example, using the StandardScaler class from Scikit-Learn), the decision frontier looks much better (in the plot on the right). If we strictly fix that all samples are outside the band and on the right side, then we get a hard margin classification. Two major prob-

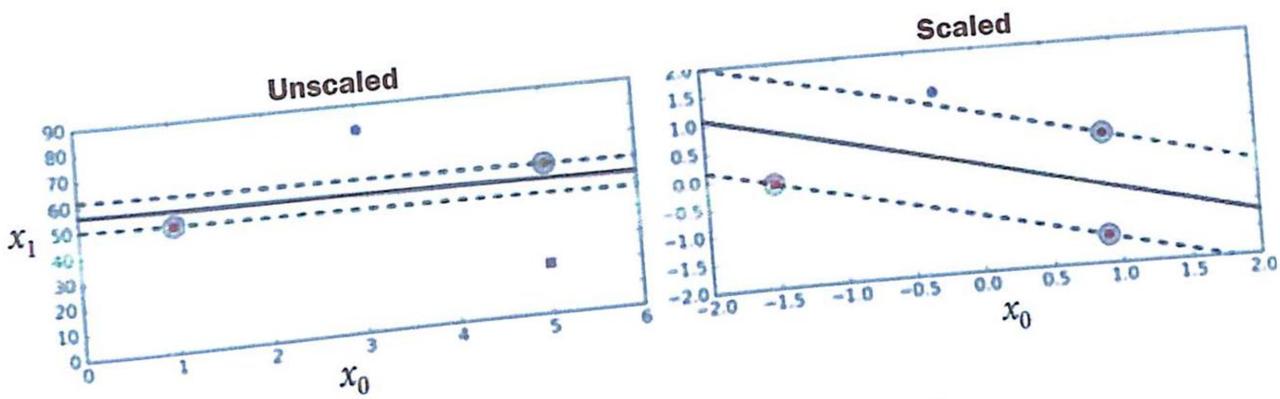

Figure 3.10: Sensitivity to feature scales

lems plague hard clearance classifications. Data must first be separable linearly for this algorithm to work. Secondly, it is quite sensitive to outliers. The figure 3.11 illustrates an iris dataset with just one additional outlier: on the left, the gap cannot be determined, whereas on the right, the decision boundary differs significantly from that shown in the previous figure. 5.1 without an outlier, and the model probably won't generalize as well. To avoid such problems, a more

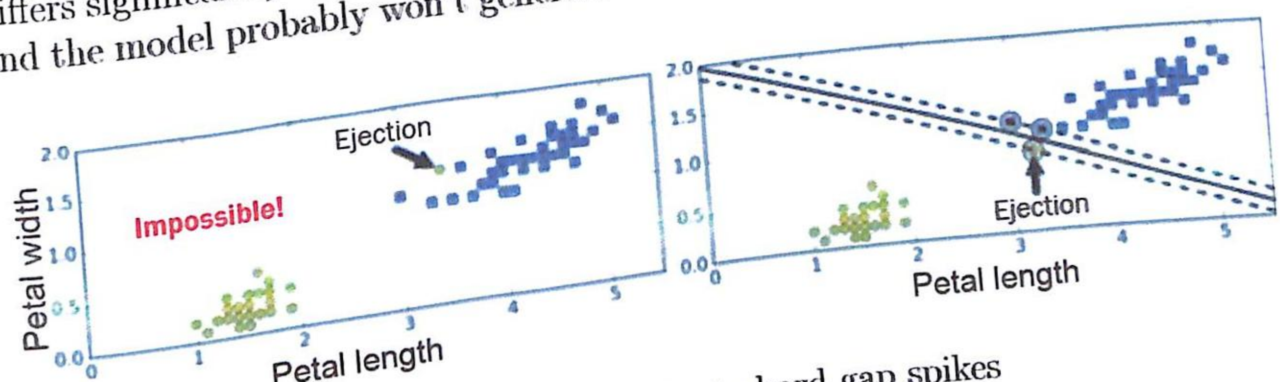

Figure 3.11: Sensitivity to hard gap spikes

flexible model is preferable. To reduce the number of gap violations (frankly, instances where the band is too narrow or is even too narrow) it is important to maintain a balance between keeping the band as wide as possible and limiting the number of instances where it is too narrow or even on the wrong side of the band. A soft margin classification is what this is called.

Scikit-Learn's SVM classes have the hyperparameter C to control the banding balance: a lower value of C results in a wider band, but more gap violations. An example of non-linearly separable data is shown in Figure 3.12 along with the decision boundaries and gaps for two soft-gap SVM classifiers. In the example on the left, the classifier has fewer gap violations when it uses a high C value, but has a smaller gap. On the right, due to the use of a low C value, the gap is much larger, but many samples fall into the band. Even on this training set, the second classifier seems to perform significantly better than the first, since most violation gaps fall within the decision boundary rather than on the incorrect side.

According to Hsieh [20], the development of SVM classifiers for two-class problems can be accomplished in three steps. The maximum margin classifier is introduced

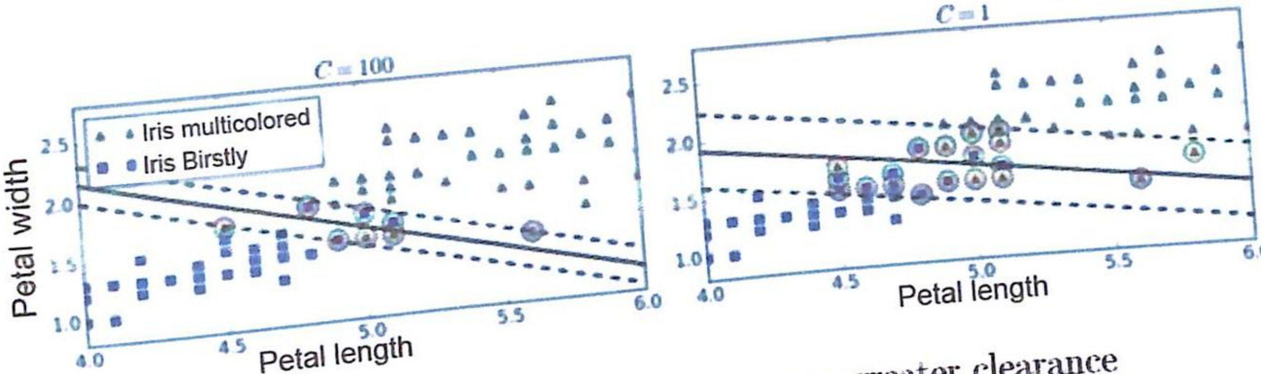

Figure 3.12: Fewer clearance violations or greater clearance

in the first step to differentiate between two classes for problems in which there is a linear decision boundary, and then the classifier is modified in the second step to allow misclassification, while the kernel classifier is used in the third step to generalize the maximum margin classifier nonlinearly [20]. (x_1, y_1) to (x_k, y_k), where x is the input space, and y is the class labels, the goal is to construct a hyperplane for the given set of data. According to [19], given a set of data (x_1, y_1) to (x_k, y_k) with $x \in R^n$ as input space and $y_i \in \{-1, 1\}$ as classification label, the goal is to define a hyperplane

$$\omega^T x + b = 0 \quad (6)$$

that, divides classes by maximising the distance between

$$\omega^T x + b = 1 \quad (7)$$

and

$$\omega^T x + b = -1 \quad (8)$$

that can be addressed by minimising $\frac{1}{2}\omega^T\omega$, resulting in the equation

$$min\frac{1}{2}\omega^T\omega + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \quad (9)$$

where ξ_i, ξ_i^* represent the upper/lower bounds of outputs, and C is regularizing constant of equation. In general,

$$K(x_i, x_j) = \phi(x_i)\phi(x_j) \quad (10)$$

is a kernel function whose purpose is to perform the mapping $x_i \in R^n$. A number of kernel functions were studied in this study, including:

RBF kernel [26]

$$K(x_i,x_j)=\exp\left(-\gamma\|x_i-x_j\|^2\right), \gamma \text{ is a width parameter } (\gamma > 0) \quad (11)$$

sigmoid kernel [43]

$$K(x_i, x_j) = \tanh(\gamma x_i x_j + \text{coefficient}),\ \gamma > 0,\ \text{coefficient} \geq 0 \qquad (12)$$

and polynomial [43]

$$K(x_i, x_j) = (\gamma x_i x_j + \text{coefficient})^{\text{degree}},\ \text{degree} \in N,\ \text{coefficient} \geq 0,\ \gamma > 0. \qquad (13)$$

All models were subjected to 10-fold cross-validation to avoid overfitting.

Chapter 4

Learning Management System (LMS)

4.1 Origin history

The compilation of special applications, allowing to make administrative changes in technological and technological terms, became an organic possible presentation of distant proposals and occurred in the second half of the 20th century in the academic century [8]. The introduction of the reason why the original LMS became a university, a sheet - a large number of students carried away huge resources for educational departments, but at the same time it became dedicated and resourced for software development. However, the prototypes of modern LMS arrests before the arrests of computers. The first documented time period of deviation dates back to 1728, when American professor Caleb Phillips received the result of a shorthand correspondence course through the Boston Gazette [15]. Already in the 1840s, distance learning became two-way - the famous British linguist Isaac Pitman began not to send assignments to students (also only from stenoblastography), but also to receive them back for verification. In 1856, the method broke out beyond the teaching of shorthand - the German Gustav Langenscheidt and the French Charles Toussaint singled out the first distance learning school for foreign languages, which turned out to be dangerous. With the advent of audio and video recording, new perspectives have opened up for distance learning. In Edward Forster's 1909 short story "The Machine Stops", an unexpected prediction of the advent of the Internet and, in particular, its use for lectures and examinations at a distance. Already in 1920, at the Ohio State University, Sidney Presley created a prototype of a mechanical "teaching machine" that allows students to be tested, [19] by 1938, twenty-five US states had centralized programs for distance learning of radio communications [15]. As electronics developed, it became possible to create an OMS within the framework of this term. In 1960, Donald Bitzer of the University of Illinois created the prototype of the first e-learning system, PLATO. By the end of the 1970s, the system had several thousand terminals around the world and more than a dozen mainframes connected in a common computer network. The emergence and development took place with the explosive growth of the LMS in the 1990s - electronic research went beyond the boundaries of universities and became the exception to the occurrence and detection of cases of suspicious activity in everyone. In 2004, the SCORM standard appeared, which allows the use of mixed materials from various courses [47].

4.2 Technical aspects

Initially, e-learning systems were developed for local hosting, in which the organization licenses the software (or, as in the case of the first systems, develops it itself) and installs it on its own servers. At the same time, for each LMS, its own course editor was created, which could not be used to write courses in another system. Most modern LMSs are web-based, and in order to ensure standardization, the content created for them (training courses) usually follows certain international formats, the most popular of which is SCORM (and its successor Tin Can). AICC and cmi5 are also common. The presence of these formats allows the use of content created using various editors in different LMS and provides the ability to integrate the LMS with other systems (for example, CRM). Currently, many LMSs are offered as SaaS (software as a service) and hosted by vendors[48].

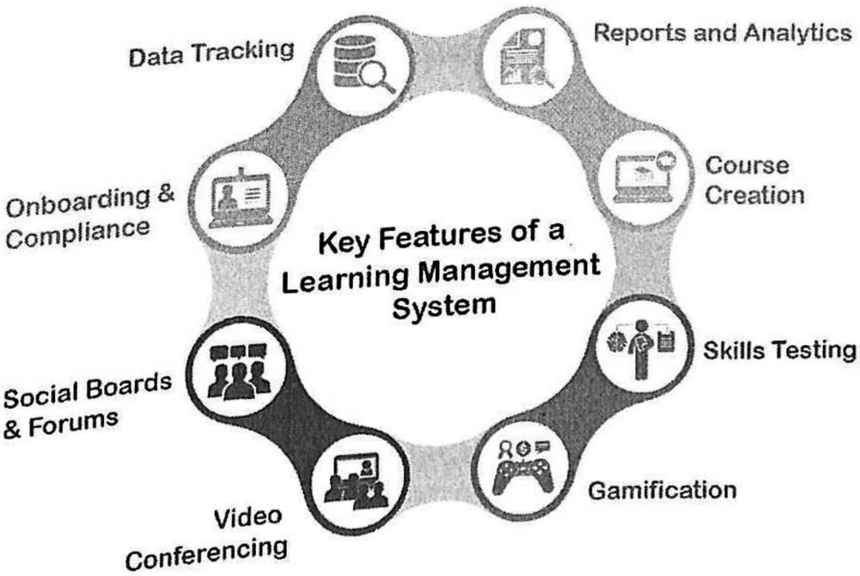

Figure 4.1: Key features of a Learning Management System

The first LMSs were created for a specific project and, most often, by the same organization (or its subsidiary), in which they were then used. With the advent of specialized companies supplying OMS (during the boom of the 1990s), universal solutions were developed that were used (with modifications and customizations) in various industries. The next stage in the development of the market was the specialization of the LMS in the areas of use, with the "sharpening" of the functionality to the relevant requirements.
Today, the following main groups can be distinguished [3] [14]:

- "Academic" - used in educational institutions;
- "Sales platforms" - online sale of B2C courses, usually associated with e-commerce systems;
- "Training platforms" - provision of training and educational services to third parties (B2B);
- "Corporate" - training of company personnel.
- "Industry" - both for students of educational institutions, and the staff of companies in a particular industry, direction.

4.3 Advantages and disadvantages

The main advantages of the LMS stem from the most accessible and differences from value:

- It is possible to study almost anywhere since students have access to the Internet. An adult student can study without interruption from the main job.
- Students can reduce their tuition fees by not reimbursing the expense of methodological literature. In addition, savings through wages, which are not necessary for educators, support educational institutions and so on.
- Flexibility in the learning process - the students and teachers can adjust the learning process as their abilities and needs dictate.
- The ability to keep one's skills and abilities up to date with modern technology and standards. E-courses also ensure the efficiency and effectiveness of the use of educational materials.
- Potentially high learning opportunities - learning becomes natural from the quality of teaching in an intensive learning process.

- The ability to determine objective criteria for assessing the perception of knowledge - the equipment has the ability to release clear characteristics, the assessment determines the knowledge gained by the student in the learning process.

The disadvantages of LMS (and e-learning in general) are, as often happens, a reflection of its merits [39]:

- The lack of direct communication between the student and the teacher makes it difficult to control the learning process and evaluate its results.
- The implementation of an LMS requires a well-built technological infrastructure. Educators should be prepared to adapt their curricula to e-learning.
- The role of the teacher's individual skill is decreasing.

4.4 Eduway.kz as LMS

In this work, the eduway.kz platform was used as a learning management system. This electronic platform has registered students studying both online and offline. The eduway.kz platform is designed to control student learning, collect and process student data in a single electronic database, introduce an electronic journal and provide the necessary resources for remote learning.

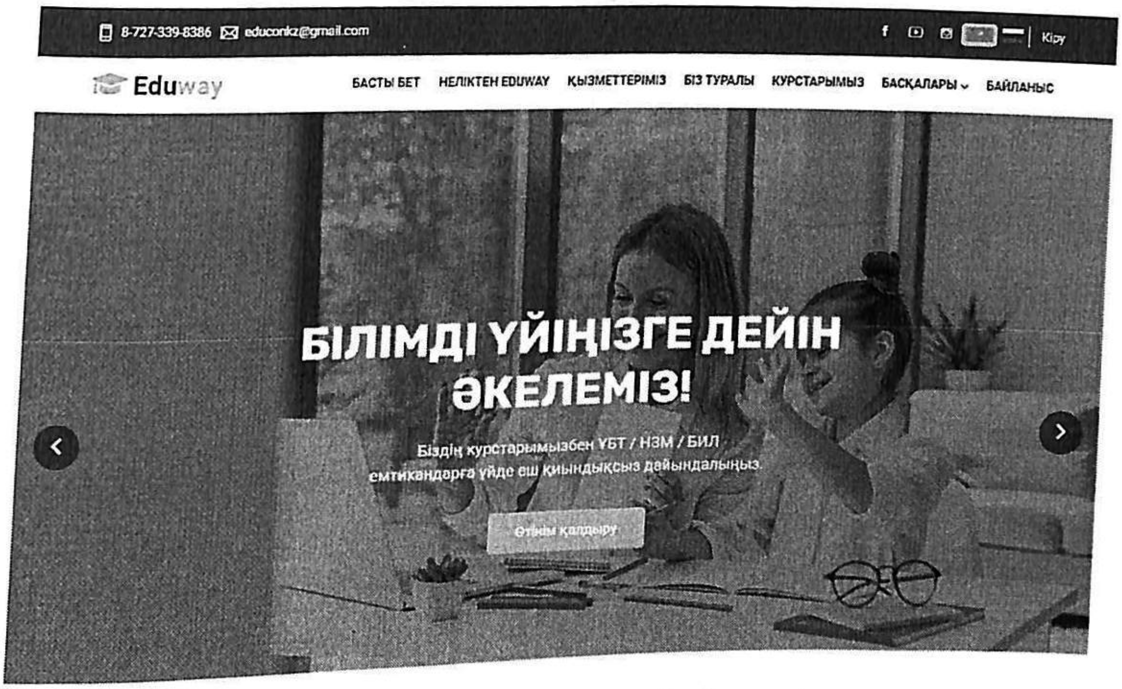

Figure 4.2: Home page of eduway.kz

There are a number of opportunities for training and management activities. If you list, then for the administration there are the following:

- Registration of students.
- Registration of teachers.
- Making timetables for students and teachers.
- Assign lessons to teachers.
- Choosing the right subjects for students.
- Automatic drawing up of an agreement between an educational organization and a client.
- Automated accounting.
- Acceptance of cash and online payments.
- Analysis of the conducted exams.

- Analysis of watching video lessons.
- Analysis of homework performance.
- Analysis of class attendance.
- Control of income and expenses.

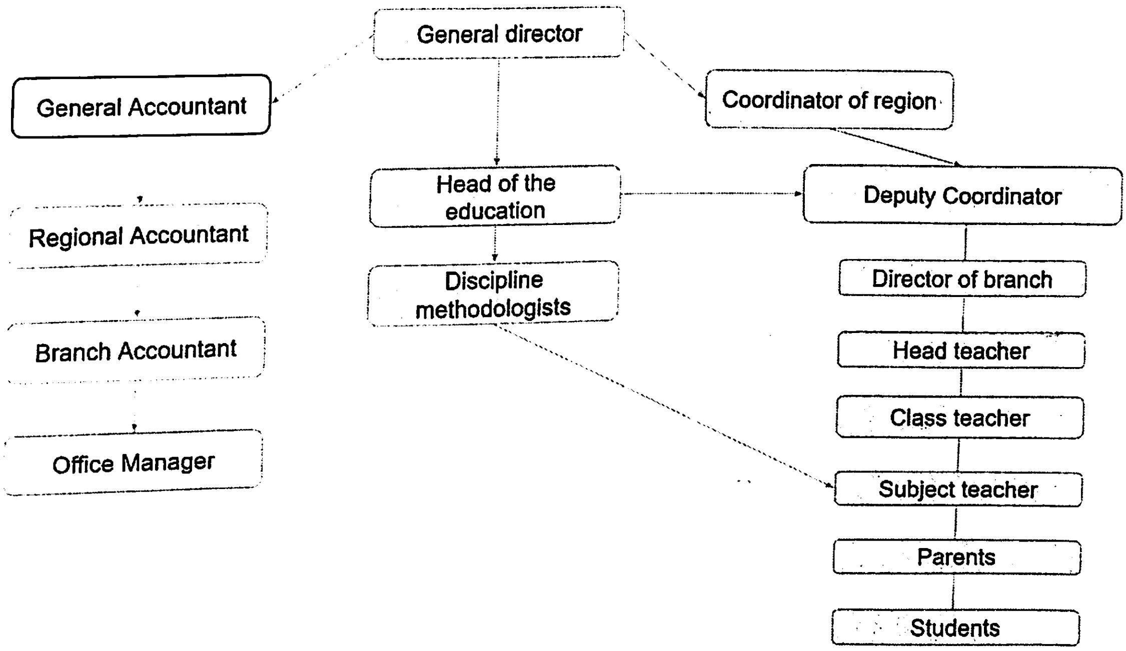

Figure 4.3: Management tree in eduway.kz.

Students have the following options:

- Students can watch online video lessons.
- Students can take tests on topics.
- Pass scheduled exams.
- Pass the questionnaire.
- Receive homework from the teacher and complete online.
- Complete classwork tasks.

This platform is used by about 60 institutions in Kazakhstan. Most of them are private educational centers. They are also used for charity purposes (for example, the iQanat Foundation was founded to educate children from villages in different parts of Kazakhstan). Eduway.kz is easy to use. Office managers in a matter of minutes will be able to register a new student and draw up an agreement with the parent. The system automatically fills in the text of the contract by typing data

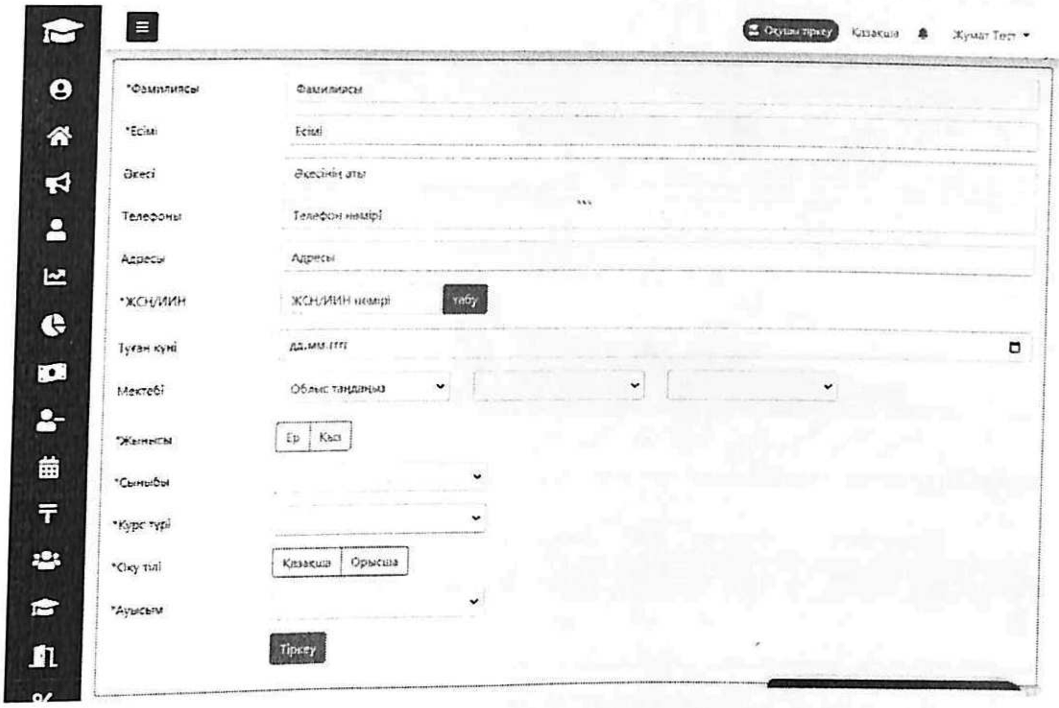

Figure 4.4: Registration of student

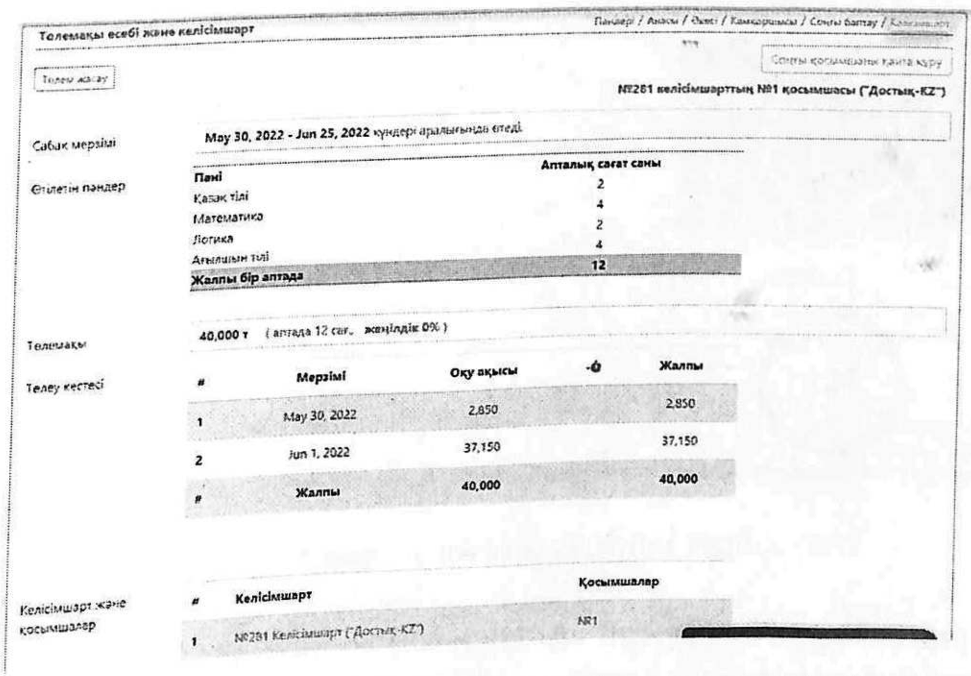

Figure 4.5: Automatically generated contract with client

from the database and prepares the document for printing. Office manager just needs to click on the download button. And also the system itself automatically calculates the monthly tuition fees. Administrators of the organization conducts monthly exams and all data is stored in a single database. Examinations are held in two forms: paper and electronic. Students take a paper-based exam directly at the school. The results of this type of exam are read using a special scanner apparatus and stored in the eduway.kz database through the application.

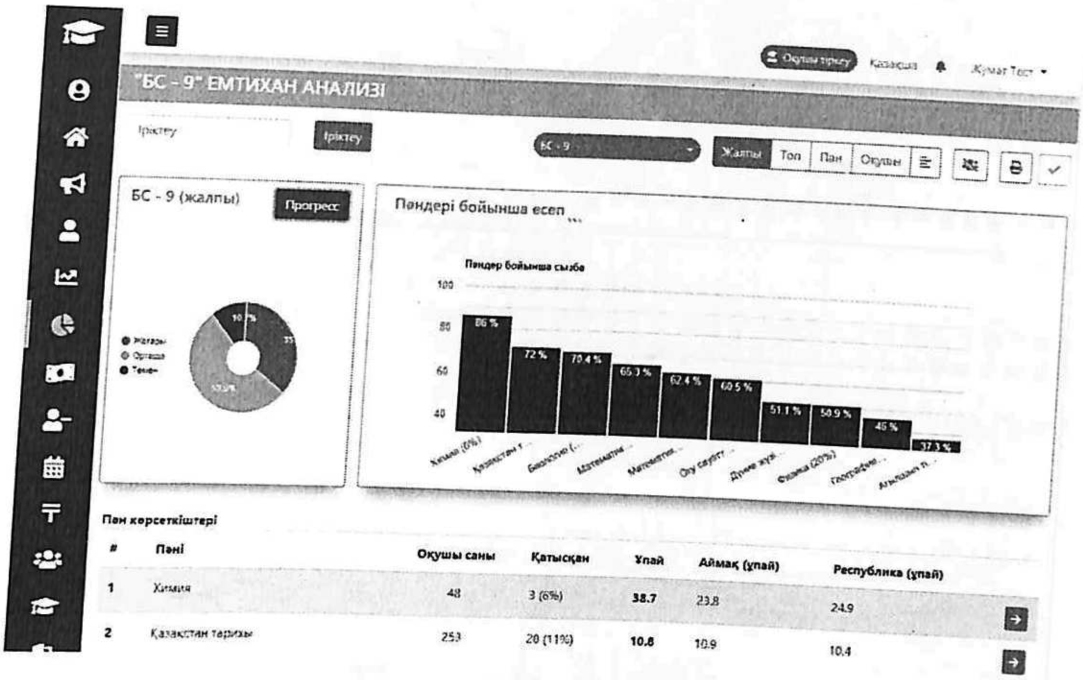

Figure 4.6: Analysis of exam

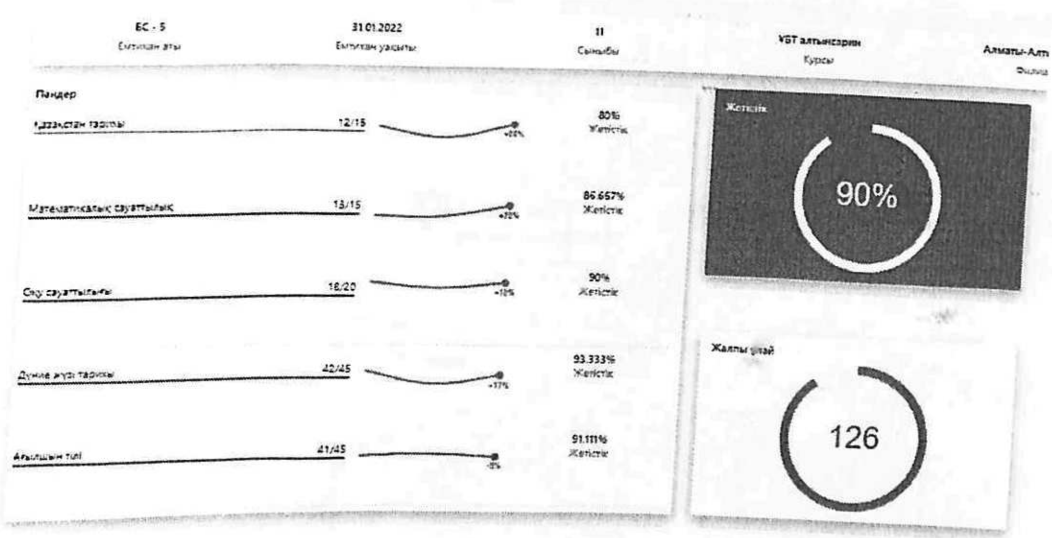

Figure 4.7: Exam student's personal report card

An electronic form of the exam is taken via the Internet. Thus, students take the exam from home using a computer, tablet or phone. Bosses and teachers will be able to watch different types of analysis for the exam:

- Comparative analysis of branches;
- Subject analyses;
- Comparative analysis among classes;
- Analyzes by topic;
- Rating analysis of students;

- Student's personal report card.

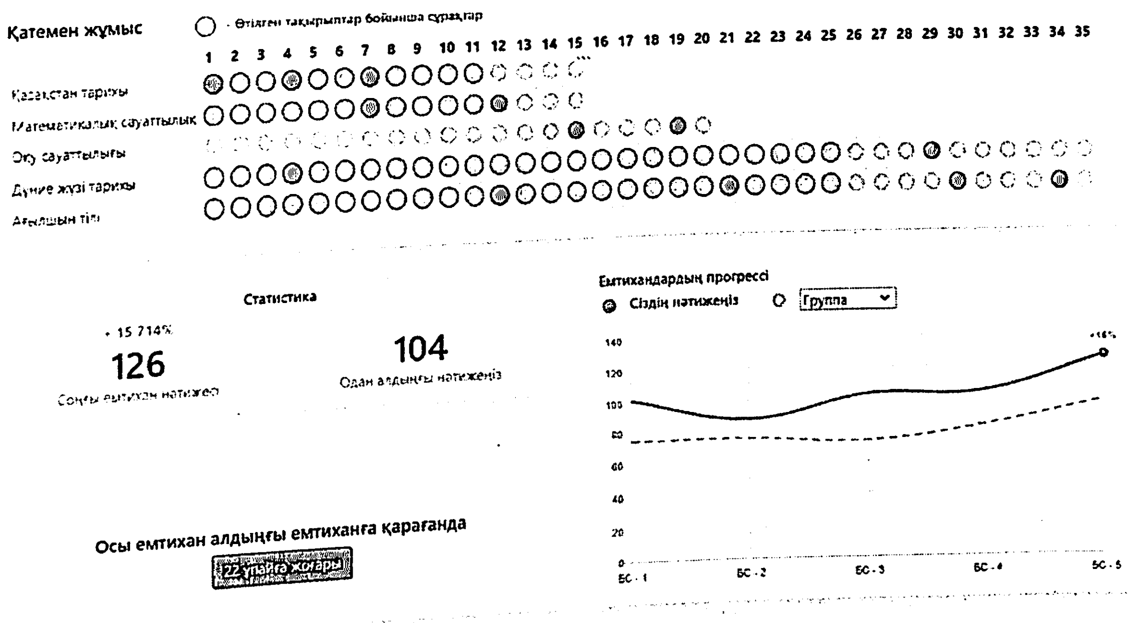

Figure 4.8: Work on the wrong answers of the exam result

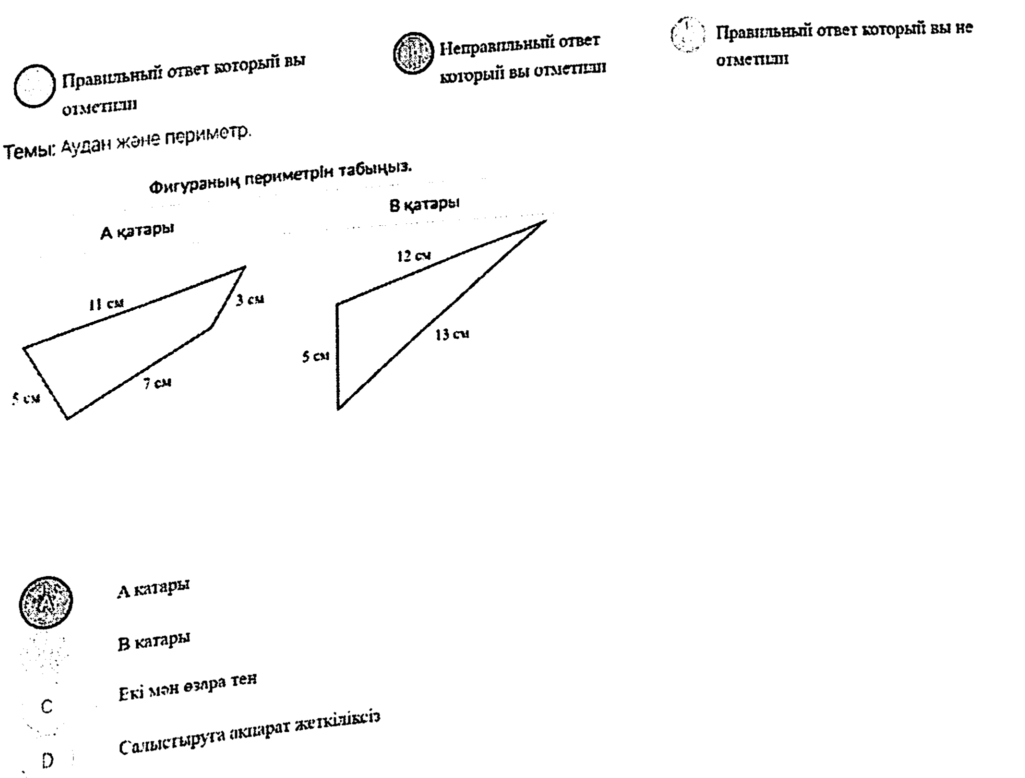

Figure 4.9: Revisiting a past exam question

Students also supplement their knowledge with the help of video lessons in their personal account. The student has access to video lessons only on a suitable subject (selected subjects at the conclusion of the contract).

Figure 4.10: Video lessons page for student

The head teacher, in turn, very easily schedules and manages educational activities. Teachers and students are always aware of schedule changes in their personal account. The teacher in his account will be able to mark students who

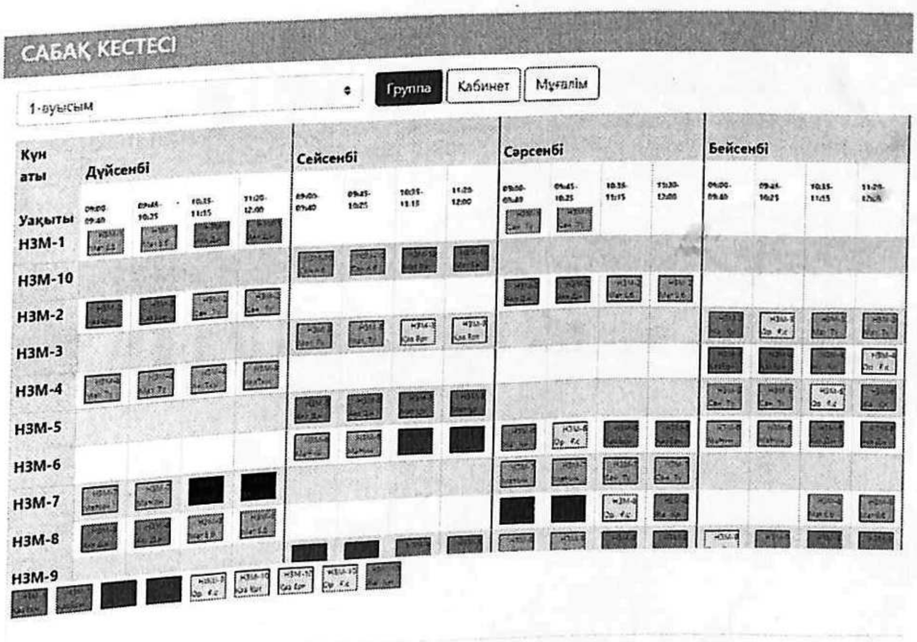

Figure 4.11: Page for scheduling lessons

are not present during the lesson and put marks in the electronic journal. Very convenient homework management for both students and teachers. It is possible to monitor all ongoing activities of the institution using a dashboard in your personal account.

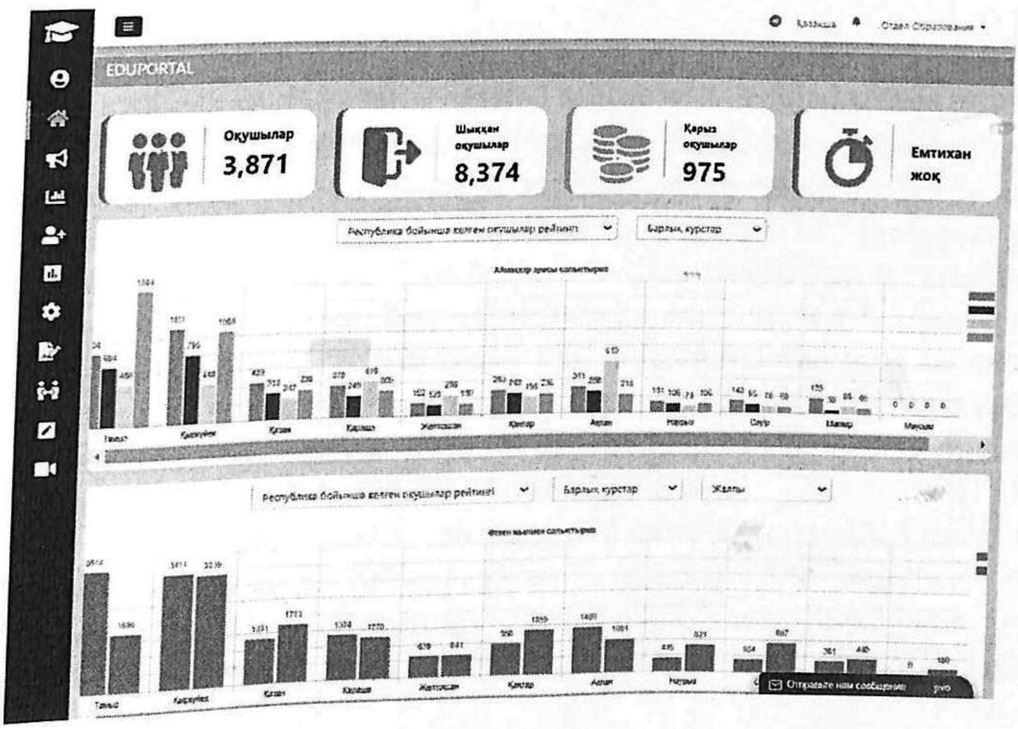

Figure 4.12: Dashboard for monitoring in cabinet

Chapter 5

Sample and research methodology

To solve the problem, a data set is needed and it was decided to use student data from the LMS platform Eduway.kz. More than 8,000 high school students are registered on the platform. Of these, about 1050 village students study under the program of the iQanat foundation. We chose the students of the iQanat program for the study. The LMS stores data on how actively a student is watching video lessons, midterm exam results, homework results and lesson attendance.
A questionnaire developed by Vallerand and colleagues was used to assess academic motivation. This questionnaire asks the participants to rate their agreement with 28 statements on a 7-point scale. With the help of this questionnaire, three levels of motivation are determined (internal, external, amotivation). Internal and external motivation separately consists of 3 subscales (internal motivation to know, internal motivation to achieve, internal motivation to experience stimulation, external motivation - identified, external motivation - introjected, external motivation - external regulation). Additionally, there are 7 subscales, including the motivation scale.
The AMS questionnaire was launched online on the Eduway.kz platform. Exactly 534 students from different regions filled out the questionnaire. Of these, 433 are girls and 101 are boys. The results (Table 5.1) show that the lowest score is on the amotivation scale (2.7), and the highest score is on "intrinsic motivation to know" (5.5). The scores for boys and girls are almost the same across all subscales. All subscales are correlated and statistically significant. And analysis shows that all subscales are valid. Cronbach's alpha was used to assess the internal consistency of the subscales. The alpha values ranged from 0.736 to 0.825 (Table 5.2), and they are close to the alpha values in the research paper by Valerand et al. (1992).

Table 5.1: Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Intrinsic motivation(IM) - to know	534	1.00	7:00	55.558	134.820	1.818
Intrinsic motivation - toward accomplishment	534	1.00	7.00	47.869	156.308	2.443
Intrinsic motivation - to experience stimulation	534	1.00	7.00	48.451	144.394	2.085
Extrinsic motivation - identified	534	1.00	7.00	52.358	148.142	2.195
Extrinsic motivation - introjected	534	1.00	7.00	44.819	164.310	2.700
Extrinsic motivation - external regulation	534	1.00	7.00	53.569	146.029	2.132
Amotivation	534	1.00	7.00	27.225	169.577	2.876

Table 5.2: Reliability Statistics

	Cronbach's Alpha	N of Items
Intrinsic motivation - to know	.791	4
Intrinsic motivation - toward accomplishment	.815	4
Intrinsic motivation - to experience stimulation	.736	4
Extrinsic motivation - identified	.804	4
Extrinsic motivation - introjected	.790	4
Extrinsic motivation - external regulation	.766	4
Amotivation	.825	4

Initially, 6 variables describing the student's primary behaviors in LMS taken into account for prediction purposes (video watched, video completion, homework point, homework completion, exam point, exam completion). Below is a correlation analysis of the data which can be seen in Table 5.3. Obtained correlation coefficients indicate that the variables are correlated either weakly or moderately, or they are not correlated (that is, negligible). Several variables were taken into account for use in the modeling, one being video completion (V1), homework point (V2), homework completion (V3), and exam point (V4), and academic motivation was treated as a dependent variable.

Table 5.3: Correlation analysis ($p < 0.05$)

	video completion	homework point	homework completion	exam point
video completion	1	0.29	0.39	0.2
homework point	0.29	1	14	0.51
homework completion	0.39	0.14	1	-0.09
exam point	0.2	0.51	-0.09	1

Table 5.4: Descriptive statistics of variables

Variable	Valid number (N)	Mean (M)	Standart deviation (SD)
video completion (V1)	534	30.01	24.29
homework point (V1)	534	61.94	18.45
homework completion (V1)	534	19.77	23.22
exam point (V1)	534	60.83	14.24

As part of the AMS, 28 items provide the participants with the opportunity to rate their agreement with the statement from does not correspond at all through exactly corresponds on a Likert scale. Amotivation, identified, introjected, external regulation, extrinsic motivation, identified, intrinsic, and extrinsic motivation - the AMS identifies amotivation, intrinsic and extrinsic levels of motivation . When a particular motivation type scores higher for each subscale, it indicates a higher level of motivation.

Despite the fact that the questionnaire was translated into another language, it showed that it is applicable to measure the scale of academic motivation and can be used for further research. Since there are several types of motivation in AMS, it is difficult to determine what scale to use for machine learning models in research. But there are other methods for determining the level of motivation. For example, AMS can be used to calculate the Self-Determination Index (SDI) [17] (Figure 5.1).

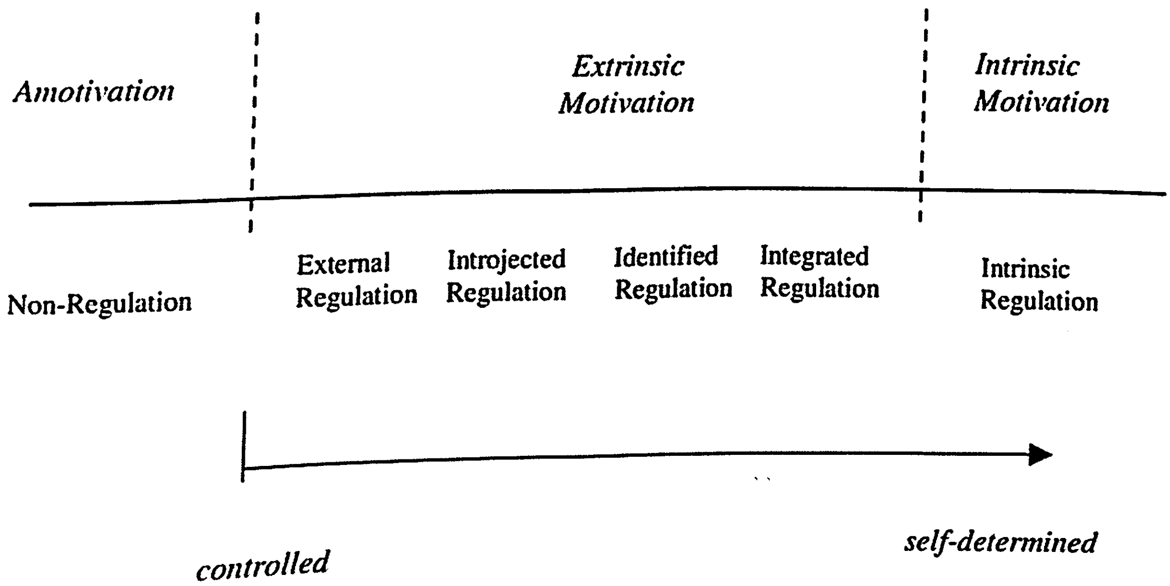

Figure 5.1: The continium of self-determination

All participants had a mean SDI score of 5.06 (SD=3.05). Those with SDI scores below this mark were considered those with a low level of academic motivation and categorized as 0. A participant whose SDI was above 5.06 was thought-out as one with an elevated motivation level in this study, and he or she was classified as 1.

$$SDI = (2 \times intrinsic + identified) - (2 \times external + introjected)$$

A number of studies used this instrument and several of them tested its concurrent validity as well as the construct validity of the instrument (see [40]). As reported in [21], the AMS subscales had a high level of internal consistency (Cronbach's alpha values ranged from 0.60 to 0.86) and a high level of temporal stability. This sample was randomly split into two subsamples with the purpose of comparing models (80% training and 20% testing). Train and selection subsamples were included in the neural network training subsample.

Table 5.5: Items of the AMS questionnaire.

№	WHY DO YOU GO TO SCHOOL ?
1	Because I need at least a high-school degree in order to find a high-paying job later on.
2	Because I experience pleasure and satisfaction while learning new things.
3	Because I think that a high-school education will help me better prepare for the career I have chosen.
4	Because I really like going to school.
5	Honestly, I don't know; I really feel that I am wasting my time in school.
6	For the pleasure I experience while surpassing myself in my studies.
7	To prove to myself that I am capable of completing my high-school degree.
8	In order to obtain a more prestigious job later on.
9	For the pleasure I experience when I discover new things never seen before.
10	Because eventually it will enable me to enter the job market in a field that I like.
11	Because for me, school is fun.
12	I once had good reasons for going to school; however, now I wonder whether I should continue.
13	For the pleasure that I experience while I am surpassing myself in one of my personal accomplishments.
14	Because of the fact that when I succeed in school I feel important.
15	Because I want to have "the good life" later on.
16	For the pleasure that I experience in broadening my knowledge about subjects which appeal to me.
17	Because this will help me make a better choice regarding my career orientation.
18	For the pleasure that I experience when I am taken by discussions with interesting teachers.
19	I can't see why I go to school and frankly, I couldn't care less.
20	For the satisfaction I feel when I am in the process of accomplishing difficult academic activities.
21	To show myself that I am an intelligent person.
22	In order to have a better salary later on.
23	Because my studies allow me to continue to learn about many things that interest me.
24	Because I believe that my high school education will improve my competence as a worker.
25	For the "high" feeling that I experience while reading about various interesting subjects.
26	I don't know; I can't understand what I am doing in school.
27	Because high school allows me to experience a personal satisfaction in my quest for excellence in my studies.
28	Because I want to show myself that I can succeed in my studies.

Chapter 6

Results

From the results obtained (see Table 6.1), the hyperbolic tangent was found to be the best activation function to use when applying MLP models (61.5%) while a RBF model (76.9%) finished with the highest overall classification accuracy. CART was used as a split-selection method to obtain the classification tree model with the highest overall accuracy. Despite the large and complex tree (see Appendix 1), the tree structure was quite straightforward. A total of 17 splits were connected to 18 terminal nodes. The accuracy of several SVM models was the same (57.69%). Using the RBF kernel, the accuracy was 63.64% for low motivated students, and with the sigmoid kernel it was 60.00 for highly motivated students when using type II SVM. Based on this research, the best SVM model was the type II model with a RBF kernel.
The t-test of proportional differences was used to compare models and examine the statistical significance of any differences in accuracy. Based on the t-test results, we found no statistically significant difference in performance between the two models. Table 6.2 summarizes those results.
For comparison and assessment, sensitivity and specificity were calculated along with positive and negative predictive values (PPV). Based on the confusion matrix (Table 6.3), the RBF model achieved the top performance on identifying less academically motivated students than the CT model (87%), while the CT model has the best performance on identifying students who are more motivated academically than the RBF model. Moreover, this model is able to detect the lowest level of error (0.13), thus making it the model with the highest level of type II error (64%). The SVM model makes the least accurate diagnosis of students with the highest levels of academic motivation (53%), along with corresponding type I errors (0.47). Students with exceptionally high levels of academic motivation can also be misclassified by this model.
Furthermore, the RBF model didn't show any characteristics of misclassifying students who had a low motivation level (0.00), and as a result it was the best at accurately identifying students who really had low academic motivation (see Table 6.4) and was the most accurate in predicting the outcome of students who were misclassified.

Table 6.1: Results of the test sample

Model	Accuracy for students with a below-average level of academic motivation - category 0 (%)	Accuracy for students with an above-average level of academic motivation - category 1 (%)	Total accuracy (%)
MLP - exponential	9.09	93.33	57.69
MLP – tangent hyperbolic	45.45	73.33	61.54
MLP - logistic	9.09	93.33	57.69
MLP - sine	9.09	86.67	53.85
MLP – identity (linear)	9.09	86.67	53.85
RBF	100.00	60.00	76.92
decision tree -CART	53.33	81.81	65.38
decision tree - discriminant-based univariate splits	60.00	63.36	50.00
SVM type I - linear	0.00	100.00	57.69
SVM type I - RBF	0.00	100.00	57.69
SVM type I- sigmoid	0.00	100.00	57.69
SVM type I- polynomial	0.00	100.00	57.69
SVM type II - linear	54.55	46.67	50.00
SVM type II- RBF	63.64	53.33	57.69
SVM type II- sigmoid	54.55	60.00	57.69
SVM type II- polynomial	81.82	6.67	38.46

Table 6.2: t-test results of differences in proportion

Hypothesis	p-value results
H_0 : RBF = CT	p =0.18
H_0 : SVM = CT	p =0.28
H_0 : RBF = SVM	p =0.07

In relation to the achieved results, it seems that it is most appropriate to use the RBF models to predict student academic motivation. In Figure 6.1 and Figure 6.2 you can see how the Box and Whisker plots reflecting the values from the Radial basis function output could be used to identify a relationship between the predicted motivation level of student and their activity in the learning management system. According to the results of this study, students with above motivation levels fewer times watched video and did homework, received lower points for homework in LMS courses than half the number of times that students

Table 6.3: Confusion matrix of models

Model		0	1
RBF	0	1.00	0.40
	1	0.00	0.60
Classification tree	0	0.36	0.13
	1	0.64	0.87
SVM	0	0.64	0.47
	1	0.36	0.53

Table 6.4: Measures of observed and evaluated models

Model	Sensitivity	Specificity	PPV	NPV
RBF	1.00	0.60	0.65	1.00
CT	0.36	0.87	0.67	0.65
SVM	0.64	0.53	0.50	0.67

with below-average levels of academic motivation.

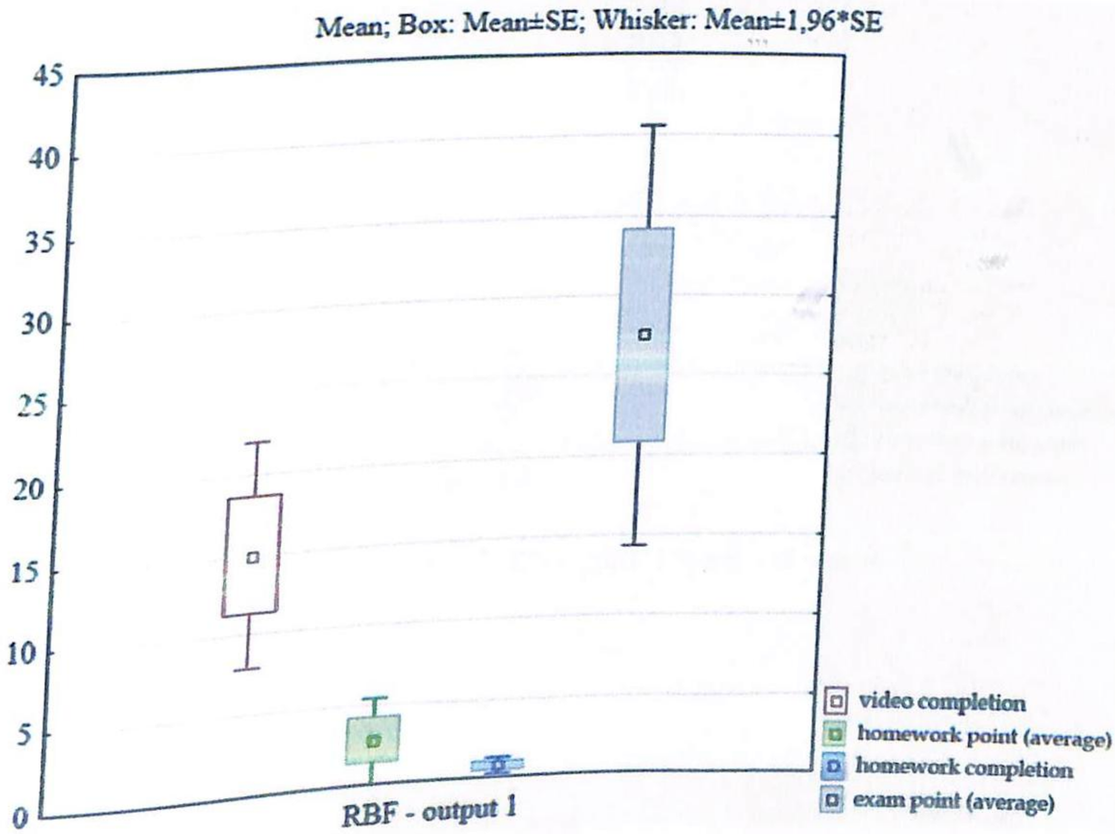

Figure 6.1: RBF box plot based on value 1

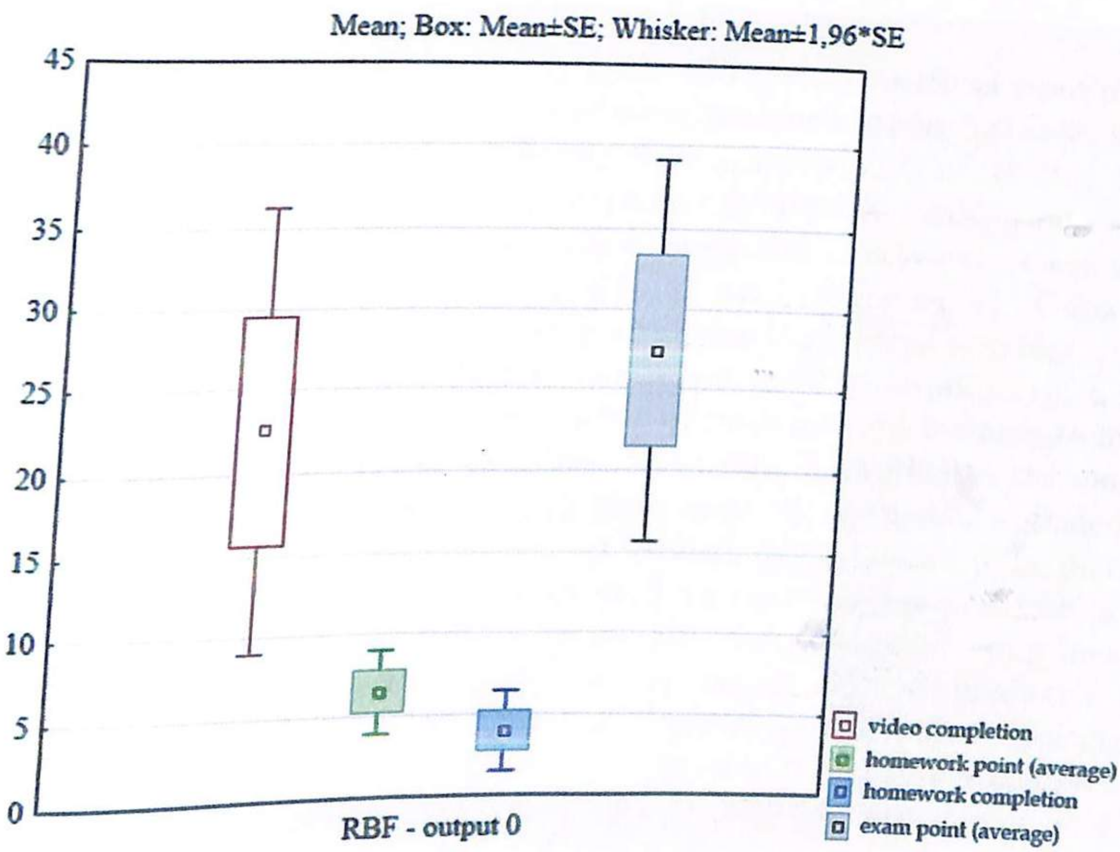

Figure 6.2: RBF box plot based on value 0

Chapter 7

Conclusion

Using students data from Learning management system course as input data and guiding our research through the use of three machine learning methods, the purpose of this study was to assess whether there is a possibility of creating a model that can effectively predict student academic motivation. Taking into account the performance of the different models, we were able to select the most suitable one by comparing their efficiency. In order to conduct the model, it was necessary to analyze the relationship between all potential predictor variables and then between potential predictor variables and target variables, resulting in a reduction of potential predictors. The application of three machine learning techniques (decision trees, support vector machines, and neural networks) to the modeling process yielded acceptable results in all three cases. In the results obtained from the most recent analysis, the RBF neural network model proved to be more efficient than the two techniques that were used for obtaining the best results. The model generated in total 76.92% accuracy and were evaluated using measures which demonstrated that this model had the highest negative predictive value of all of the tested models (100%) and the highest accuracy of recognizing less academically motivated students (100%) compared to the two other models investigated. The positive predictive value of this model (65%) was decent, although not the best. In the t-test to compare the proportions between the observed models, we were not able to detect a statistically significant difference in performance at a 5% level of significance.

The results of this study confirm that machine learning methods are effective in producing a model to predict student academic motivation even though different predictors, algorithms, and methods were used to detect student academic motivation. Although, it was found that the ANN model gave the most accurate results. Research conducted in this study could have a significant impact on scholars and other educators, especially when it comes to identifying students with below-average or above-average motivation. Early identification of students with the lowest motivation level will help prevent poor outcomes. For example, a teacher may give group projects to students, mixing amotivated students with the best students. In this way, educators can actively involve them in learning and ensure the best possible academic performance.

Since the results from this research are limited, they should be extended in future work. Other input variables and machine learning techniques may also be included in future research, and a combination of them may improve the model.

References

[1] S. Areepattamannil. "Relationship between academic motivation and mathematics achievement among Indian adolescents in Canada and India." In: *The Journal of general psychology* 141.3 (2014), pp. 247–262.

[2] S. Areepattamannil, J. G. Freeman, and D. A. Klinger. "Intrinsic motivation, extrinsic motivation, and academic achievement among Indian adolescents in Canada and India." In: *Social Psychology of Education* 14.3 (2011), pp. 427–439.

[3] Josh Bersin. *The Talent Management Software Market Surges Ahead.* URL: `https://www.forbes.com/sites/joshbersin/2014/06/26/the-talent-management-software-market-surges-ahead/`. (accessed: 26.02.2019

[4] E. W. Black, K. Dawson, and J. Priem. "Data for free: Using LMS activity logs to measure community in online courses." In: *The Internet and Higher Education* 11.2 (2008), pp. 65–70.

[5] M. Cocea and S. Weibelzahl. "Can log files analysis estimate learners' level of motivation?" In: *LWA. University of Hildesheim, Institute of Computer Science* (2006).

[6] M. Cocea and S. Weibelzahl. "Disengagement detection in online learning: validation studies and perspectives." In: *IEEE Transactions on Learning Technologies* 4.2 (2011), pp. 114–124.

[7] M. Cocea and S.. Weibelzahl. "Eliciting motivation knowledge from log files towards motivation diagnosis for Adaptive Systems." In: *International Conference on User Modeling* (2007), pp. 197–206.

[8] Beth Davis, Colleen Carmean, and Ellen D. Wagner. "The Evolution of the LMS: From Management to Learning". In: *The eLearning Guild* (2009). URL: `https://web.archive.org/web/20170829051138/http://www.blackboard.com/resources/proed/guild-lmsreport.pdf`.

[9] E. L. Deci et al. "Motivation and education: the self-determination perspective." In: *Educational psychologist* 26.3-4 (1991), pp. 325–346.

[10] K. L. Du and M. N. S. Swamy. *Neural networks and statistical learning.* Springer Science & Business Media., 2013.

[11] N. L. Ennen, E. Stark, and A. Lassiter. "The importance of trust for satisfaction, motivation, and academic performance in student learning groups". In: *Social Psychology of Education* 18.3 (2015), pp. 615–633.

[12] Amanda J. Fairchild et al. "Evaluating existing and new validity evidence for the Academic Motivation Scale". In: *Contemporary Educational Psychology* 30.3 (2005), pp. 331–358.

[13] Pelletier L. G. et al. "Toward a New Measure of Intrinsic Motivation, Extrinsic Motivation, and Amotivation in Sports: The Sport Motivation Scale (SMS)." In: *Journal of Sport and Exercise Psychology* 17.1 (1995), pp. 35–53.

[14] *Global Learning Management System (LMS) Market Trends, Applications, Analysis, Growth, And Forecast: 2018 To 2027*. URL: `https://marketresearch.biz/report/learning-management-system-lms-market/`. (accessed: 26.02.2019).

[15] *Going the Distance*. URL: `https://web.archive.org/web/20190213224631/http://bear.warrington.ufl.edu/tutorials/going-the-distance/`. (accessed: 13.09.2019).

[16] F. Guay et al. "Academic self-concept, autonomous academic motivation, and academic achievement: medating and additive effects." In: *Learning and Individual Differences* 20.6 (2010), pp. 644–653.

[17] M. Hackett. *The Impact of Racial-Ethnic Identity on Academic Motivation of African American High School Students (Doctoral dissertation)*. 2017. URL: `http://www-cs-faculty.stanford.edu/~uno/abcde.html`.

[18] S. Haykin. *Neural networks and learning machines*. Upper Saddle River, NJ, USA, Pearson., 2009.

[19] Ernest Hilgard. *Learning & the technology of instruction*. New York, Appleton-Century-Crofts, 1966.

[20] W. W. Hsieh. *Machine learning methods in the environmental sciences: Neural networks and kernel*. Cambridge university press, 2009.

[21] Vallerand R. J. et al. "The Academic Motivation Scale: A measure of intrinsic, extrinsic, and amotivation in education." In: *Educational and psychological measurement* 52.4 (1992), pp. 1003–1017.

[22] Donald Knuth. *StatSoft, Classification Trees*. URL: `https://www.tibco.com/products/data-science`.

[23] M. Komarraju, S. J. Karau, and R. R. Schmeck. "Role of the Big Five personality traits in predicting college students' academic motivation and achievement". In: *Learning and individual differences* 19.1 (2009), pp. 47–52.

[24] M. Komarraju, S. Musulkin, and G. Bhattacharya. "Role of student–faculty interactions in developing college students' academic self-concept, motivation, and achievement." In: *Journal of College Student Development* 51.3 (2010), pp. 332–342.

[25] K. Kularbphettong and C. Tongsiri. "Mining educational data to analyze the student motivation behavior." In: *World Academy of Science, Engineering and Technology* 68.2012 (2012), pp. 1256–1259.

[26] H. Labjar et al. "Support vector machines for modelling phosphocalcic hydroxyapatite by precipitation from a calcium carbonate solution and phosphoric acid solution." In: *Journal of Taibah University for Science.* (2015).

[27] E. J. Lauría et al. "Mining academic data to improve college student retention: an open source perspective." In: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (2012), pp. 139–142.

[28] E. Lee. "The relationship of motivation and flow experience to academic procrastination in university students." In: *The Journal of Genetic Psychology* 166.1 (2005), pp. 5–15.

[29] S. Marsland. *Machine learning: an algorithmic perspective.* CRC press, 2015.

[30] A. R. Matthews et al. "Academic motivation in calculus." In: *Canadian Journal of Science, Mathematics and Technology Education* 13.1 (2013), pp. 1–17.

[31] M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of machine learning.* MIT press, 2012.

[32] S. K. Murthy. "Automatic construction of decision trees from data: A multidisciplinary survey." In: *Data mining and knowledge discovery* 2.4 (1998), pp. 345–389.

[33] C. Orsini et al. "Psychometric validation of the academic motivation scale in a dental student sample." In: *Journal of dental education* 79.8 (2015), pp. 971–981.

[34] Mary L Phillips et al. "Neurobiology of emotion perception II: implications for major psychiatric disorders". In: *Biological Psychiatry* 54.5 (2003), pp. 515–528.

[35] M. Próspero and S. Vohra-Gupta. "First generation college students: motivation, integration, and academic achievement." In: *Community College Journal of Research and Practice* 31.12 (2007), pp. 963–975.

[36] L. Rokach and O. Maimon. *Data mining with decision trees: theory and applications.* World scientific, 2015.

[37] C. Romero, S. Ventura, and E. García. “Data mining in course management systems: Moodle case study and tutorial.” In: *Computers & Education* 51.1 (2008), pp. 368–384.

[38] C. Romero et al. “Web usage mining for predicting final marks of students that use Moodle courses.” In: *Computer Applications in Engineering Education* 21.1 (2010), pp. 135–146.

[39] Judith Schoonenboom. “Using an adapted, task-level technology acceptance model to explain why instructors in higher education intend to use some learning management system tools more than others”. In: *Computers & Education.* 71 (2014), pp. 247–256.

[40] K. J. Smith, J. A. Davy, and D. L. Rosenberg. “An empirical analysis of an alternative configuration of the academic motivation scale.” In: *Assessment in education: Principles, policy & practice* 19.2 (2012), pp. 231–250.

[41] C. W. Struthers, R. P. Perry, and V. H. Menec. “An examination of the relationship among academic stress, coping, motivation, and performance in college.” In: *Research in higher education* 41.5 (2000), pp. 581–192.

[42] S Suthaharan. *Machine learning models and algorithms for big data classification.* Boston: Springer, 2016.

[43] T. L. B. Tseng et al. “E-quality control: a support vector machines approach.” In: *Journal of Computational Design and Engineering* 3.2 (2016), pp. 91–101.

[44] E. L. Usher and D. B. Morris. “Academic motivation.” In: *Encyclopedia of the Sciences of Learning* (2012), pp. 36–39.

[45] S. Valsamidis et al. “A Clustering Methodology of Web Log Data for Learning Management Systems”. In: *Educational Technology & Society* 15.2 (2012), pp. 154–167.

[46] Т.О. Гордеева. “Базовые типы мотивации деятельности: потребностная модель”. In: *Вестник Московского университета* 14.3 (2014).

[47] *Что такое электронная система дистанционного обучения (СДО)?* URL: https://www.ispring.ru/elearning-insights/chto-takoe-sistema-distantsionnogo-obucheniya.

[48] *Что такое электронная система дистанционного обучения (СДО)?* URL: https://elearningindustry.com/saas-learning-management-system-lms-truly-saas. (accessed: 05.02.2019).