Eötvös Loránd University

Faculty of Information Technology

Software Engineering Project Report

Student's Performance Prediction

Prepared by:

Jaber Seif

Charra Zakaria

El Mahdaouy Zahira

Olga Strukova

Tshifaro Justin

Mentor:

Yuping Yan

Oct. 2021

Problem

Many students can't seem to identify their weaknesses when they are studying, they can't really determine why they are not good at some subjects, is it because they don't understand it? Or they don't understand the pre-requisites? most of them seek guidance from their Academic mentor, and others just shy away from asking for help. The issue with Academic mentors is that they don't always know the student's history grades, which is why the advice will almost always be generic. Even if the Mentor took some time to understand the student's grade history, he/she is still not able to match this student's case with other previous similar students to come up with an accurate advice based on past experiences. Universities will always look for ways to improve their student's performance as it is essential for their ranking and their revenue, but they don't always come up with successful plans regarding improving the performance.

Solution

We should utilize the existing student's data that universities are storing, we can apply Machine Learning to their data and start predicting each student's future performance, this way the student can see his/her predicted grades for the upcoming semester and start preparing before it's too late. Such software might eliminate the need of an Academic Mentor, but even if a student still needed a mentor for guidance, the mentor will then be able to understand the student's weaknesses and provide an accurate advice. The predicted grade will be based on the student's previous grades with the pre-requisites having the highest effect.

Business Ideas

The main target of this project are students. Therefore, the software can be used by different universities. Revenue can be generated through subscriptions, it could be monthly or yearly subscriptions, or better yet it could be per semester. The revenue can also be generated by using this software to provide a service for universities, we can provide reports and insights to universities about at-risk students and how to help them and charge them for these insights.

Target Users
- Students
- Universities

Planned way of working

We need to be as flexible as possible because of the short time frame that we have for completing this project, we are mainly going to work according to a pattern with some aspects and practices of agile methodologies (**eXtreme Programming**) instead of working according to a whole methodology.
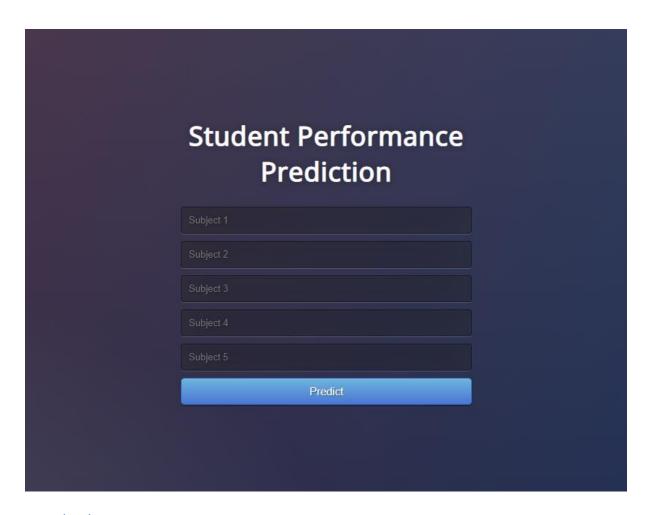
- **Behaviour Driven Development** to increase the productivity of the team, every team member will work on what he knows best and make his own decisions, the data scientist can decide what models and which algorithm to use for the data filtering and cleaning, and the machine learning expert will decide how and which models to use in order to obtain the most accurate results.

- **Shared Responsibility:** all team members are collective owners of the source code Information and Knowledge will be sheared at all costs, and since the team members are from different backgrounds thus the written code should follow certain guidelines in order to be understandable and less time consuming.

- **Extensive Code Reviews** so the information will be shared among all the team members and having the possibility of spotting problems. We will work with google Collab so all the team members will keep track of the code and give feedback.

- **Continuous Integration:** each team member will try to integrate his code after succeeding the tests with the other's code.

- **XP Testing (Test Driven Development)** to guarantee that we are not over working, and we are doing just what we need to achieve our solution.

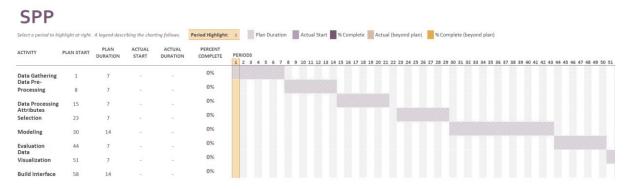We also might use pair programming at some point in the project.

Development

We will develop the solution using Python, we have not yet determined what machine learning model we will be using, we do know we will be using more than one model to test the accuracy, and we will eventually apply the model with the best accuracy. We will eventually use Power BI for data visualizations, we will also be using excel as a data source. The predicted results should be based on the student's previous grades and the similarity between students, the algorithm should be able to detect which previous students are like the current student and use their data to predict the current student marks. Once we develop the prediction model, we will build a user interface using Flask framework for a better usability. At first the student will have to sign in and then can enter his/her data for prediction. We will use the Sign In as our form of security. The prediction interface could look like the following image but will be even better. The student will enter the courses he/she wishes to register for, and the results will be predicted.

Work Plan

Scheduled Deadlines

ACTIVITY	PLAN START	PLAN DURATION	ACTUAL START	ACTUAL DURATION	PERCENT COMPLETE
Data Gathering	1	7	-	-	0%
Data Pre-Processing	8	7	-	-	0%
Data Processing	15	7	-	-	0%
Attributes Selection	23	7	-	-	0%
Modeling	30	14	-	-	0%
Evaluation	44	7	-	-	0%
Data Visualization	51	7	-	-	0%
Build Interface	58	14	-	-	0%

Details

1. Data Gathering: we will kick off this project by gathering our data, and this could be tricky to find a suitable data.
2. Data Pre-Processing: in this step, we will perform a Data Cleaning, and Missing Values Imputation.
3. Data Processing: we will perform feature encoding and Normalization, the main goal of this process is to have a data that is ready for modeling.

4. Attributes Selection: this process is connected to the previous step, its good to have a connected process just in case we wanted more time with the data processing. However, the main goal here is to determine which attributes should be selected for modeling, what are the dependent and independent variables.
5. Modeling: we will try multiple models, eventually we will choose the model with the best accuracy.
6. Evaluation: this is connected to the previous step, the main target here is achieve a good and accepted result.
7. Data Visualization: we will visualize the predicted data for students. We will also provide some insightful statistics to the university decision makers.
8. Build Interface: we will build an interface for students to interact with and will be used as a form of security.

Individual Efforts

Individual efforts will be determined for each step at a time. When we work on Data Gathering step, we will assign each other tasks and then combine our work.

Phases

1. Planning
   - The team creates a project plan based on the required features
2. Designing
   - Creation of a simple design and structuring of the system
   - Sharing responsibilities
3. Coding
   - Extensive Code Reviews
   - Continuous integration and refactoring
4. Testing
   - Unit tests
   - Acceptance tests with customer
5. Feedback listening
   - Improve our work based on the Mentor's feedback

Milestones

1. Data Gathered
2. Data Processed
3. Model Applied
4. Results are accepted
5. Data is visualized
6. Interface is built