Grade Prediction API

Group name: ShallowMind

Zachary Frank
Department of Computer Science
Hanyang University
Seoul, South Korea
zfrank135@gmail.com

Cédric Chauvet
Department of Information Systems
Hanyang University
Seoul, South Korea
cedricjchauvet@gmail.com

Emyl van der Kooi
Department of Nuclear Engineering
Hanyang University
Seoul, South Korea
emylrs@gmail.com

1. Abstract

With this project we will create an AI that will be able to predict what grade a student will receive in a course. Users will access a website and create a profile using information about themselves that will be considered when predicting a grade. The AI will have a general prediction model for all classes, but will become more accurate over time by giving users the option of submitting their grade anonymously to bolster the size of the training dataset. Over time, each course will have an individual AI model to make predictions with rather than using a general model.

2. Role Assignment

Roles	Name	Task description
User	Zachary Frank	Provide feedback on updates to the project.
Customer	Zachary Frank	Give the software developer and development manager requirements based on user feedback and needs.
Software developer	Emyl van der Kooi	Lead the development and make software design decisions.
Development manager	Cédric Chauvet	Project management and managing customer requirements.

3. Introduction

As students, we wanted to create something that was relevant to us that also made use of artificial intelligence. One of the most known websites among students is Rate My Professor. We wanted to create something similar; however, instead of using a professor rating to determine how well you would perform in a class, you can use our website to determine a grade based on factors such as study time per week, internet access, home location, and more. These predictions can be used by students to help them pick courses where they'll excel and help universities advise and design a curriculum that accommodates their students.

Similar software:

There are many published research papers on the topic of predicting student grades. Nearly all of them are conducted with the purpose of providing insight to educators. *Machine Learning Based Student Grade Prediction: A Case Study* aimed to develop methods of grade prediction in an effort to improve student retention in higher education. If this was in the hands of academic advisors, it would give them an invaluable tool for directing students in their first year who have little guidance. This is an asset both to the universities and the students.

In another research paper, *Using Data Mining to Predict Secondary School Student Performance*, researchers used data from secondary school students in Portugal to learn what factors affect students' performance. As an outcome, this research will allow for more efficient student prediction tools to be developed that will enhance the quality of education and school resource management.

4. Requirements

4.1 Keras Training and Prediction

4.1.1 Initial Training
ID: 101
There must be an initial training using a large dataset to start. Each class will have a separate AI models trained specifically for predictions for that class.

4.1.2 Continued Training
ID: 102
Using new data that the system collects from users, it will continue to evolve and train itself to make predictions using the most recent data.

4.1.3 Predictions
ID: 103
For each class, there must be a way to query the AI model for predictions using their own data.

4.2 User Interface

4.2.1 Login
ID: 201
In order to access the features of the site, students must log in to an account. Login credentials will not compromise the privacy of any users.

4.2.2 Profile Data
ID: 202
Each user will need to submit information about themselves that can be used as input for predictions. This information will be part of the account creation process.

4.2.3 Class Search
ID: 203
After users create an account and give their own data, they will be given access to search for classes and request predictions for their grade. This page should also give the students the opportunity to submit their grade for the course if they have completed it to add data for training and help improve the accuracy of the model.

4.2.4 Profile
ID: 204
Users will be able to access their profile and view their profile data or make changes to it by sending a request to the web server.

4.3 Web Server

4.3.1 AI Model
ID: 301
The client will be able to access the AI model functions via a Web API. The web service should format data for input/output accordingly.

4.3.2 Class Search
ID: 302
The client will be able to query the web service for a class and retrieve the address to access that class' AI model and functions.

4.3.3 User Account
ID: 303
The client will be able to access user account data using correct credentials and read or update fields in the database.

4.4 Database

4.4.1 Users
ID: 401
User information such as account credentials and profile data will be stored in a database. This will make the maintenance of large amounts of user data feasible.

4.4.2 Classes
ID: 402
Each class will have data associated with it to facilitate class searching and giving the client access to a class AI model in minimal time.

4.4.3 Dataset Storage
ID: 403
Datasets for training will be maintained in a database for easy maintenance and quick access.

5. Development Environment

5.1 Platform Choice

This will run in a web environment so that it can be accessed by many users through a web browser. This will also relieve the heavy processing loads from the client machine and put them on the web server.

5.2 Languages

5.2.1 Python 3.6

Provide a language that has access to necessary libraries required to realize all requirements effectively. Will be used to develop web server.

5.2.2 Java 13

Has the capabilities to perform business logic for preparing and processing database resources properly for storage and reading. Will be used to develop database server.

5.3 *Cost Estimation*

Jetbrains All Products Pack $64.90 / user / month
3 users, 2 months
$389.40

MySQL Standard Edition
$2,000.00

Amazon Web Services - EC2 t3a.xlarge
$71.84 / month after first year

5.4 *Development Environment Information*

5.4.1 Django 2.2.6

Create a browser based web application that will provide users with an easy to use user interface and access to backend services.

5.4.2 Maven 3.6.1

Manage package dependencies with database server.

5.4.3 Spring Boot 2.1.9

Create a REST API that will answer requests to the database server.

5.4.4 MySQL 8.0.17

Database management system for the database server.

5.4.5 Apache 2.4.41

Host a server responsible for storing files unrelated to the operation of the web server.

5.4.6 Ubuntu 19.10

Cloud computing platform.

5.4.7 Jetbrains PyCharm 2019.2.3

5.4.8 Jetbrains IntelliJ IDEA 2019.2.3

5.4.9 Keras 2.3.0

A machine learning library that is well maintained and provides a library of tools for training and testing machine learning and AI applications.

5.5 Commercial Cloud Platform

With enough resources, the service will be hosted on an Amazon Web Services server. This will host the web server and database server components.

5.6 Software in use

Using algorithms and techniques laid out by existing research papers such as *Machine Learning Based Student Grade Prediction: A Case Study* and *Using Data Mining to Predict Secondary School Student Performance,* we will be able to develop the system much more quickly. This research will help guide us as we work to build an AI model that is accurate and efficient.

6. Specifications

6.1 Keras Training and Prediction

6.1.1 Initial Training

ID: 101

Using a large dataset from Kaggle, the system will begin training a general model. This model will be used in the absence of sufficient training data for any one class. This model should give an approximate prediction, once trained, for any class based on several variables given by the user. The data points given by the dataset include:

Sex: Male or Female
Age: 15 to 22+
Address: Urban or Rural
Weekly Study Time: 1-2 hours, 2-5 hours, 5-10 hours, or 10+ hours
Failures: 0-4
Extra-Curricular Activities: Yes or No
Internet Access at Home: Yes or No
Romantic Relationship: Yes or No
Free Time: 1 (very low) - 5(very high)
Grade: 0-100

6.1.2 Continued Training

ID: 102

The system should periodically continue training AI models with new data it has collected from users. If a class has been using the general AI model and has recently collected 100+ data points from users, it can train using its own data and create its own model. For classes already using their own model, they will retrain on 1 week cycles if they have acquired 10 or more data points since the last retraining.

6.1.3 Predictions

ID: 103

Each class will be able to predict a student's grade in a course using a student id to reference the database with. Once the student's profile data has been retrieved, it can be fed into the AI model as input. Prediction results should contain predicted grade as well as other metadata on the current class and AI model.

6.2 User Interface

6.2.1 Login

ID: 201

Before a user can access the class search function, they must have an account. The login page will consist of a single line text field for ID and another single line text field for a password. The password field should not show the passwords as users type them. One the correct ID and password combination have been entered, there will be a submit button they may press to authenticate. If the ID and password are correct, the user will be directed to the main page of the site; if they are not correct, it will clear the password field and display a message, "There is no ID and password combination that matches. Please try again." If a user does not have login credentials, there will be a sign up button below the submit button. This button will show a sign up dialogue box that will ask the user for an ID and password combination. Once the user submits this information, the system will validate that the ID does not already exist and save the credentials to the database. If the ID already exists, the dialogue box will be reset and a message, "ID already exists" will be displayed.

6.2.2 Profile Data

ID: 202

Immediately after account creation, users will be directed to a user profile page where there will be several fields they will be required to fill out in order to access the site. These fields will be data points about the user so that predictions can be made. These fields will include:

Sex: Male or Female
Age: 15 to 22+
Address: Urban or Rural
Weekly Study Time: 1-2 hours, 2-5 hours, 5-10 hours, or 10+ hours
Failures: 0-4
Extra-Curricular Activities: Yes or No
Internet Access at Home: Yes or No
Romantic Relationship: Yes or No
Free Time: 1 (very low) - 5(very high)

6.2.3 Class Search

ID: 203

Once users have created an account with the relevant profile data or logged in to an existing account, they will be given a search bar where they may search for a class. Similar classes from different schools will be considered the same class. For example, Calculus from NYU and MIT will be considered the same class. This will allow for larger datasets being used for each class allowing more accurate predictions, generally speaking. Once a user searches for a class, it will list all results of the search in a list. Each element in the list can be used to access a dynamic web page based on the class selected. The page should automatically request a prediction for the students grade upon arrival and display number of data points for the class, average grade across all data entries, and a field to enter a grade if the user has already completed the course. All grade submissions are made anonymously.

6.2.4 Profile

ID: 204

In the top right corner of each page of the site will be a button to access the user's account profile. This page will allow the user to alter their profile data. Profile data will be listed in a label and text field format. At the bottom of the page will be a save button so that the user can save their changes and return to the class search page.

6.3 Web Server

6.3.1 AI Model

ID: 301

Client will be able to access AI Model functions via a REST API. This will mainly include querying the model for predictions. For integration with other backend services, the AI Model must also have an interface that includes functions for training and creation as well.

6.3.2 Class Search

ID: 302

The client will integrate with a class search interface through a REST API that will accept a string input as a search parameter and return a list of search results. It will search through a database of all classes and return the class' data entry in the database. This will include the address that the client can use to access the class' AI model as well as any other relevant metadata.

6.3.3 User Account

ID: 303

Creating, reading, updating, and deletion of user accounts and user account data will be handled by a single user account interface. This will be handled via a REST API that will accept a student id as input and perform functions on that user account. It will access the user database and read or make changes according to the request.

6.4 Database

6.4.1 Users

ID: 401

All user data will be stored in a single table in a database. This will include account login credentials as well as data related to training and predictions. These fields are:

Sex: Male or Female
Age: 15 to 22+
Address: Urban or Rural
Weekly Study Time: 1-2 hours, 2-5 hours, 5-10 hours, or 10+ hours
Failures: 0-4
Extra-Curricular Activities: Yes or No
Internet Access at Home: Yes or No
Romantic Relationship: Yes or No
Free Time: 1 (very low) - 5(very high)

Through an account interface, the client will be able to make REST API requests to create, read, update, and delete entries in the user table. All fields will be required as it is critical to the operation of the system that they all be present for predictions.

6.4.2 Classes

ID: 402

Classes will be stored in a database for easy searching and maintaining. The database will store the class name, a class id, and related table name for this classes dataset. Class ids will be used as a primary key as well as a means to select an AI model for use during predictions.

6.4.3 Dataset Storage

ID: 403

As students submit grades for classes, these grades must be stored for training in the future. These grades will be stored along with the students profile data in a table dedicated for a single class. During training for the class' AI model, it will use this dataset as training and testing data. In addition, the dataset must also reference the student id or user id and create a relation to the users table.