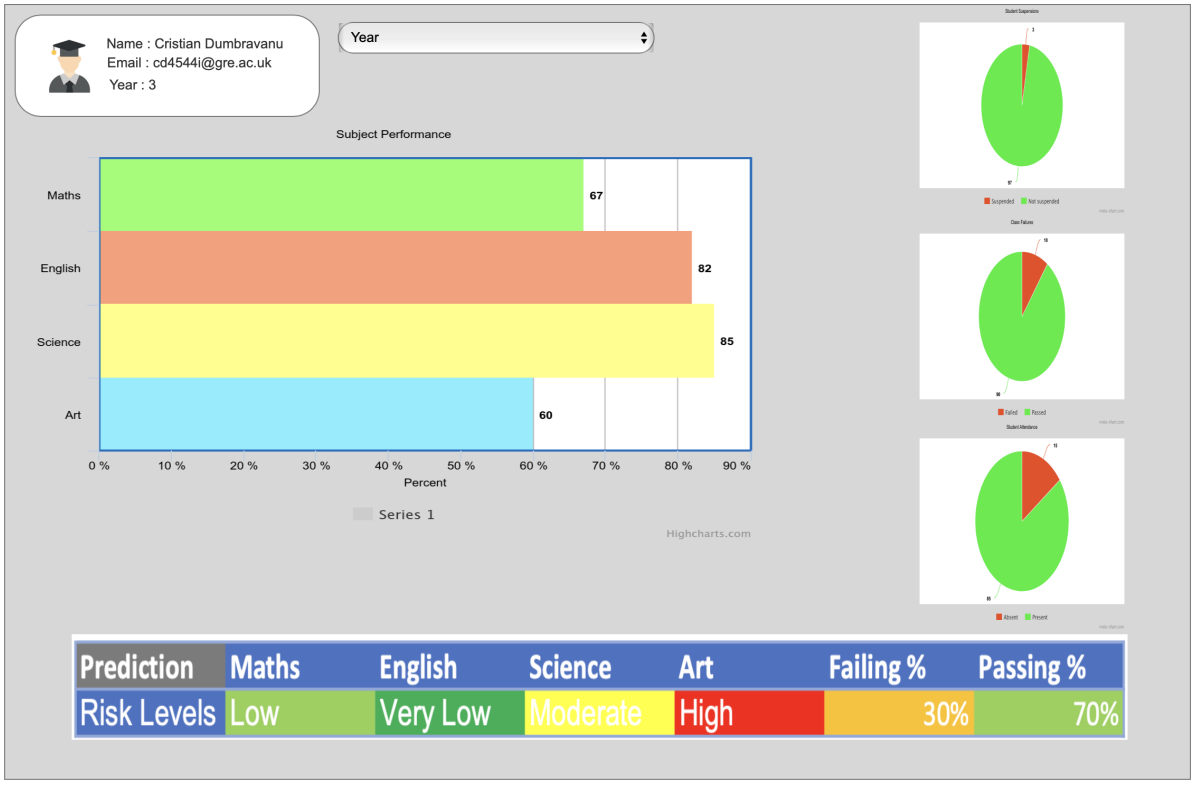
Roadmap

▼ How the project will look like :

👤 Inspiration

First Draft :



▼ How does it work :

The project will be used by : teachers, students, parents and academic institutes.

1. Teachers / Lecturers :

- They can use the models to identify students that are risk of failing behind in their academic career. They can then prevent this by providing a specific intervention in the form of extra tutoring, advising and developing a learning plan that is optimised to mitigate the risk of failure.
- The models can help develop personal instructions for each student. The model can help find strengths and weaknesses, and knowing those weaknesses it can allow the teacher to communicate the student on how to specifically improve on those weaknesses.
- The teachers then will understand how to allocate their time more efficiently on students that need it the most. For instance, if the model identifies a group of students who are likely to excel in a particular subject, the teacher can focus on challenging them with advanced material.

2. Students :

- Getting an insight into academic performance can allow for self awareness which presents potential areas for improvement by motivating the student to set goals, seek the necessary help and to develop necessary study habits.
- Awareness of potential academic performance allows for course planning by making informed decisions about course selection and potential study habits.
- If a student is aware of their strengths and weaknesses they can allocate time appropriate and effectively.

3. Parents :

- It can also help parents understand their kids strengths and weaknesses to provide the support they need efficiently for example specific tutoring sessions or extracurricular activities that work on the weaknesses.
- The insight into student performance can allow for meaningful conversations with teachers. This can encourage strategic communication that helps the educational journey.
- Having an understanding of how your kid performs can enable the parents to take into account their mental health.

4. Academic Institutions :

- Universities and schools can use the models to allocate resources efficiently whether its teaching staff, educational resources, infrastructure or support services.
- It can enable institutes to developed a extra curriculum activities to maximise learning outcomes of students by targeting the most common areas of weakness between students.
- An understanding of how students perform can enable to change the strategic approach of teaching a course. An example could be introducing new academic support programs, mentoring, the rate at which the information is presents or the ratio between assignments and teaching intensity.

▼ How will it be demonstrated :

How I will showcase the functionality and effectiveness of project.

1. Presentation where I will explain the objectives of project, the algorithmic methodologies used, database, database preprocessing, metrics, UI, its functionality, metrics, results and discussion.
2. Live demonstration where the Machine Learning model is running and the results.
3. Graphs and charts to indicate key patterns and findings with the code and data to make the results and conclusions more understandable to the teachers and potential future users.
4. The evaluation metrics that will be used to evaluate and compare the models are :

- Confusion Matrix
  - True Positives
  - True Negatives
  - False Positives
  - False Negatives
- Accuracy

- Precision
- Recall
- F1 Measures
- Area under the precision recall curve

5. A user interface prototype to ensure the project is user friendly and intuitive to use while also being able to present the results of the ML models in a user friendly manner. A prototype can also be used as a demonstration and a proof of concept of the functionality and capabilities of the being able to predict the student performance.

▼ Brain Storm of project :

1. Problem and objectives : The objective is to predict student performance by finding patterns and understanding the factors influencing academic success and prevent student failing or dropping out.
2. Data : the data will be open public academic data which will contain information about student demographic : age and sex, attendance records, grades in : exams, quizes, assignments and homework, any extra curricular activities, socioeconomical position.
3. Data Preprocessing : fill in missing values (if any), handle outliers or inconsistencies in the database. Feature scaling : normalisation (fit the data into a specific range between 0 to 1 to prevent any biasses in datatypes) or standardisation (rescaling to have the mean of the data as 0 and the standard deviation of 1)
4. Testing and Evaluation :

- Unit testing to each individual component to ensure the quality.
- Integration testing of testing the entire system to ensure all components work together seamlessly.
- Performance evaluation using metrics by comparing the predictions to the existing targets of features.

5. Evaluation and discussion of all the models to then select the most optimal and accurate model to ensure the accuracy of predictions.

▼ Ethical, Social, Professional and Legal considerations :

1. Ethical Considerations : there is always a concern about the student privacy, data protection and anonymity when it comes to collecting data to train and test the model and the potential lack of transparency.
    - The data I will be using is open source databases which comply with the regular regulations that prioritises anonymity as no real name, address or other personal data is mentioned.
    - The functionality and the predictions made by the models will be explained to ensure the transparency.
2. Social Consideration :
    - Predicting that a student might fail may created a negative label that can lead to stigmatisation and self droughts. This is why the project aims to handle predictions sensitively in order to prevent any sort of further complication and severity of the risk of failure.
    - There will be no inequality and discrimination because the prediction will be made with fairness on marginalised or disadvantaged students.
3. Professional Consideration :
    - With a well performing model, teacher may follow and rely the program unconsciously to make judgements and impressions of students. The project needs to emphasise that the predictions are only probabilistic and that profesional judgements and intuition should still be considered when allocating any academic resources.
4. Legal Consideration
    - The Machine Learning model application will comply with the relevant regulation laws that govern data privacy, anonymity and discrimination against its users or any of the database records.

▼ Potential Sub Problems at various stages of the project development :

1. Project documentation :

- Choosing relevant database features that I require or any data that I need to require.
- Following a scheduled timeline and not missing on deadlines
- Reaching the target functionality and performance measurements that follow the determined quality plan
- Do literature review on relevant articles, books or reports

2. Database quality and data preprocessing :
    - Selecting relevant features and scaling them correctly to improve the prediction accuracy of the model.
3. Algorithmic model training and development :
    - Choosing the most appropriate supervised Machine Learning algorithm for predicting student performance while addressing any imbalances.
    - Tuning of hyper parameters to maximise the positive measurements and increase the speed.
4. Evaluation and explanation of the models :
    - Have a white box explanation of the models and understand how each model works and how the models arrived to certain predictions.
3. Evaluation of the predictions :
    - Choosing the appropriate metrics and cross validation methods to assess the performance.
4. Social, Legal , Ethical and Bias Considerations :
    - Identify potential biases in data (e.g. race or gender) to ensure fairness in predictions
    - The biggest legal concern would be safeguarding personal and sensitive student information to comply with the privacy regulations and laws of the UK.
    - The biggest social concern would be that the project would encourage academic institutes to make decisions based on the model and ignore

their profesional experience or intuition.

6. Deployment, Implementation and prototypes
    - Ensure that the model can work efficiently with realistic large volumes of data and function in a real environment.
    - Being able to integrate the project into any existing academic platforms or systems (e.g. Moodle or any university portal)
    - Having an intuitive, visually clean and optimised user interface that displays the information produced by the model.

▼ Moving the idea of predicting student performance forward :

DATABASE :

- Most of the already existing models use databases with features that are very general and don’t take into consideration other factors that can influence student’s performance. Most of the features used are :
    - Demographic information : age and gender
    - Academic History : academic performance, grades in previous courses, class attendance
- In order to improve the accuracy of predicting student performance, I believe that there are other factors that can influence the academic performance of a student. I would like to find an open source database with the basic features of demographic information and academic history, and then modify it with additional features of :
    - Engagement in class
    - Participation in extracurricular activities (wether its sports, arts, etc...)
    - Interaction with other students and teachers (asking for help or collaborating)
    - Family structure
    - Parental education
    - Parents job industry

    - Average income
    - Available student resources
    - Health of the student and parents
    - Access to modern technology / internet
    - Amount of time spent studying
    - Discipline
    - Motivation of learning

METHODOLOGY :

- The most commonly used supervised machine learning algorithms for predictive models in previous projects are :
    - Logistic regression (LR)
    - k-nearest neighbour (KNN)
    - support vector machine (SVM)
    - random forest (RF)
    - artificial neural network (ANN)
    - BayesNet
- In my project I would also like to analyse and test :
    - Linear Regression
    - Decision Trees
    - Neural Networks
    - Naive Bayes
        - In addition to the most commonly used models :
            - Random Forest
            - K-nearest neighbours
            - Support Vector Machine
            - Logistic regression

METRICS :

- Most common metrics that are used to assess and compare the models in previous projects are :
    - Accuracy
    - Precision
    - Recall
    - F1 Measure
- In addition to the previously stated metrics, to gain more insightful information on how the models perform and compare to each other I will employ :
    - Mean absolute error
    - Precision Recall Curve
    - Area Under the PRC
    - Confusion Matrix
    - Mean Absolute Error
    - Mean Square Error
    - Root Mean Squared Error
    - R squared
    - Cross Validation score
    - Root Mean Square Logarithmic Error
    - Mean Square Logarithmic Error
    - Mean Bias Deviation