Work Placement Classification of Students using C4.5 algorithm

User Manual

Version 1.0

March 2019

Table of Contents

1. Introduction

Work Placement is a web application that classifies a student's possible type of work suited for them using their college grades. The application provides graphical representation of decision tree based on the training data and then classifies a student work placement based on the generated decision tree. This user manual is intended to provide step-by-step procedures in using the system as well as some troubleshooting for admin users.

2. Overview

Work Placement is a web application that classifies a student's possible type of work suited for them using their college grades. The application provides graphical representation of decision tree based on the training data. It has graphical user interface (GUI) for both admin and client.

The system has 3 steps in using: importing data, generating decision tree, and then classifying students.

1. Importing Data

   There are two types of data that must be imported into the system: the training dataset and the student data. The training dataset contains sample grades and corresponding class. The student data contains their actual grades that will be needed in classifying work placement.

2. Generating Decision Tree

   A training data will be used to generate decision tree. Decision tree will be needed to classify student work placement.

3. Classify Work Placements

   Student data will be used in classifying work placements based on their grades and the generated decision tree.

3. Getting Started

3.1 Set-up Considerations

Work Placement uses 3rd Party applications such as Anaconda 3 (Python 3.7). To fully access the system, the following should be installed on your computer:

1. Python 3.7.1 (Anaconda 3 or newer version).
2. Apache
3. PHP 7.2.4
4. MySQL
5. Sklearn (Scikit-learn)

   To install Sklearn, open the Anaconda Prompt then run the command:

   ```
   pip install scikit-learn
   ```

6. Graphviz

   To install Graphviz, open the Anaconda Prompt then run the command:

   ```
   conda install python-graphviz
   ```

7. Pydot

   To install Pydot, open the Anaconda Prompt then run the command:

   ```
   pip install pydot
   ```

8. Google Chrome Browser

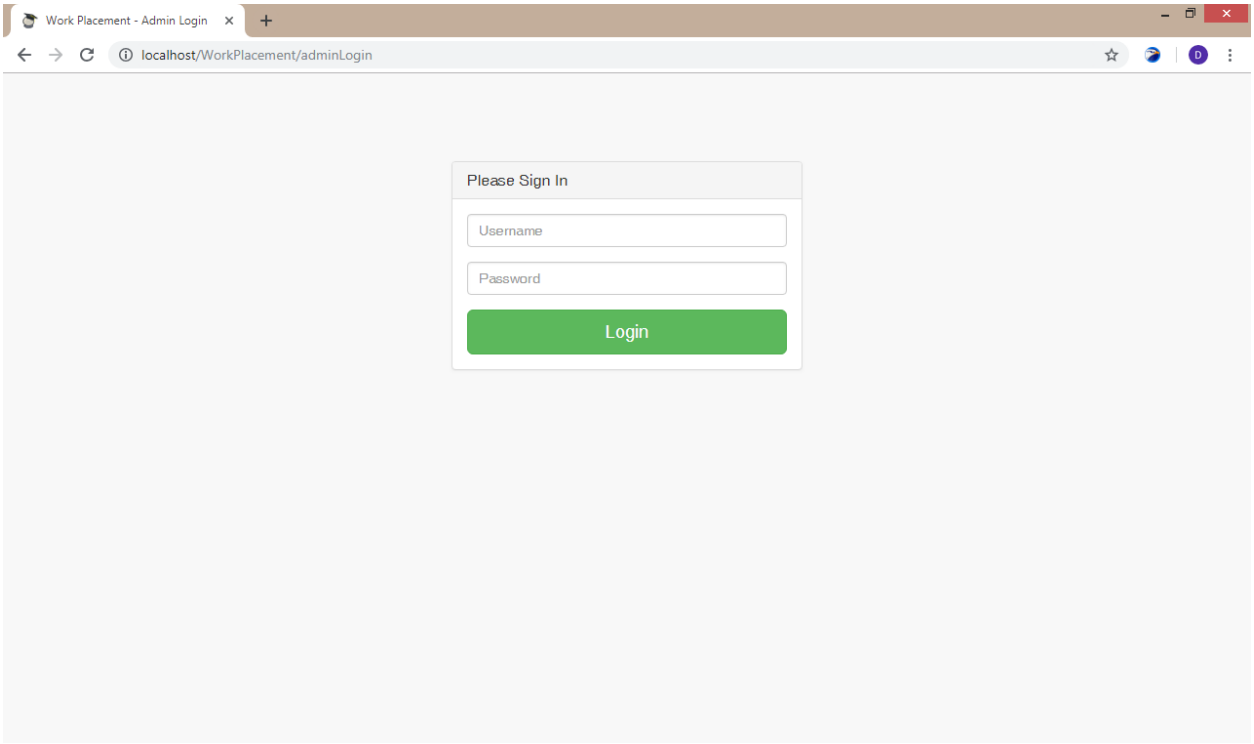
3.2 User Access Considerations

Every student can use application using their ID Numbers and an admin must enter username and password to use the system.

3.3 Accessing the System

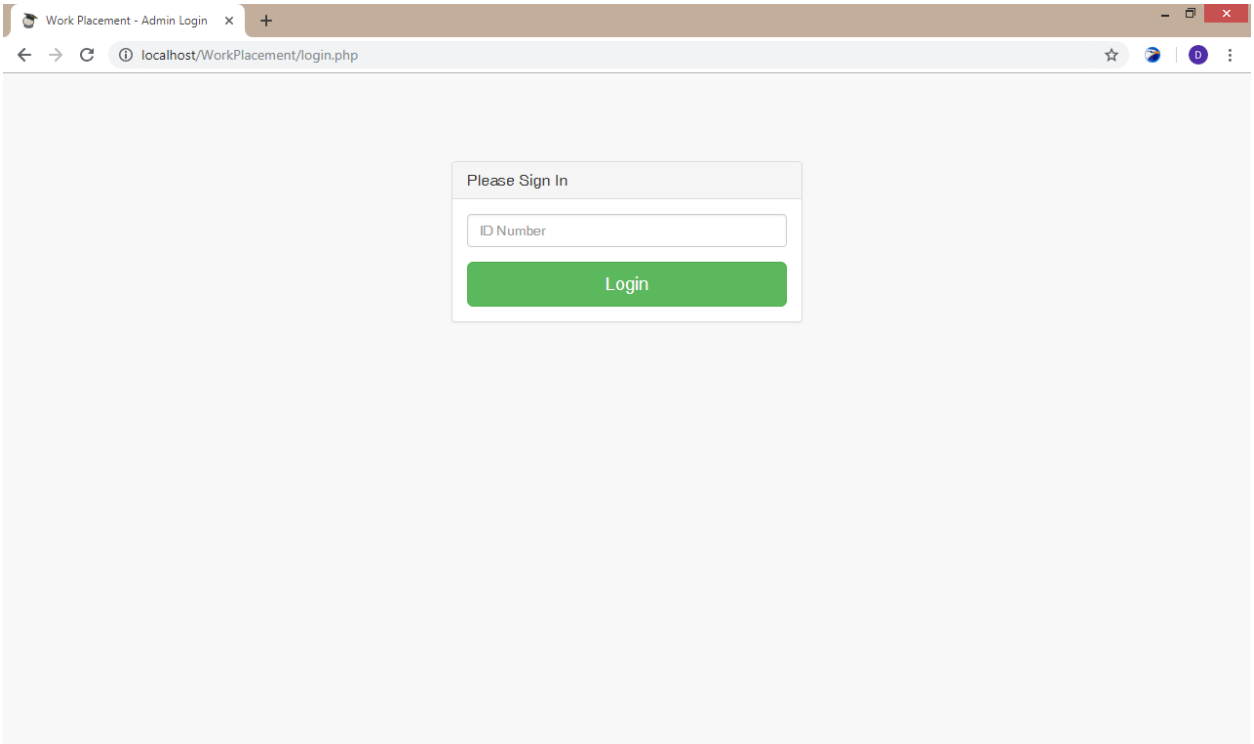
To access the system, an admin account must be created first by a legitimate database administrator in the database “studentgrade” and table “admin_user”. This is to prevent unauthorized registration of admin users.

If an admin account is created, you must type the url “localhost/WorkPlacement/adminLogin” to login as admin, otherwise, you will be redirected into the user’s login page.



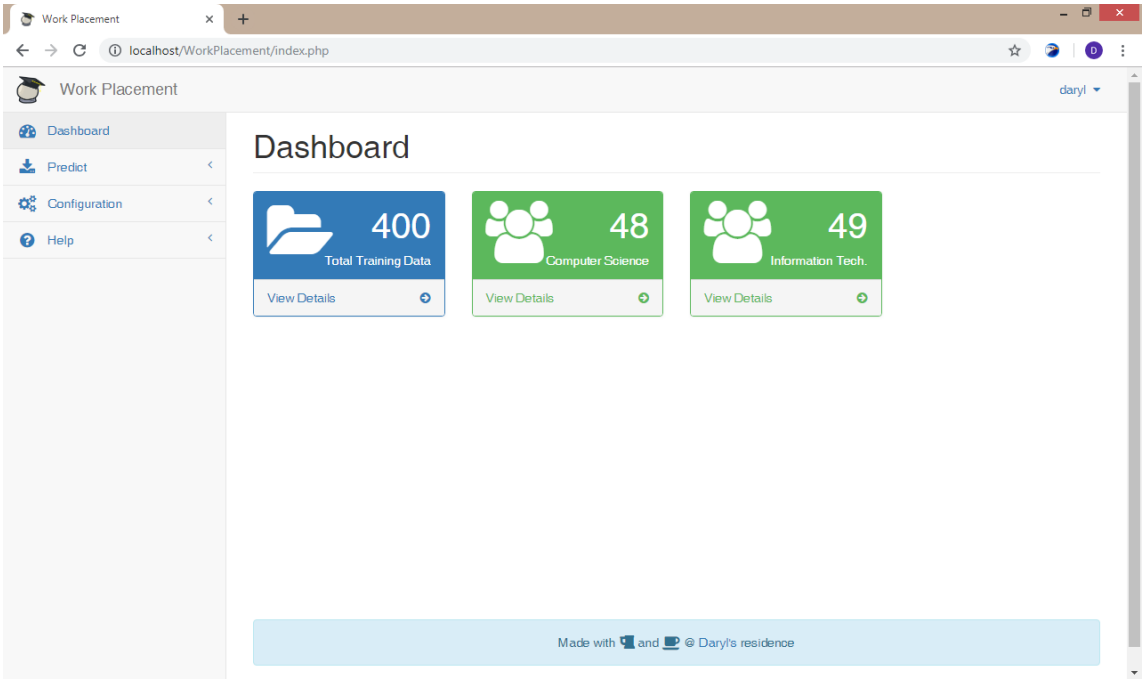
An admin's credentials can only be changed in the database by a database administrator. An admin can add student account by importing student data in the system. A student's credential is only their student number for the system.

A student can access the system by typing the url "localhost/WorkPlacement". He/She will be redirected to the login page.

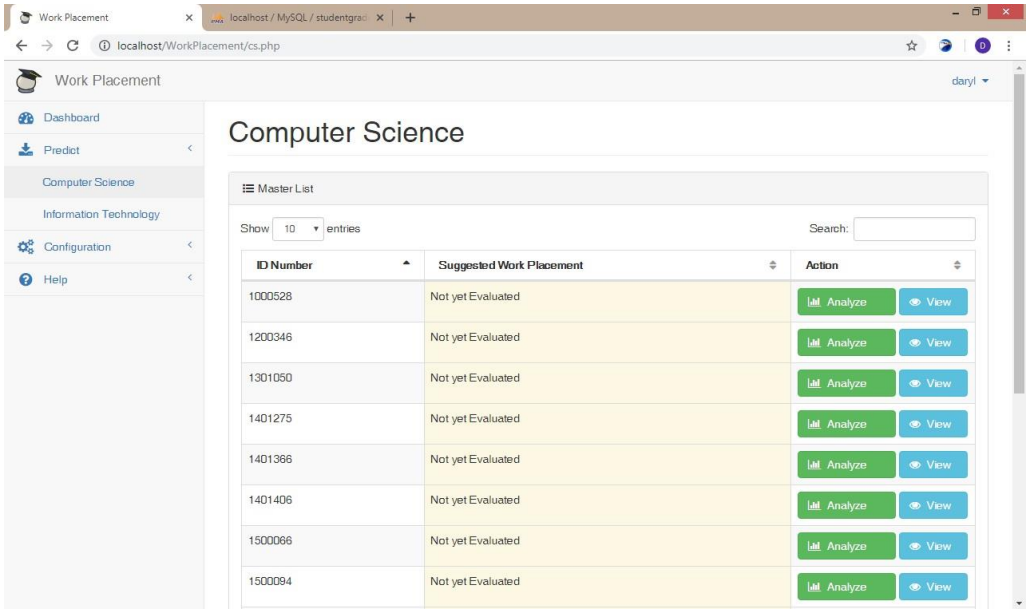


3.4 System Organization & Navigation

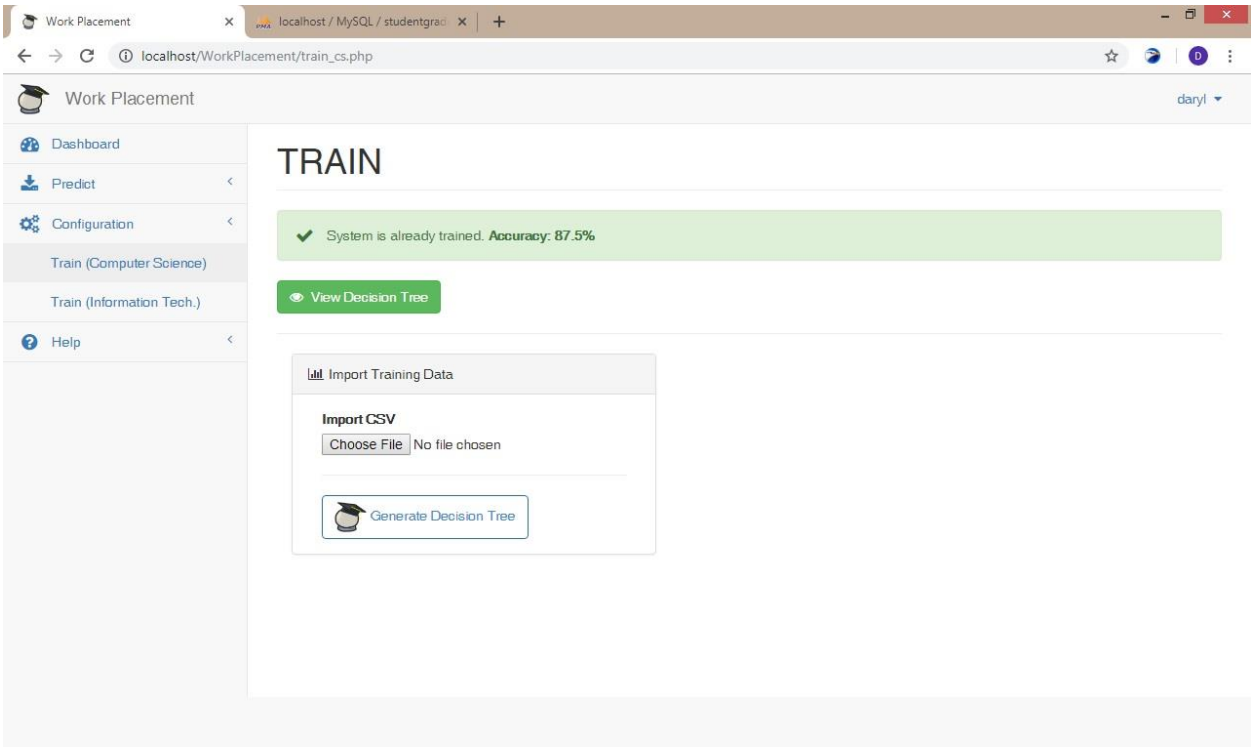
System Homepage (currently logged in as admin):



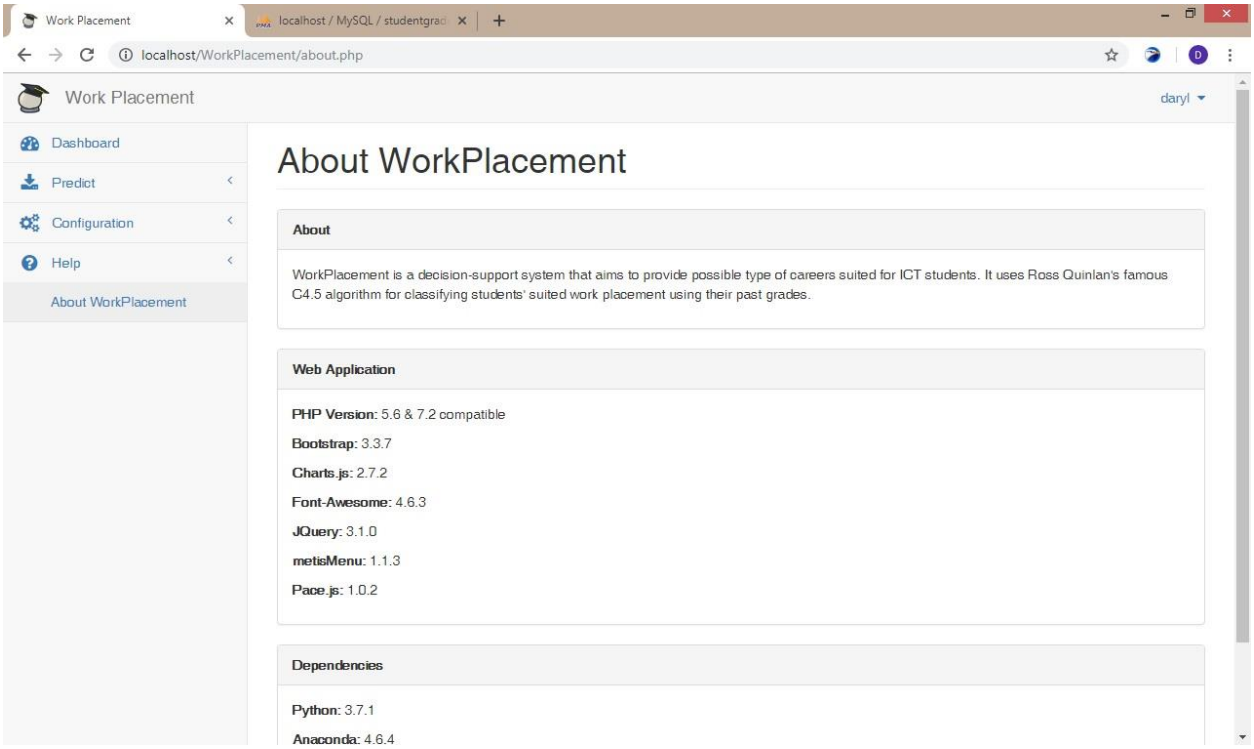
Predict/Import Student



Create Decision Tree Model

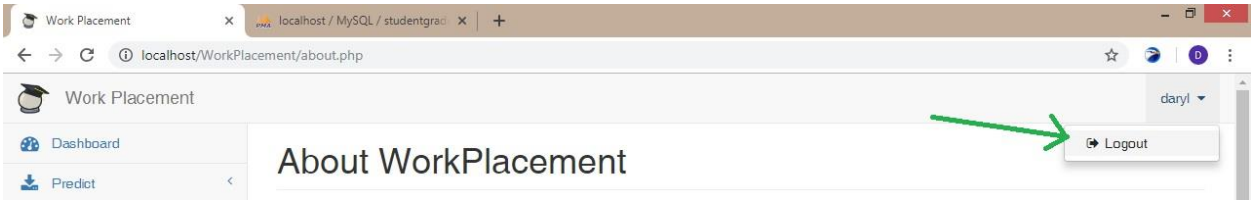


About the System



3.5 Exiting the System

To exit the system, you must logout by clicking your username in the upper right part then clicking ‘Logout’. Once you’re back in the login page, you can close the browser window.

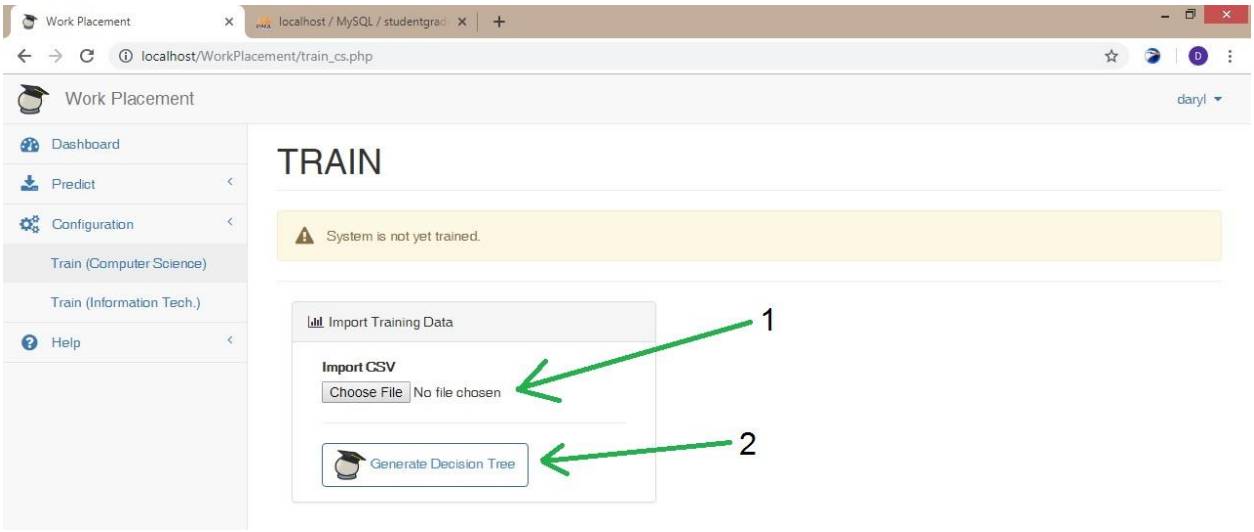


4. Using the System

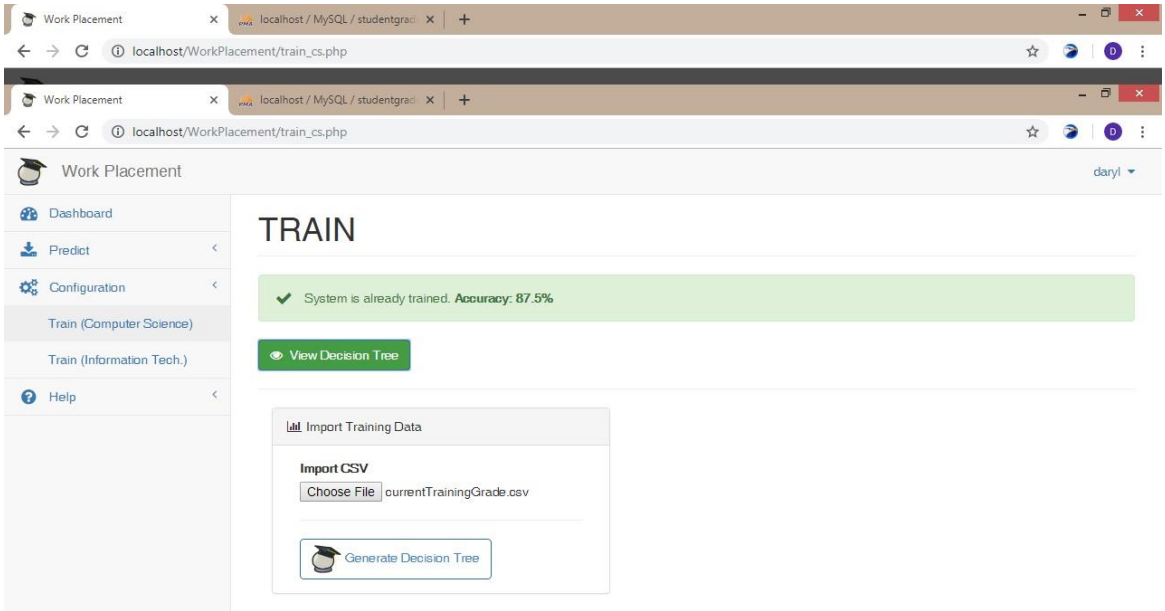
4.1 Importing Training Data & Generating Decision Tree

The system needs a training data in order for it to function. A training data contains sample datasets of grades and a corresponding class/job placement. For example, for a 'Web Development' class, you can create a dataset that has a high grade in 'CS122' or 'Web Development and Programming' subject.

The system provides a training data with 200 datasets and 10 classes with a filename of 'currentTrainingGrade.csv' and you can use it if you want. This dataset produces an accuracy of over 85%. You must follow the format of the given csv file if you wish to change the dataset and save the file inside the system folder.



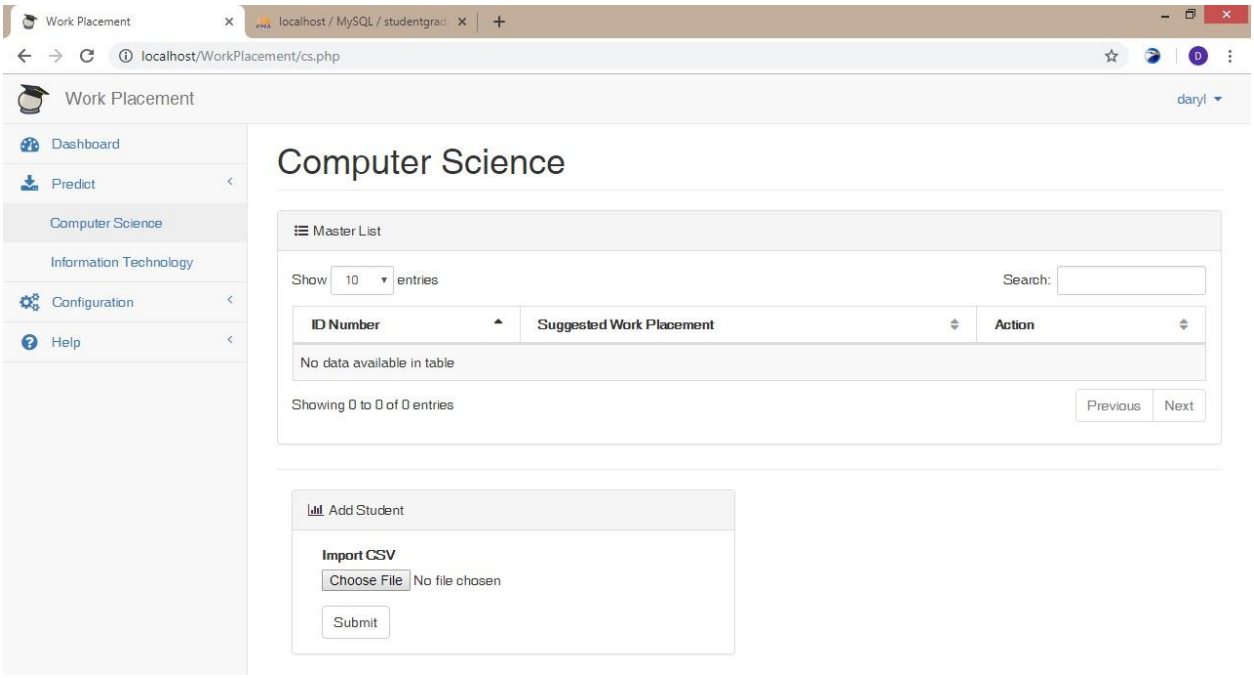
After selecting the csv file of dataset (1), you can click the 'Generate Decision Tree' button (2) to create a decision tree model.



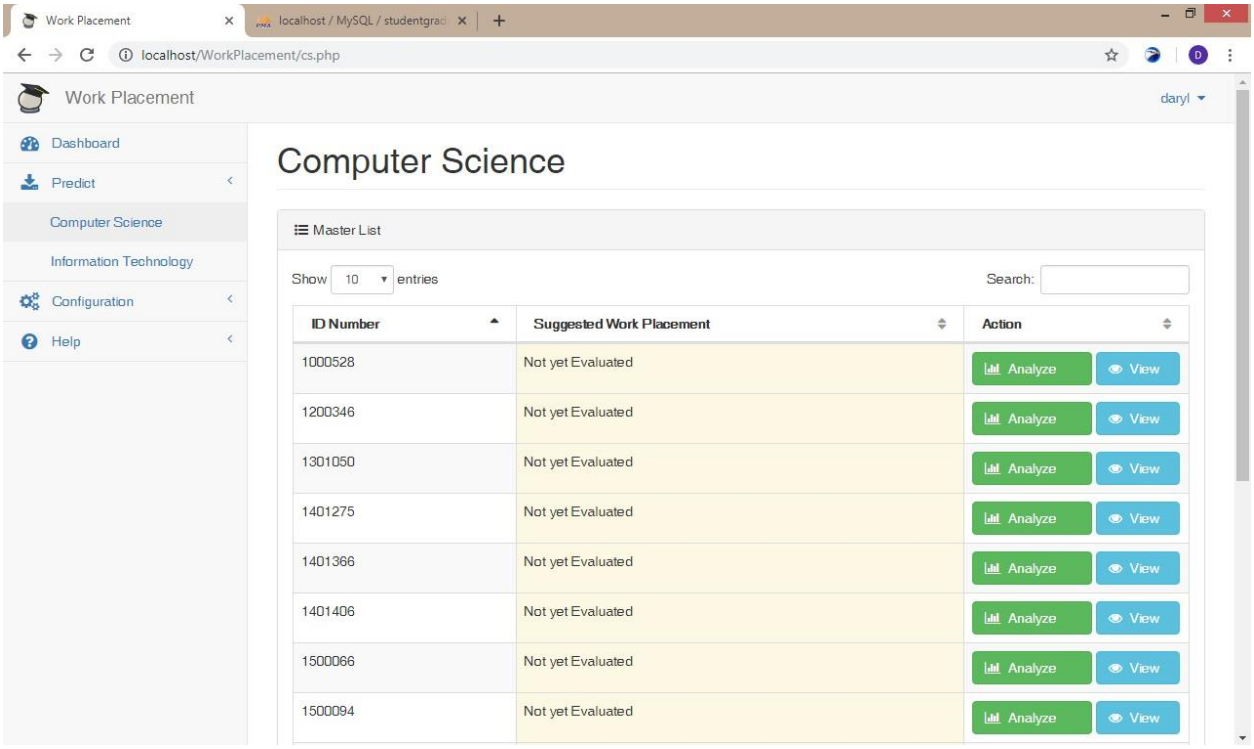
After generating decision tree, you can view its graphical representation by clicking the 'View Decision Tree' button. This will open a new tab that displays the decision tree image. You can now import actual student data and classify their work placements by clicking the tab Predict -> Computer Science/Information Tech.

4.2 Importing Student Data

The system includes a csv file for Computer Science student record with a filename of 'CSGrade50.csv' and a csv file for Information Technology student record with a filename of 'ITGrade50'. Please follow the format of the given file and then save it in the system folder.



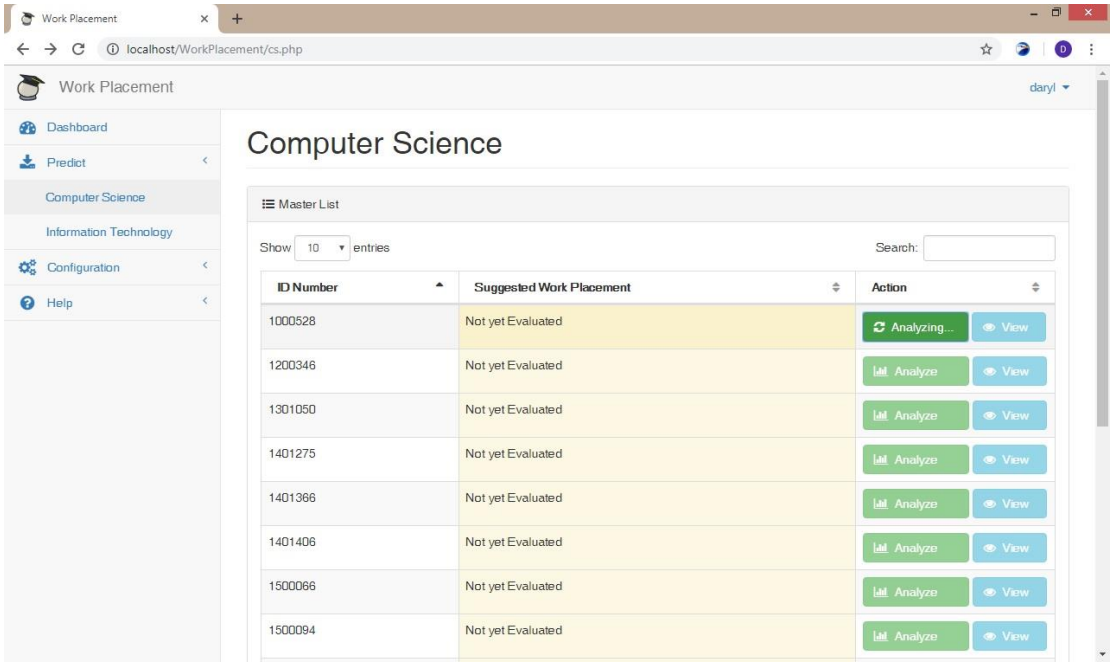
Click 'Choose File' and then click 'Submit' after selecting the csv file of student grade.

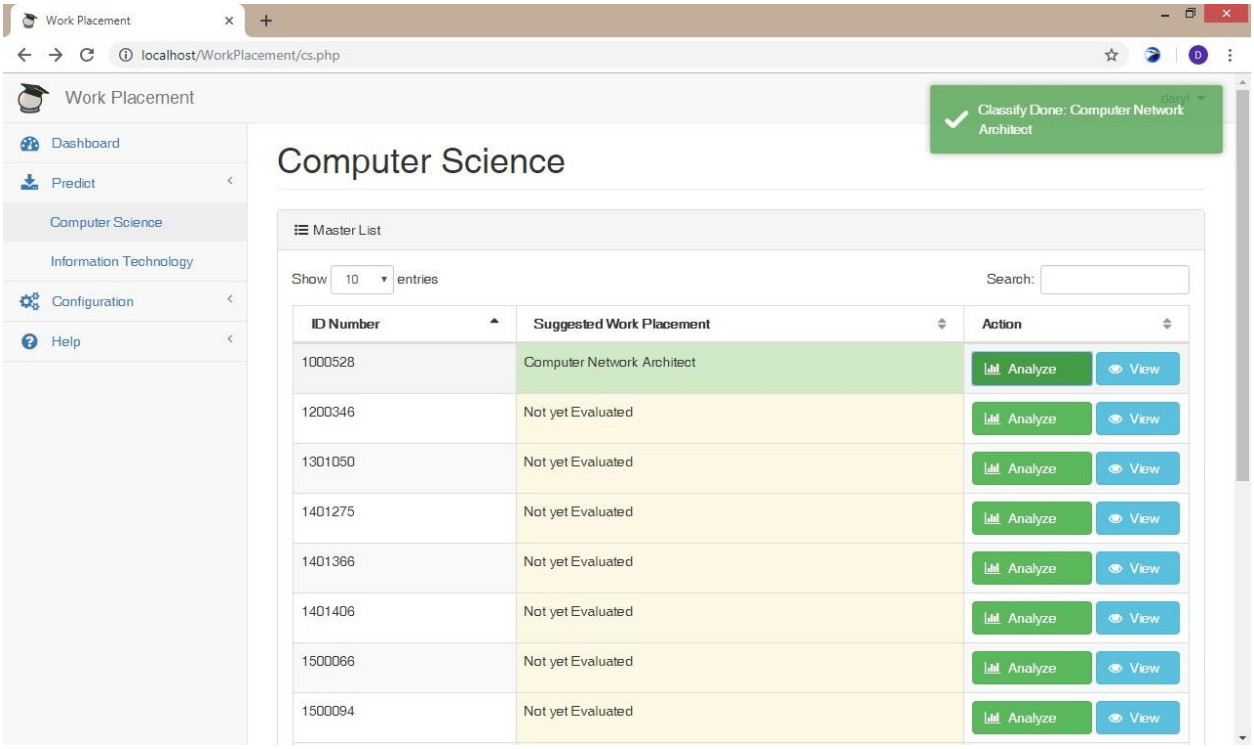


The new student records will be displayed after importing.

4.3 Classifying Work Placements

After importing student grades, you can now classify their work placements by clicking the 'Analyze' button.




Work Placement
localhost/WorkPlacement/cs.php
Work Placement
Classify Done: Computer Network Architect
Dashboard
Predict
Computer Science
Information Technology
Configuration
Help
Computer Science
Master List
Show 10 entries
Search:
ID Number
Suggested Work Placement
Action
1000528
Computer Network Architect
Analyze
View
1200346
Not yet Evaluated
1301050
Not yet Evaluated
1401275
Not yet Evaluated
1401366
Not yet Evaluated
1401406
Not yet Evaluated
1500066
Not yet Evaluated
1500094
Not yet Evaluated

5. Troubleshooting

5.1 Special Considerations

The decision tree image might appear the same even after changing the training dataset and re-generating new decision tree. In this case, you must clear the browsing data (cached images and files) in your browser.