Tampere University

Ara Jo

VISUALIZATION AND ANALYSIS OF STUDENT DATA USING MACHINE LEARNING AND STATISTICAL METHODS

Faculty of Information Technology and Communication Sciences (ITC)
Master's thesis
Examiners: Prof. Tarmo Lipping, University lecturer Jari Turunen
November 2023

Abstract

Ara Jo: Visualization And Analysis Of Student Data Using Machine Learning And Statistical Methods
Master's thesis
Tampere University
Master's Degree Programme in Information Technology
Major: Signal Processing and Machine Learning
Examiners: Prof. Tarmo Lipping, University lecturer Jari Turunen
November 2023

The foundation of Finnish education is based on the principle of equality. As data analysis techniques evolve, they provide diverse approaches to understanding educational data and developing an enhanced learning environment. Using machine learning and statistical methods, this thesis seeks to provide useful insights into the current educational status and offers guidance for improved student support. This study presents findings without bias toward any school or student.

A significant portion of this study is dedicated to data visualization and trend discovery. By utilizing Power BI visualizations, analyzing features by different groups, such as native language, gender, class, and support groups, becomes easier, allowing users to interactively derive meaningful insights.

In this research, data from students in classes 1 to 9 were analyzed. Among two language groups, native Finnish speakers had a higher rate of absenteeism but also achieved higher average grades. When evaluated by gender, females achieved a higher average grade. Within support classifications, students in the special support group exhibited the highest absenteeism and the lowest average grades.

In terms of predicting student academic performance, both the LSTM model and the Markov state model were explored. The LSTM model's prediction accuracy is calculated with a mean squared error.

Through the application of data mining techniques and data visualization, this study provides deeper insights into trends among different student groups, paving the way for customized student support.

Keywords: Machine Learning, Education Data Analysis, Data Visualization, LSTM, Academic Performance Prediction.

The originality of this thesis has been checked using the Turnitin Originality Check service.

Preface

This thesis study was conducted at the Department of Signal Processing and Machine Learning at Tampere University during the year 2023.

First and foremost, I deeply appreciate the guidance and support I received from my supervisors, Prof. Tarmo Lipping and University Lecturer Jari Turunen. Their invaluable guidance, support, and patience were paramount throughout this study. It has been a privilege to conduct research under their teaching.

I also want to thank my friends who have consistently been there for me throughout my study journey. Their unwavering support and trust have kept me motivated. They stood by me during both the good and challenging times, for which I am profoundly grateful.

Tampere, 12 November 2023
Ara Jo

Usage of AI-based tool

This section outlines how AI tools were employed in this thesis. In accordance with Tampere University's policy on AI tool use, it is important to highlight proper scientific methods when using such tools. The AI tool's responses were checked and mostly used to enhance text clarity and correctness.

The primary reason for using AI tools, such as the Language Model, in this work was to polish the grammar and make the text more readable. The Chat GPT version 3.5 tool was utilized to identify and correct English grammar errors. Additionally, Chat GPT assisted in refining word choices and improving the flow of sentences and paragraphs.

Contents

List of Figures

1 Introduction

Since 2001, many have come to view Finland's educational framework as a global benchmark, intriguing international educational professionals. This interest is evident in the growing trend of 'PISA tourism'. [27]

Educational progress has always been crucial for education professionals. This study focuses on the Finnish educational system, specifically based on data from Pori. By examining academic performance metrics and utilizing machine learning and data visualization techniques, the study seeks to uncover insights into Finnish academic achievements.

By analyzing academic performance metrics and leveraging advanced machine learning and data visualization techniques, the objective is to discover the contributing features behind Finnish academic accomplishments. Over the preceding three years, Pori has dedicated to advance its educational standards. For example, Pori adopted and upgraded the educational data collecting system.

In collaboration with the city, this research aims to extract valuable insights beneficial to education specialists. The primary objective is to provide better student support and present a transparent overview of the findings. Instead of focusing on individual students or schools, the research examines comprehensive trends that might influence student outcomes across primary and secondary education. To ensure confidentiality, schools are referenced using letter aliases from A to R.

Another key aspect of this thesis is its focus on displaying Pori's data visually. Using the latest data visualization tools, this study provides users with interactive charts and graphs. These visuals not only summarize the main findings but also encourage users to explore the data themselves. By interacting with these visuals, users can dive deeper and find more insights, giving them a complete and engaging view of the data.

The research aims to answer specific research questions (RQs) to better understand Pori's student data:

- RQ1: Does a student's native language impact their absenteeism and academic results?
- RQ2: Does a student's academic grade correlate with absenteeism rates, and are specific grade levels more prone to higher absenteeism?
- RQ3: Are there noticeable differences across three distinct academic years?
- RQ4: How do absenteeism rates and academic grades differ among schools?

- RQ5: Can the number of absences serve as an indicative metric to anticipate a student's need for special support?

The structure of this thesis is systematically organized as follows: Chapter 2 delves into the background of the study, providing an overview of the education systems in Nordic countries and examining existing studies related to student academic performance. Chapter 3 explains the methods employed in the study. This encompasses a brief introduction to machine learning (ML), its techniques, and the statistical models used. Chapter 4 is dedicated to visualization. It outlines the data preprocessing steps, the tools, and logic applied for effective data visualization. Chapter 5 presents the analysis and results. The findings are illustrated using a variety of charts and graphs. Moreover, this chapter offers visual comparisons across support groups, gender, and schools to discover patterns and trends. Lastly, Chapter 6 concludes the thesis by summarizing the main findings and offering suggestions for future research.

2 Background

This chapter aims to establish a foundational understanding of the educational contexts and academic dynamics. This chapter is structured into three sections: the Finnish education system, the Nordic education system, and existing research on academic performance. Each section will provide clarity on the educational structures and academic performance study outcomes that are relevant to the scope of this study. For clarity within this research, 'class' will denote 'Year group,' and 'grade' will be used in the context of 'Academic grade.'

2.1 Finnish education system

The Finnish educational system is structured around the principle of equality, where education is universally accessible and state-funded. By legal standards, Finnish municipalities are responsible for delivering comprehensive education from ages 7 to 15. Any exceptions from this custom require explicit state authorization. Generally, the education system divides students between primary schools (age 7-12) and lower secondary schools (age 13-15). The education system in Finland is illustrated in Figure 2.1. In primary school, students typically have a single class teacher for their entire study. However, in the lower secondary phase, specialized teachers instruct different subjects. [28]

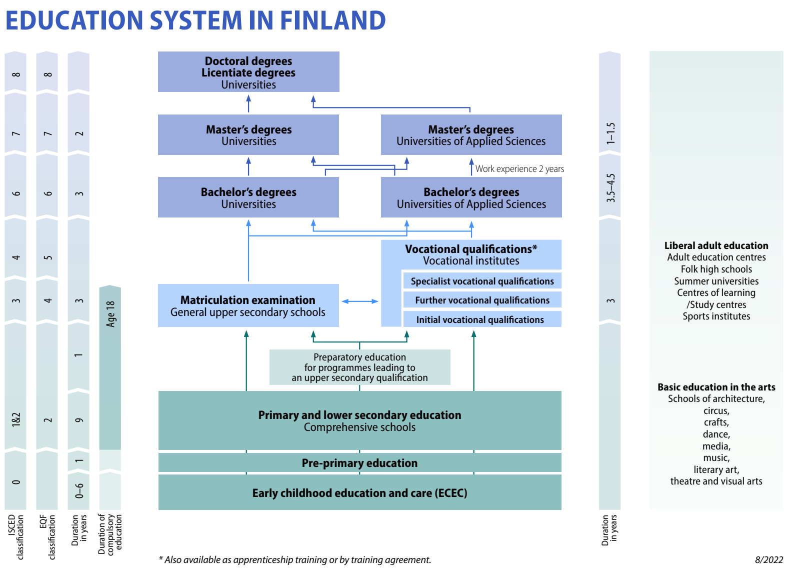


Figure 2.1 *General structure of the Finnish education system borrowed from [7]*

While municipalities have comprehensive policies, there are variations in practices across cities regarding how students are allocated to schools and how much families can influence this allocation. In the majority of cities, each school has a distinct district or designated enrollment zone for both primary and lower secondary schools, with a student's residence primarily determining their school placement. Some municipalities even retain the flexibility to adjust these zones annually to meet academic and financial objectives. [28]

Transitioning from primary (6th class) to lower secondary schools (7th class) in most cities follows predefined routes, ensuring a seamless progression from a primary to a lower secondary school. This practice is consistent, irrespective of whether the initial school allocation was based on a specific zone or a broader area. Some comprehensive schools offer education from class 1 through 9, enabling students to remain at the same institution. While students are primarily assigned to a local school, they also retain the option to apply elsewhere, depending on seat availability. Moreover, while schools adopt different criteria for admissions, they are mandated by the Basic Education Act to ensure fairness in these processes. [28]

2.2 Nordic education system

In this chapter, the educational systems of Nordic countries, specifically Norway, Sweden, and Denmark, will be assessed.

Swedish mandatory education is divided into four phases: förskoleklass (often referred to as the 'preschool year' or year 0), lågstadiet (encompassing years 1–3), mellanstadiet (covering years 4–6), and högstadiet (spanning years 7–9). Additionally, children aged six to thirteen have the option of attending out-of-school care before and after regular school hours. In 2011, the grading mechanism was revised to a six-grade scale ranging from A to F. Grading from A to E denote passing grades, while F indicates failure. This structure aligns closely with the European Credit Transfer and Accumulation System (ECTS), a prevalent grading model in European higher education. Furthermore, from 2012 onwards, students receive grades starting from year 6. [14]

In Norway, the standard duration for primary and secondary schooling is 13 years. The primary and lower secondary education spans year 1–10, and upper secondary education covers the year 11–13. The Norwegian Directorate for Education and Training supervises the quality of this educational system. The Norwegian education system divides into primary education, which lasts from the year 1–7, and lower secondary education, from the year 8–10. Generally, children begin their primary schooling in the year they turn six and finish lower secondary education by the age of sixteen. [22]

In Norway, the majority of schools are municipally managed, with the local government overseeing their operation and administration. Primary and lower secondary education, which is mandatory and free, emphasizes inclusiveness and tailored learning for all students within a comprehensive system. In primary schools, students in lower secondary schools receive grades twice a year. Upon finishing this level, students receive a certificate presenting their grades and are eligible for three years of upper secondary education. [22]

The Danish educational system, as established by the constitution (Grundloven), provides free schooling at primary and lower secondary levels. Though education is compulsory, attending a physical school is not. Students can learn through public schools (Folkeskolen), private institutions, or homeschooling, with the education aligning with public school standards. Compulsory education spans ten years, starting when a child turns six, with an optional 10th class available. The Folkeskole, averaging 21.6 students in one classroom, aims not only to teach subject-specific skills but also to prepare students for a democratic society. A 2021 agreement introduced regular assessments in reading and math from 2nd to 8th classes. The 9th class culminates in comprehensive exams, encompassing various subjects, along with a week-long project. [12]

2.3 Studies in academic performance

Since 1993, Educational Data Mining (EDM) methods have been employed to address challenges in educational research. [23]

One research study analyzed student academic performance using a decision tree methodology. The primary goal was to forecast the final grade of students pursuing graduate degrees in engineering. The dataset contained 524 instances, each with 18 distinct attributes. This research focused on students' academic performance from high school to the end of their university Engineering program to predict their final graduate grade. Using the decision tree, classification rules were derived by tracing the path from the root node to every leaf node. The final graduate grade labels were divided into four categories: A, B, C, and F. The J48 decision tree algorithm was found to be the most accurate, with an overall accuracy rate of 80.15%. This means 420 out of 524 students were accurately categorized based on their grades. [21]

Another research team studied predictive models for student academic performance using decision trees, neural networks, random forests, and support vector machines (SVM). This study covered subjects in the urban and geographical programs. Data, comprising scores from 20 essential courses, was collected from 123 undergraduates at the School of Urban and Geographical Sciences over four years. The final grades were calculated from the credits and grades of each subject and categorized from A to C. The study revealed significant correlations between courses and prediction results. [11]

The research introduced the AdaBoost ensemble algorithm as a significant tool to improve student classification predictions. AdaBoost, a boosting algorithm, combines multiple weak classifiers trained on a single dataset to create a strong classifier for precision. To validate the AdaBoost method, models were developed using multiple regression, decision tree, neural network, random forests, SVM, and AdaBoost. Once the optimal model for assessing student learning quality was determined, the study highlighted the importance of each feature in the model to enhance student learning. The results highlighted the correlation between courses to improve the final grades. [11]

The continuous development of Educational Data Mining methods has advanced, particularly with contributions from machine learning and deep learning. As research in this domain continues to grow, it provides educators, administrators, and academic institutions with a more tailored guide to academic assistance, curriculum planning, and methods for engaging students.

2.4 Data collection and overview

Over the last three years, Pori has been collecting data using its academic recording system, making changes to improve its utility and effectiveness.

In this study, a total of 18 schools from Pori were analyzed and assessed. Among these institutions, 10 are exclusively primary schools, dedicated to offering foundational education. The remaining 8 schools operate as comprehensive schools, providing both primary and secondary education to their students.

The academic assessment for students begins from Class 4. From this class onward, students receive grades that reflect their academic proficiency and understanding. The grading system employed is structured as follows:

- (10) - Excellent
- (9) - Very good
- (8) - Good
- (7) - Satisfactory
- (6) - Moderate
- (5) - Adequate
- (4) - Fail

This comprehensive grading system aims to capture the diverse range of student capabilities and provide a detailed understanding of their academic journey.

3 Theoretical Foundations of Analysis

This section explores essential computational techniques and their applications in data analysis and machine learning. It covers the fundamentals of machine learning and various optimization strategies such as Stochastic Gradient Descent (SGD), AdaGrad, RMSprop, and Adam, and examines their role in refining machine learning algorithms.

The focus then shifts to deep learning concepts, including Recurrent Neural Networks and Long Short-Term Memory models. Additionally, key statistical methods for analysis, such as Markov state models, correlation, and heatmap visualization, are discussed. The section concludes by highlighting the Power BI application, which is used to derive meaningful insights from complex datasets

3.1 Machine Learning

Artificial intelligence is the logical process for computers to act like humans. Machine learning is a subset of artificial intelligence. The 21st century is often regarded as the "Big Data" age. A question arises: do discernible patterns exist within this vast data? Can these patterns be encoded into computers to derive meaningful and advantageous insights? [1]

Achieving a full understanding of certain processes might remain very difficult, however, it is plausible to develop an accurate approximation. Such an approximation might not explain every reason but could potentially interpret significant portions of the data. The objective of machine learning is to uncover inherent patterns or consistencies. These discovered patterns could enhance our understanding of the fundamental process or be used for making predictions. Assuming that upcoming events will resemble past occurrences described in the data, predictions based on these patterns are highly accurate. [1]

In machine learning, there are two primary methods of learning: supervised and unsupervised. In supervised learning, algorithms are trained by generalizing from defined output examples. Users train the model with both inputs and the corresponding desired outputs. Given new inputs, the algorithm can generate outputs on its own. These models are termed "supervised" because they rely on the provided outcomes for each training example. Constructing a dataset of input-output pairs can be demanding, but the advantages are evident: supervised learning techniques are well-established, and their outcomes can be easily evaluated. [20]

Conversely, unsupervised learning operates when only the input data is available without output labels. Algorithms in this category transform the data into a format that might be more interpretable, either to humans or subsequent algorithms. A common application of unsupervised learning is clustering, which categorizes extensive data into more concise groups of features. [20]

Regardless of the learning approach, whether supervised or unsupervised, it is essential to structure the data in a way that aligns with machine learning libraries.

3.2 Optimization

In the following section, a detailed discussion on optimizers used for machine learning models is presented. Additionally, the section includes an explanation of Stochastic Gradient Descent (SGD), AdaGrad, RMSprop, and Adam.

3.2.1 Stochastic gradient descent

Stochastic gradient descent (SGD) and its variations are primary optimization techniques for machine learning, especially in deep learning. While earlier perspectives often dismissed gradient descent as inefficient, particularly for non-convex optimization tasks, contemporary findings demonstrate its effectiveness in training various machine learning models. Despite no guarantees of reaching an optimal solution rapidly, SGD often achieves satisfactory results in minimal time. This method is particularly beneficial for training large linear models on substantial datasets. The computational demands for each SGD update remain independent of the dataset's size, allowing models to potentially achieve optimal performance even before the entirety of the data is sampled. [8]

3.2.2 Adagrad

The AdaGrad algorithm uniquely adjusts the learning rates for each model parameter. This adjustment is based on the inverse proportion to the square root of the sum of their past squared values. In simpler terms, parameters that significantly influence the loss function experience a faster reduction in their learning rates, while those with a lesser impact see a smaller decrease. This approach effectively facilitates more progress in areas of the parameter space that are less steep. [9]

While AdaGrad shows promising theoretical attributes in convex optimization, its practical application in training deep neural networks has its limitations. The continuous accumulation of squared gradients from the start of the training can lead to an overly rapid reduction in the effective learning rate. As a result, although AdaGrad is efficient in some deep learning scenarios, it doesn't universally excel across all such models. [9]

3.2.3 RMSprop

The RMSProp algorithm enhances AdaGrad for better performance in non-convex settings. It does so by updating the gradient accumulation process into an exponentially weighted moving average, rather than using the entire history of squared gradients as AdaGrad does. AdaGrad, initially designed for convex functions, can reduce the learning rate excessively when applied to non-convex functions in neural network training. However, the issue arises because the learning path in such scenarios passes through various structures before potentially reaching a locally convex area, leading to an undesirably small learning rate. [9]

RMSProp addresses this by discarding older gradient information, allowing it to adapt more quickly once it reaches a convex-like region. This adjustment makes RMSProp resemble of AdaGrad. In practice, RMSProp has proven to be a highly effective and practical algorithm for optimizing deep neural networks. [9]

3.2.4 Adam

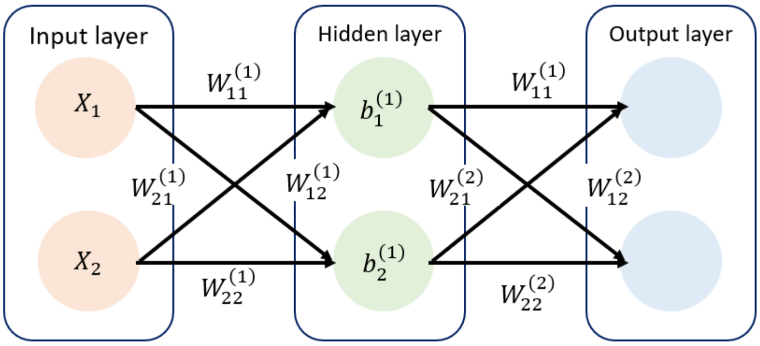
The Adam method is an efficient approach requiring only first-order gradients with minimal memory usage. This technique calculates unique adaptive learning rates for each parameter based on estimations of the gradients' first and second moments, a process termed "adaptive moment estimation." Adam integrates the strengths of two recent optimization methods AdaGrad and RMSProp. [15]

This algorithm is a straightforward yet effective tool for optimizing stochastic objective functions, especially suitable for machine learning problems with large datasets or high-dimensional parameter spaces. This method combines AdaGrad's ability to manage sparse gradients with RMSProp's flexibility in changing environments. It is easy to implement and does not require much memory, making it practical for various applications. [15]

3.3 Deep learning

Deep learning involves advanced artificial neural networks for enhanced machine learning tasks. This chapter focuses on the basics of simple neural networks. Each element that holds data is termed a 'neuron'. While there may be multiple output neurons, forming the output layer, several hidden layers can also be situated between the input and output layers. Neurons within these hidden layers are connected to the next layer, but neurons within the same layer remain isolated from each other. Figure 3.1 illustrates neural network model. In this explanation the sigmoid function serves as the activation function. [25]

Each connection between neurons carries a weight denoted as $w_{ij}^{(k)}$. 'i' sequences from the previous layer, while 'j' represents the sequence in the next layer. 'k'



$$Z_1^{(2)} = W_{11}^{(1)} * X_1 + W_{21}^{(1)} * X_2 + b_1^{(1)} \qquad a_1^{(2)} = \sigma\left(Z_1^{(2)}\right)$$

$$Z_2^{(2)} = W_{12}^{(1)} * X_1 + W_{22}^{(1)} * X_2 + b_2^{(1)} \qquad a_2^{(2)} = \sigma\left(Z_2^{(2)}\right)$$

Where activation function is $\sigma(x) = \frac{1}{1+e^{-x}}$

$$Z_1^{(3)} = W_{11}^{(2)} * a_1^{(2)} + W_{21}^{(2)} * a_2^{(2)} + b_1^{(2)} \qquad a_1^{(3)} = \sigma\left(Z_1^{(3)}\right)$$

$$Z_2^{(3)} = W_{12}^{(2)} * a_1^{(2)} + W_{22}^{(2)} * a_2^{(2)} + b_2^{(2)} \qquad a_2^{(3)} = \sigma\left(Z_2^{(2)}\right)$$

Figure 3.1 *Simple neural network diagram with formula*

represents an index of the layer. These weights define the proportion of the original value to be transferred to the next neuron. In the Figure 3.1, the inputs are labeled as X_i, with weights as $w_{ij}^{(k)}$. Each neuron has an adjustable bias, represented as $b_i^{(k)}$, which is added to the previous sum. The result is typically denoted by $z_i^{(k)}$. This combined value is processed by the activation function, in this instance, the sigmoid. The process continues to subsequent layers until it reaches the output layer. Final outputs are labeled as $a_i^{(k)}$. Deep learning prominently employs the backpropagation technique. Backpropagation is the primary mechanism by which a neural network adjusts its weights. It evaluates the loss function and then implements gradient adjustments. [25]

To refine the predefined weights and biases, a loss function is used to calculate the difference between the output value and the labeled value. A common form of the loss function is the squared mean error, defined as:

$$L = \frac{1}{2} \sum_{n \in train} (a^n - \hat{a}^n)^2$$

where L represents loss function, $\hat{a}^n$ represents the outputs from the training phase, and a^n corresponds to the expected outputs. [26]

To minimize the squared mean error, it is crucial to determine the gradient of the loss function with respect to the weights and biases between the hidden layer and the output. Similarly, understanding the gradient with respect to the weights and biases from the input to the hidden layer is vital. This relationship is expressed as:

$$\frac{\partial L}{\partial W^{(n)}} = \frac{\partial L}{\partial \hat{a}^{(n)}} * \frac{\partial \hat{a}^{(n)}}{\partial Z^{(n)}} * \frac{\partial Z^{(n)}}{\partial W^{(n)}}$$

With the incorporation of the sigmoid activation function, the derivative takes on the following form:

$$\frac{\partial L}{\partial \hat{a}^{(n)}} = -(a^{(n)} - \hat{a}^{(n)})$$

$$\frac{\partial \hat{a}^{(n)}}{\partial Z^{(n)}} = \hat{a}^{(n)} \times (1 - \hat{a}^{(n)})$$

Based on the weight across the layers, the formulas can be generalized as follows:

$$\frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial \hat{a}^{(n)}} \times \frac{\partial \hat{a}^{(n)}}{\partial Z^{(n)}} \times \frac{\partial Z^{(n)}}{\partial w_{ij}} = -\frac{\partial Z^{(n)}}{\partial w_{ij}} \times \hat{a}^{(n)}(1 - \hat{a}^{(n)})(a^{(n)} - \hat{a}^{(n)})$$

The procedure for adjusting weights is systematic, known as the general weight update rule:

$$w_{ij}^{new} = w_{ij}^{old} - \eta \frac{\partial L}{\partial w_{ij}^{old}}$$

In this formula, η represents the learning rate, generally varying between 0.1 and 0.00001. [26]

3.3.1 Recurrent Neural Network

Neural networks require longer sequences when there is an increased number of inputs. This leads to a heavy network since each input is interconnected with all others. As a result, the network comprises numerous connections, signifying a multitude of weights. The Recurrent Neural Network model is one way to optimize this densely connected network.

Recurrent Neural Networks (RNNs) were first mentioned in the 1980s [24]. Over time, many scholars have enhanced their designs, driven by advancements in deep learning and computational capabilities. The core principle of RNNs is that the model processes each sequence element individually, retaining its memory for subsequent elements. This mechanism resembles human reading habits, where individuals discern words by sequentially recognizing letters. Similarly, RNNs can correlate a word with a particular outcome by remembering the sequence of incoming letters. [19]

Figure 3.2 illustrates a recursive cell on the left. The model on the right reveals a sequential flow of results. The outcome of X_{t-2} serves as input to the subsequent stage, which is X_{t-1}. As this process unfolds, the RNN emits an output and modifies

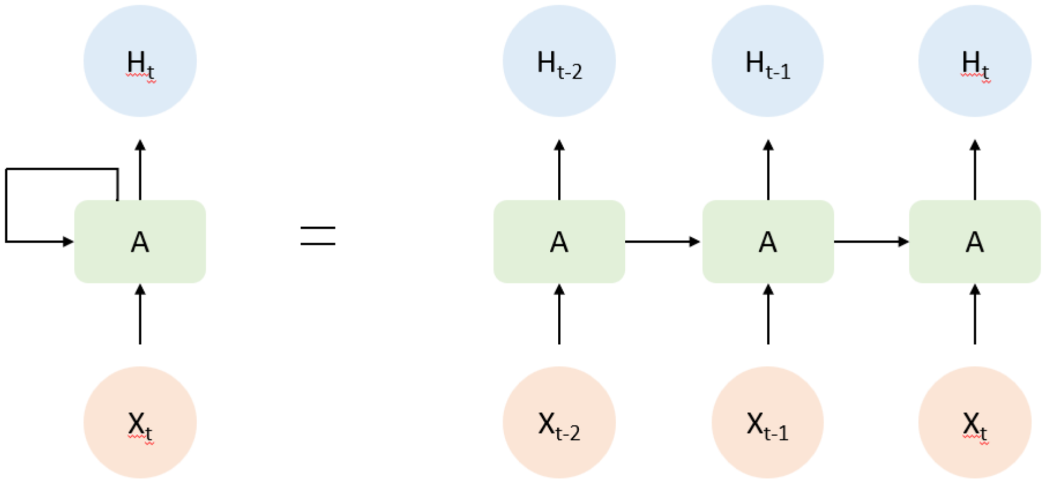


Figure 3.2 *RNN Structure*

its internal parameters. By adjusting its internal parameters, the unit learns from the data it receives and from the memory of previous data. The sum of this learning is the state of the RNN cell. [19]

3.3.2 Long Short-Term Memory model

The LSTM (Long Short-Term Memory) architecture was introduced by Hochreiter and Schmidhuber in 1997 [13]. The LSTM aims to enable RNNs to differentiate between short-term and long-term states. This "state" refers to the cell's memory, which is divided into short-term memory and long-term memory. The short-term memory directly integrates with the incoming sequence data. On the other hand, long-term memory selectively captures elements from the short-term memory that need to be stored for a long time. Moreover, the long-term memory channel consists of smaller sets of parameters for tuning. Its design is relatively straightforward, primarily involving arithmetic operations on the elements from short-term memory. [18]

Figure 3.3 illustrates an LSTM unit. This unit consists of multiple gates, enabling summation, multiplication, and the use of activation functions to regulate the internal flow of information. These adjustments allows the gate to either retain, amplify, or dismiss incoming data from both short and long-term memory channels. [18]

The LSTM operates in the following manner: initially, the existing short-term memory (depicted as the hidden state in Figure 3.3) integrates with the new input, resulting in an initial derivation. This combined memory, which now contains both the previous and current input, is channeled through the "forget gate" towards the long-term memory (represented as the Cell state in Figure 3.3). The "forget gate" determines which pieces of information will be discarded before they reach

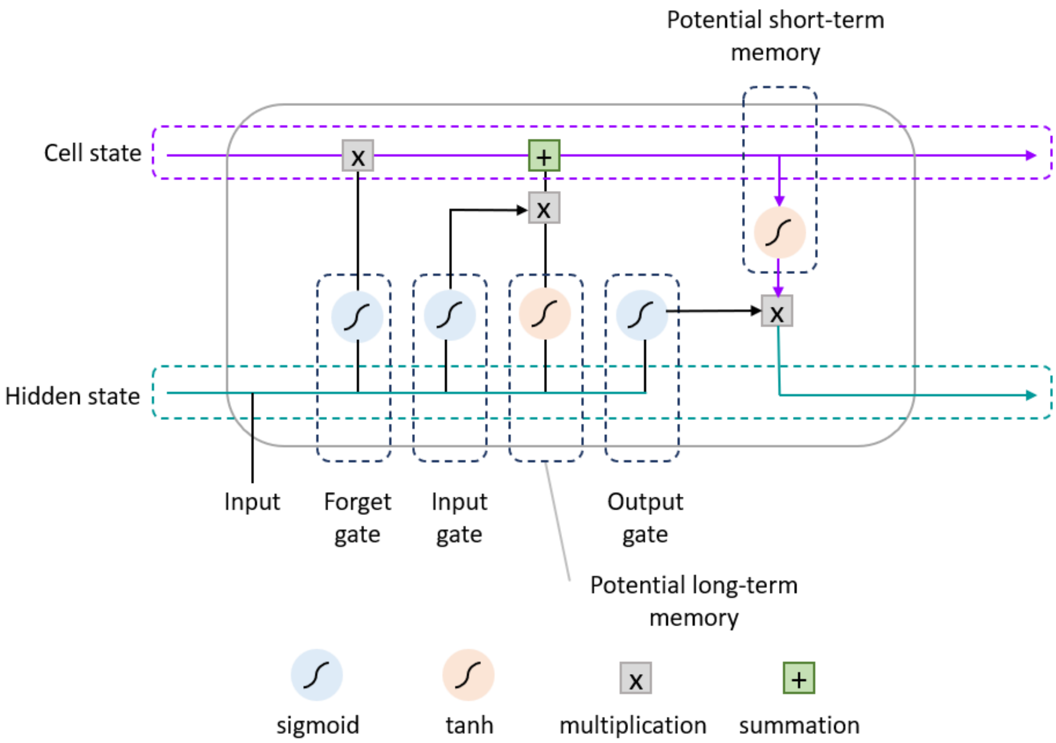


Figure 3.3 *LSTM unit diagram borrowed from [18]*

the long-term memory. Using the sigmoid activation function, the gate filters out irrelevant signals. Subsequently, the values proceeding through the activation stage are multiplied with the existing long-term memory values. [18]

The input also interacts with the input and output gates. Within the input gate, data undergoes processing via both the sigmoid and tanh functions. The products of these functions are multiplied and then added to the long-term memory. The newly revised long-term memory values are then multiplied with the outcomes from the output gate, leading to an update of the short-term memory values. [18]

In LSTMs, both sigmoid and tanh activation functions are significant as activation functions. While tanh ensures the normalization of input values between -1 and 1, the sigmoid constrains them between 0 and 1. Consequently, tanh preserves inputs within an operable range, and sigmoid has the capability to diminish weaker signals. In other words, the sigmoid function enhances relevant signals and suppresses irrelevant ones. [18]

3.4 Statistical method for analysis

This chapter focuses on the essential tools of Exploratory Data Analysis (EDA), highlighting approaches such as mean, median, standard deviation, p-values, and correlation. In this study, mean, median, and standard deviation were evaluated to discern trends in the data. The mean provides an average value of the dataset, while the median indicates its central value. Variance is defined as the sum of squared deviations from the mean, divided by n, where n represents the total number of data points. The standard deviation, which is the square root of variance, conveys the dispersion of the dataset.

Statistical significance is a measure used by researchers to determine if the results of an experiment or analysis differ significantly from what might be expected by chance alone. When results deviate substantially from expected random outcomes, they are considered statistically significant. The p-value plays a crucial role in determining this statistical significance. Operating under the null hypothesis, which assumes no effect or difference, the p-value calculates the probability of observing results as extreme as, or more extreme than, those actually observed. [2]

In other words, the p-value describes the likelihood that results occurred by mere chance. A smaller p-value is preferable, indicating a stronger case against the result being coincidental. In the field of data science, the p-value serves as a valuable metric. It aids scientists in determining whether their findings have potential significance. [2]

3.4.1 Markov state model

The Markov model has been employed in numerous studies to investigate molecular interactions in proteins and oligomers [4], as well as transitions between DNA proteins [29]. The Markov Model is a mathematical representation of a system characterized by defined states, wherein transitions can transpire from one state to another. A fundamental characteristic of this model is that any current state, symbolized as $X(t)$, can impact subsequent states, represented as $X(t + \Delta t)$, assuming Δt is a positive value. Over time, states within this model evolve, guided by probability. The evolution of states following a time span, $X(t + \Delta t)$, is exclusively influenced by its antecedent state, $X(t)$. [10]

State	1	2	3
1	0.7	0.2	0.1
2	0.5	0.3	0.2
3	0.4	0.2	0.4

Figure 3.4 *Markov transition probability matrix*

Figure 3.4 illustrates a simple markov model with 3 states. For illustrative purposes, transitions were assumed to happen at consistent intervals, with the probability of these transitions being consistent over time. The probabilities of transitions between these three states within a specified time frame are captured in a 2D matrix. This matrix is often referred to as the transition probability matrix (TPM). Within this matrix, the term p_{ij} signifies the likelihood of a shift from state i to state j. For Markov State Models (MSM), the matrix is normalized by row. [10]

3.4.2 Correlation

The correlation coefficient is calculated using the formula mentioned above. The correlation coefficient provides a standardized measure of the relationship between two variables, always presented within a consistent scale. Pearson's correlation coefficient is calculated by multiplying the deviations from the mean for each variable and then dividing by the product of their standard deviations. The formula of Information gain is given by:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where x_i are the values of the x-variable in a sample, $\bar{x}$ is the mean of the x-variables, y_i are the values of the y-variable in a sample, and $\bar{y}$ is the mean of the y-variables. Correlation value can range between +1, indicating a perfect positive relationship, and -1, signifying a perfect negative relationship. A value of 0 indicates no relationship. This coefficient can be influenced by outliers, hence, many software tools offer robust alternatives to the traditional correlation coefficient. [3] [16]

The Correlation Matrix displays variables on both its rows and columns, with each cell indicating the correlation value between the corresponding variables. Conversely, a Scatterplot visually plots values of one variable on the x-axis, while values of the other variable are mapped on the y-axis. [3]

3.4.3 Heatmap

A heatmap serves as a graphical representation of data, where each value found in a matrix is represented as colors. It is an efficient method to provide a visually comprehensive overview. The use of a heatmap is particularly effective when attempting to plot the correlations among different factors. [5]

In this study, the Seaborn library was utilised to produce the result. The Seaborn library is highly regarded for its ease of use and high-quality visualizations. By using Seaborn's heatmap function, a color-coded matrix was produced, where the intensity of the color represented the strength of the correlation between the variables. [30]

In the Seaborn visualization library, the default color scheme illustrates correlation strength as follows: strong positive correlations are represented in red, strong negative correlations appear in blue, and the absence of correlation is indicated by a light orange color. Adjacent to each heatmap, a color bar provides a reference for interpreting these colors.

3.5 Power BI Application

Power BI is a user-friendly tool designed for data visualization, allowing users to analyze data without in-depth technical expertise. The ”BI” in Power BI stands for business intelligence, highlighting its importance in assisting businesses in understanding their data effectively. As the data analysis field grew, Microsoft Excel became widely used. However, with the advent of big data, it faced challenges in handling vast amounts of information. This realization prompted a shift towards more robust tools like Power BI. In this study, Power BI Desktop is utilized, a widely approved free application. This tool connects to a multitude of external sources and offers comprehensive analytics. [17]

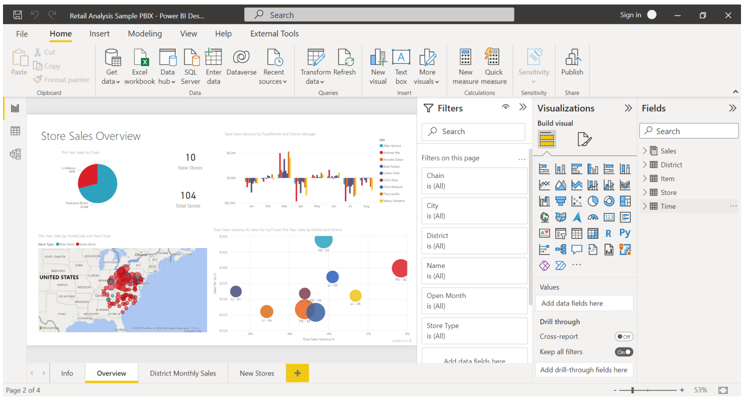


Figure 3.5 *Example of power BI report borrowed from [17]*

In this study, various visualization components such as line and clustered column charts, donut charts, tables, bar charts, and the key influencer components were utilized. The key influencer visual assists in understanding the primary factors affecting a chosen metric by ranking and displaying them. Internally, the tool leverages ML.NET to conduct a linear regression, assessing how the outcome field varies based on the explanatory factors. In case of insufficiency of data points for certain characteristics, a Wald test is applied, with a p-value threshold of 0.05 set for determination. The key influencer tool in Power BI provides a systematic approach to explore the impact of various factors on a selected attribute, showcasing those with the most significant relationships to the outcome. [6]

The top segment tab in Key influencer components uses ML.NET to create a decision tree, which helps identify important subgroups based on specific metrics. This decision tree examines different factors and tries to best categorize the data. After its first categorization, the tree continues to refine the data into even more specific groups. With each new group formed, the system checks if there is enough data to make the group valid. A statistical method, with a specific criteria (p-value of 0.05), checks the validity of these groups. Once finished, the decision tree turns these groups into Power BI filters. These filters are then presented as a segment in the visual. [6]

Power BI has its unique approach to calculating statistical metrics like average, standard deviation, and variance. Notably, it performs these calculations on a row-wise basis. This implies that each row in a dataset, irrespective of its origin or relation to other rows, is treated as an individual data point. In this study, a dataset with 11 distinct features was used to visualize. In a bar chart showing the relationship between subjects and the average of absent lessons, Power BI aggregates all rows corresponding to each subject and then divides the sum by the total row count.

While the visuals provide a comprehensive understanding of averages, they do not address another critical aspect of data analysis: understanding data dispersion. The 'error bar' feature in Power BI was utilised to display the data's variability. An error bar can represent data variations using fields, percentages, percentiles, or standard deviations. In this study, the percentile error bar, capturing both the 25th and 75th percentiles, was chosen. These indicators closely resemble the insights that a box plot would offer.

4 Data Acquisition, Preprocessing, and Analytical Techniques

In this chapter, the primary objective is to describe the implementation of the program. The chapter is divided into four sections: Data Description, Data Preprocessing, Power BI Application Description, and Python Program Description.

Initially, in the 'Data description' section, a thorough examination of the original data is undertaken. The types of information and the structure of the data are presented, establishing a solid foundation for the analytical methodologies used in the later stages.

In the 'Data preprocessing' section, the procedures applied to the data are introduced sequentially, emphasizing technical precision and a methodical approach to ensure data integrity and usability. This section underscores the importance of data preprocessing for enhancing the quality of data analysis and model building.

In the 'Power BI application' section, a detailed exploration of Power BI pages is provided. These pages seamlessly transition from presenting basic demographic distributions to offering intricate insights into absenteeism and academic performance. The distinctive features and functions are utilized to create a visualization and analytical interface for interactive data analysis.

Finally, in the 'Data analysis methods' section, methods for correlation analysis and prediction analysis are highlighted. These libraries have been selected to analyze and illustrate the correlation among subjects and to predict students' academic performance from the previous year.

4.1 Data description

The City of Pori initiated the collection and analysis of student data. The city has been collecting data on under-aged students since 2019. A significant development occurred in 2020 with the enhancement of data collection systems in Pori, which made the data accumulation process more efficient. The data is obtained from 2020, spanning three academic years, and specifically targets students from grades 1 to 9. Due to the structure of the original data, it was divided into four separate CSV files before being shared. Moreover, the student IDs were hashed to ensure anonymity, preventing the identification of individual students. These four files encompass various data fields such as 'Student hash,' 'School name,' 'Support type,' 'Finnish as a second language,' 'Gender,' 'Class,' 'Birth quarter,' 'Academic year,' 'Subject,' 'Total absent time,' 'Number of absences,' 'Grade,' 'Special support,' and 'Support started date.' In this study, the term "Class" refers to the year group, while ”Grade” denotes the academic grade. Figure 4.1 illustrates the structure of the initial four files. Shared attributes such as class, school name, class code, and student hash are present across the files, creating a linked data framework. Within the figure, the Grade file, Absent file, and Support file are connected to the Student file, highlighting a central point of data association among them.

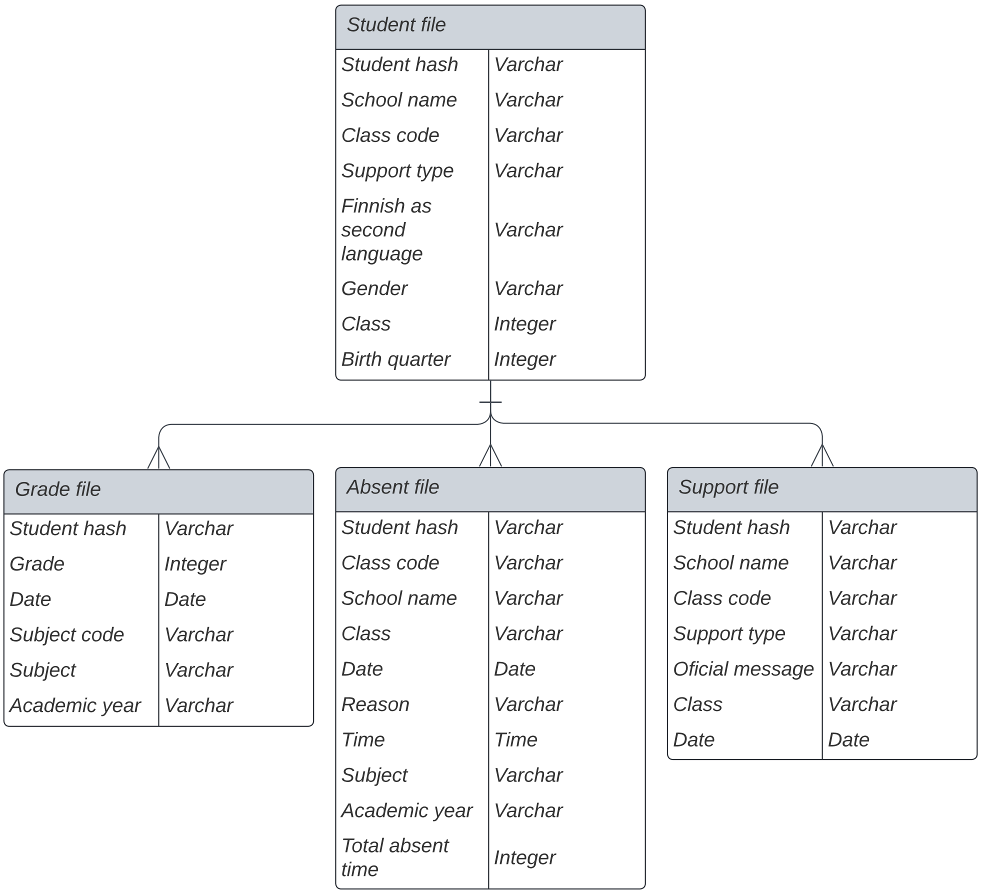


Figure 4.1 *File diagram*

4.1.1 Student File

The ’Student File’ mainly contains a wide range of student-related information, acting as a basis for further analysis. In the dataset, the ’Gender’ field refers only to the students’ biological gender. The data reveals a unique student count of 6638, including 19 schools, with three distinctive support types. Furthermore, there are Finnish as a second language, gender, and birth quarter, each divided into two, two, and four categories, respectively. Lastly, the class or year group ranges from 1 to 9. This large volume of data highlights the chance to gain meaningful insights and understand various factors affecting the academic performance of the students.

4.1.2 Grade File

The 'Grade File' primarily serves as an academic performance indicator, showing the grades obtained by students. The 'Grade' field explicitly indicates the student's academic performance in a range of 4-10.

The 'Academic year' represents academic periods. With a total of 134,721 recorded grades, the 'Grade File' provides vast data for analytical exploration. Furthermore, the grade data reveals a unique student count of 4483, with 19 subjects over three academic years. This dataset gives the possibility to examine academic performance trends, subject-wise competency, and subject correlations.

4.1.3 Absent File

The 'Absent File' represents a comprehensive collection of absence records, capturing a variety of details such as class, reasons for absence, duration, and the subject.

The 'Absent File' consists of durations of absenteeism through the 'Absent minute' field. This absence data reveals 1,211,215 rows with a unique student count of 6635, encompassing 19 school names, five different reasons for absences, 20 different subjects, over three academic years. The 'Absent File' provides meaningful conclusions which could be useful for academic policies to better support student attendance and overall academic achievement.

4.1.4 Special Support File

The 'Special Support File' contains students who have received special support. With a total of 400 rows, each corresponding to a unique student across 18 different schools, their classes range from 1 to 9. This 'Special Support File' presents a dataset for specialized support within the academic environment. This data presents an opportunity for detailed analysis, which could be vital in improving the understanding and management of special support across the classes.

4.2 Data Preprocessing

Figure 4.2 shows the final structure of the cleaned and organized data. The task of preparing the dataset for comprehensive analysis required a structured approach.

Two preprocessing methods were adapted: merging and data cleaning. The initial phase involved integrating four separate files into a single file. This integration aimed at centralizing the data for analysis covered in this study. This step is crucial for removing any inconsistencies, inaccuracies, or duplications found in the data, thereby enhancing its quality. Data cleaning involved various activities, such as rectifying any discrepancies in the class and academic year columns, and condensing the extensive data from the 'Absent File'.

Preprocessed data	
Student hash	*Varchar*
School name	*Varchar*
Support type	*Varchar*
Finnish as second language	*Integer*
Gender	*Varchar*
Class	*Integer*
Birth quarter	*Integer*
Academic year	*Varchar*
Subject	*Varchar*
Total absent time	*Integer*
Number of absent	*Integer*
Grade	*Integer*
Special support	*Varchar*
Special support started	*Integer*
Total absent lesson	*Integer*

Figure 4.2 *Cleaned data diagram*

The merging process was applied systematically to ensure minimal loss of data. Prior to merging, the 'Absent File' was cleaned. Distinct from the other three files, the 'Absent File' had a considerable number of 1,211,215 rows, making it complicated for insightful examination. The aggregation was applied to the file based on the 'Subject' field and by summing the 'Absent time' for each subject. Subsequently, rows lacking 'Subject' data were removed from the file. Subsequently, the outer join operation was utilized to merge the 'Grade File' with the 'Absent File'.

Following this, the left join operation was used to merge the file with the 'Student File'. This ensured that all student records were kept intact, and the corresponding grade details were added without compromising data quality. Lastly, the 'Special Support File' was integrated using the same left join method.

The result of these systematic merging steps was a single, comprehensive file. This file captures students' academic achievements, absences, and special support details. The clear and organized structure of this merged file makes future analysis easier and more effective.

The cleaning process of the dataset was applied using two methods to ensure the accuracy and usability of the information. The initial step involved eliminating data due to insufficient student data.

The second method was applied to remove inconsistencies between the 'Class' and 'Academic year' fields. The 'Class' column represents the year group of students. Nevertheless, the column contained the final class data for each student, resulting in a mismatch with the 'Academic year' field. To rectify this, a Python script was executed to align the 'Class' field with the appropriate 'Academic year'.

4.3 Power BI Application

Three distinct pages were implemented within the Power BI application: General Information, Absence Analysis, and Grade Analysis. These pages were structured to visualize the complex features of data, allowing for deeper exploration and interpretation.

The General Information page serves as an introduction page. It displays a variety of fundamental data points which include the count of students by school, gender, proficiency in Finnish, and birth quarter. This page delivers demographic distributions of the data.

The Absence Analysis page provides a deeper dive into the patterns and trends of absenteeism. Specifically, the analyses are divided by different support types, gender distribution by subject, and the pattern of absenteeism over various academic years. The objective of this page is to study potential correlations and insights regarding absenteeism.

The Grade Analysis page is designed to examine academic performance. It features various charts focused on grade distributions across different support groups, genders, and different language groups. This detailed analysis aims to spot trends in academic performance.

By dividing the analysis into these three separate pages, the Power BI application is equipped to provide an intensive understanding of the data, transitioning from a general overview to more concentrated analyses on absenteeism and academic performance.

4.3.1 General Information

Within the general information pages, there are five charts presented. Two bar charts illustrate the student counts by school and the percentage distribution within each support group. The remaining three pie charts serve as filters for Native Finnish speakers, gender, and birth quarter. Figure 4.3 illustrates the general information page.

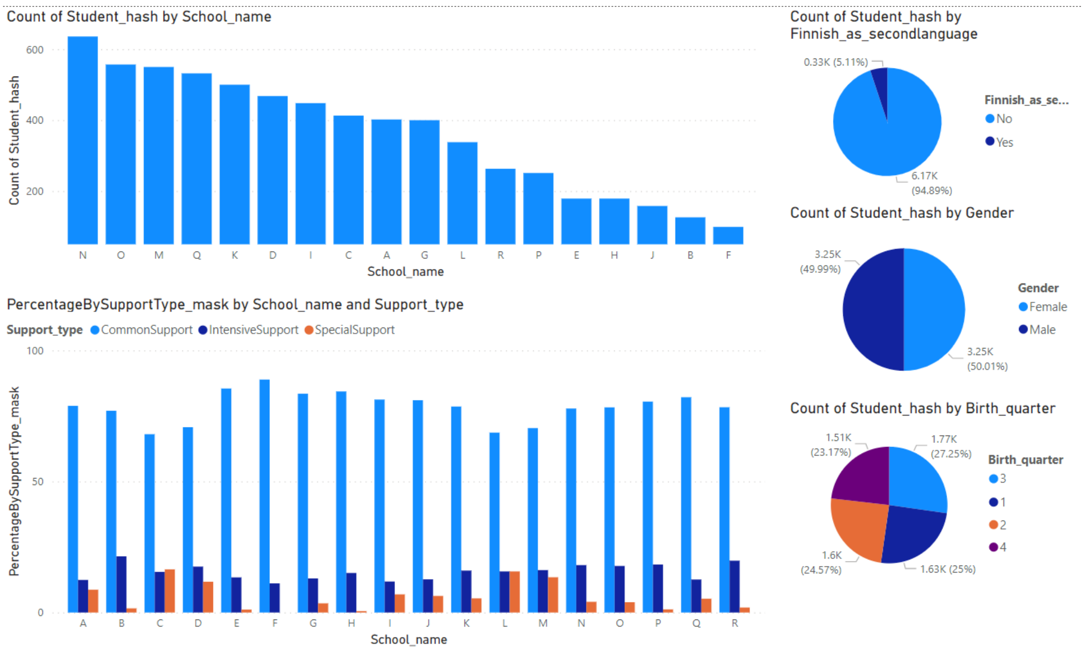


Figure 4.3 *General Information page*

4.3.2 Absent Analysis

The absence analysis page showcases five bar charts and key influencer components. Each chart displays the 25th and 75th percentiles. Due to Power BI's limitations with multi-dimensional data, each chart utilizes a modified data frame. Figure 4.4 illustrates the Absent Analysis page.

In the chart representing the average number of absent lessons by academic year, absences are averaged for each student per academic year. In the subject-based analysis, absences are averaged for each student by subject. When examining absences by school, they are initially averaged for each student per academic year, then further averaged across academic years. For grade and class-based analyses, absences are averaged for each student by grade and class, respectively. The key influencer utilizes absences averaged by class, subject, grade, and individual students.

4.3.3 Grade Analysis

The grade analysis page similarly presents five bar charts and key influencer components. Each chart is highlighted with the 25th and 75th percentiles. Due to the constraints of Power BI in processing multi-dimensional data, a tailored data set is used for each chart. Figure 4.5 illustrates the Grade Analysis page.

Figure 4.4 *Absent Analysis page*

In the chart detailing the average grades by academic year, grades are determined by averaging each student's performance per academic year. For subject-specific analysis, grades are averaged for each student by subject. In the school-based analysis, grades are initially averaged per academic year for each student and subsequently averaged across the academic years. For the class-based chart, grades are averaged for each student by class.

4.4 Data Analysis Methods

In this section, the data analysis methods, including heatmap analysis and prediction analysis, are presented.

Figure 4.5 *Grade Analysis page*

4.4.1 Heatmap analysis

In this study, a heatmap was used to explore the relationships between various factors such as grade, total absent time, support type, proficiency in Finnish as a second language, gender, class, and birth quarter. The objective was to highlight any potential relationships or trends that might exist among these variables. Additionally, an analysis of the correlation between subjects by each grade was conducted to uncover any subject-specific trends or relationships relevant to academic performance.

4.4.2 Prediction analysis

In conducting the prediction analysis, two different types of prediction methods were applied. The analysis focused on predicting students' academic performance for the final academic year, which is significant for understanding and potentially enhancing educational outcomes.

To enhance the clarity and analysis of academic performance, not only were students with three academic years of data selected, but also the seventeen subjects were grouped into three distinct categories. This grouping aimed to bring together similar subjects, creating a better structure for prediction. The first category, titled Social Sciences, included the subjects of 'History', 'Ethics', 'Finnish Language', and 'Social Studies'. The second category, named Arts and Physical Education, encompassed the subjects of 'Art', 'Music', 'Handicraft', 'Cooking Studies', 'Physical Education', and 'Health Education'. The third category, referred to as the Science domain, comprised the subjects of 'Mathematics', 'Environmental Studies', 'Biology', 'Geography', 'Chemistry', 'Physics', and 'Information Technology'. By averaging the grades under each subject category, more organized data was obtained.

Initially, a machine learning approach was used, utilizing a Long Short-Term Memory (LSTM) model. This model is well-suited for handling time-series data, making it adequate for analyzing academic performance over a period. Simultaneously, a statistical modeling approach was also utilized to enhance the prediction. The Markov model was employed in this aspect. Markov models are recognized for their ability to model sequential data, making them a suitable choice for analyzing the sequential data statistically. By using both machine learning and statistical modeling approaches, a more comprehensive and potentially more accurate prediction of students' academic performance could be achieved.

The combination of machine learning and traditional statistical modeling allowed for a richer analysis. This integrated approach aimed at achieving accurate predictions and understanding the factors influencing academic performance. Through these analytical techniques, a significant contribution was made towards understanding and potentially improving academic outcomes. In conclusion, the prediction analysis provided an insightful forecast of students' academic performance for the last academic year.

5 Analysis And Results

The objective of this chapter is to explore the analysis and results of the visualization. The chapter consists of five segments: student analysis, absent analysis, grade analysis, heatmap results, and academic performance prediction results. This chapter delivers significant trends as well as valuable insights about underage students and their academic performance. The various charts and graphical displays provide a clear visual understanding of the observed data trends and patterns. In the depicted charts, the Interquartile Range (IQR) is represented by bars. The IQR is defined as the difference between the 75th percentile and the 25th percentile, representing the middle 50% of the scores [31]. A broader IQR indicates a higher variability in data.

The primary objective of the student analysis section is to understand general information, which includes the number of students as well as the distribution of support groups across various schools. Subsequently, the absent analysis focuses on the examination of absenteeism, emphasizing its trends and its association with academic performance. The grade analysis section is devoted to revealing the distribution and trends in academic grades.

In the heatmap results section, analyses are created using the seaborn library. The heatmap depicts color-coded correlations among subjects within each grade, as well as correlations among each feature. The final section presents academic performance prediction results. Utilizing machine learning libraries and statistical models, this section aims to predict the academic grades of students based on past and present data.

5.1 Student analysis

Figure 5.1 provides a comprehensive overview of student demographics across various schools. This visualization displays key information such as the distribution of students by support type, gender, Finnish language proficiency, and birth quarter. Schools N, M, and O have the highest student counts. In Pori, approximately 5% of students have Finnish as their second language. The gender distribution among students is nearly even between females and males. Birth quarters also exhibit a balanced distribution. In the support chart, the data reveals a higher proportion of students in Schools C, L, and M receiving special support.

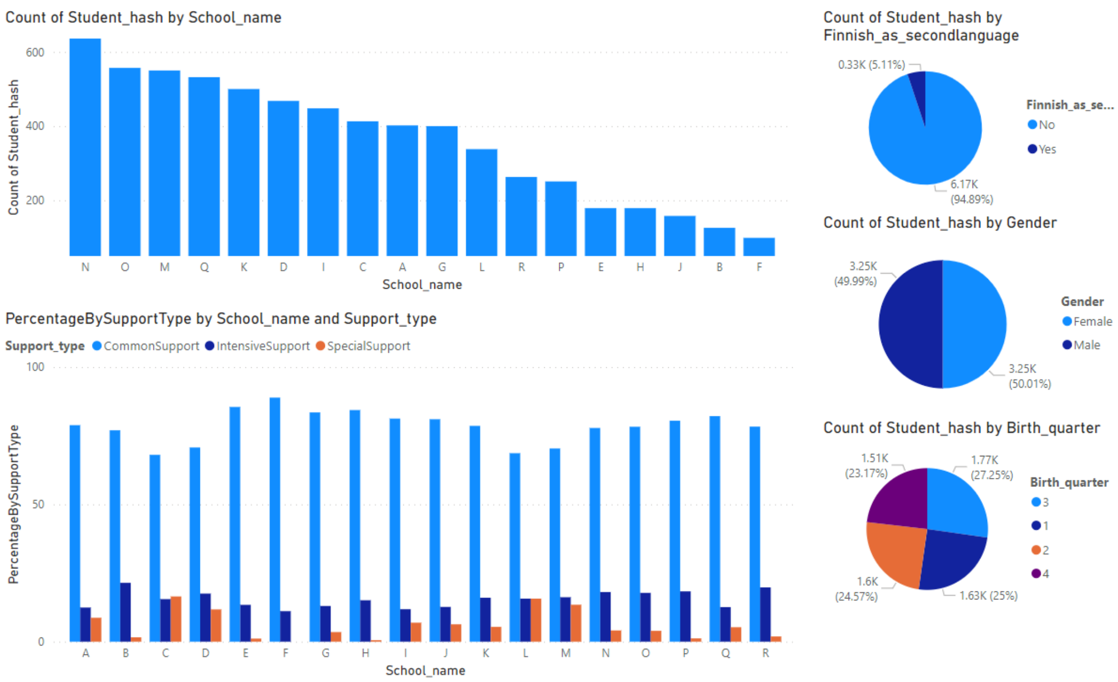


Figure 5.1 *Student Demographics and Support Types Across Schools*

5.2 Absent analysis

In Figure 5.2, two distinct patterns of absenteeism among students are observed. The percentage is calculated based on all absences, not on the number of lessons. The chart on the left reveals that absence is more prevalent during the morning hours compared to the afternoon. Conversely, the chart on the right highlights a lower absenteeism rates on Fridays than other days of the week. Lower absenteeism in the afternoon and on Fridays is partialy due to fewer learning hours

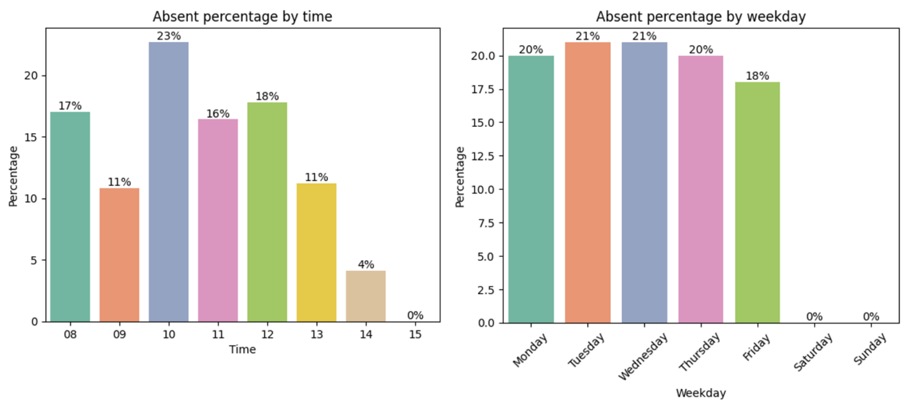


Figure 5.2 *Absenteeism by weekdays and time*

In Figure 5.3, the chart provides an overview of absenteeism percentages for each month across three academic years. Significantly, March stands out with the highest rate of student absenteeism compared to other months. In June, the percentage of absences is remarkably low as well due to the fewer learning hours compared to other months.

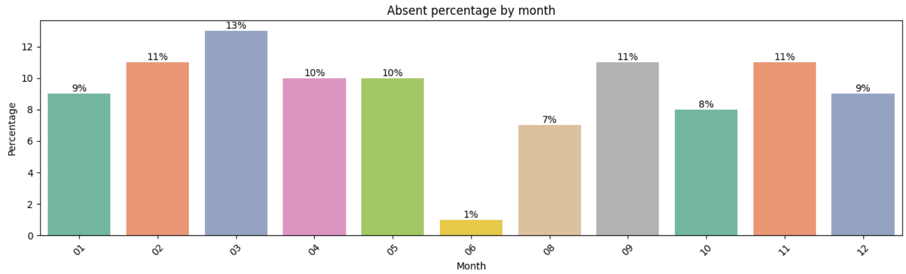


Figure 5.3 *Absenteeism by month*

In Figure 5.4, the chart depicts a discernible trend in absenteeism over three academic years. The data are averaged for each student per academic year. In the academic year 2020-2021, the average absent lesson per student for a subject stood at 3.97. Due to COVID-19, online learning was prevalent during the academic year 2020-2021. This figure records 6.49 in academic year 2021-2022 and 6.63 in 2022-2023. Over this period, it is evident that the average absences per student tend to converge around the value of 6.5 per subject.

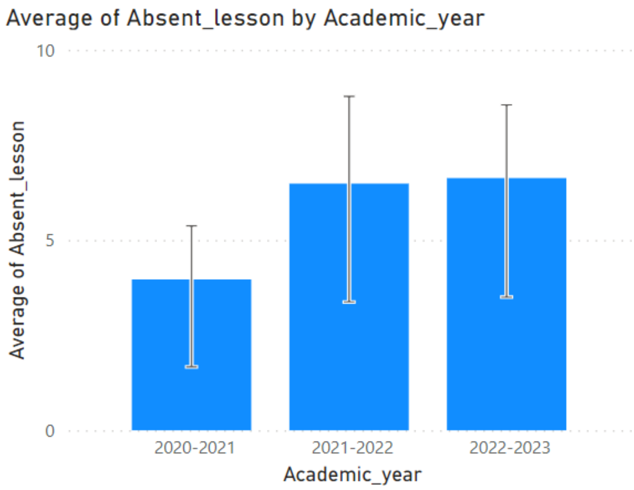


Figure 5.4 *Average absences per student for a subject by academic year*

In Figure 5.5, the average distribution of absent lessons for various subjects during one academic year per student is presented. Finnish Language, Mathematics, and Physical Education exhibit the highest number of absent lessons across the assessed subjects. Conversely, Ethics, Swedish, and Health Education disclose the lowest number of absent lessons. Averages were calculated across students, and the results are not normalized according to the number of lessons per week for each subject, as this data was not available. A more detailed analysis based on Finnish language background unveils a distinguishable trend among non-native students.

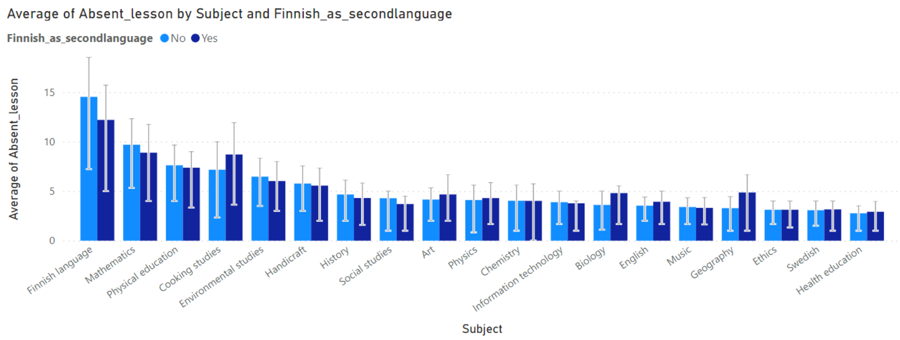


Figure 5.5 *Annual average absences by subject and Finnish language group*

Within the non-native speaker group, subjects such as Cooking Studies, Biology, and Geography are significantly associated with a higher rate of absent lesson. The notable difference in absenteeism among non-native students in these subjects reflects underlying factors, which might involve greater language challenges compared to other subjects.

Figure 5.6 presents a distinctive trend in the absent lessons among various support groups across different schools. Unfortunately, School F did not have any data for a special support group. The groups that receive intensive and special support exhibit a higher number of absent lessons compared to the common support group. Schools M, G, and N record the highest number of absent lessons. A closer examination reveals that the special support group, in particular, exhibits a high number of absent lessons, especially in schools B, P, and D. In schools G, J, and D, a higher standard deviation is observed among special groups.

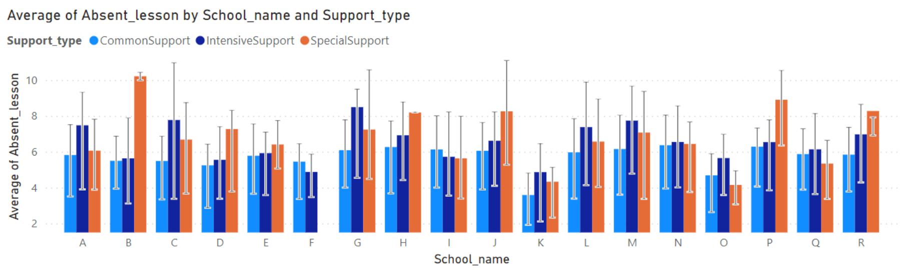


Figure 5.6 *Annual average absences per student by support group across schools*

In Figure 5.7, a trend is depicted regarding the average number of missed lessons according to the academic grade. A closer examination of the data also reveals that the common support group exhibited a higher rate of absenteeism to achieve a Grade 4. For the intensive support group in Grade 4, the absenteeism rate for the 75th percentile is below average. This indicates that beyond this percentile, the absenteeism rate is significantly higher compared to the rest of the 75% of individuals.

In Figure 5.8, a clear trend is illustrated regarding the average number of missed

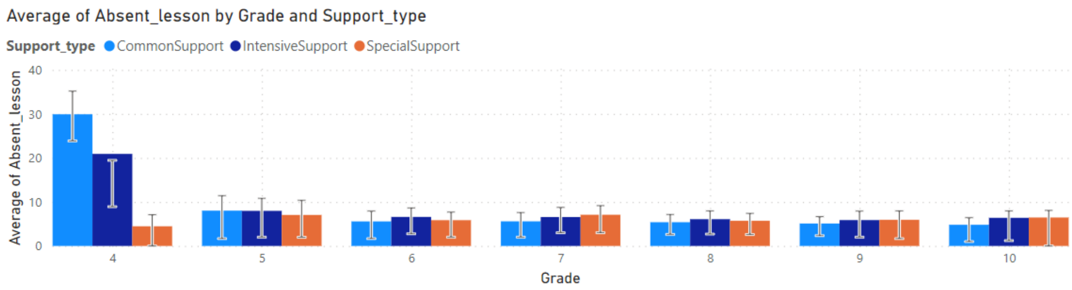


Figure 5.7 *Annual average absenteeism by grade and support type*

lessons from Grade 8 to Grade 9, with the average number of absent lessons increasing. Furthermore, a gender-based analysis of absenteeism reveals distinct patterns. Up until Grade 8, female students are observed to have a lower absenteeism rate compared to male students. However, this trend reverses from Grade 8 onwards, suggesting a shift in absenteeism patterns between genders as they advance in their studies.

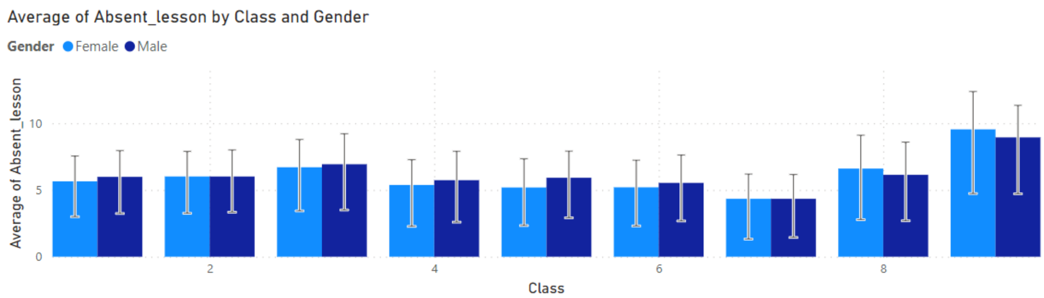


Figure 5.8 *Annual average absenteeism by class and gender*

Figure 5.9, the Power BI key influencer analysis shows the factors contributing to the growth of total absent lessons. The analysis pinpoints four primary factors that lead to an increase in absent lessons. First, for the subject Finnish Language, an average increase of 7.73 lessons is observed. Second, for the subject Mathematics, an average increase of 5.7 absent lessons is observed. Third, the subject of Physical Education shows an increase in the average number of absent lessons by 3.68 lessons. Lastly, classes higher than 8 demonstrate a significant rise, with 3.42 additional absent lessons. Averages were calculated across students, and the results are not normalized according to the number of lessons per week for each subject, as this data was not available.

5.3 Grade analysis

Figure 5.10 illustrates the average grade achieved by students for a subject, categorized by academic year. The academic grades are averaged for each student per academic year. For the academic year 2020-2021, the average grade per subject

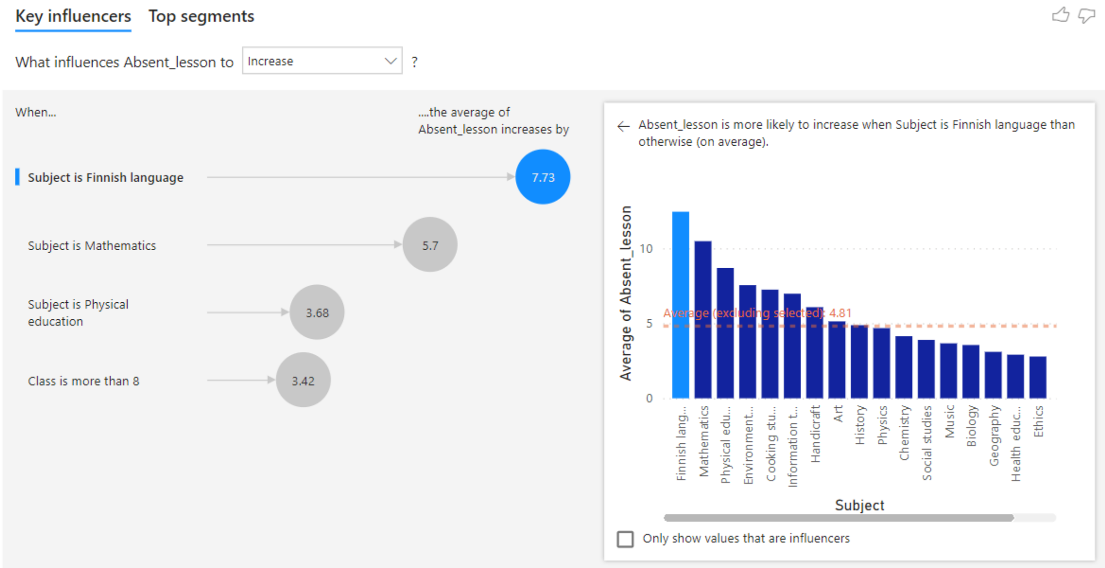


Figure 5.9 *Key factors influencing absenteeism to increase*

stands at 8.07. Subsequent years display a slight decrement, with recorded averages of 8.00 in the academic year 2021-2022 and 7.98 in the academic year 2022-2023.

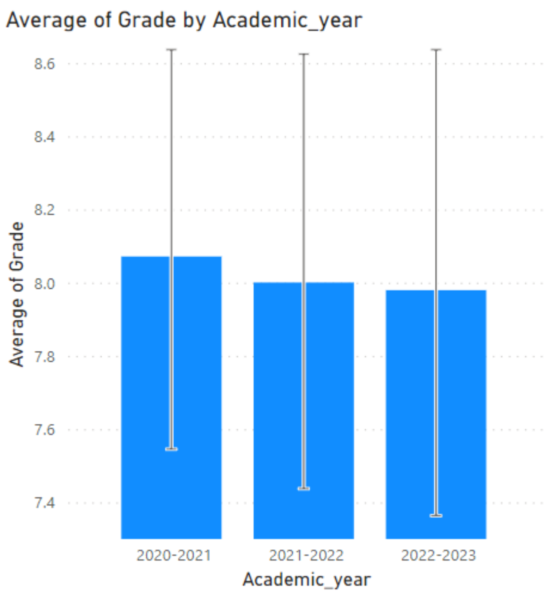


Figure 5.10 *Average grade per student for a subject by academic year*

In Figure 5.11, the average academic grade of students for various subjects during one academic year is presented. The data reveals that students perform well in Physical Education, English, and Art, obtaining the top average grades. On the other hand, subjects such as History, Biology, and Physics record lower average grades in comparison.

A pattern is observed when evaluating the performance of non-native Finnish

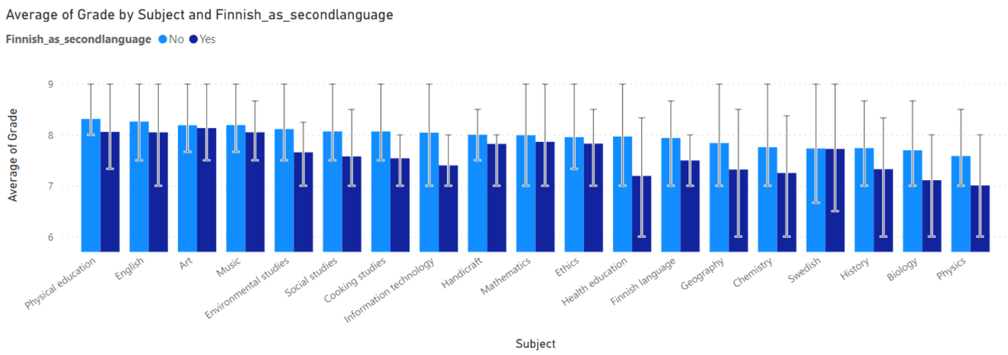


Figure 5.11 *Annual average grade by subject and Finnish language group*

speakers: their average grades tend to be slightly lower than for those of native Finnish students. The most grade differences between these groups is observed in Health Education, Biology, and Physics, with respective disparities of 0.78, 0.59, and 0.57. Furthermore, the subjects of Geography, Swedish, and Chemistry display the most considerable variances in grades, as indicated by the extensive range between the 25th and 75th percentiles.

Figure 5.12 provides an analysis of the yearly average grades across various subjects by gender. Women predominantly outperform men in the majority of subjects, with exceptions such as Physical Education. The subjects of Art, Health Education, and Cooking Studies show differences of 0.97, 0.69, and 0.65, respectively.

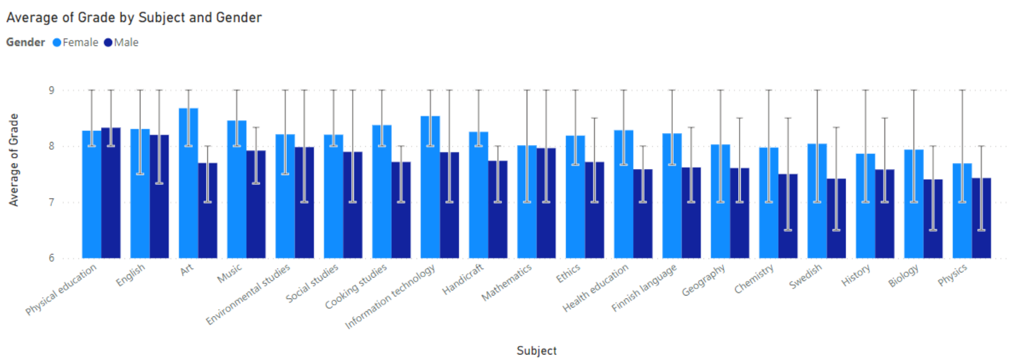


Figure 5.12 *Annual average grade by subject and gender*

Additionally, subjects such as Environmental Studies, Social Studies, Mathematics, and Chemistry display higher variation in student grades.

In Figure 5.13, the chart presents the yearly distribution of average grades among different schools. The data reveals that Schools E, R, and A are the top three schools with the highest average grades. Meanwhile, schools G and C show the most significant variation in grades among all support groups. It is generally observed that the special support group had a lower average grade. However, the special support group in schools P, H, B, and I achieve higher grades than the intensive support group. Particularly, schools J, K, and G have the lowest grades in the special support group, with scores of 6.54, 6.84, and 6.91, respectively.

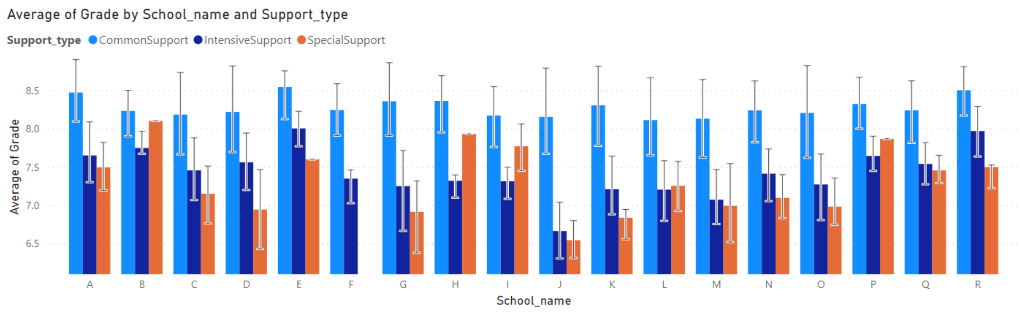


Figure 5.13 *Annual average grade by school and support type*

In Figure 5.14, the chart displays the yearly average grades by class and Finnish language groups. The results suggest that native Finnish speakers consistently achieve higher average grades throughout all classes. Interestingly, the greatest average difference between native and non-native Finnish speakers is observed in class 7. As students advance in their studies, this disparity gradually diminishes, reaching its smallest margin in class 9.

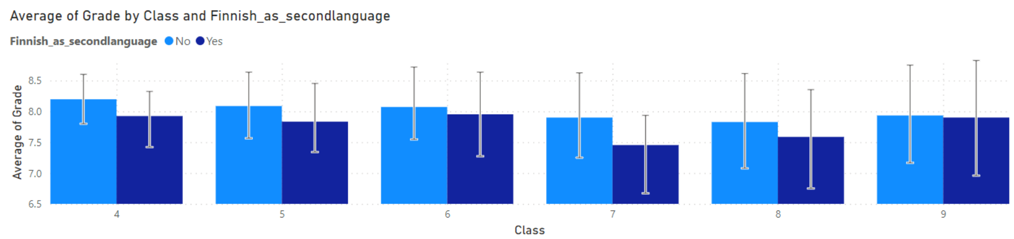


Figure 5.14 *Annual average grade by class and Finnish language group*

The Power BI key influencer displays several factors contributing to the decrease of average grade in Figure 5.15. The analysis highlights four predominant factors. Firstly, there is a pronounced decrease in average grade when students are categorized under the special support type, leading to an average decline of 1.06 in the grade. Secondly, the intensive support type is associated with a decrease of 0.8 in the average grade. Thirdly, male students, on average, exhibit a decrease of 0.41 in their grades compared to female students. Lastly, for classes higher than 4, decrement was observed by 0.22.

5.4 Heatmap results

Some values in the heatmap are missing because the p-values below 0.05 were filtered out to eliminate points with significantly low number of data points or lacking statistical relevance.

Figure 5.15 *Key factors influencing average grade to decrease*

In Figure 5.16, a heatmap illustrates the relationships between different features in the dataset. To assess correlations, features such as Support Type, Finnish as a Second Language, and Gender have been converted to numeric values. Table 5.1 describes the mapping values. The heatmap distinctly highlights a strong correlation between Gender and grade, inferring that females often secure higher grades. There is also a discernible negative correlation between Support Type and grade, suggesting that students receiving special support tend to score lower. Similarly, Total Absent Time and Finnish as a Second Language negatively correlate with grade, indicating that students with fewer absences and native Finnish speakers often outperform in grading.

Support type		**Finnish as a second language**		**Gender**	
CommonSupport	1	No	0	Male	0
IntensiveSupport	2	Yes	1	Female	1
SpecialSupport	3				

Table 5.1 *Numeric Encoding of Categorical Variables*

The heatmap, illustrated in Figure 5.17, displays the correlations among subjects for class 4. In class 4, it is noted that most students take nine subjects. Particularly, the heatmap results show that the subject of Finnish Language demonstrates a high correlation with other subjects, indicated by a redder row in the heatmap. On the contrary, Physical Education shows the lowest correlations with most other subjects, exhibiting a lesser association with other subjects.

Further analysis reveals that the top three correlated subject pairs are as follows: Finnish Language and Environmental Studies with a correlation coefficient of 0.62, Finnish Language and Mathematics at 0.61, and Finnish Language and Social

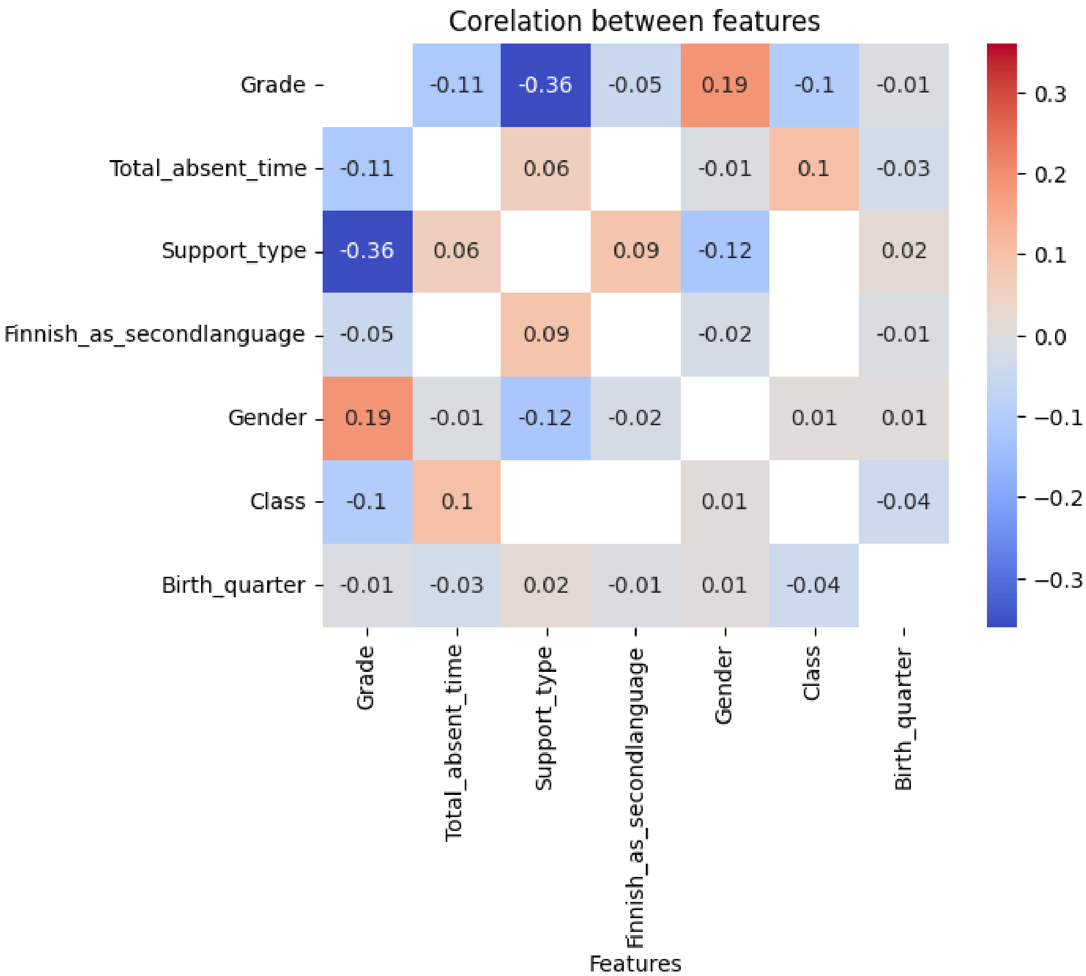


Figure 5.16 *Heatmap of correlations between features*

Studies also at 0.61.

The heatmap, illustrated in Figure 5.18, displays the correlations among subjects for class 5. In class 5, it is found that most students take 10 subjects. Particularly, the heatmap results show that the subjects of Finnish Language and Environmental Studies demonstrate a high correlation with other subjects. On the contrary, art subjects show the lowest correlations with most other subjects.

Further analysis reveals that the top three correlated subject pairs are as follows: History and Environmental Studies with a coefficient of 0.73, Finnish Language and Environmental Studies with a correlation coefficient of 0.72, and History and Ethics also at 0.7.

The heatmap, illustrated in Figure 5.19, displays the correlations among subjects for class 6. In class 6, it is found that most students take 9 subjects. Particularly, the heatmap results show that the subject of History demonstrates a high correlation with other subjects. On the contrary, art subjects show the lowest correlations with most other subjects.

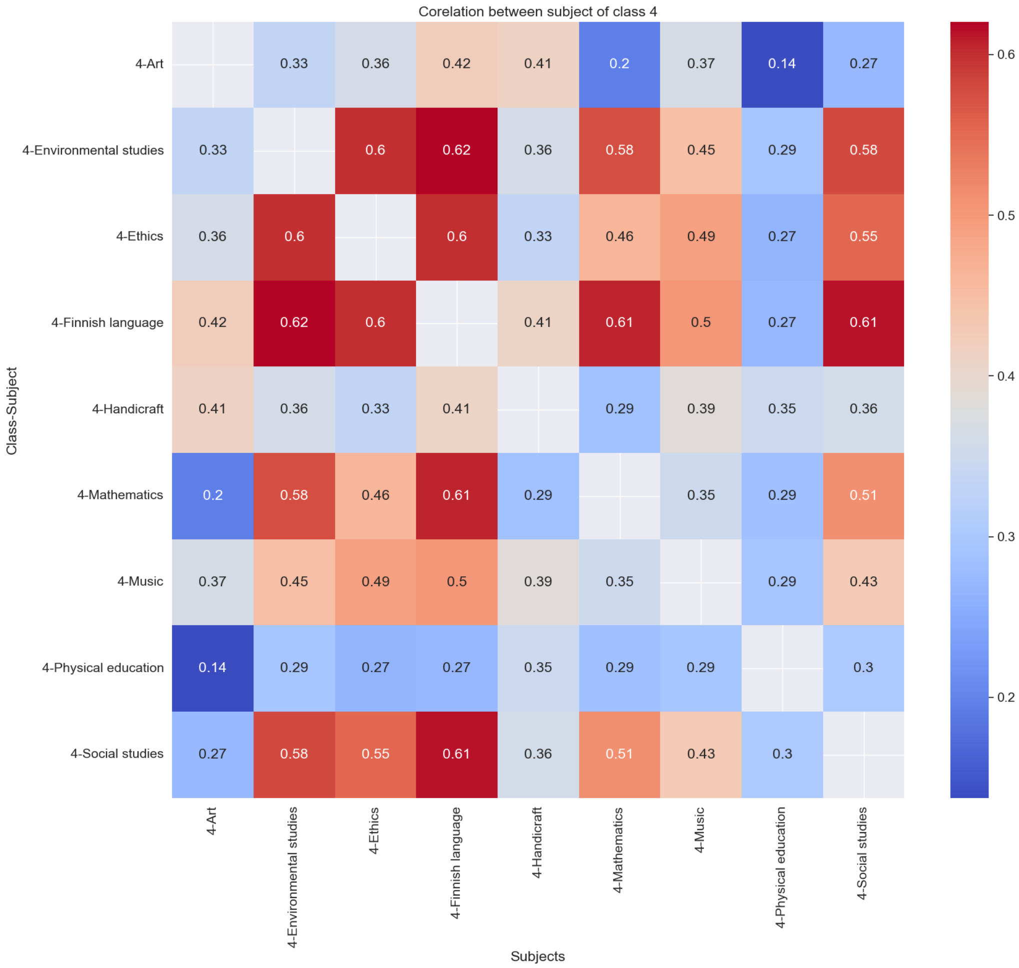


Figure 5.17 *Heatmap of subjects in class 4*

Further analysis reveals that the top three correlated subject pairs are as follows: History and Environmental Studies with a coefficient of 0.79, History and Ethics with a correlation coefficient of 0.78, and History and Finnish Language also at 0.73.

The heatmap, illustrated in Figure 5.20, displays the correlations among subjects for class 7. In class 7, it is found that most students take up to 14 subjects. Particularly, the heatmap results show that the subjects of History and Physics demonstrate a high correlation with other subjects. On the contrary, art subjects show the lowest correlations with most other subjects.

Further analysis reveals that the top five correlated subject pairs are as follows: History and Ethics with a correlation coefficient of 0.78, Physics and Chemistry at 0.77, Mathematics and Physics also at 0.79, Geography and Biology at 0.76, Mathematics and Physics at 0.75, and History and Biology at 0.74.

On the other hand, the subject pairs with the least correlation are identified as Art and Physical Education, and Handcraft, with a correlation coefficient of 0.29. The next pair is Physical Education and Chemistry, with a correlation coefficient of

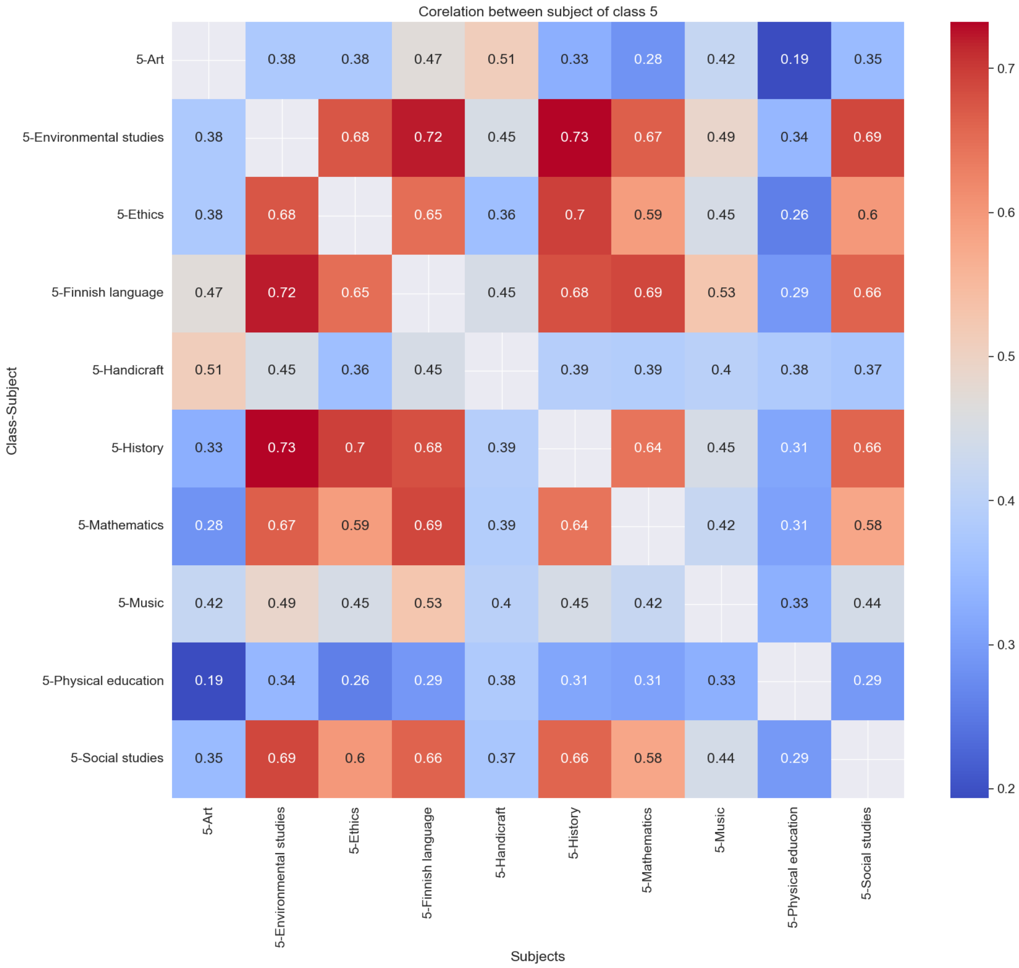


Figure 5.18 *Heatmap of subjects in class 5*

0.37. Finally, Cooking Studies shows a correlation coefficient of 0.38 with Physical Education.

The heatmap, illustrated in Figure 5.21, displays the correlations among subjects for class 8. In class 8, it is discerned that most students take 15 subjects. Particularly, the heatmap results show that the subjects of Chemistry and Biology demonstrate a high correlation with other subjects. On the contrary, Physical Education shows the lowest correlations with most other subjects.

Further analysis reveals that the top three correlated subject pairs are as follows: Finnish language and Environmental studies with a correlation coefficient of 0.62, Finnish language and Mathematics at 0.61, and Finnish language and Social studies also at 0.61.

On the other hand, the subject pairs with the least correlation have been identified. Information Technology and Physical Education have a correlation coefficient of 0.25. Information Technology and Handcraft have a coefficient of 0.33. Lastly, Information Technology and Cooking Studies are correlated at 0.36.

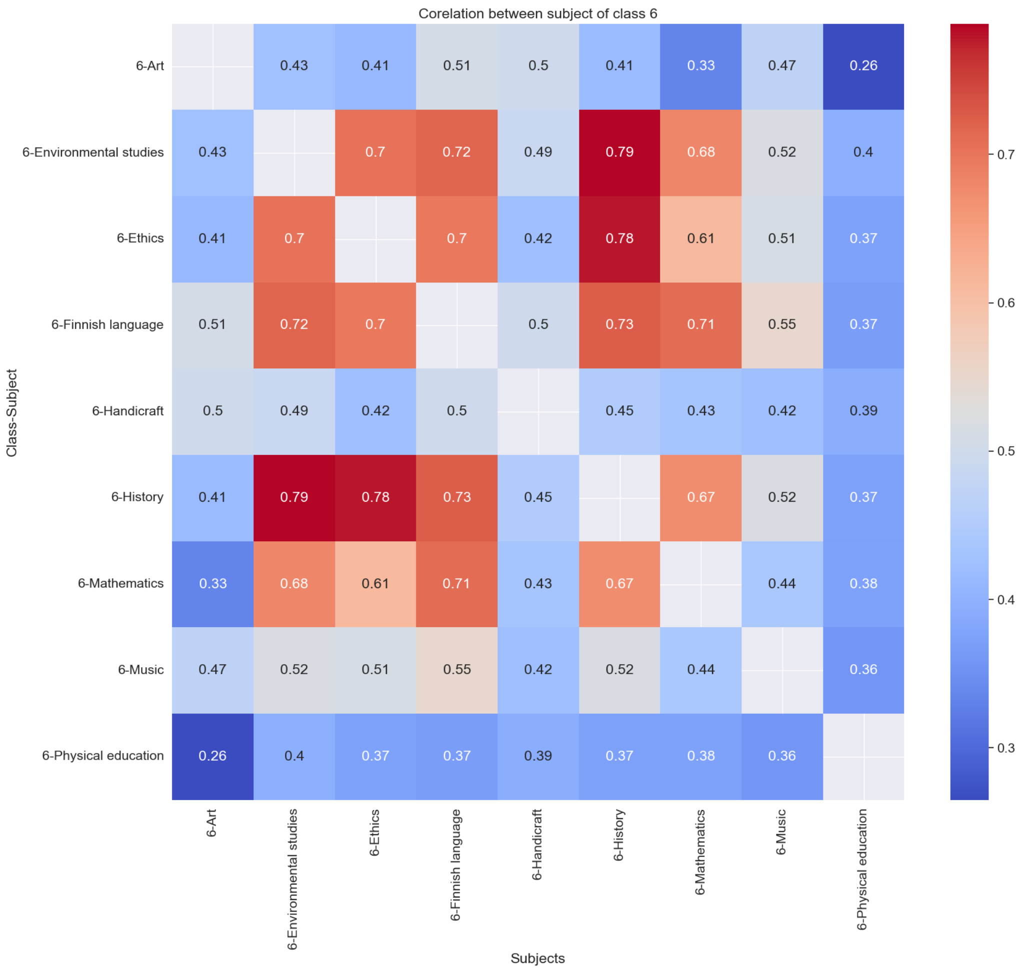


Figure 5.19 *Heatmap of subjects in class 6*

The heatmap, illustrated in Figure 5.22, displays the correlations among subjects for class 9. In class 9, it is discovered that most students take up to 14 subjects. Particularly, the heatmap result shows that the subject of social studies demonstrates a high correlation with other subjects. On the contrary, Physical Education and Music subjects show the lowest correlations with most other subjects.

Further analysis reveals that the top five correlated subject pairs are as follows: Biology and Geography with a correlation coefficient of 0.86, Health Education and Biology at 0.79, Mathematics and Physics also at 0.79, and Physics and Biology at 0.79.

On the other hand, the subject pairs with the least correlation are identified as Physical Education and Art, with a correlation coefficient of 0.33. Physical Education and Cooking Studies, as well as Music and Finnish Language, both show a correlation value of 0.33. Meanwhile, Music and Geography have a correlation of 0.34, and Music and Ethics, 0.36.

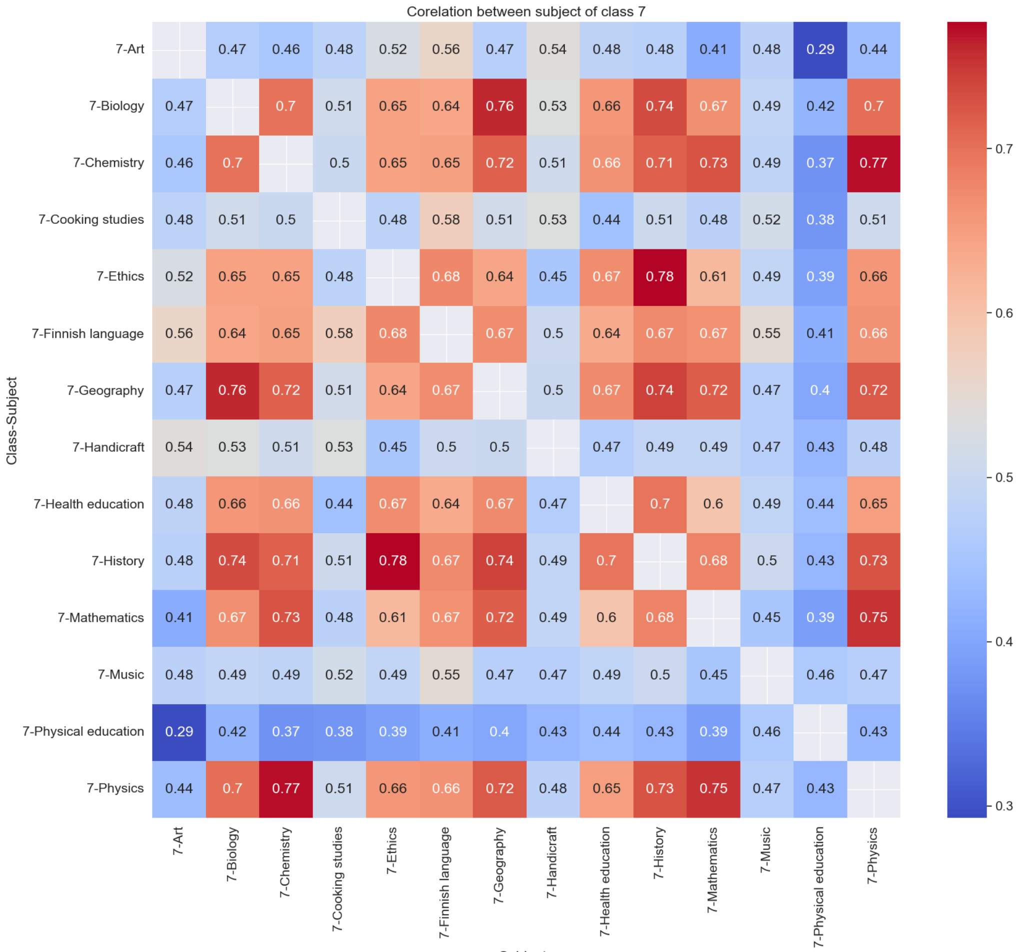


Figure 5.20 *Heatmap of subjects in class 7*

5.5 Academic performance prediction results

In this section, the results of academic performance prediction using the LSTM and Markov state models are presented.

5.5.1 LSTM results

To predict students' academic performance, the LSTM model was adopted. Figure 5.23 showcases the Long Short-Term Memory (LSTM) diagrams. Each LSTM layer is labeled with numbers in order; however, there is no difference between the numbers. Designing an LSTM model using the Keras framework involves various decisions. Including selecting an appropriate activation function, determining the number of epochs, choosing an optimizer, and specifying the batch size. With only data spanning three academic years available, the selection for predictive models was limited. Given the prominent effectiveness of LSTMs with smaller datasets, this method was the most suitable option. Through this methodical approach, the

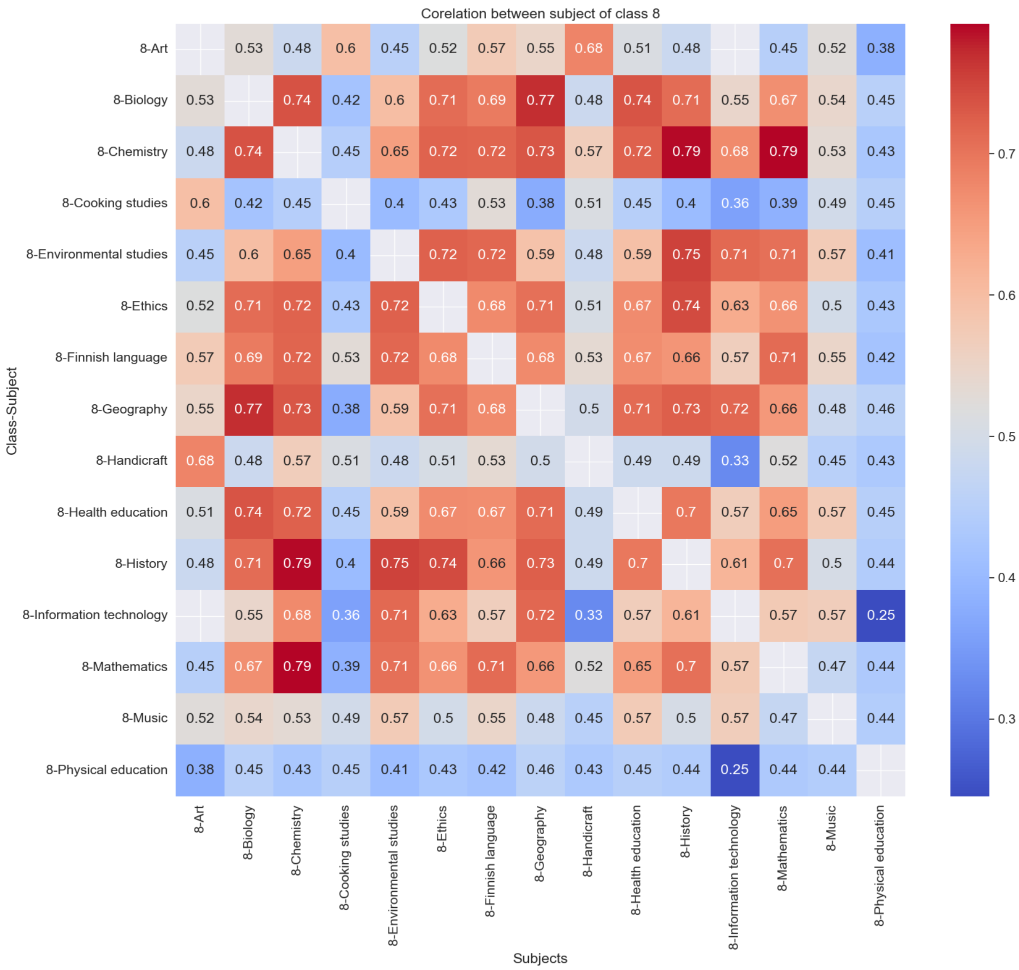


Figure 5.21 Heatmap of subjects in class 8

model aims to enhance the probability of accurately predicting students' academic performance within a three-year timeframe.

The variables selected for this analysis included 'Student Hash,' 'Class,' 'Support Type,' 'Finnish as a Second Language,' 'Gender,' 'Birth Quarter,' 'Academic Year,' 'Total Absent Time,' 'Grade,' and 'Subject Category.' These variables were considered relevant for speculating the academic performance of students. To ensure data consistency and comparability, the Min-Max scaler from TensorFlow was used to scale all data to a range of 0 to 1. This scaling process was especially significant for the Long Short-Term Memory (LSTM) layers, as it markedly enhanced the performance of the analysis.

The data was organized into two separate layers based on the academic year in chronological order. Every student included in this LSTM analysis had records for three academic years: 2020-2021, 2021-2022, and 2022-2023. The LSTM model architecture is described in 5.23. The academic grades from the year 2020-2021 were assigned to the first LSTM layer, which had 9 features across 3 different rows,

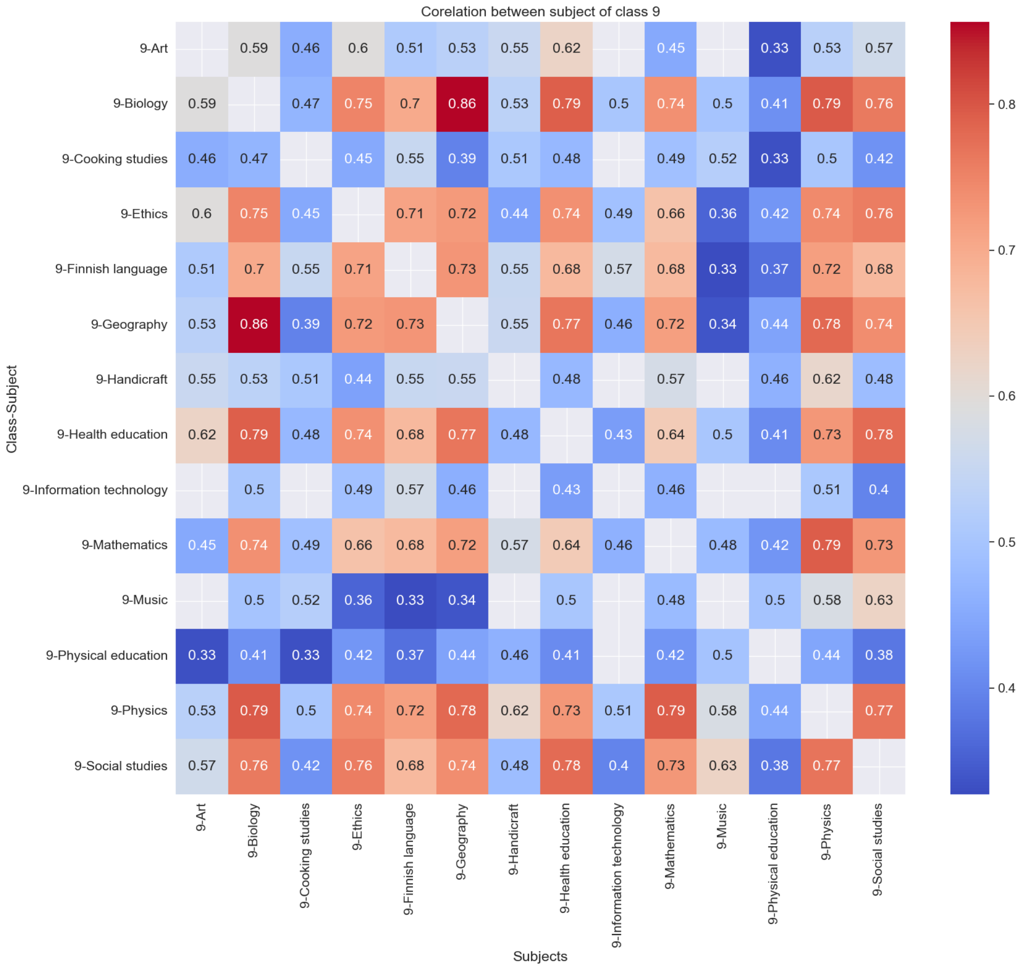


***Figure 5.22** Heatmap of subjects in class 9*

representing the 3 different subject categories. Similarly, the academic grades from the year 2021-2022 were assigned to another LSTM layer on the same level, also containing 9 features across 3 different rows.

The results from these layers were then concatenated; this means the LSTM outputs from each layer were combined into one layer. This combined data was then passed to another LSTM layer to further process the information. Subsequently, a dense layer was used for the final output, utilizing a sigmoid activation function since all the data was scaled within the 0 to 1 range.

Throughout the model's tuning phase, several optimizers 3.2 such as SGD, Adam, and Adagrad are evaluated. For the data at hand, the Adam optimizer outperformed the other models. In the context of activation functions, the sigmoid function is selected due to its output range of 0 to 1, compatible with the dataset's scaled values.To design a model to avoid both underfitting and overfitting, the selection of epochs and batch size is crucial. Upon iterative testing, a configuration of 30 epochs paired with a batch size of 40 proved to be optimal for this dataset. Following the

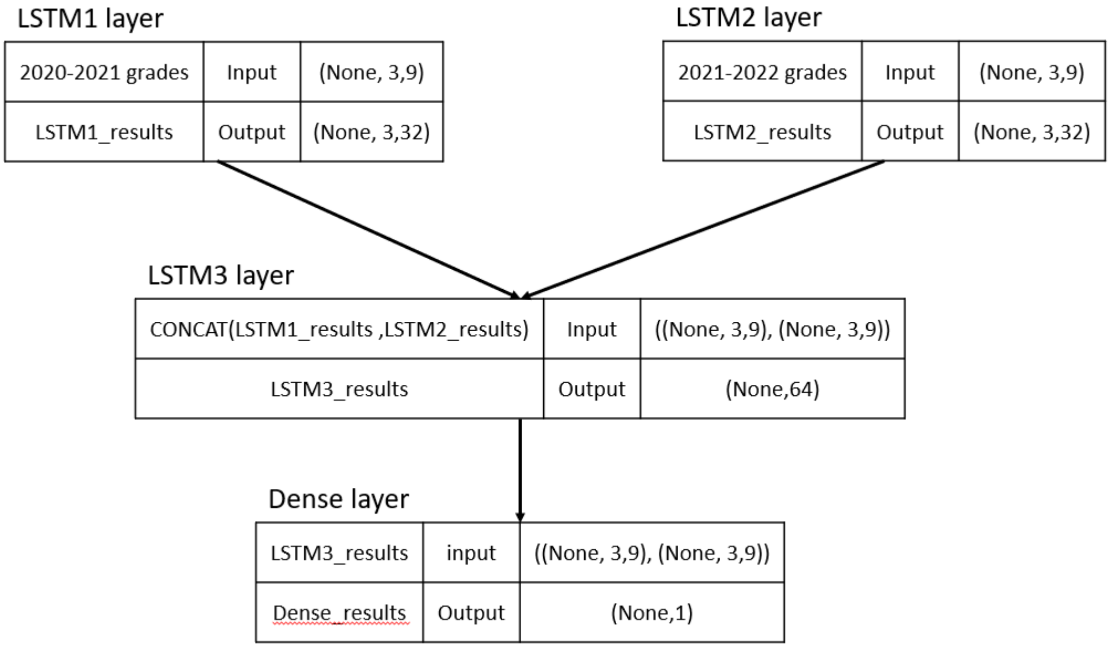


Figure 5.23 *Architecture of LSTM model*

training phase, the model give a mean squared error (MSE) of 0.005 for the training data and 0.0013 for the testing data.

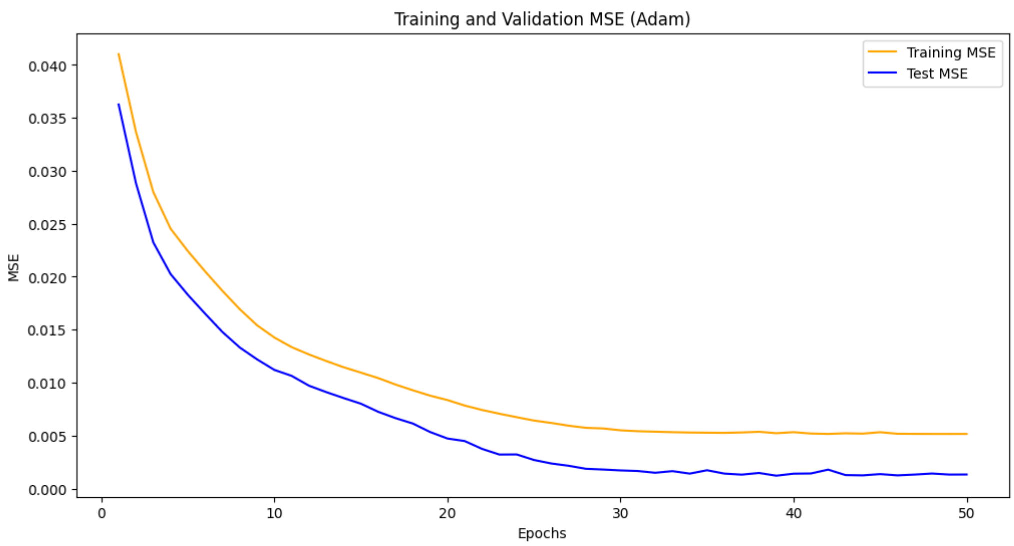


Figure 5.24 *LSTM model result using Adam*

The Adam optimizer is renowned for its effectiveness. Adam demonstrated the best results in this study, achieving an MSE of 0.0013 after 50 epochs. This performance is depicted in Figure 5.24. The Adam optimizer utilized a learning rate of 0.0001, distinct from other optimizers considered in the research, to minimize the fluctuation of the MSE plot. A significant reduction in MSE was observed during the first 30 epochs, with an MSE of 0.0016, as clearly illustrated in Figure 5.24, highlighting Adam's capability for efficient error minimization.

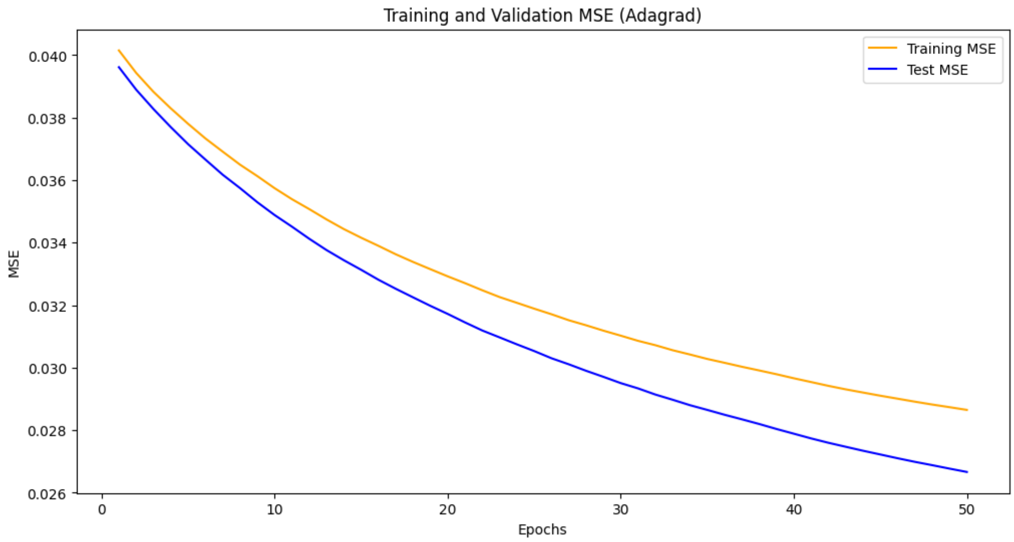


Figure 5.25 *LSTM model result using Adagrad*

Figure 5.25 provides the MSE trend across 50 epochs of the Adagrad optimizer's performance. The Adagrad optimizer was configured with a learning rate of 0.001. After 50 epochs, the MSE was recorded at 0.026. The MSE trend resembled a rational function. However, identifying the most significant epoch in this progression proved challenging, as indicated in the chart.

Figure 5.26 provides the MSE trend across 50 epochs of the SGD optimizer's performance with momentum. The SGD with a momentum of 0.7 implies that 70% of the previous update vector will be added to the current update vector, which could significantly enhance the speed of the optimizer in the relevant direction. After 30 epochs, the error became 0.0209; after 50 epochs, the error became 0.0195. This optimizer significantly lowers the MSE until epoch 10, after which it starts to slow down in reducing the error rate.

5.5.2 Markov results

The structure of the Markov table model is illustrated in Figure 5.27. This simplicity leads to a more straightforward analysis of data, especially in evaluating academic performance. Among the chosen group of students, it was necessary to sort the academic grades into three clear categories: 'low', 'medium', and 'high' to create

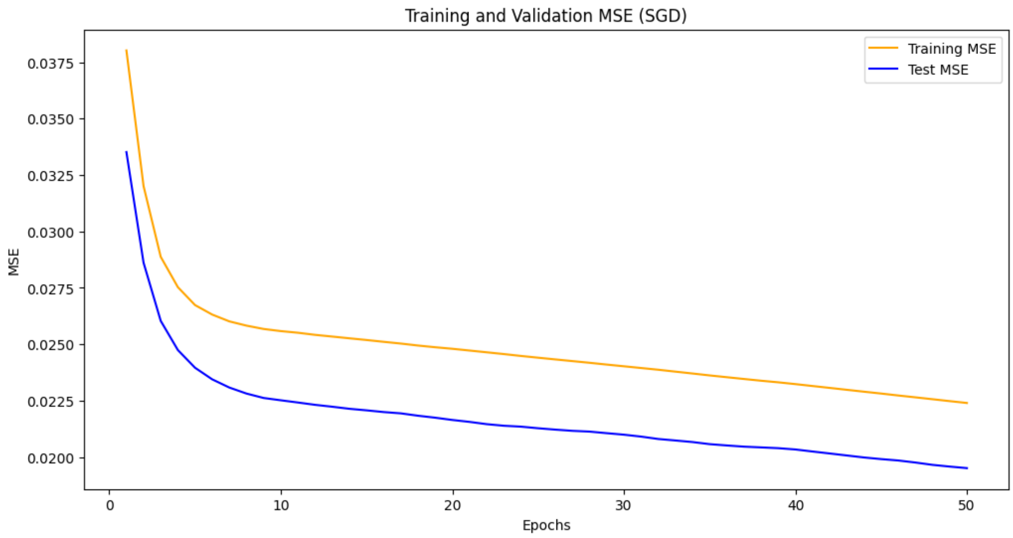


Figure 5.26 *LSTM model result using SGD with momentum*

the Markov state table. The grading scale was set as follows: a grade between 4 to 5 is labeled as 'low,' a grade between 6 to 7 is considered 'medium,' and a grade between 8 to 10 is regarded as 'high.'

Following this categorical arrangement, a 3x3 matrix was created, where each column represents the mapped academic grade. In this matrix, the rows represent the previous year's academic grade, while the columns represent the current year's grade. This design captures the shift in academic performance over the academic period.

The next procedure involved counting the numbers of students whose academic grades either transitioned from one category to another or remained within the same category from the previous year to the current year. This counting process was conducted for each categorized group: Social Sciences, Arts and Physical Education, and the Science field, hence providing a separate analysis based on the academic categories.

After obtaining the data for each grade and subject, a normalization process was conducted on each of the matrices by each row. Subsequently, the matrices for each academic domain were subjected to an element-wise multiplication process. This step was important as it helped in understanding the shifts in academic performance across different subject categories over classes.

The results from the Markov State Model are displayed in Figure 5.28. Three tables are presented, each pointing to a specific subject category. The primary objective is to ascertain the probability of students, who are in class 4, sustaining their grade category (high, medium, or low) as they advance to class 9. The probabilities in the tables are expressed up to four decimal places for precision.

In the Art and Physical Education category, 90% of students who secured high

Next year / Current year	High grade	Medium grade	Low grade
High grade			
Medium grade			
Low grade			

Figure 5.27 *Structure of Markov table*

Arts and Physical Education	High grade	Medium grade	Low grade
High grade	0.9069	0	0
Medium grade	0.1446	0.1072	0
Low grade	0	0	0

Science domain	High grade	Medium grade	Low grade
High grade	0.7078	0.0012	0
Medium grade	0.0125	0.4286	0
Low grade	0.1481	0.0740	0

Social Sciences	High grade	Medium grade	Low grade
High grade	0.6984	0.0013	0
Medium grade	0.0277	0.3146	0
Low grade	0.1250	0.1250	0

Figure 5.28 *Markov state model result*

grades in class 4 are anticipated to uphold their high-grade status until class 9. Similarly, 10% of students in the medium grade are expected to remain within this category through class 9.

However, in the Science domain, a more diverse pattern is discovered. 70% of students who achieved high grades in class 4 are expected to persist with their high academic standards in class 9. Concurrently, 42% of students from the medium grade students are anticipated to stay in the medium grade status. The findings further reveal that 14% of students initially graded as low have the potential to elevate to high grade status during class 4 to 9. Furthermore, 7% of these students might progress from a low to a medium grade during class 4 to 9.

In the Social Science domain, 69% of students who attained high grades in class 4 are forecasted to persist with their high-grade performance through to class 9. Concurrently, 31% of those initially in the medium grade range are predicted to stay consistent in this category. Moreover, 12% of students who are initially in the low-grade tier have the potential to proceed to high grade status between class 4 to 8, while an equivalent 12% might advance to a medium grade.

6 Conclusion

By leveraging advanced statistical techniques, machine learning, and deep learning methodologies, the study objective was not only to describe but also to predict student outcomes. A thorough examination of the student, grade, absentee, and special support data revealed the factors influencing academic success. Features such as attendance trends, grade distributions, and differences between support groups provide key insights that address the research questions.

RQ1: Influence of Native Language on Absenteeism and Academic Performance

In our assessment of absenteeism patterns based on native language, it was observed that students who are non-native Finnish speakers generally reported fewer absences than the native Finnish-speaking group. Interestingly, there were certain subjects, such as cooking studies, biology, and ethics, where this group noted a higher average of missed lessons. When analyzing academic grades, it became evident that non-native Finnish-speaking students typically attained lower grades than the native Finnish-speaking group. The difference in academic grade was most noticeable in Class 7. By Class 9, however, this difference narrowed considerably, indicating an improvement in performance or adaptation by non-native speakers over time.

RQ2: Correlation Between Absenteeism and Academic Performance

An inverse relationship was identified between the average number of absences and academic performance. Essentially, as the academic grades increased, the average number of absent lessons decreased. This suggests that regular attendance may positively influence a student's academic performance or vice versa.

RQ3: Yearly Trends in Absenteeism

Evaluating the absenteeism trends over the past three academic years, the 2020-2021 academic year displayed the lowest average of missed lessons. The subsequent years, 2021-2022 and 2022-2023, showed almost comparable averages in absenteeism, with no significant deviation between them.

RQ4: School-wise Analysis of Performance and Absenteeism

On examining different schools, Schools E, R, and A stood out with the highest average grades. In contrast, Schools M, L, and C reported the lowest. When looking at absenteeism patterns, Schools M, G, and N had the highest average missed lessons, whereas Schools K, O, and F showcased the least absences.

RQ5: The Role of Absenteeism in Identifying Support Needs

While the number of absences can offer some insights into a student's potential need for special support, it is crucial to consider other correlated factors, such as the specific subject or class in question. Interestingly, the results suggested that grades might be a more indicative factor for determining the type of support a student may require, compared to average absent lessons.

Beyond the analysis results, machine learning and deep learning models, such as LSTM and Markov state models, have shown significant potential in predicting academic performance. The LSTM model's prediction mean square error was 0.0016 using an Adam optimizer.

In conclusion, the analysis offers comprehensive insights into the patterns of absenteeism and academic performance among students, highlighting the significance of various influencing factors. It is essential for educators to understand these patterns and provide better support accordingly.

Future Research

While the current study provided valuable insights into factors influencing academic success and the potential of machine learning and deep learning models in predicting academic performance, there are several paths for future research to explore.

Future studies could benefit from a broader and more diverse dataset. By spanning a larger time frame and including data from multiple regions, various academic periods for each student, and different educational systems, a more holistic understanding of academic success could be attained. If the dataset tracks the same students over several years, this longitudinal perspective could offer invaluable insights into various crucial factors. For example, the long-term effects of consistently needing support can be assessed, or how grades change and evolve over time can be observed. Moreover, with such an enriched dataset, prediction models would likely provide more precise and reliable results.

Exploring other frontline machine learning and deep learning models might yield even more accurate predictions or identify subtle patterns missed by the current methodologies. This includes newer architectures and hybrid models that combine the strengths of multiple approaches.

In conclusion, while the present research sheds light on the complicated factors impacting academic success and the potential of predictive models, continued interdisciplinary collaboration between educators and data scientists is crucial to harness the full potential of these insights for the betterment of education systems.

References

[1] Ethem Alpaydin. *Introduction to machine learning*. eng. Third edition. Adaptive computation and machine learning series. Cambridge, Massachusetts: MIT Press, 2014, pp. 1–2. ISBN: 0-262-32575-6 (cit. on p. 8).

[2] Peter Bruce and Andrew Bruce. *Practical statistics for data scientists : 50 essential concepts*. eng. 2nd edition. Beijing, [China: O'Reilly, 2017 - 2017, pp. 106–110. ISBN: 978-1-492-07294-2 (cit. on p. 15).

[3] Peter Bruce and Andrew Bruce. *Practical statistics for data scientists : 50 essential concepts*. eng. 2nd edition. Beijing, [China: O'Reilly, 2017 - 2017, pp. 30–32. ISBN: 978-1-492-07294-2 (cit. on p. 16).

[4] Yang Cao, Xuehan Jiang, and Wei Han. "Self-Assembly Pathways of -Sheet-Rich Amyloid- (1–40) Dimers: Markov State Model Analysis on Millisecond Hybrid-Resolution Simulations". eng. In: *Journal of chemical theory and computation* 13.11 (2017), pp. 5731–5744. ISSN: 1549-9618 (cit. on p. 15).

[5] Winston Chang. *R graphics cookbook : practical recipes for visualizing data*. eng. Second edition. Beijing: O'Reilly, 2019. Chap. 13. ISBN: 1-4919-7859-7 (cit. on p. 16).

[6] Microsoft Corporation. *Introduction to Power BI*. Accessed: 2023-10-16. 2023. URL: `https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-influencers?tabs=powerbi-desktop` (cit. on pp. 17, 18).

[7] Minnistry of education and culture. *Finnish Education system*. Accessed: 2023-10-17. 2023. URL: `https://okm.fi/en/education-system` (cit. on p. 4).

[8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. eng. Adaptive computation and machine learning. London, England: The MIT Press, 2016, pp. 150–152, 294. ISBN: 9780262035613 (cit. on p. 9).

[9] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. eng. Adaptive computation and machine learning series. Cambridge, MA: MIT Press, 2017 - 2016. ISBN: 978-0-262-03561-3 (cit. on pp. 9, 10).

[10] Ke-Li Han, Xin Zhang, and Ming-jun Yang. "Application of Markov State Models to Simulate Long Timescale Dynamics of Biological Macromolecules". eng. In: *Advances in Experimental Medicine and Biology*. Vol. 805. Advances in Experimental Medicine and Biology. Switzerland: Springer International Publishing AG, 2014, pp. 29–66. ISBN: 9783319029696 (cit. on pp. 15, 16).

[11] Meimei Han et al. “Application of Ensemble Algorithm in Students’ Performance Prediction”. eng. In: *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*. IEEE, 2017, pp. 735–740. ISBN: 1538606216 (cit. on pp. 6, 7).

[12] Ministry of Higher Education et al. eng. Ministry of Higher Education and Science, 2022. Chap. 1, pp. 6–7. ISBN: 978-87-93706-81-1 (cit. on p. 6).

[13] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-Term Memory”. eng. In: *Neural computation* 9.8 (1997), pp. 1735–1780. ISSN: 0899-7667 (cit. on p. 13).

[14] Swedish Institute. *The Swedish school system*. Accessed: 2023-10-17. 2023. URL: `https://sweden.se/life/society/the-swedish-school-system` (cit. on p. 5).

[15] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014) (cit. on p. 10).

[16] David Lane et al. *Introduction to statistics*. Citeseer, 2003, p. 177 (cit. on p. 16).

[17] Alexander Loth. *Teach Yourself Visually Power Bi*. eng. Teach Yourself VISUALLY (Tech) Series. Hoboken, New Jersey: John Wiley & Sons, Inc., 2023. Chap. 1. ISBN: 9781119903796 (cit. on p. 17).

[18] John Paul Mueller and Luca Massaron. *Deep Learning for Dummies*. eng. 1st ed. Newark: John Wiley & Sons, Incorporated, 2019. Chap. 11. ISBN: 9781119543046 (cit. on pp. 13, 14).

[19] John Paul. Mueller and Luca Massaron. *Deep learning*. eng. 1st edition. For dummies. Hoboken, New Jersey: For Dummies, 2019. Chap. 11. ISBN: 1-119-54303-7 (cit. on pp. 12, 13).

[20] Andreas C Müller and Sarah Guido. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. eng. 1st ed. Sebastopol: O’Reilly Media, Incorporated, 2016, pp. 2–4, 17, 145. ISBN: 1449369413 (cit. on pp. 8, 9).

[21] Mrinal Pandey and Vivek Kumar Sharma. “A Decision Tree Algorithm Pertaining to the Student Performance Analysis and Prediction”. eng. In: *International journal of computer applications* 61.13 (2013), pp. 1–5. ISSN: 0975-8887 (cit. on p. 6).

[22] Norwegian Agency for Quality Assurance in Education. *General information about education in Norway*. Accessed: 2023-10-17. 2023. URL: `https://www.nokut.no/en/norwegian-education/general-information-about-education-in-norway/` (cit. on pp. 5, 6).

[23] Cristóbal Romero and Sebastián Ventura. "Educational data mining: a review of the state of the art". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews)* 40.6 (2010), pp. 601–618 (cit. on p. 6).

[24] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. "Learning representations by back-propagating errors". eng. In: *Nature (London)* 323.6088 (1986), pp. 533–536. ISSN: 0028-0836 (cit. on p. 12).

[25] Sandro Skansi. *Introduction to Deep Learning: From Logical Calculus to Artificial Intelligence*. 1st. Springer Publishing Company, Incorporated, 2018, pp. 79–81. ISBN: 3319730037 (cit. on pp. 10, 11).

[26] Sandro Skansi. *Introduction to Deep Learning: From Logical Calculus to Artificial Intelligence*. 1st. Springer Publishing Company, Incorporated, 2018, pp. 89–90. ISBN: 3319730037 (cit. on pp. 11, 12).

[27] Martin Thrupp et al. *Finland's Famous Education System: Unvarnished Insights into Finnish Schooling*. eng. 1st Edition 2023. Singapore: Springer Nature, 2023, pp. 1–2. ISBN: 9811982414 (cit. on p. 2).

[28] Martin Thrupp et al. *Finland's Famous Education System: Unvarnished Insights into Finnish Schooling*. eng. 1st Edition 2023. Singapore: Springer Nature, 2023, pp. 4–5, 198–199. ISBN: 9811982414 (cit. on pp. 4, 5).

[29] He Wang et al. "Markov State Models Underlying the N-Terminal Premodel of TOPK/PBK". eng. In: *The journal of physical chemistry. B* 126.50 (2022), pp. 10662–10671. ISSN: 1520-6106 (cit. on p. 15).

[30] Michael L. Waskom. "seaborn: statistical data visualization". In: *Journal of Open Source Software* 6.60 (2021), p. 3021. DOI: `10.21105/joss.03021`. URL: `https://doi.org/10.21105/joss.03021` (cit. on p. 16).

[31] R.S. Witte and J.S. Witte. *Statistics*. Wiley, 2017, pp. 76–77. ISBN: 9781119254515 (cit. on p. 28).