



---

# MACHINE LEARNING PREDICTION TOOL

---

USER GUIDE



PHD CANDIDATE: KHALID ALALAWI  
SUPERVISORS  
DR. RUKSHAN ATHAUDA  
ASSOC. PROF. RAYMOND CHIONG

## **Machine Learning Prediction Tool (MLPT)**

We introduce a Machine Learning Prediction Tool (MLPT) that develops machine learning (ML) prediction models to predict students' performance based on students' assessment scores. This tool aims to help academics to detect at-risk students early on during the term allowing academics to identify and intervene for better learning outcomes.

Firstly, academics will create ML prediction models using historical student assessment data of the course. Next, they will use these prediction models to predict performance of students in the current cohort. This tutorial explains how academics can use this tool to create their own machine learning (ML) prediction models and utilize it to predict their students' performance.

This tool has seven main features as follows:

1. Creating a Course
2. Creating Assessments for the Course
3. Uploading Dataset to Create ML Models
4. Creating ML Models
5. Perform Prediction
6. Delete Course and Model
7. Update Existing Models

This tutorial will explain, with examples, how academics can use this tool to create models and predict students' performance.

### **Logging into MLPT**

First, each academic will be provided with their own username and password to login to the system. The URL for the prediction tool is:

<https://studentprediction.pythonanywhere.com>.

Figure 1 shows the login page that needs users' authentications to access this tool. When username and password are correctly entered, users can access the main page, which has all features available (see Figure 2). In Figure 2, all seven features appear in the left-side menu. This tutorial shows how academics can use all of these features with examples. In our example, Joe will be the user who will use the prediction tool.

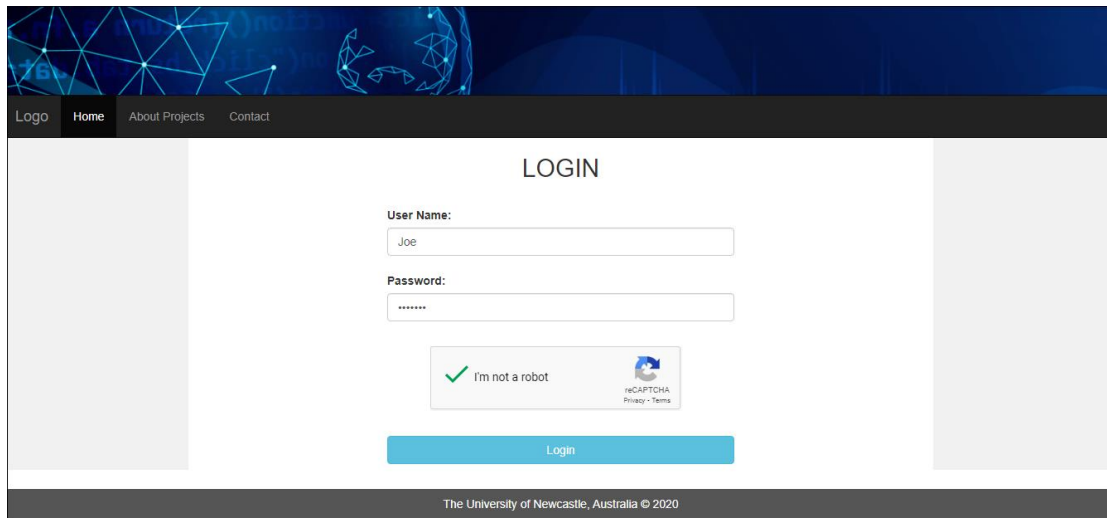


Figure 1: Login Page

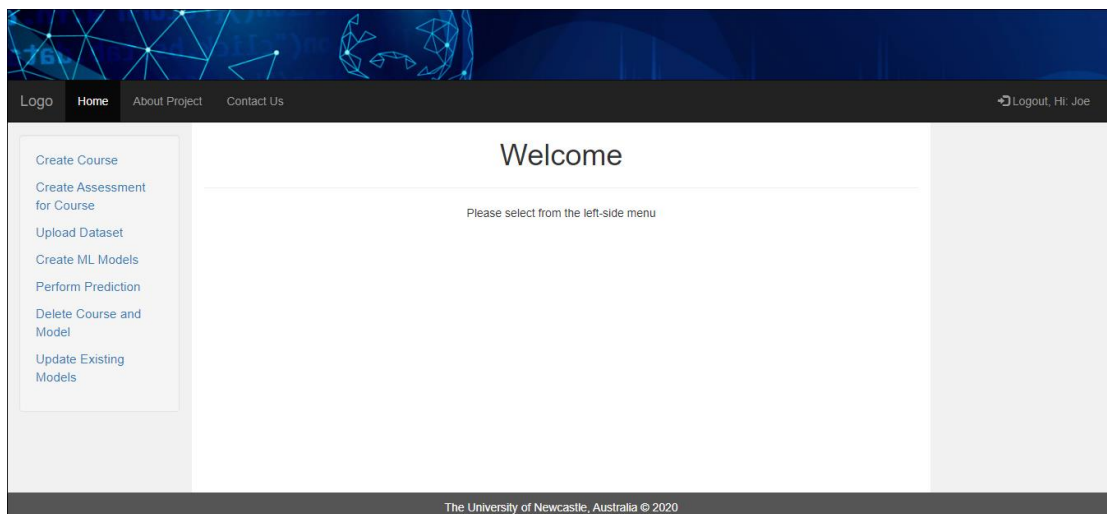


Figure 2: Main page

### 1- Create a Course:

This feature allows academics to create a course in which they want to create ML models for. Figure 3 shows how an academic creating a course by entering: Course ID, Course Name, and Course Description (optional). When academics click on *Insert* button, they simply create their course. A message will confirm if the course was correctly created.

Figure 3: Create a course page

## 2- Create Assessment:

After you create your course, the next step is to add all the assessments in your course. You need to select your course first. If have created many courses, select the appropriate course you have created from the drop down list. Then add the names, weightage of marks for the final grade for each assessment, and the mark out of (see Figure 4). Ensure that the assessments are added in the chronological order that they are assessed throughout the semester. After adding each assessment, a message is displayed at the top to confirm if the assessment is inserted correctly or not. Besides, you can see how many assessments you inserted so far in the bottom message box. In this example, four assessments are added in this course.

Figure 4: Creating assessments page

## 3- Upload Dataset:

After creating a course and adding all assessments in your course, you need to add historical data from the same course that will be used to create ML prediction models. The prediction models are created using the most popular ML algorithms. The data that the prediction tool needs is only students historical marks for each assessments out of 100. See the example in Figure 5. There are a few conditions that need to consider when uploading historical assessment data as discussed below:

- 1- You need to upload a CSV file, not a .xlsx Excel file.
- 2- The order of the columns in your CSV file must match the order of the assessments you added in the previous step when you created assessments. For example, in Figure 4, user, Joe, added five assessments: Practical Test 1, Assignment 1, Assignment 2, Practical Test 2, and Final Exam. Thus, when he wants to upload his historical data, the CSV file need to be like Figure 5.
- 3- The first row in the CSV file should contain assessments' title
- 4- The second row should be the marks out of which the assessment was marked (see Figure 5).
- 5- The third row is the weightage of the assessment out of 100%. That is, the percentage of the assessment mark that counts towards the final grade. For example, in Figure 5 Practical Test 1 is marked out of 50 but counts 20% towards the final grade.
- 6- Last two columns must be **Total** column that has students' total mark of all assessments out of 100. **Grade** column that contains students' Grade such as HD, D, P, ...
- 7- In the data set, we ignore the first row which contains the headings and also the last assessment task (which is usually the final exam) because this assessment occurs at the end of the term and is too late to predict student performance by then.

Note: If the assessments weightage changes in the current course offering which you are predicting compared to the historical data, then this could have an impact on the accuracy of prediction model results.

	A	B	C	D	E	F	G
1	PracticalTest 1	Assignment 1	Assignment 2	PracticalTest 2	Final Exam	Total (out of 100)	Grade
2	50	20	25	50	80		
3	20	10	10	20	40		
4	47.5	14	25	48.75	69.6	90	HD
5	42.5	19	17.5	16.25	46	63	P
6	37.5	18	15	43.75	43.6	69	C
7	40	12	12.5	43.75	42.4	66	C
8	43.75	18	22.5	48.125	56.4	83	D

Assessment's Title

Assessment's Marked out of

Assessment's Weight (% out of 100)

Figure 5: CSV file containing students' historical data

Figure 6 shows an example of uploading historical data. If your CSV file columns do not match with the above conditions, you will get an error. Finally, a message will confirm if the data is uploaded correctly.

Note that you can repeat this step to upload all historical assessment data for previous cohorts of students.

Figure 6: Upload historical dataset page

#### 4- Create ML Model:

After uploading all your historical datasets for your course, you simply need to click on the Create button as in Figure 7. What will happen is five prediction models using popular ML algorithms will be created. The ML algorithms used to create your prediction models are: Logistic Regression (LR), Support vector machine (SVM), Decision Tree (DT), k-Nearest Neighbors (kNN) and Naive Bayes (NB).

Binary and multiclass classification models are created. The binary models can identify students who might pass or fail the course. While the multiclass classification will detect students who will pass, fail, and students in the 'borderline' as well. Borderline students are student whose predicted final grades fall between 45 and 55 of the overall grade.

An example of creating models is shown in Figure 7. The process of creating models may take a few minutes, so please wait until it finishes. After creating ML models, the performance results of your models will appear. It shows how accurate your models are (see the example in Figure 8). The evaluation metrics used to evaluate ML models are: Accuracy, Recall, Precision, and F-measure.

Accuracy is a measure of how many correct predictions the model made to the total dataset. Recall is the measure of how many true positives get predicted out of all the positives in the dataset. Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. F-measure is the weighted average of Precision and Recall.

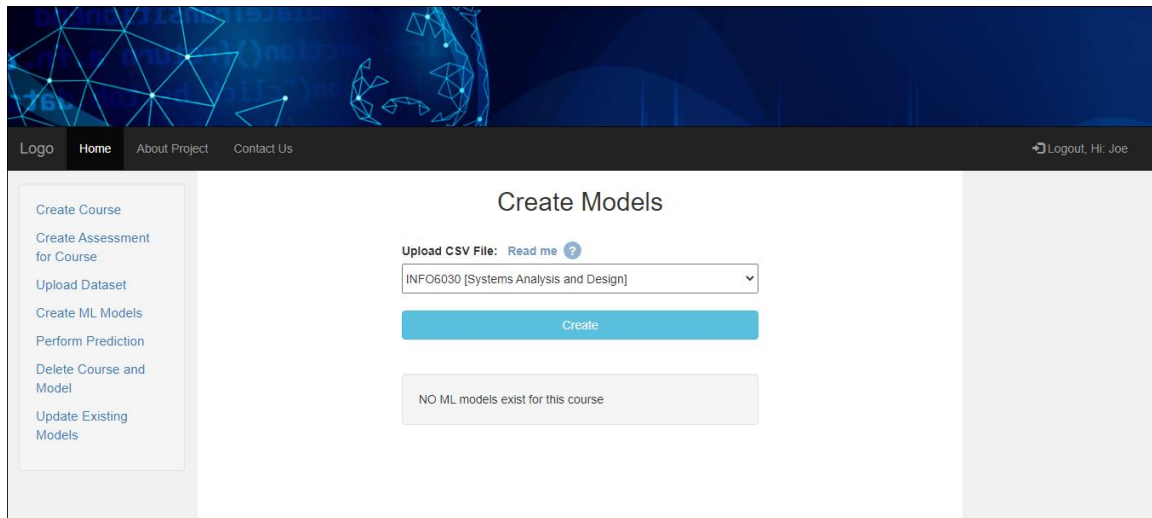


Figure 7: Create Models page

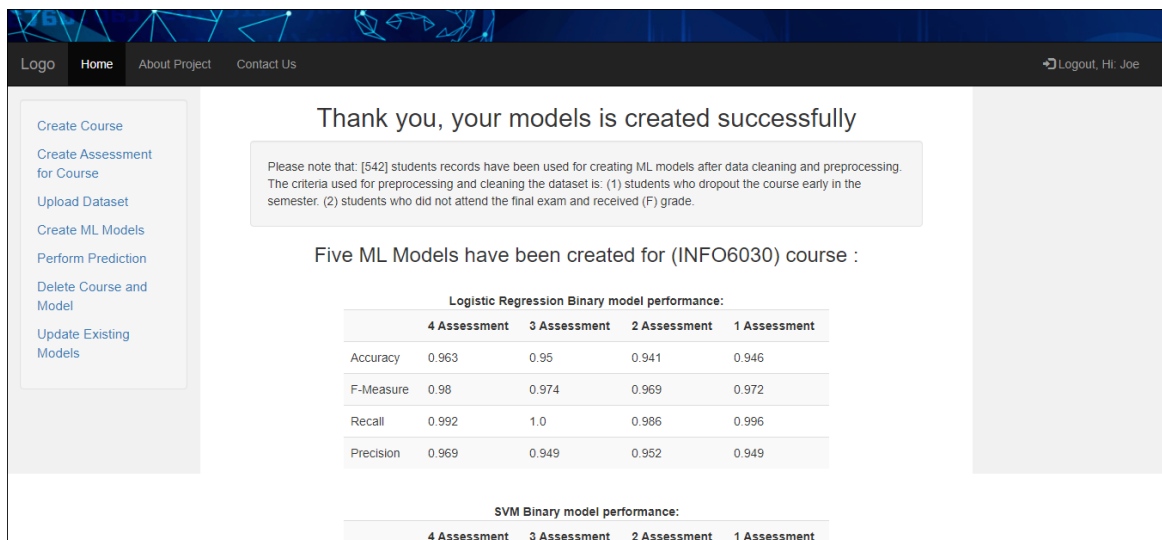


Figure 8: ML prediction models results

## 5- Perform Prediction:

Now that you have created your own ML models, you can use your models to predict students' performance in the current cohort and after each assessment (except for the Final Exam) during the term. In our example, the user Joe created models for his course "INFT6030" which has five assessments. Thus, Joe can predict his students' performance after each assessment, such as Practical Test 1 or after Assignment 1. Figure 9 shows how you can perform prediction after the second assessment as an example in Joe's course.

Figure 9: Perform prediction page

There are some conditions that need to be made to perform prediction:

- 1- You need to upload CSV file, not a .xlsx Excel file.
- 2- Your CSV file must have at least three columns. The first column is the students' name. The second column is the students' ID. Then students mark in their assessments. See the example in Figure 10 where user Joe tries to perform prediction after the third assessment which is "Assignment 2" using: Students Name, Students ID, first assessment 'Practical Test 1', the second assessment 'Assignment 1', and the third assessment 'Assignment 2'.
- 3- The second row must have the marks out of which the assessment was marked and third row the assessments weightage as a percentage. (see Figure 10)

Note that the first row in the csv file is the columns' heading which will be ignored.

	A	B	C	D	E
1	Studnet Name	Studnet ID	PracticalTest 1	Assignment 1	Assignment 2
2			50	20	25
3			20	10	10
4	AJ	123	47.5	14	25
5	JHONE	456	42.5	19	17.5
6	PHILEP	789	37.5	18	15
7	BELE	147	40	12	12.5
8	NANCY	258	43.75	18	22.5
9	ALIX	369	40	18	22.5
10	NEE2	951	38.75	16	22.5

Assessment's

Assessment's Marked out of

Assessment's Weight (% out of 100)

Figure 10: Example of CSV file to predict students performance

Basically, the tool will choose the best models, both the binary and multiclass models, to perform prediction in order to give you the most accurate results. See the prediction results in Figure 11



(a) and (b). If you want to choose different algorithms for your prediction, you can click on the 'Advanced Option' link on the perform prediction page.

The prediction results predict students' predicted performance at the end of the term based on their assessment marks so far based on the prediction model. The binary results predict whether a student is predicted to pass or fail the course (0: Fail, 1: Pass). While the multiclass models predict if a student is predicted to fail, pass, or be in the borderline (at-risk, pass, and borderline). These results can help you to identify and intervene early on to improve their performance. You can also export the results as PDF documents or Excel files.

\*Note that these results are predicted results not actual and should not be used in any way for marking assessments.

\*\* Note that the predicted results' accuracies may be impacted if the weightages of the assessment differ from the current cohort from the historical data records which has been used to create the prediction models.

The screenshot shows a web application interface for performing predictions. On the left is a sidebar with navigation links: 'Create Course', 'Create Assessment for Course', 'Upload Dataset', 'Create ML Models', 'Perform Prediction', 'Delete Course and Model', and 'Update Existing Models'. The main content area is titled 'Perform prediction'. It contains a 'Choose a course:' dropdown menu with 'INFT2012 [Application Programming]' selected. Below this is a link for 'Advanced Options'. There is an 'Upload CSV File:' section with a 'Choose file' button and the text 'No file chosen'. A link 'See an example of a CSV file format' is also present. A blue 'Submit' button is at the bottom of the form. Below the form, the text 'Results of the SVM Binary model with 97.3% accuracy:' is displayed above a table.

Student Name	Student ID	Prediction Result (Pass [1] /Fail [0])
AJ	123	1
JHONE	456	1
PHILEP	789	0
BELE	147	1

(a)

Results of the LR Multiclass model with 94.1% accuracy:		
Student Name	Student ID	Prediction Result (Pass/Borderline/At-Risk)
AJ	123	borderline
JHONE	456	pass
PHILEP	789	at_risk
BELE	147	pass
NANCY	258	pass
ALIX	369	at_risk
NEE2	951	borderline

15/5/2021

Export Excel

Export pdf

(b)

Figure 11: Prediction results

#### 6- Delete Course and Model:

In this feature, you can delete your course and your models as well. Simply choose the course you want to delete and click on Delete.

#### 7- Update Existing Models:

Finally, this feature allows you to update and modify your models. It is very useful when you have more historical data, and you want to improve your models. First, you need to upload your new data the same way you uploaded your data in step 3 'Upload Dataset'. Then you simply click on the Update button. The results of your new models will appear and show you how your updated model performs. This process may take a few minutes to complete. Please wait until it finishes.

Thank you for reading this tutorial. If you have any comment or question, you are welcome to contact me via email or phone and am happy to be of any assistance. Also, I am happy to have a personal Zoom session or face-to-face meeting to help you develop the prediction models for your course.

**Khalid Alalawi**

PhD. Candidate at University of Newcastle

[Khalid.Alalawi@uon.edu.au](mailto:Khalid.Alalawi@uon.edu.au)

Phone: 0481175286