MGI

Mestrado em Gestão de Informação
Master Program in Information Management

Predicting Success of University Applicants Based on Subjects' Preferences as an Extra Tool for Admission Considerations

Predictive Analytics Approach

Simão Caridade Menezes

Project Work Proposal presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Knowledge Management and Business Intelligence

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

PREDICTING SUCCESS OF UNIVERSITY APPLICANTS BASED ON SUBJECTS' PREFERENCES AS AN EXTRA TOOL FOR ADMISSION CONSIDERATION – A PREDICTIVE ANALYTICS APPROACH

by

Simão Caridade Menezes

Project Work Proposal presented as partial requirement for obtaining the master's degree in Information Management, with a specialization in Knowledge Management and Business Intelligence

Advisor/Co-Advisor: Roberto Henriques

November 2021

ABSTRACT

This study uses a dataset of student performance indicators and psychological patterns associated with each individual to examine the prediction efficiency of psychological traits on academic results, more specifically grade point average (GPA). We propose building a classification machine learning model that predicts GPA performance, dividing the students into the top and bottom performers. Several features were used in the modelling, namely, student's previous performance, such as GPA, course progression (how close the student master is related to previous academic courses), and personality traits obtained by surveying 319 students and recent graduates with a quiz developed by Association Better Future based on the RIASEC model for type theory of personality.

It is widely accepted that psychological characteristics can impact student churn and performance (Costa and McCrae, 1992). Furthermore, numerous papers have found that GPA can be predicted by multiple factors, including past performance, intelligence coefficient (IQ), demographic background, previous area of studies, but, to increase the model's accuracy, psychological factors are recommended for future works (Abele and Spurk, 2009).

Whilst past performance and, to a lesser extent, IQ are currently evaluated in university admissions, psychological traits are yet to have a place in selecting the best candidates. In this study we propose that, although IQ and past performance are good indicators of student performance, the predictive power of psychological traits, when combined with these classical indicators, increases the predictability accuracy of the machine learning model.

With this in mind, we used the performance of past and current university students, measured in GPA, analysed it against the collected psychological indicators and developed multiple machine learning models to predict the student GPA based on the collected indicators. These were divided into 3 groups: psychological traits only, GPA and age only, and a combination of both.

Four types of models were used: neural networks, Support Vector Machines (SVM), decision forests and decision trees. Decision forests, for the problem at hand, consistently outperformed neural networks, SVM and decision trees both in accuracy and Area Under the Curve (AUC), the curve being the Receiver Operating Characteristic (ROC).

From the database with 176 entries, comparing the models created with the GPA and age-based dataset with the ones based on the full dataset that includes psychological variables, decision forests were the model with higher fitness to the training model, and with the higher AUC against the validation set, with values of 0.717 and 0.790, respectively. The models based on the full dataset, including psychological variables, consistently outperformed the models based solely on the classical GPA predicting metrics.

We further propose and discuss that the model can be used as an extra indicator for the admission process.

KEYWORDS

University Admission; Predictive Modeling; Student Performance Prediction; Student Performance Indicators

INDEX

TABLE OF TABLES

TABLE OF FIGURES

1. Introduction

1.1. Background

When considering and evaluating Masters, students find many universities with numerous courses with ever-growing levels of discipline granularity. A common approach taken by the students is to narrow down universities and disciplines based on rankings or recommendations from peers. Applicants often do not entirely understand the lectured topics of each course justifying some of the student churn seen today. This ad hoc approach is hardly a good indicator of fit to one's objectives or potentials.

On the flip side, it is up to the admissions office to choose the ideal candidates from a vast array of resumes and motivational letters. In this very manual process, teachers and administrative personnel do the screening. This lengthy process hinders a timely response and diminishes the capacity of schools to screen candidates.

Unlike undergraduate courses, acceptance in master courses is not as standard since the candidates applying have more diverse backgrounds, areas of study, and different levels of career experience. The background of the students is, thus, very diverse, including, for instance, a range of undergraduate courses completed. It is not uncommon to have medical students applying to business master courses. The same behaviour is not frequently found in the access process to undergraduate programmes. The high school experience is highly standardized, and applicants rely solely on their GPA for the admissions process. We can also see demographic differences in both processes. In master programmes, age varies greatly across candidates as well as work experience, especially when compared to the undergraduate population.

Regardless of this, there is still no single clear indicator of future graduation success in the Master environment. However, it is known that previous GPAs play an important essential role for admissions, as do motivation letters and career experience. Little attention is being paid to the role psychology, and human traits play in a student's motivation and performance. This paves the way to a series of potential biases and non-standardized methods of acceptance. Whilst some factors such as past performance are easily compared, others are hard to compare. Professors choosing the students based on their CVs will, inevitably, add bias into the process. How to compare a 40-year-old student with a 25-year-old student, early in his/her career?

Filtering down applications at universities consists of a case-by-case manual selection process, that is arduous and time consuming with little automation. This leads multiple universities to set their acceptance terms on a first come first served basis. Prioritizing students that apply in more timely manners rather than the best overall candidates. Even though there has been a growing adoption of intelligent algorithms in some private companies, this approach is still to cascade down to the education sector.

1.2. Problem Identification

The school's masters have a high number of applicants to a limited number of seats per course. Due to different backgrounds, demographics, and experience levels, it is difficult for the university to distinguish between a high and low-performance candidate, especially when looking at smaller variances. This metric comparison process leads to a weigh in of previous course GPA and work experience seldom unbiased from the background and preferences of the evaluator. These do not always end up corresponding to the expected performance at admission.

Student psychological traits represent a significant proportion of school performance variance in college students (Costa and McCrae, 1992). It is also known that objectivity in the entrance processes is hard to achieve with indicators like GPA and age, frequently criticized as their predictability efficiency does not explain smaller variances (Pantages and Creedon, 1975). The usage of psychological traits in conjunction with GPA and age will be discussed ahead.

Furthermore, when a student applies for multiple masters at once at the same school, a common practice, the university has no way of understanding which course will reveal itself as the best fit for the individual. Nor does the applicant. It is important to provide the admissions personnel with the necessary tools to achieve a more unbiased and valuable process while simultaneously making the same set of tools available to the students to facilitate their decision-making process when applying to higher education by identifying the weight of a previous course, say bachelors, has on a master.

1.3. Motivation

This is an interesting scientific challenge with space for new knowledge creation: In the one hand, for the understanding of the importance of psychological traits in predicting student performance in conjunction with the other classical variables. On the other hand, for the comprehension of the best methodologies for data capture and posterior machine learning algorithm selection. Both issues are still not fully explored by the scientific community. This study has the potential to benefit universities interested in improving their student admission systems by including psychological traits in their admissions process.

1.4. Project Objectives

1.4.1. Main Goal

This project focuses on a proactive approach to the university admissions sector more specifically, on the metrics used to filter down applicants. To introduce personality traits, we resourced to a questionnaire from *designed the future* allowing for data collection, analysis and integration onto this process. This questionnaire is introduced later on the data collection chapter and is available in the annex.

GPA is measured or was converted to a scale of 0 to 20. Taking this as the dependent variable, predictive analytics were used to predict the GPAs of applicants while utilizing the responses of the psychological questionnaire as independent variables. As the goal is not to calculate the actual GPA but is to predict the overall performance of the students, the variable is divided into two buckets: top performers and low performers.

The goal is to provide a complementary metric, developed as a forecaster of student performance, transversal to all applicants, that the administrative body can rely on to identify the candidates with the highest potential. This metric will be tested to include, not only GPA and age as inputs, but also personality traits. This is to test what is the improvement to classical models achieved by introducing psychological variables. Microsoft Azure Machine Learning was the tool used for the development of the various supervised machine-learning models.

Testing the model with the remainder of the school's body is vital to ensure the reliability of the model. This is not tackled in this project and is further discussed during the conclusion chapter.

To construct the model, the understanding of the data available and its quality was fundamental and represented a major challenge of this work. Throughout this process, entries, outliers, and missing values were analyzed and either corrected or discarded.

1.4.2. Specific Objectives

The scope of the project can be broken down into five specific objectives that complement each other to accomplish the main goal.

Developing the Quiz. The questionnaire was the single source of psychological data and demographic data. It was not edited throughout the project to keep data integrity between the various responses. Considerations about the questions used and some improvement recommendations are done in the concluding stages of this project. Besides psychological traits, other demographic information was also gathered to use as extra indicators for the machine learning algorithm and provide a better analysis

Gather enough data. As part of the machine learning process there was a need to partition the data into training and testing dataset, from the initial response subset. This means that the number of responses captured had a direct impact on the performance of the algorithm. Due to the sheer dimension of the Psychologic questionnaire with 69 entries, the amount of data ultimately dictates how many indicators could be used before overfitting and the curse of dimensionality kicked in. It was also important that the information gathered spanned across different courses as we are aiming to study the impact of the students' background as a predictor to each course or area of study.

Identify the main traits responsible for higher grades. Using the collected data, identify the traits that correlate strongly with higher performance. There was an expectation that these would vary from model to model, but it was anticipated that common indicators across the tests would be found.

Develop predictive analytics supervised machine learning model with GPA as the dependent variable. As stated above, this part of the study consisted of data quality assessment and posterior execution and development of the set of models in Microsoft Azure Machine Learning. Understanding the best models to use was achieved initially by researching articles and later by testing empirically on the data available for this study. In this stage, besides understanding the predictive efficiency of psychological variables, it was also important to measure the impact these variables have when complementing a traditional model focused on more classic indicators of academic success, where a lot of the academic research found in the literature focuses.

Testing on a considerable population of the school community. Being able to test the model will provide enough confidence in the outputs developed. It might also pose an opportunity to capture more data focused on testing the model. Moving forward, it will be possible to test the model on actual applicants, follow their academic performance and observe its accuracy.

1.4.3. Limitations

As with any machine learning model, the quality of the output will directly reflect the quality of the inputs. A high number of independent variables was collected which implies that either a large dataset needed to be collected or the number of features would need to be trimmed down to prevent overfitting the models and decreasing the value of the model.

There is a possibility that psychological traits and preferences might change after entering the course and that the answers collected might vary from those that the students would have answered during application, meaning that the outputs of our model would be skewed. A good way to test this is to apply the questionnaire to a wave of applicants, and then follow their school progress, and only then, building the model. While this approach would ensure the data would be more accurate, due to time constraints, the data available is the one we can survey today. After the project is applied, a database can be created, and the model trimmed to improve its accuracy and usefulness. This is discussed in later stages of this project in the future work section.

It is also important to note that the output will represent a value of course fitness and expected outcome based on preferences and traits. Regardless of the quality of the model, we must still evaluate the other aspects of applications, and this is not to be the single indicator of student performance or the single topic of admission until empirical testing is taken.

1.5. Expected Outcomes

Three machine learning model experiences will be performed. One based on the classical indicators of student performance, one with only personality traits and one with a combination of the two. Each of the experiences will have multiple machine learning models, neural networks, SVMs, decision trees and forests. From these, according to literature, it is anticipated that SVMs will outperform decision trees. There are no expectations for the neural networks performance as there was no clear performance indication in the literature.

An analysis will be made to the outputs of the models to understand the importance of the psychological variables. It is believed that the addition of these indicators will improve the performance of the classification model but there are no reasonable beliefs on whether these variables alone can accurately predict students' performance.

Lastly, it is thought that these models will allow us to infer the value of each variable as a predictor of student success.

1.6 Document Structure

This project is divided into 5 chapters. The first chapter, introduction, where the problem is identified, and its context is explained, an approach is proposed, objectives are set and the expectations at the beginning of the project are described.

In chapter 2, literature around the admissions processes to university is summarized. Student success indicators are identified and summarized. Literature was studied to understand the most important factors that predict student success, including GPA and personality traits. Lastly, there is a discussion around the machine learning models that should better suit the problem.

The methodology, in chapter 3, is used to explain the CRISP-DM methods that were followed during the empirical process.

The later, in chapter 4, comprised of an analysis and preparation of the data, followed by the creation and evaluation of the models.

During the conclusion, chapter 5, the insights acquired during this work are summarized and the goals set up in the introduction are discussed again in light of the new knowledge. In the end, future work possibilities are discussed and recommendations are given.

2. Literature Review

2.1. Traditional admission processes

In the past, colleges and universities relied on admission programs based on processes and criterions dependent on thresholds to select the future students from a pool of candidates and filter out the ones to be left out (Beecher and Fischer, 1999). The objective of this type of selection is to narrow down the size of the pool. This process should, ideally, identify top performers and admit them to the courses (Wendy, 1996). On the flipside, the process should also eliminate those who are not likely to succeed. However, this is not often the case as traditional admission processes are heavily reliant on a small number of indicators that do not account for smaller variances in student performance. Even though traditional processes dependent on Scholastic Assessment Test (SAT) and GPA scores have been thoroughly studied throughout the years, other indicators are still to be looked at (Berger and Sireci, 2002).

Traditional admission processes have been found to have limited value in predicting student's degree completion rates. They fail especially regarding fairness, efficiency, consistency, transparency, and equitable access (Epstein, 2010). GPA and SAT scores are the main factors behind the university admission decision making process in the United States of America. Whilst they have an overall good predictive value in identifying significant variances, subjective reader ratings of application letters sometimes beat them. Much of the modest value of these two indicators comes from the fact that institutions neutralize their effect by selecting from a narrow band of the score distribution; this is especially true as the predictive value is higher for students substantially above or lower than the average (Epstein, 2010).

There is evidence that GPA and SAT scores have little predictability of academic achievement, grades, and degree completion beyond the first year (Nora, Cabrera, Serra Hagedorn and Pascarella, 1996). In fact, replacing high school GPA with class ranking or a restrictive GPA with only a subset of specific academic courses is more effective in predicting college GPA (Mulvenon, Stegman, Thorn and Thomas, 1999). It has been highlighted that the use of such traditional measures has the potential to increase the bias of standardized tests against some students' demographics (Pribbenow, Phelps, Briggs and Stern, 1999).

Student success is also influenced by expectations from the academic staff and from the expectations each student has. The consistency of these expectations is a condition for student success, and low expectations turn out to be self-fulfilling prophecies (Tinto, 2012) This effect compounds the impact of a flawed admissions process as the students are seen as the most promising have access to better resources, more professor time and are also kept motivated by the staff.

2.2. Defining Student Success

Success at an institutional level is often measured in terms of degree completion rates. This contrasts with the individual success that tends to take the form of GPA measurements (Abrica, 2018). These reports, focusing on the institutional level, tend to assess the status of the entire sector of higher education of a specific institution, with the number of aggregate samples ranging from the low thousands to the hundreds of thousands, thus relying on churn indicators rather than on GPA (Mullin, 2012). Focusing on the individual, most studies measure success by GPA or by honors retention rates.

(Mould and B. DeLoach, 2017). At an individual level, GPA is thus a better indicator of student performance than churn.

Looking solely at GPA as a single indicator of student success is not necessarily a correct approach either. Students that have the highest GPA in each semester are not the same as the ones determined by the leadership of the school to be success examples (Abrica, 2018). However, a Korean study found out that high achievement, when measured as GPA, is highly influenced by some generally accepted successful characteristics. With data obtained by surveying 1111 students, notetaking capacity, using motivational regulation strategies rather than motivation itself and managerial traits with regards to their emotions, physical condition, time management and cognition, appear to be critical to achieve higher GPAs (Lee and Lee, 2012).

While there are only a few literature instances where GPA as an indicator of success is questioned, for most scholar articles, GPA and student success are used interchangeably as if GPA is the sole indicator of success, especially when looking at the individual level. Furthermore, attempts to predict student success tend to focus on GPA as a sole indicator of performance. (Slim, Heileman, Kozlick and Abdallah, 2014). It has also been shown that GPA has a positive influence in early career success, having the influence diminish with time. (Abele and Spurk, 2009).

The case for studying GPA as an indicator of academic achievement seems to be supported by literature, especially when looking at the individual. When looking at the institutional level the most agreed upon success indicator is college dropout rates. Thus, for the purpose of this project it is important to understand what are the factors that predict academic achievement, or, in other words, what predicts a high GPA.

2.3. Factors that predict student GPA

Most research focused on predicting student performance, having GPA as the indicator, focus on IQ levels of students and high school GPA or SAT scores (Abrica, 2018), (Power et al, 1987]), (Pascoe et al, 1997), (Jensen, 1998), (Frey & Detterman, 2004).

Findings range from attributing most of the predictability power to one of these factors to completely disproving their usefulness in specific scenarios. In general, IQ and previous GPA scores are good indicators of future performance but are insufficient in predicting smaller variances.

Research has repeatedly shown support for the predictability of one's academic success based on previous academic performance. In one study (Power et al, 1987), the correlation between secondary school grades and GPA at university was found to be at 50%. It also shown that the predictive capacity varied greatly from student to student as the secondary school grades predictive capacity diminishes on mature aged students. In short, the bigger the time span since the previous course completion the least the predictive capacity the GPA level of that course has on future academic success (Power et al, 1987). It has also been found that the ease of entry at a specific university affects the predictive capacity of previous GPA levels (Pascoe et al, 1997).

One study by psychologist Arthur Jensen showed a 0.50 to 0.70 correlation between IQ performance and GPA levels of high school students (Jensen,1998). Inversely, scores on measures of academic achievement accurately estimate IQ levels (Frey & Detterman, 2004). Whilst evidence has shown a strong link between cognitive capacity and academic performance, GPA variance is not accounted for by IQ alone in anywhere between 51 to 75% (Rohde and Thompson, 2007).

Another academic performance influencer was put forward in a 1975 paper by Pantages and Creedon and later revisited in 1992 by Abbott-Chapman, Hughes, and Wyld. The likelihood of churn and academic adjustment problems is highly influenced by one's study habits, especially when looking at the transition from High school to university.

According to the study, integration is essential for a higher commitment to the university. In a study in 1978, Terenzini and Pascarella showed that predicting student attrition and churn with previous academic performance accounted for only 4 percent. In contrast, academic and social integration variables accounted for most of the variance.

It seems, according to previous literature, that whilst GPA and exam scores are put forward for success prediction, they are insufficient as indicators of student churn and academic failure. In these cases, studies focus more on adaptability and social integration. The common indicators put forward for predicting both success and failure tend to reside in psychological traits and personal will and study skills.

This can be seen in Rickinson and Rutherford's 1995 study, which found dissatisfaction with the course and social support as the main reasons for student churn. This, along with financial hurdles, are the main indicators for students abandoning their studies (Kaufman & Rousseeuw, 1983). Looking at GPA predictors besides intelligence coefficients and other cognitive capability indicators puts a clear career orientation as a preventive indicator of churn and as an indicator of higher academic results, as shown by Himelstein in 1992. Himelstein showed that low self-expectations are correlated with lower performances. Inversely, self-efficacy has the opposite effect; this is, a belief that one will perform well is correlated with higher performances.

The study of age as a predictor of academic success has seen two different approaches over time. The first relates the predictive value of cognitive indicators mentioned above (previous GPA performance and cognitive abilities) with the age of the individuals or the time since last graduation and, in general, the takeaway is that the predictive value of these indicators diminishes with the span of time since graduation.

Alternative studies have been inconsistent in trying to measure the relationship between age and academic success. In some instances, age and academic achievement are inversely correlated (Clark & Ramsay, 1990) while others found age and maturity provided a clearer career orientation and lower integration needs thus increasing the chance of greater academic success (McInnis et al., 1995). Furthermore, employment status does also influence student churn. In the study mentioned above by Pantages and Creedon in 1975, it was found that 15 hours of employed work per week was the threshold at which the relationship between work and churn inverted. Students working more than 15 hours a week on a full-time course are more likely to withdraw that those working fewer hours.

GPA and student success are used as synonyms whenever analyzing student performance. The most accepted predictors of GPA and thus of student success are IQ, previous GPA results corrected by the time passed since course completion, study habits, course, and self-oriented expectations as well as career orientation. This contrasts with the predictors of churn and course completion where dissatisfaction, financial stress, career orientation, expectations and employment responsibilities are the main factors.

Even the most accepted student performance indicators like GPA and IQ are often criticized as their predictability efficiency does not explain smaller variances. Since these indicators are responsible for around 50 percent of student performance, the smaller variance must lay within other indicators not as often studied.

Studies of personality traits and psychology analysis like the Big Five traits, put together by Costa and McCrae in 1992, have seen increased attention when trying to predict student success. These traits account for nearly 20% of GPA variance in college students. If we narrow down to the impact of each of the five traits, we see that conscientiousness and openness explain nearly 17% of the variance in intrinsic motivation. These, together with neuroticism and agreeableness, explain 14% of GPA variance. Results indicated that extraversion seems not to predict student performance. When it comes to intrinsic motivation to accomplish things, its explanative value amounted to 5% of GPA results (Komarraju, Karau and Schmeck, 2009).

In short, regardless of the predictability values of psychological traits, schools are yet to adopt them on their admission processes, as these still rely primarily on GPA and other cognitive performance indicators. Incorporating psychological traits and course preferences into admission process requires more than a resume or a motivational letter. Studies and attempts to add psychological metrics to the admission processes have focused on predictive machine learning models trimmed to each course.

2.4. Data Mining on Educational Systems

There are a number of papers applying predictive analytics to educational systems. In some studies, test grades, midterm grades and student class attendance are used to build decision trees and compute the likelihood of a certain student failing the course (Ahmed and Elaraby 2014). K-means clustering algorithms were also used in studies by Bhise in 2013 to forecast student success. Decision tree classifiers have also been used to predict student churn. In a particular article, a model implementation was used to develop a new teaching plan that was able to reduce student dropout by 14% when compared to previous semesters (Burgos et al. 2018).

There is a space for the use of data mining tools as predictors of student performance. In a study by Abele and Spurk, 2009 , four different mathematical processes were used to predict student academic performance using GPA and grades in selected courses. Multiple Linear Regression model (MLR), the multilayer perception network model, the radial basis function network model, and the Support Vector Machine model (SVM) were used and compared to find the best suit for the admission criteria problem. The analysis showed that the type of model had only a small effect on the prediction accuracy. Combining the multiple models, however, seemed to increase the percentage of accurate predictions highly.

In this study (Abele and Spurk, 2009), they attempted to answer the question of what the best model would be to predict the average academic performance of a particular class. For this instance, the MLR model sufficed. On the other hand, if we were to predict each student's performance, then the SVM model yielded the best results of the bunch.

As a final remark, the authors noted that, to increase model's accuracy, psychological factors should be considered.

In 2018, a study used machine learning models to understand the predictability of academic performance based on engagement with online classes, using four different classification algorithms. Decision trees and gradient-boosted trees returned high recalls on the classification model (Hussain et al. 2018).

When it comes to predicting individual performance based on previously obtained indicators, support vector machines are put forward as one of the models with the best results. As of today, not much literature has delved into the accuracy of neural network models for this topic. Studying the results obtained with neural networks comparing them with SVMs and decision trees will be critical in the present study, as neural networks can achieve higher complexity levels when compared with SVM. In turn, SVM might prove helpful as, for well-separated classes, the number of observations required to train the model is not high.

It is also important to note that SVMs tend to require less computing power than neural networks. This will need to be considered further down the line to understand this impact on the use case.

3. Methodology

This chapter provides an outline of the research methodology used to develop the predictive model. The model creation will be developed with using Azure Machine Learning software. All the necessary analysis and mining, from data preparation to model simulation, can be done in this single tool.

The CRISP-DM (CRoss Industry Standard Process for Data Mining) reference model for data mining provides an overview of the life cycle of a data mining project, including the different phases, tasks, and outputs (Shafique et al. 2014)[36]. The life of a data mining project is broken down into six phases, which are not strict in sequence. These are Data and Business Understanding, Data Preparation, Modeling, Evaluation and Deployment.

Although widely accepted in data mining as a de-facto standard, CRISP-DM does not cover the processes regarding data acquisition (Huber et al. 2019). The data used for this project was acquired using a questionnaire developed by Design the Future aimed at understanding individual personal interests and individual passions. This data is analyzed in the Data Preparation stage.

3.1. CRISP-DM use case

Data mining methodologies were introduced to give structure to the knowledge discovery process beyond machine learning algorithms. Their goal is to define a workflow, beginning with creating questions summarising the problem, leading to targeted processing of the raw data and ultimately creating new knowledge (Huber et al. 2019). That said, where possible the workflow is aligned with the steps of CRISP-DM.

3.1.2. Business Understanding

The project objective was the development of a model capable of predicting the relative academic success of an individual based on easily obtained information from the well know and studied GPA to less studied yet very relevant psychological traits.

To achieve this, 3 experiments were run, the first had the goal of creating a model including both the psychological variables and also measures of performance like GPA and age. The second, focused only on the traditional measures of performance as inputs. The third had, as inputs, only the psychological variables.

This allowed us to test the impact psychological traits can have in measuring student performance, when combined with GPA and age. The three experiments are compared in the evaluation chapter.

4. Empirical Process

4.1. Data Understanding

Data understanding drives the focus to collect, identify and analyze the acquired data sets that are used to accomplish the project goals. It is in this step that the quality of the data is assessed. This process consists of four phases: data collection, data description, data exploration and data quality verification.

4.1.1. Data Collection and Description

Throughout this project, data was acquired via a Quiz shared with the student body of Nova University and other students across Portugal via social media platforms like Facebook, LinkedIn and Whatsapp. This quiz was built by *design the future* a platform that is part of the Association Better Future in Portugal. The questionnaire asks participants 69 questions (table 2) that have historically been used to indicate potential career paths based on psychological traits. This quiz is based on the RIASEC model for type theory of personality, proposed by Dr. John Holland in the 1970's. The full questionnaire, including the additional demographic entries and other questions can be seen in the annex of this project.

A dataset comprising 318 individual responses answering a questionnaire with the demographic and academic performance questions depicted in Table 1 was collected. Our dependent variable is represented by the question: *What is your current GPA on a scale of 0-20?*

Question	Data Type	Missing/Invalid Values
What year were you born?	Integer	0
To which gender do you most identify?	Categorical	3
What is the level of your current course?	Categorical	2
Which area of study better suit your current course?	Categorical	14
What is your current GPA on a scale of 0-20?	Integer	129
Prior to the course you answered about, what was the level of the last course you graduated from?	Categorical	0
Which area of study better suit the past course?	Categorical	2
What was your GPA on that course (0-20) Scale?	Integer	3
How many years ago did you graduate from it?	Integer	2

Table 1 - *Questionnaire - Demographic Questions*

The respondents were also asked to select the psychological traits and preferences they identify with from the subset of questions depicted in Table 2. All the data entry points are Boolean.

Questions	Number of entries
I listen, with interest, to recommendations to improve my skills and fix my flaws	215
I recognize my qualities and flaws	202
I have moments of reflection on what I do, say, or feel	197
I can put myself in another one's shoes and can understand what they like, how they think and what they feel	177

I listen closely to words, expressions and information and try to use them when I speak	161
I actively participate in conversations and debates	158
I recognize emotions that I feel and their reasons	155
I efficiently communicate with gestures, movements, ideas, questions, and emotions	141
I interpret and identify numbers in multiple contexts (transports, dates, prices, discounts, locations, schedules, economic information, historical information...)	137
I remember faces, colours, sizes, objects, and scenarios	132
I correctly interpret maps and plans	122
I am good at handling materials when the goal is to count, compare, order, weigh, and measure	117
Study and resolve scientific problems	108
I solve riddles and mathematical quizzes	106
Help others acquiring new skills and knowledge	102
I solve problematic situations and am fast at mind calculations	99
Lead projects with societal impact	98
lead the activities of a company	98
I can write letters, histories and transmit adequately ideas, emotions, and experiences	98
Social service activities and volunteering	96
Listen, advice and support people	96
Talk, debate, and publicly present your ideas	92
Help others making decisions	92
Developing efficient work methods	92
The others show interest and desire my company	92
I am a good storyteller	90
Lead people and groups	86
Dedicate high amounts of time to reading	86
In games, sports, and dances I can perform the adequate movements	86
I use illustrations to help remembering information	85
Solve problems related to people's lives	83
I easily copy other people's expressions and movements	83
Organize events, concerts, group travels and other leisure activities	82
Plan and establish work related goals	82
I mingle with ease in situations that allow me to meet and relate with others	82
Talk fluently about any subject matter	81
I am, generally, seen as a natural leader	79
Play a musical instrument	67
I can reproduce and remember with ease any melody I listened to	67
I assemble and disassemble objects with ease	66
Persuade others to modify their behaviour	61
Manage the work tasks of others	58
Plan budgets	56
With instruments and my voice, I am able to interpret melodies and songs, with emphasis in execution, lyrics and rhythm	55
Creating marketing campaigns	54
Advice companies on financial topics	50
Perform economic analysis of national and international topics	49
I sing and move with spontaneity whilst I realize a plethora of tasks	48
Work with animals and nature	47
Develop mobile applications and other tech	44
Take care of a garden or forests	43
I analyze art considering quality criteria such as perspective, proportion, use of colour, sizes, etc.	40
Manage and organize a company's archives	37

Sing in a choir or band	36
Heal and treat the sick	36
I identify musical instruments and can create new ones out of ordinary objects	34
Write to a newspaper	31
Evaluate lab samples and experiment	29
Control Industrial Machine usage	28
Participate and compete in sporting events	27
Paint, design and sculpt	23
Writing romances	21
Perform lab experiments with disease and treatment	21
Plan and project buildings and parks	20
Develop activities related to theatre and dramatic expression	19
Choose the art pieces to be displayed in a gallery	18
Create and design clothes	18
Professional athleticism	12
Partake in defence and patriotic activities	5

Table 2 - *Questionnaire - Psychological Traits and Preferences*

4.1.2. Data Exploration

During this phase, data was queried to understand the relationships among the data and whether the present data can verify the business question. Variables with low representation can be discarded to avoid problems with overfitting.

Of the 318 entries, 129 did not contain information regarding the dependent variable. These rows were discarded. By removing extra 13 dependent variable invalid values there are 176 valid entries for the classification model development. This removal mitigated the rest of the missing or invalid datapoints available and there was, thus no more missing value treatment needed.

29 of the 69 psychology questions had below 20% representativity and were removed, leaving 40 of these variables available for the project's next steps.

4.1.3. Data Quality

Correlation is one of the most common statistics. Using one single value, it describes the "degree of relationship" between two variables. The Pearson Correlation Coefficient is effective as an optimization criterion to derive different optimal noise reduction filters (Benesty, et al. 2009).

The Pearson Correlation was used leveraging the *Filter Based Feature Selection* node in Azure ML and filtering down to 12 features, see figure 1. Pearson's correlation coefficient is computed by taking the covariance of two variables and dividing by the product of their standard deviations. For each level in the categorical column, the model computes the conditional mean of the numeric column and correlates the column of the conditional means with the numeric columns (Microsoft, 2021).

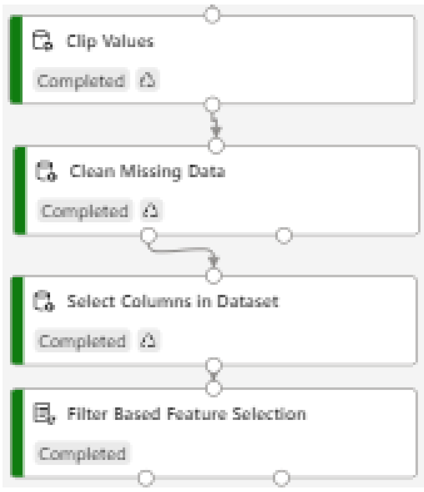

Figure 1 - *Filter Based Feature Selection Node*

The Pearson Correlation Coefficient outputted 12 independent variables that correlate strongly with the dependent variable. This new information was used to run models with fewer variables parallel to the raw data, as depicted in the following chapters. Aligned with literature, age, GPA and years since graduation correlate with the dependent variable at 0.31, 0.17 and 0.11, respectively. Psychological preferences are also highly correlated with GPA in this dataset, as shown in table 3.

Independent Variable	**Correlation Coefficient**
Age	0.31
GPA on previous course	0.17
Others show interest and desire my company	0.17
Help others acquiring new skills and knowledge	0.16
I recognize my qualities and my flaws	0.15
Listen, advice and support people	0.13
Years since graduating from previous course	0.11
Dedicate high amounts of time to reading	0.10
Social service activities and volunteering	0.08
Study and resolve scientific problems	0.07
Organize events, concerts, group travels and other leisure activities	0.07
Lead the activities of a company	0.07

Table 3 - *Pearson Correlation Results*

4.2. Data Preparation

4.2.1. Select Data & Clean Data

Missing data was removed and corrected in the previous steps. Data was selected after new datapoints were constructed.

4.2.2. Construct and integrate Data

Two variables were created. The first, a Boolean, represents students whose current course and the previous course are the same or from the same field of study (Engineering to engineering-related or Medicine to a Medical-related masters are two examples) and is identified as *same_course.* The second was created to represent the decrease in GPA predictability as years since course completion elapse. It was computed by dividing GPA on the previous course by the years elapsed since conclusion and is identified as *previous_GPA_value*.

The independent variable was partitioned into two bins, the top and bottom half. The distribution was not evenly balanced, with 99 data points as low performers and 77 data points as high performers. This results from the data type of our independent variable, integer, as the median GPA was 16 and mean 16.01. In effect, all the GPA scores of 16 ended up below the threshold, on the bottom of the class.

One more binning process was done, dividing the previous GPA variable into four equal weight bins.

4.2.3. Format Data

Metadata was edited to ensure *Current_GPA* was set as the dependent variable and the rest as independent variables with the accurate data types.

4.3. Modelling

It is in this phase that the machine learning algorithms are created. This process is based on literature and on empirical evidence obtained from the set of data being utilized. Modelling is divided into model selection, test design, model development and model assessment.

4.3.1. Model Selection

Multiple models were selected and trained for this project.

Decision forests and decision trees were trained iteratively. Decision trees were chosen for their ability to explain the results obtained, despite being, generally, weak. They have two elements, nodes and branches. At each node, one of the features of our data is evaluated in order to split the observations. These trees are constructed by iteratively evaluating the features and identifying the one that reduces entropy by the largest amount.

Decision trees can be combined by bagging. This is an ensemble method that relies on sampling with replacement from the original set. Some observations may be repeated in each new training data set. From each new set, a decision tree is formed and, in the end, a majority vote, by the created models, is taken into consideration. The results of the vote increase the accuracy of the model. This ensemble model, is the decision forest we are also training.

Support vector machines were modelled as per literature recommendations for tackling both the prediction problem and the dimensionality issue. It functions by finding an hyperplane in an N-dimensional space, where N are the number of features, that maximizes de margin around the separating hyperplane and the samples, classifying the datapoints. This means that, for two variables the hyperplane is a line, for three it is a plane and so on.

Neural networks were tested too due to the lack of literature reportability on its efficacy. This model is designed to recognize patterns in complex data. It consists of neurons (or nodes) spread across 3 layers (input, hidden and output) , the connections between neurons (weights) and biases connected to each neuron. Neural networks learn by backpropagation. The output the network produces is compared against the output it was meant to produce, using the difference between them, the weights of the connections are modified.

To determine the optimum hyperparameters for the different models the *Tune Model Hyperparamenters* node in Azure ML was used. This tool tunes the parameters by trying and testing different parameters iterating multiple times, measuring the AUC, until it stops improving (Microsoft, 2021). The AUC plots the true positive rate against the false positive rate and ranges from 0 to 1. The value of 1 represents a model that predicts perfectly. The accuracy, specificity and sensitivity values were calculated based on the confusion matrix of the model. Building a confusion matrix requires a threshold value to determine whether the output probability level gets attributed to a case or a control. Accuracy will be the metric used to optimize the threshold value.

All the models were tested with both the full variable set and with the 12 variables trimmed set and the latter had consistently better performance for all the models.

4.3.2. Test Design

Data was split into training and testing sets using a hold-out method. This is a simple and fast method to evaluate a classifier without using too much resources. The data was partitioned into two sets containing 80% (141 entries) and 20% (35 entries) of the entries respectively. The split was stratified ensuring both the test and training sets had the same proportion of high "performers", 44% on the training set and 43% for the testing set.

After the partition, the same process for normalization was followed for both datasets. Two variables were normalized: *Previous_GPA_value* and *Age.* For both, Z-score normalization was used to scale the variables in such a way that their standard deviations amount to 1 and mean to 0. Z-score was used over Min-Max as outliers, such as people that have been in the workforce for decades without having completed a course during this time span, should be reflected in the model. Min-Max suppresses the effects of outliers, and this is not wanted for a data set where outliers are expected and represent a particular demographic. This ensures that 20 and 10 years of work experience are treated differently.

In the Azure ML model, it is represented by the nodes shown in figure 2: *Split Data, Edit Metadata and Select Columns in the dataset.*

Prior to the model development, the columns in the dataset were selected. To separate the data for the experiments, three different selections were created (

Figure 3

Figure 3 - *Column selection for the different experiments*). The first was based on the Pearson correlation matrix evaluated above and on the new variables created; this experiment includes both the psychological traits and the classical indicators as inputs. The second one is a control set comprised only of well documented features that influence GPA: *Same_course, Previous_GPA, time since last course* and *Previous_GPA_value*. The third one is only comprised of psychological variables.

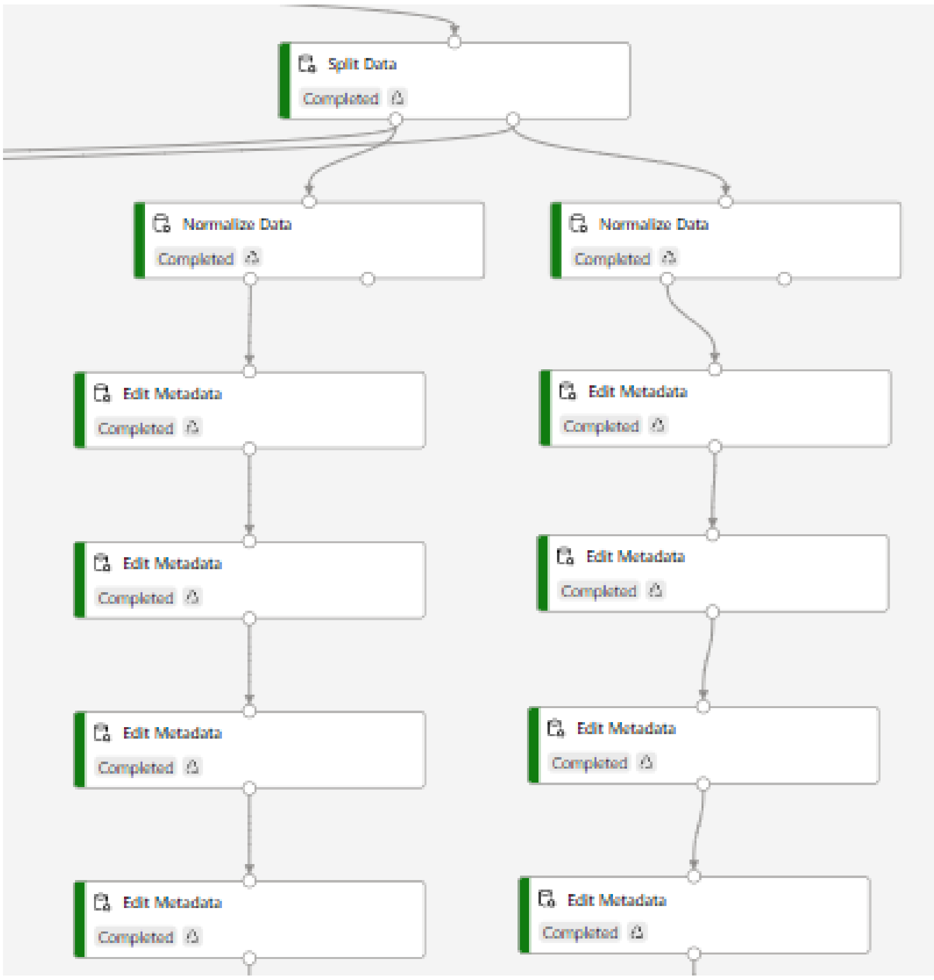

Figure 2 - *Stratification, Normalization and Column selection*

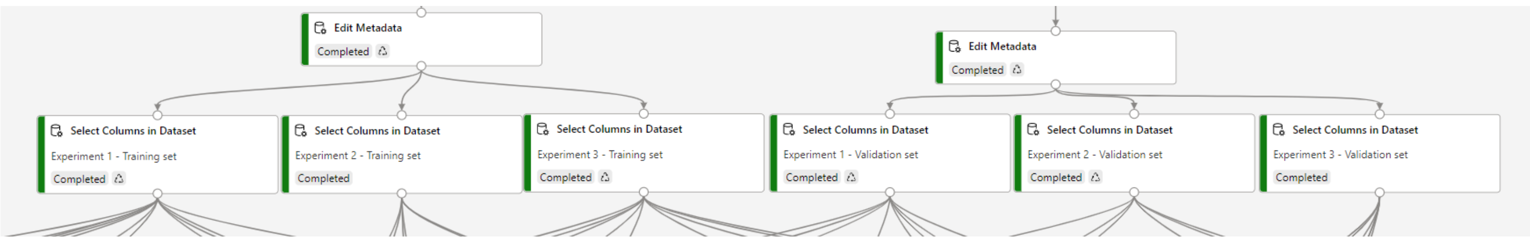

Figure 3 - *Column selection for the different experiments*

4.3.3. Model Development

The same method for model development was followed for the different models: Support vector machines, decision forests, decision trees and neural networks.

Data was split into two stratified sets, data was normalized, columns were selected for three different runs, the first run based on the Pearson Correlation, the second one on considering the three GPA related features and the third one taking only the psychologic variables as inputs. When deciding the parameters for the different models, the tune model hyperparameters node was set to try all the combinations possible (GridSearch) and return the parameters responsible for the best accuracy.

4.3.3.1. Pearson Matrix Data Set – Experiment 1

Support Vector Machines were trained with 100 iterations and L1 weight for regularization of 0.0001. This weight sets the trade-off between the number of mistakes on the training data and the margin of the data separation, in short, it is the degree of importance that is given to misclassifications. The smaller the value the higher the tolerance for misclassifications which allows for model development for non-separable problems.

Decision forests were trained with 8 decision trees with pre-pruning at a depth of 16 and with a minimum of 16 samples per leaf node. The ensemble follows a Bootstrap aggregation. In this method, each tree is grown on a new sample, created by randomly sampling the original dataset with replacement until the original dataset size is reached. The outputs of the models are combined by *voting*. Each tree in a classification decision forest outputs an unnormalized frequency histogram of labels. The aggregation is to sum these histograms and normalize to get the "probabilities" for each label. In this manner, the trees that have high prediction confidence will have a greater weight in the final decision of the ensemble.

Decision trees were pre-pruned, leveraging the values outputted by the tune model hyperparameters node, with a maximum depth of 12 and a minimum number of samples per leaf node of 10.

As for neural networks, the same approach was taken, relying on the sweep to identify the set of hyperparameters that optimized the accuracy for the training set. The learning rate was set to 0.1. It determines the step size for a model to reach the minimum loss function. The loss function is a prediction error function for the neural net. A higher rate makes the model learn faster but may miss the minimum loss function, a lower learning rate makes the model more resource intensive, risks finding a local minimum instead of the global minimum but can be more precise in finding this minimum. The number of hidden layers was set to 1 (the maximum allowed in Azure ML today) and the number of neurons on the hidden layer was set to 20 as per the results of the sweep, optimizing accuracy. The number of learning iterations was set at 20 and the momentum (weight to apply during learning to nodes from previous iterations) was set at 0.15.

4.3.3.2. Control Data Set – Experiment 2

Support Vector Machines were trained with 10 iterations and L1 weight for regularization of 0.001.

Decision forests were trained with 32 decision trees with pre-pruning at a depth of 64 and with a minimum of 16 samples per leaf node. The ensemble follows a Bootstrap aggregation.

Decision trees were pre-pruned, leveraging the values outputted by the tune model hyperparameters node, with a maximum depth of 16 and a minimum number of samples per leaf node of 7.

As for neural networks the learning rate was set to 0.2. The number of hidden layers was set to 1 (the maximum allowed in Azure ML today) and the number of neurons on the hidden layer was set to 40 as per the results of the sweep, optimizing accuracy. The number of learning iterations was set at 80 and the momentum (weight to apply during learning to nodes from previous iterations) was set at 0.15.

4.3.3.3 Psychologic Data Set – Experiment 3

Support Vector Machines were trained with 10 iterations and L1 weight for regularization of 0.001, the same as the control set.

Decision forests were trained with 8 decision trees with pre-pruning at a depth of 16 and with a minimum of 4 samples per leaf node. The ensemble follows a Bootstrap aggregation.

Decision trees were pre-pruned, leveraging the values outputted by the tune model hyperparameters node, with a maximum depth of 16 and a minimum number of samples per leaf node of 7.

As for neural networks the learning rate was set to 0.2. The number of hidden layers was set to 1 (the maximum allowed in Azure ML today) and the number of neurons on the hidden layer was set to 100 as per the results of the sweep, optimizing accuracy. The number of learning iterations was set at 80 and the momentum (weight to apply during learning to nodes from previous iterations) was set at 0.2.

4.3.4. Model Assessment

During the model assessment phase, the models are tested against classification accuracy, sensitivity, specificity, false positive rate, and precision. These parameters are computed by analyzing the confusion matrix. It represents counts from predicted and actual values. It outputs "TN", True Negatives, showing the number of accurately classified negative examples. "TP", True Positives, shows the number of accurately classified positive examples. "FP", False Positives, representing the number of negative examples misclassified as positive. "FN", False Negative, shows the number of positive examples misclassified as negative.

Accuracy is computed by takes the proportion of correctly identified inputs over the total number of inputs. As per the following formula:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Sensitivity (also known as recall) shows the ratio of correctly classified positives divided by the total number of actual positives as can be seen below:

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity (also known as precision) shows the ratio of correctly classified negatives divided by the total number of actual negatives as depicted in the formula below:

$$Specificity = \frac{TN}{TN + FP}$$

The classification threshold (also called decision threshold) is an important tuneable value. It defines the score necessary for the model to classify the value as positive or negative. In effect, this number is changed empirically to ensure the model achieves the best fit.

4.3.4.1. Experiment 1

In the training set, Decision Forests yielded higher accuracy (0.936) than Decision trees (0.853), Neural Networks (0.730) and SVMs (0.667). This is depicted on Table 4.

Model	Accuracy	Specificity	Sensitivity
Decision Forests	0.936	1	0.855
Neural Networks	0.730	0.731	0.613
Decision Tree	0.853	0.800	0.857
SVMs	0.667	0.653	0.516

Table 4 - *Training Model Assessment - Experiment 1*

The models were scored against the validation set, Decision Forests yielded higher AUC (0.790) values than decision trees (0.677), SVMs (0.650), and Neural networks (0.573).

Comparing the training results with the validation set results of the neural networks, for the dataset at hand, we conclude that these are not good models to solve this problem. Decision trees, but also decision trees and SVMs showed better performances with accuracies of 0.714, 0.714 and 0.657 respectively (table 5).

Model	Threshold	Area Under the Curve	Accuracy	Specificity	Sensitivity
Decision Forests	0.45	0.790	0.714	0.647	0.733
Neural Networks	0.5	0.573	0.629	0.75	0.200
Decision Tree	0.6	0.677	0.714	0.727	0.553
SVMs	0.5	0.650	0.657	0.588	0.667

Table 5 - *Model Assessment - Experiment 1*

The ROC curve and the confusion matrix for the decision forest, the best model for the experiment 1, are depicted in figures 4 and 5 below.

Taking a better look at the decision forest confusion matrix, the sensitivity value outperforms the specificity value. Of the 35 values, 11 were classified as TP, 6 as FP, 14 as TN and 4 as FN.

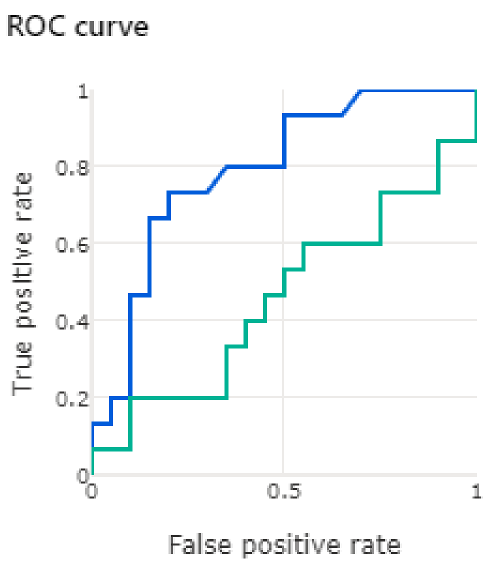

Figure 4 - *Decision Forest ROC curve (in blue) – Experiment 1*

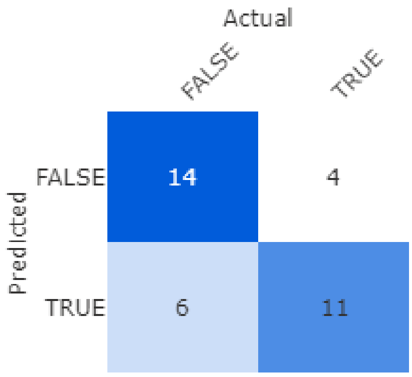

Figure 5 - *Decision Forest Confusion Matrix - Experiment 1*

Azure ML leverages a Python SDK to measure feature importance for interpretability illustrating the magnitude and direction of each independent variable, also called global understanding. The SHAP (Shapley Additive exPlanations) Tree Explainer is a game-theoretic approach to explain the output of machine learning models (Lundberg, Scott M And Lee, Su-In, 2017). The dataset top 5 features for the Decision Forest, based on their importance are *age* (0.137), *Help others acquire new skills* (0.115), *previous_GPA_value* (0.097), *Help others make decisions* (0.054) and *Lead projects with societal impact* (0.048).

Analyzing the decision tree, the most relevant features are: *help others acquire new skills, Previous_GPA* and *Previous_GPA_value, I actively participate in conversations and debates, Talk fluently about any subject matter, lead projects with societal impact, I listen closely to words, expressions and information and try to use them when I speak, same_course, previous_GPA_value, years since graduation,* with the split gains visible on table 6 and the tree design in figure 6. The split gain or information gain is calculated by subtracting the weighted entropies of each branch from the original entropy. Entropy is calculated using the equation below where p_i is the probability of randomly picking an element of class i where c is the number of classes. In short, higher information gains indicate higher feature importance for the model.

$$E = -\sum_{i}^{c} p_i \log_2 p_i$$

Feature	Information Gain
Help others acquire new skills and knowledge	0.161
I actively participate in conversations and debates	0.143
Lead projects with societal impact	0.184
Talk fluently about any subject matter	0.118
Previous_GPA	0.112
I listen closely to words, expressions and information and try to use them when I speak	0.066
Same_Course	0.03
Previous_GPA_value	0.09
Years since graduating from previous course	0.118

Table 6 - *Decision Tree Information Gain*

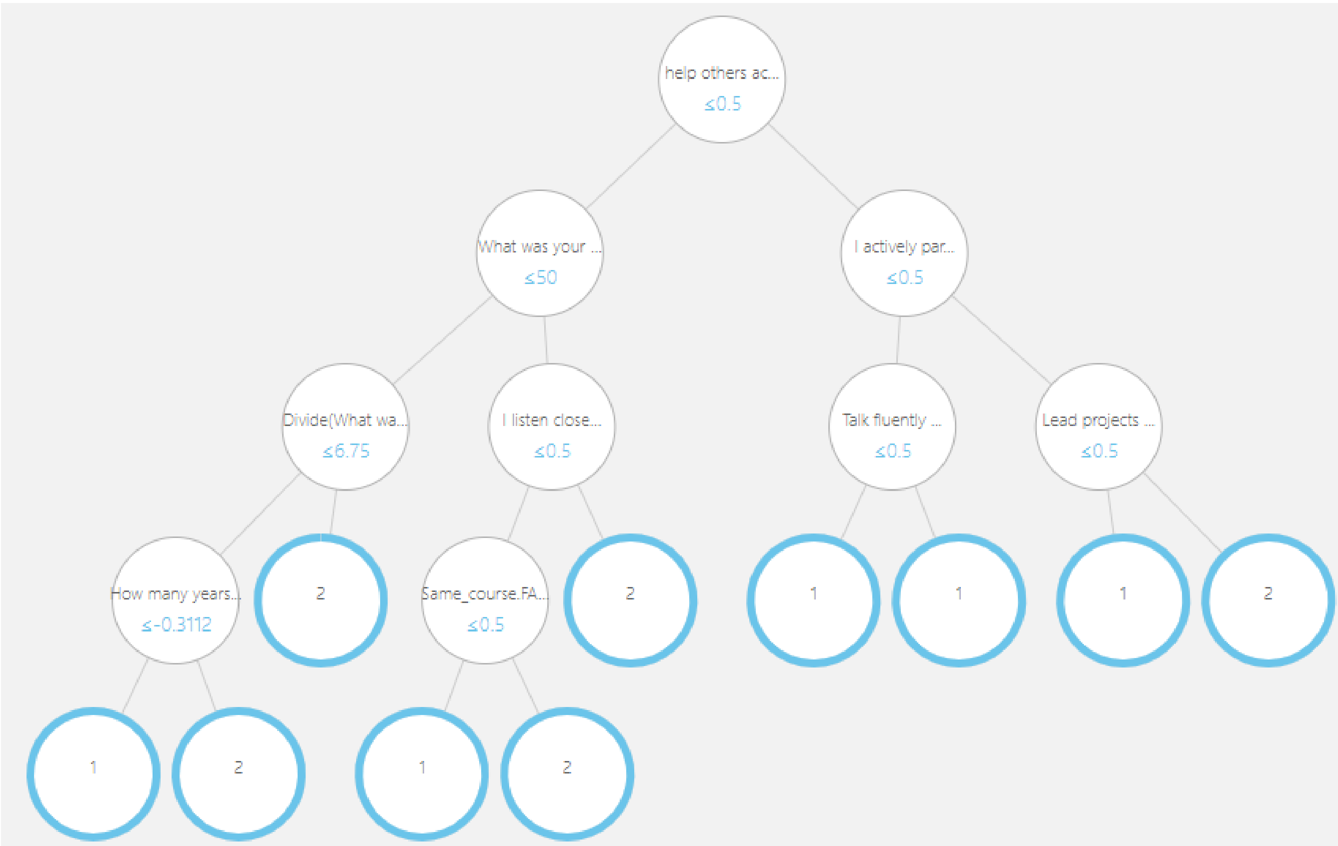

Figure 6 - *Decision Tree visualization – Experiment 1*

From the decision tree and decision forests feature analysis we get *Help others acquire new skills, previous_GPA_value* and *Lead projects with societal impact* as common important independent variables

4.3.4.2. Experiment 2

In the training set, Decision forests yielded higher accuracy (0.759), followed by decision trees (0.674), neural networks (0.660) and SVMs (0.638). Decision forest also presented the higher specificity (0.733) and sensitivity (0.710) values as seen in table 7.

Model	Accuracy	Specificity	Sensitivity
Decision Forests	0.759	0.733	0.710
Neural Networks	0.660	0.684	0.419
Decision Tree	0.674	0.617	0.692
SVMs	0.638	0.677	0.339

Table 7 - *Training Model Assessment – Experiment 2*

The models were scored against the validation set, decision forests yielded higher AUC (0.717) values than the decision tree (0.702), neural networks (0.435) and SVMs (0.398).

Comparing the training results with the validation set results of both neural networks and SVMs, for the dataset at hand, are not good models to solve this problem. Decision trees and decision forests on the other hand showed better performances with accuracies of 0.743 and 714 respectively (table 8).

Model	Threshold	Area Under the Curve	Accuracy	Specificity	Sensitivity
Decision Forests	0.5	0.717	0.743	0.800	0.533
Neural Networks	0.5	0.435	0.486	0.333	0.200
Decision Tree	0.5	0.702	0.714	0.692	0.600
SVMs	0.5	0.398	0.486	0.286	0.133

Table 8 - *Model Assessment – Experiment 2*

The ROC curve and the confusion matrix for the decision forest, the best model for the experiment 2, are depicted in figures 7 and 8 below. Taking a better look at the decision forest confusion matrix, the specificity value outperforms the sensitivity value. Of the 35 values, 8 were classified as TP, 2 as FP, 18 as TN and 7 as FN.

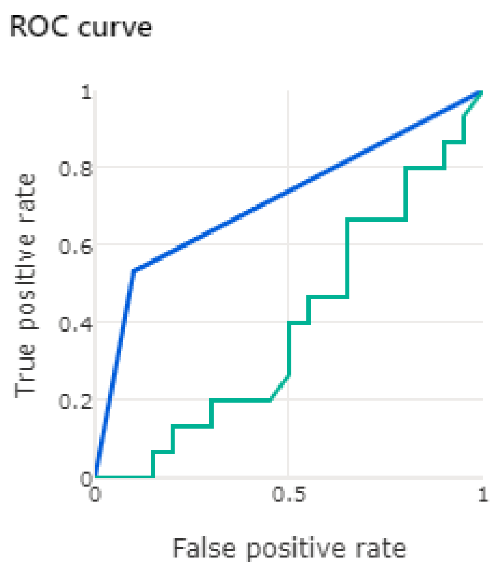

Figure 7 - *Decision Forest ROC curve (in blue) – Experiment 2*

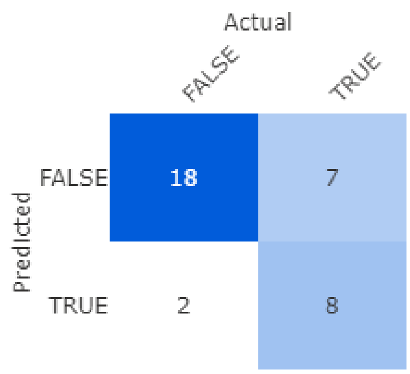

Figure 8 - *Decision Forest Confusion Matrix – Experiment 2*

The experiment 2 relied on only 5 features. Their importance, based on the SHAP Tree Explainer (table 9), outputs *Age* and *GPA* as the most important features for predicting GPA for this control group. Whether the course the student is pursuing is aligned with his background is of least importance.

Feature	**Importance**
Age	1.968
Previous GPA	1.282
Previous_GPA_Value	0.983
Same_Course	0.581
Years since graduating from previous course	0.415

Table 9 - *Feature importance – Experiment 2*

4.3.4.3. Experiment 3

The same assessment was made for the models developed exclusively with the psychological features. SVMs yielded higher accuracy (0.667) than both decision forests and decision tree (both 0.657), the three models had the same specificity value (0.667) and SVMs obtained sensitivity of 0.5 comparing to 0.4 of decision forests and decision tree (table 10). Neural networks obtained an accuracy of 0.517. In any case, a good model fitness was not possible to achieve with the dataset at hand.

Model	**Accuracy**	**Specificity**	**Sensitivity**
Decision Forests	0.657	0.667	0.4
Neural Networks	0.517	0.406	0.277
Decision Tree	0.657	0.667	0.4
SVMs	0.667	0.667	0.5

Table 10 - *Training Model Assessment – Experiment 3*

In the test set, the AUC values ranged from 0.510 and 0.560, accuracies ranged between 0.486 and 0.600. The low AUC and accuracy values show that the psychological features alone in the collected dataset are not robust predictors of academic success. These values are summarized in table 11.

Model	**Threshold**	**Area Under the Curve**	**Accuracy**	**Specificity**	**Sensitivity**
Decision Forests	0.45	0.560	0.571	0.500	0.533
Neural Networks	0.5	0.510	0.571	0.500	0.400
Decision Tree	0.5	0.540	0.600	0.529	0.600
SVMs	0.5	0.520	0.486	0.414	0.800

Table 11 - *Model Assessment – Experiment 3*

The ROC curve and the confusion matrix for the decision forest, the best model for the experiment 3, are depicted in figures 9 and 10 below. Taking a better look at the decision forest confusion matrix, of the 35 values, 8 were classified as TP, 8 as FP, 12 as TN and 7 as FN, with an accuracy of 0.571.

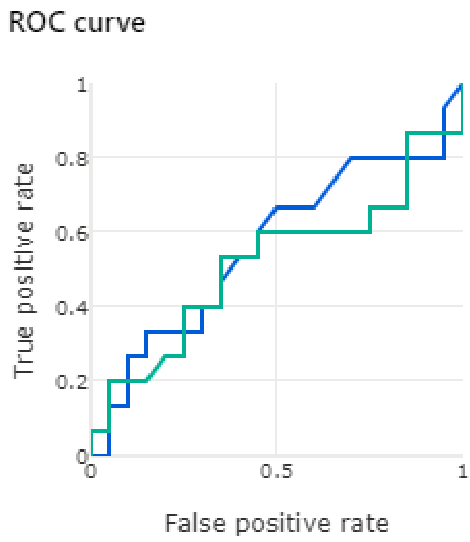

Figure 9 - *Decision Forest ROC curve (in blue) – Experiment 3*

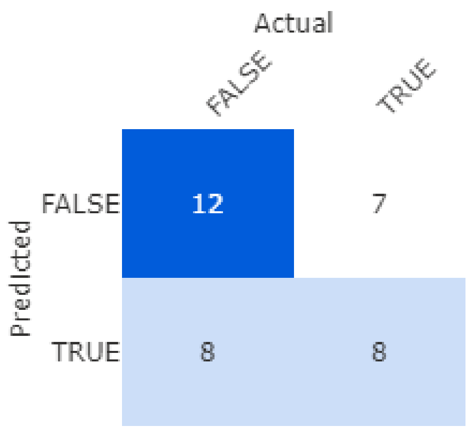

Figure 10 - *Decision Forest Confusion Matrix - Experiment 3*

4.4. Evaluation

4.4.1. Model Evaluation Considerations

For both the control dataset and the experiment 1 decision, forests were the model with higher fitness to the training model and higher AUC with the validation set 0.717 and 0.790, specificities of 0.800 and 0.647 and sensitivities 0.533 and 0.733, respectively. Looking at both ROC curves, the one correspondent to the control dataset only has 3 points in the plot, full positive prediction, full negative prediction, and the value represented by the confusion matrix in figure 9. This model cannot be artificially made, by adjusting the threshold value, more specific or sensitive, to adapt the output to our problem. This means that in a case where a false negative prediction is costly, the model cannot be tuned to accept more false positives to overcompensate the cost of an FN, in effect increasing the specificity in detriment of its sensitivity is not possible. This, however, contrasts with the ROC curve from the Pearson matrix-based model. Adjusting the model threshold to 0.45 maximized the accuracy but sensitivity is maximized (1) when the threshold is 0.51 and specificity is maximized (1) with a threshold at 0.40.

This means that, while the control model has a higher specificity, it is outperformed by experiment 1 model in its AUC, and, by shifting the threshold to 0.44 its specificity matches the one from model 2 at 0.800 and the sensitivity beats it at 0.6.

Focusing on the experiment 3, all the models proved inaccurate with the best AUC of 0.560 for the decision forest. As expected, this dataset is not a good predictor of student success when used in isolation. Taking into consideration that the dependent variable had a proportion of top performers at 44% and 56% for the low performers, an accuracy of 0.571 is marginally better than a random predictor.

4.4.2. Feature importance considerations

Feature importance varied between models, the value of importance computed by SHAP is not absolute and thus importance values should not be compared across models and only in between the models' features (values summarized in table 12). The control set is highly impacted by age and previous GPA, as well as the created metric of *Previous_GPA_Value*. In short, age, previous GPA and the time since completion are good predictors for this model.

Looking at experiment 3, willingness to help others, participation in conversations and debates as well as the ability to listen, advise and support people are the most important features. These, while part of a bad classifier model, can still give us some insights on the psychological features that have a higher predictive power.

When analyzing experiment 1 we see a mix of these features, in fact, only one of the top 5 features of this model is not present in the top 5 features of the others: *Help others make decisions.* The others are present in experiment 2 (*Age* and *Previous_GPA_Value*) and in experiment 3 (*Help others acquire new skills and knowledge* and *Lead projects with societal impact*). Interestingly, the fact that the masters pursued is or not in the same area of studies as the bachelors does not seem to greatly impact future performance when psychological variables are taken into consideration. The addition of psychological variables into the control model seems to be well accepted and improve the overall classification.

Feature	Experiment	Importance
Age	1	0.137
Help others acquire new skills and knowledge		0.115
Previous_GPA_Value		0.097
Help others make decisions		0.054
Lead projects with societal impact		0.048
Age	2	1.968
Previous_GPA		1.282
Previous_GPA_Value		0.983
Same_Course		0.581
Years since graduating from previous course		0.415
Help others acquire new skills and knowledge	3	0.052
I actively participate in conversations and debates		0.029
Listen, advice and support people		0.028
Lead projects with societal impact		0.019
Lead the activities of a company		0.003

Table 12 - *Feature Importance summary*

4.4.3. Dataset size considerations

The small size of the dataset negatively impacted the model development process. From the 319 datapoints initially gathered only 176 (55%) could be used, mostly caused by the high number of responses from people not currently doing any level of secondary education.

A direct approach to students in universities should improve this ratio. It is expected that the results from the current study can be improved upon with a bigger dataset. This impacted the neural networks models as these tendentially require larger datasets to enhance their performance. Nevertheless, the decision forests models 1 and 2, with different rates of success, proved capable of predicting the performance of a student.

In line with the literature, features such as previous GPA, age and time since last course have a high predictive value. The model showed that while psychological traits alone are not good predictors of student success, they can be added to the traditional set of features to improve the predictability of the models.

4.4.4. Azure ML considerations

The entire project was developed on Azure ML, the tool proved useful by providing a cloud computing unit with 2 cores, 14 GB RAM and 28GB of disk space, relieving the personal device for exploratory work while running the model in parallel on a browser window. This also meant the model could be run overnight without the need for a device to be turned on during the process.

It provided the entire nodes necessary, from clipping for dealing with outliers, data binning, data normalization, data split, math operations, missing data cleaner, metadata editor, multiple models prebuilt with the ability to change hyperparameters, the tune model hyperparameters discussed before and evaluation and testing nodes. For the more complex activities it also allowed for the use of a SQL node for SQL transformations and a Python node to run scripts.

There were just a few inconsistencies when exporting the ROC curves, the confusion matrices and the feature importance data as the evaluation preview model is still under preview and sometimes didn't provide any result.

Once a model is completed, Azure ML allows the user to automate the pipeline by creating an endpoint externally accessible.

5. Conclusions and future work

5.1. Conclusions

This project aimed to provide a complementary metric to assist in student performance predictions. Based on the literature, it is widely accepted that GPA and time since course completion are important indicators in determining the future GPA of a student, up to 50% correlation with future grades. Another factor commonly referred, based on its predictive value, is career orientation, influenced by the age of the individual. It is also accepted across the field that these indicators alone are insufficient in accurately predicting academic success.

While psychological traits are put forward as indicators of student quality, not many studies in academia have focused on its predictive power in isolation nor on the benefits it can add to a model including the other indicators mentioned above.

This project shows, for a relatively small dataset, that it is possible to build a classification model based on the integration of psychological characteristics with GPA, age and time since course completion, that outperforms a model based solely on the traditional classifiers identified as control throughout this project.

We can also conclude that neural networks and support vector machines were not effective models to predict GPA taking into consideration the number of features and the small dataset presented. These two models consistently underperformed decision forests. That said, it is likely that these two models achieve better performance when fed with a larger dataset. However, the set of conclusions obtained in this exploratory work are still relevant to develop the knowledge on how qualitative psychological measures of performance can be merged with quantitative student performance indicators.

In short, in this study a decision forest was created, capable of identifying, with 71,4% accuracy the top 44% of students in a given dataset based on the combination features mentioned above. The goal of providing a complementary metric, developed as a forecaster of student performance, is met with the decision forest model depicted in experience 1. This is the most accurate model developed in this study and, as expected, outperforms the models calculated with no psychological variables used as inputs.

Another proposed goal was to identify the main psychologic traits responsible for higher academic performance. As per experience 1, these are: *Help others acquire new skills and knowledge*, *Help others make decisions* and *Lead projects with societal impact*.

5.2. Future work

Confirming the model creation to be possible and fairly accurate, and knowing it increases performance, opens the doors to more research, practical testing, and deployments.

Starting with research, it is important to reproduce the experiment with a more sizeable dataset and compare the accuracy of the model with a human being. Understanding the performance against a professional is vital to access the usefulness of the classification.

Ideally, this testing will pass by running a wave of applicants through the model and creating a dataset with the predicted score classification while simultaneously running a normal admissions process with human caretakers. After the first year passes it is possible to compare the student GPA with the classification results from admission to understand if the model accurately identified the top students, using this instance as an opportunity to improve the algorithm.

Consequently, a decision can be made to include the classification model as part of the admission process to automate some processes or provide indicative recommendations to the humans in charge.

An unsupervised machine learning model can be created on top of the data collected for the present study to cluster different psychological profiles and understand how they perform in school. This can lead to a better understanding of how particular traits, together, might lead to an increase or decrease in performance.

When adopting machine learning to life altering situations such as a job admissions process for an applicant or a university admission for a student, a look into bias, fairness, and discrimination is especially important. Ensuring the model does not discriminate against demographics is imperative prior to its deployment. To do so it is necessary to collect those features from the participants for testing purposes. This was not done in this project as it was focused on showing the relevance of psychological traits to the admissions process of a university.

In fact, the scope of the project in terms of areas of study was limited to the *same_course* metric. This leaves the door open to deeper studies on how different personality traits might have different impacts in different courses.

6. Bibliography

Beecher, M. and Fischer, L., 1999. High School Courses and Scores as Predictors of College Success.”. Journal of College Admission, pp.4-9.

Wendy, C., 1996. A Focus on Process Improves Problem-Based Learning Outcomes in Large Classes. American Journal of Pharmaceutical Education, pp.408–16.

Berger, J. and Sireci, S., 2002. STEMTEC Evaluation Report For Year 5 (Fall 2001/Spring 2002). STEM Education Institute.

Epstein, J., 2010. Behind the SAT-Optional Movement: Context and Controversy. Journal of College Admission.

Nora, A., Cabrera, A., Serra Hagedorn, L. and Pascarella, E., 1996. Differential impacts of academic and social experiences on college-related behavioral outcomes across different ethnic and gender groups at four-year institutions. Research in Higher Education, 37(4), pp.427-451.

Mulvenon, S., Stegman, C., Thorn, A. and Thomas, S., 1999. Selection for College Admission: Refining Traditional Models. Journal of College Admission, (n162), pp.p20-27.

Pribbenow, C., Phelps, L., Briggs, D. and Stern, D., 1999. New College Admission Procedures: Implications for Career-Related Learning in High School. Western Illinois University.

Tinto, V., 2012. Enhancing student success: Taking the classroom success seriously. The International Journal of the First Year in Higher Education, 3(1).

Abrica, E., 2018. How to Measure Student Success? Toward Consideration of Student Resilience as a Metric of Success in Institutional Accountability Frameworks. Community College Journal of Research and Practice, 42(7-8), pp.569-573.

Mullin, C., 2012. Student Success. Community College Review, 40(2), pp.126-144

Mould, T. and B. DeLoach, S., 2017. Moving Beyond GPA: Alternative Measures of Success and Predictive Factors in Honors Programs. Journal of the National Collegiate Honors Council

Lee, H. and Lee, J., 2012. Who gets the best grades at top universities? An exploratory analysis of institution-wide interviews with the highest achievers at a top Korean University. Asia Pacific Education Review, 13(4), pp.665-676

Slim, A., Heileman, G., Kozlick, J. and Abdallah, C., 2014. Predicting student success based on prior performance. 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM),

Abele, A. and Spurk, D., 2009. How do objective and subjective career success interrelate over time?. Journal of Occupational and Organizational Psychology, 82(4), pp.803-824.

POWER, C., ROBERTSON, F., & BAKER, M. (1987). Success in higher education. Canberra: Australian Government Publishing Service.

PASCOE, R., MCCLELLAND, A., & MCGAW, B. (1997). Perspectives on selection methods for entry into higher education in Australia. Canberra: Australian Government Publishing Service.

R.Jensen, A., 1998. The g factor: the science of mental ability. Choice Reviews Online, 36(04), pp.36-2443-36-2443.

Frey, M. and Detterman, D., 2004. Scholastic Assessment or g?. Psychological Science, 15(6), pp.373-378.

Rohde, T. and Thompson, L., 2007. Predicting academic achievement with cognitive ability. Intelligence, 35(1), pp.83-92.

PANTAGES, T. J., & CREEDON, C. F. (1975). Studies of college attrition: 1950–1975. Review of Educational Research, 48, 49–101.

ABBOTT-CHAPMAN, J., HUGHES, P., & WYLD, C. (1992). Monitoring student progress: A framework for improving student performance and reducing attrition in higher education. Hobart: National Clearinghouse for Youth Studies.

TERENZINI, P. T., & PASCARELLA, E. T. (1978). The relation of students' precollege characteristics and freshman experience to voluntary attrition. Research in Higher Education, 9, 347–366.

RICKINSON, B., & RUTHERFORD, D. (1995). Increasing undergraduate student retention rates. British Journal of Guidance and Counselling, 23, 161–171.

LECOMPTE, D., KAUFMAN, L., & ROUSSEEUW, P. (1983). Search for the relationship between interrupted university attendance of first year students and some psychosocial factors. Acta Psychiatrica Belgica, 83, 609–617

HIMELSTEIN, H. C. (1992). Early identification of high-risk students: Using non-cognitive indicators. Journal of College Student Development, 33, 89–90.

CLARK, E. E., & RAMSAY, W. (1990). Problems of retention in tertiary education. Education Research and Perspectives, 17, 47–57.

MCINNES, C., JAMES, R., & MCNAUGHT, C. (1995). Diversity in the initial experiences of Australian undergraduates. Committee for the Advancement of University Teaching (On-line). Available: www.online.anu.edu.au/caut/com … e/FYEfront.html.

Costa, P. and McCrae, R., 1992. NEO PI-R: Professional manual: Revised NEO PI-R and NEO-FFI. Psychological Assessment Resources, Inc, Florida,.

Komarraju, M., Karau, S. and Schmeck, R., 2009. Role of the Big Five personality traits in predicting college students' academic motivation and achievement. Learning and Individual Differences, 19(1), pp.47-52.

Ahmed, A. B. E. D., & Elaraby, I. S. (2014). Data mining: A prediction for Student's performance using classification method. World Journal of Computer Application and Technology, 2, 43–47.

Bhise, R., Thorat, S., & Supekar, A. (2013). Importance of data mining in higher education system. IOSR Journal Of Humanities And Social Science (IOSR-JHSS), 6, 18.

design the future. (2021). Design the Future. Retrieved May 12, 2021, from https://designthefuture.pt/

Burgos, C., Campanario, M. L., Peña, D. D. L., Lara, J. A., Lizcano, D., & Martínez, M. A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. Computers and Electrical Engineering, 66, 541–556.

Abele, A. and Spurk, D., 2009. How do objective and subjective career success interrelate over time? Journal of Occupational and Organizational Psychology, 82(4), pp.803-824.

Hussain, M., Zhu, W., Zhang, W. and Abidi, S., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores. Computational Intelligence and Neuroscience, 2018, pp.1-21.

Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1). London, UK: Springer-Verlag.

Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications–a holistic extension to the CRISP-DM model. *Procedia Cirp*, *79*, 403-408.

Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson Correlation Coefficient. *Noise Reduction in Speech Processing*, 1–4.

Microsoft (2021, November 4). *Filter Based Feature Selection: Component reference - Azure Machine Learning*. Microsoft Docs. https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/filter-based-feature-selection#technical-notes

Microsoft(2021, November 15). *Tune Model Hyperparameters - Azure Machine Learning. Microsoft* Docs. https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/tune-model-hyperparameters

Lundberg, Scott M And Lee, Su-In. (2017). A Unified Approach to Interpreting Model Predictions. Http://Papers.Nips.Cc/Paper/7062-a-Unified-Approach-to-Interpreting-Model-Predictions.Pdf.

7. Annex - Psychologic and Academic Questionnaire

1. I declare that I am 18 or over 18 and agree to participate in this research. I declare that I was informed that my participation in this study is voluntary and that I can leave this survey at any time without penalty, and all data is confidential. I understand that I will evaluate responses and that this study does not offer serious risks. *
   - I agree to participate in this survey
   - I do not agree to participate in this survey
2. What year were you born? ex: 1997 *
3. To which gender do you most identify? *
   - Female
   - Male
   - Gender Variant/Non-Conforming
   - Other
4. Are you currently studying or finished a degree in the last 2 years? ie. Bachelors, Masters, PhD, Post-Graduation, etc. *
   - Yes
   - No
5. Is this course taught at Nova IMS University *
   - Yes
   - No
6. What is the level of your current course? *
   - High School
   - Bachelors
   - Masters
   - PhD
   - Post Graduation
   - Other
7. Which of the areas of study below better suit your current course? *
   Business
   Medicine
   Chemistry
   Biology
   Pharmacology
   Humanities and Social Sciences
   Law
   Physics
   Engineering
   Mathematics
   Computer Science
   Marketing
   Economics
   Finance
   Information Systems
   Information Management
   Geo Informatics

Sports

Other

8. What is your current GPA in a scale of 0-20? *

QUESTIONS ON PREVIOUS ACADEMIC ACHIEVEMENTS

9. Prior to the course you answered about, what was the level of the last course you graduated from? *
   - High School
   - Bachelors
   - Masters
   - PhD
   - Post Graduation
   - Other

10. Which of the areas of study below better suit the past course? *

    Business
    Medicine
    Chemistry
    Biology
    Pharmacology
    Humanities and Social Sciences
    Law
    Physics
    Engineering
    Mathematics
    Computer Science
    Marketing
    Economics
    Finance
    Information Systems
    Information Management
    Geo Informatics
    Sports
    Other

1 1. What was your GPA on that course (0-20) Scale? *

12. How many years ago did you graduate from it? *

Psychologic Assessment

13. From the following activities, which ones do you enjoy partaking or believe will like to experiment in the future.

Control industrial machine usage
Write to a newspaper
Writing romances
Social service activities and volunteering
Talk, debate and publicly present your ideas
Advice companies on financial topics
Plan and project buildings and parks
Study and resolve scientific problems
Choose the art pieces to be displayed in a gallery
Organize events, concerts, group travels and other leisure activities
Plan and establish work related goals
Take care of garden or forests
Evaluate lab samples and experiment
Sing in a choir or band
Heal and treat the sick
Persuade others to modify their behaviour
Manage and organize a company's archives
Work with animals and nature
Perform lab experiments with disease and treatment
Create and design clothes
Listen, advice and support people
Lead projects with societal impact
Manage the work tasks of others
Professional athleticism
Develop mobile applications and other tech
Paint, design and sculpt
Help others making decisions
Creating marketing campaigns
Developing efficient work methods
Partake in defence and patriotic activities
Talk fluently about any subject matter
Play a musical instrument
Solve problems related to people's lives
Lead people and groups
Perform economic analysis of national and international topics
Participate and compete in sporting events
Dedicate high amounts of time to reading
Develop activities related to theatre and dramatic expression
Help others acquiring new skills and knowledge
Lead the activities of a company
Plan budgets

14. From the following characteristics, select the ones you believe you possess and the ones you identify the closest to.

    I actively participate in conversations and debates
    I am a good storyteller
    I listen closely to words, expressions and information and try to use them when I speak
    I can write letters, histories and transmit adequately ideas, emotions, and experiences
    I am good at handling materials when the goal is to count, compare, order, weigh, and measure
    I interpret and identify numbers in multiple contexts (transports, dates, prices, discounts, locations, schedules, economic information, historical information...)
    I solve riddles and mathematical quizzes
    I solve problematic situations and am fast at mind calculations
    I remember faces, colors, sizes, objects, and scenarios
    I analyze art considering quality criteria such as perspective, proportion, use of color, sizes, etc.
    I correctly interpret maps and plans
    I use illustrations to help remembering information
    I easily copy other people's expressions and movements
    I assemble and disassemble objects with ease
    I efficiently communicate with gestures, movements, ideas, questions, and emotions
    In games, sports, and dances I am able to perform the adequate movements
    I can reproduce and remember with ease any melody I listened to
    I sing and move with spontaneity whilst I realize a plethora of tasks
    With instruments and my voice, I am able to interpret melodies and songs, with emphasis in execution, lyrics and rhythm
    I identify musical instruments and can create new ones out of ordinary objects
    I mingle with ease in situations that allow me to meet and relate with others
    I am, generally, seen as a natural leader
    The others show interest and desire my company
    I can put myself in another one's shoes and can understand what they like, how they think and what they feel
    I have moments of reflection on what I do, say or feel
    I recognize emotions that I feel and their reasons
    I recognize my qualities and flaws
    I listen, with interest, to recommendations to improve my skills and fix my flaws