Predicting Final Grades Using Machine Learning: White Paper on the Introduction of an AI-Based Score-Predicting Feature for Educational Institutions.

Executive Summary

This white paper aims to introduce an algorithmic score-predictor designed for educational institutions worldwide to address common challenges faced by them when it comes to traditional grade predictions. These include, but are not limited to, inherent bias, favoritism, and inaccurate predictions. The feature proposed is designed to minimize bias and favoritism by predicting grades objectively and accurately through the use of modern statistical and machine learning algorithms. As a result, the feature will serve as an aid to teachers and enhance the school's image as a technology-savvy institution.

Introduction

Predicted grades have always been a source of controversy across major educational institutions worldwide. In a traditional sense, they are usually issued by teachers in schools to not only gauge the student's level of academic preparedness for their final exams but also as a proxy for those final examination grades when applying for colleges and universities. This is especially relevant for educational systems that heavily rely on final examinations to assess students, such as the International Baccalaureate, A-Levels, and the majority of European baccalaureates (including the Spanish Baccalaureate, the Italian Maturita, and the actual European Baccalaureate). While there are certain groups of individuals and organizations that vehemently advocate for the application of predicted grades across a major scale, other groups beg to differ and instead claim that predicted grades can have detrimental effects that ultimately lead to more harm than good for the students. To elaborate, these polarizing points of view create an issue where on one side we have groups arguing that predicted grades create a solid basis for students when applying for colleges, whereas on the other side we have groups arguing that predicted grades ultimately hurt the student due to several issues such as inherent biases and inaccuracy.

Problem Definition

The key challenges that lie within predicted grades are inherent biases, inaccuracy, generosity, and guesswork. These challenges are not only based on anecdotal experiences from a wide variety of educators across the world - such as from this opinion piece written by a teacher based in the UK - but also on concrete data. In fact, according to statistics pulled from the Universities and Colleges Admissions Service (UCAS) in 2019, only 21% of 18-year-olds met or exceeded their predicted A-level grades, signaling the presence of generous grade predictions and inherent inaccuracies. This trend is not only present in 2019, but also across previous years. This is exemplified according to a 2016 report presented by the University and College Union in the UK, where the report highlighted that “only 16% of applicants achieved the A-level grade points that they were predicted to achieve, based on their best three A-levels”, with “the vast majority (75% of applicants) being over-predicted”. We posit that intelligent, objective solutions based on a data-driven and algorithmic approach can be a great way to address the aforementioned challenges, thereby enabling teachers to make more informed, and objective predictions.

Problem Solution (High-Level Overview)

As Focus Project, we propose the following solution to address the aforementioned challenges faced by educational institutions worldwide: An algorithmic score-predicting feature embedded within the core Focus Project product that predicts a student’s final grade based on past student data - including past grades, demographic, social, and school-related factors. Using past data, the statistical model that this predictor is based on will predict the score of the individual student with a high level of confidence. Teachers will have the option to calibrate which kind of scores they would like to be predicted. For instance, if teachers want a tool that will help them predict if a certain student will simply pass or fail, the score-predictor can easily be modified such that it predicts a PASS/FAIL output for a certain student given the student’s inputted data (e.g. their past grades). Likewise, the score-predictor can also be modified to provide more specific grades, such as grades corresponding to the ECTS grading system - a widely used grading system between educational systems across Europe. Before we dive into the details as to how this model exactly works, we would like to reiterate that the overall objective of the implementation of this feature is not to replace teachers, but rather act as an aide to them when it comes to predicting scores. This will greatly simplify the score-prediction process and eliminate major issues present such as bias. Furthermore, an algorithmic score-predictor may not always predict the final score with perfect accuracy, but rather it is designed to predict the final score with much higher accuracy when compared to human-predicted grades. The best solution is **a combination of an algorithmic score-predictor and a human teacher to later verify if the predicted score by the model is indeed commensurate with the student’s potential to achieve a certain grade.**

Problem Solution (Details)

To verify the applicability of this solution and see whether an algorithmic grade-predictor can yield accurate results, we built a basic prototype version of a score-predictor using various Machine Learning algorithms programmed in the language of Python. This

prototype model was based on a real-life dataset that was constructed and compiled by two Portuguese researchers - Paulo Cortez and Alice Silva (see references for more information). The data was collected via student reports and questionnaires compiled from two Portuguese secondary schools. The variables in the dataset can be divided into two main categories - independent and dependent. The independent variables refer to the predictors - features which we used to predict the final grade, and the dependent variable refers to the final grade itself (the variable we were trying to predict). Specifically, in this example, the final grades correspond to student performance in the following two subjects: Mathematics and Portuguese. The independent variables included features such as past student grades (in Mathematics and Portuguese), number of absences, and other academic-related factors (we excluded demographic data to avoid inherent bias). While the dataset was indeed extremely robust in its features, we found out through our analysis that the final grades achieved by students were highly influenced by their **past grades**, although other factors such as the number of absences did play a minor role when it came to predicting the score. Before we implemented any Machine Learning algorithms, we needed to perform a few key steps. We first encoded all of our categorical variables since Machine Learning algorithms require both input and output variables to be numbers. After encoding our categorical variables, we then split the dataset into two parts: the training set (containing roughly 80% of the data) and the test set (containing roughly 20% of the data). The training set is where the “Learning” in
Machine Learning takes place - where the given algorithm is applied to the training data. Once our algorithm is trained, we then model it on new, unseen data, which in this case is the test set. Once we pre-processed the data, we then proceeded to the modeling stage - where the actual Machine Learning happens.

The data were modeled under the following tasks:

1. Binary Classification: Predict PASS/FAIL
2. Five-Level Classification based on the letter grade: Predict from A - very good to F - insufficient.
3. Regression: Predict a numeric output of the final grade from zero (0%) to twenty (100%)

For the classification tasks, we used the following algorithms: Decision-Tree Classification, Random-Forest Classification, and Gradient Boosted Trees Classification. On the other hand, for the regression task, we used a Linear Regression model. While the purpose of this white paper is not to dive too deep into the technical aspects of how each algorithm works, it can be useful for the reader to provide a very high-level overview of each algorithm to further understand how the prototype score-predictor works.

First of all, regarding Linear Regression, we can best describe it as being based on a linear approach of modeling the relationship between a dependent variable (the final grade) and a set of independent variables (the explanatory variables such as previous grades, number of absences, etc). Specifically, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated using data. Decision Tree Algorithms are based on a Decision Tree branching structure, which represents a set of 'IF-THEN' rules. At each step, the algorithm verifies if a certain criterion is passed, and if it is passed, it performs a certain operation, with the end goal of predicting a certain outcome (e.g. PASS/FAIL). A Random Forest is essentially an ensemble of multiple Decision Trees. It establishes the outcome based on the predictions of these decision trees. Specifically, it predicts by taking the average or mean of the output from various trees. Gradient Boosted Trees act similarly, in the sense that they are based on an ensemble of decision trees. However, what distinguishes them is the fact that the learner trees are trained sequentially where the "weak" learner trees are combined in some way to get a "stronger" classifier (with a higher predictive power).

In our prototype, we implemented all of the aforementioned algorithms and tested their performance using certain metrics. For our Linear Regression model, we used the coefficient of determination (which determines how well the regression predictions of our final scores approximate the real data) as a performance metric and found out that it was approximately 0.85 - with 0 implying no predictive power and 1 implying the best possible score. This meant that our model indeed had strong predictive power and that our predictions of the students' final scores well approximated the actual final scores. For the classification algorithms (Decision Tree, Random Forest, and Gradient Boosted Tree), we used accuracy as our scoring metric. The accuracy, in this case, refers to the percentage of correct classifications. All three algorithms had similar accuracies after being modeled on the data. For binary level classification (predicting whether the student passed or failed), we found that all three algorithms had an accuracy of about ~92%, whereas, for the five-level classification (predicting the student's grade on a letter scale), all three algorithms had an accuracy ranging from 73% - 79%. In both instances, the Gradient Boosted Trees algorithm outperformed the Decision Tree and Random Forest algorithms, albeit by small margins.

Business Benefits

The implementation of such a feature within the Focus Product has numerous effects that can be greatly advantageous from a strategic point of view for an educational institution such as a school. Firstly, as pointed out in previous sections, the score-predictor can be a great aid for teachers when it comes to predicting grades. An algorithmic score-predictor will reduce the overall workload for teachers, especially in classrooms with a large number of students. As a result, this will allow them to optimize their time by focusing on other tasks, allowing them to be more efficient which will ultimately prove to be beneficial for the students and the school in general. Secondly, an algorithmic score predictor will also instill a greater sense of confidence amongst parents who may be worried about possible unfair bias when it comes to their child's predicted grades. The score-predictor is designed to be as objective as possible and takes into account a multitude of factors that may impact a student's grade, with the highest focus on their past performance. While the score-predictor may not fully eliminate bias (there is an argument that machine learning models can never truly eliminate bias since after all they are based on real-life data, which may be slightly biased to an extent), it will drastically reduce it, thereby leading to an improvement in decision making, transparency, and accountability. Furthermore, a score-predictor based on an algorithmic approach can also enhance the school's image as a more data-driven, technologically advanced institution. A school that embraces technology can further attract the next generation of parents, who are mainly millennials - a generation known for being technology-savvy and increasingly data-centric. Lastly, while an algorithmic score-predictor has numerous advantages for teachers, it can also be greatly advantageous for the students involved. Students can feel greater at ease knowing that their predicted grades, which are crucial for their college applications, are mostly devoid of bias and favoritism. A happy student population is indeed crucial from a strategic point of view for a school.

Summary

In summary, this score-predicting feature is designed to be an aide for teachers that will allow them to make smarter, more objective predictions. It is not designed to replace teachers, but rather complement teachers when it comes to predicting final grades. Ultimately, the teacher will still be the final decision maker. Furthermore, its main advantages include drastically reducing bias and favoritism, enhancing the school's image as a data-driven and technologically enhanced institution, reducing the workload for teachers, attracting a more technology-savvy generation of parents, and increasing confidence amongst the student population.

Call to Action

To find out more about how education institutions can gain objective, smart predictions from our data-driven algorithmic score predictor, contact us at +34 648985172 or +91 6355396358. You can also email us at contact@focusproject.es or request a free demo by visiting https://focusproject.es/.

References

K. Lancaster "An Argument in Favour of Results by Algorithm." *Philosophy of Education Society of Great Britain*, 1 Mar. 2021, www.philosophy-of-education.org/an-argument-in-favour-of-results-by-algorithm/.

P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7.