

MODEL BUILDING

Training and Testing the Model

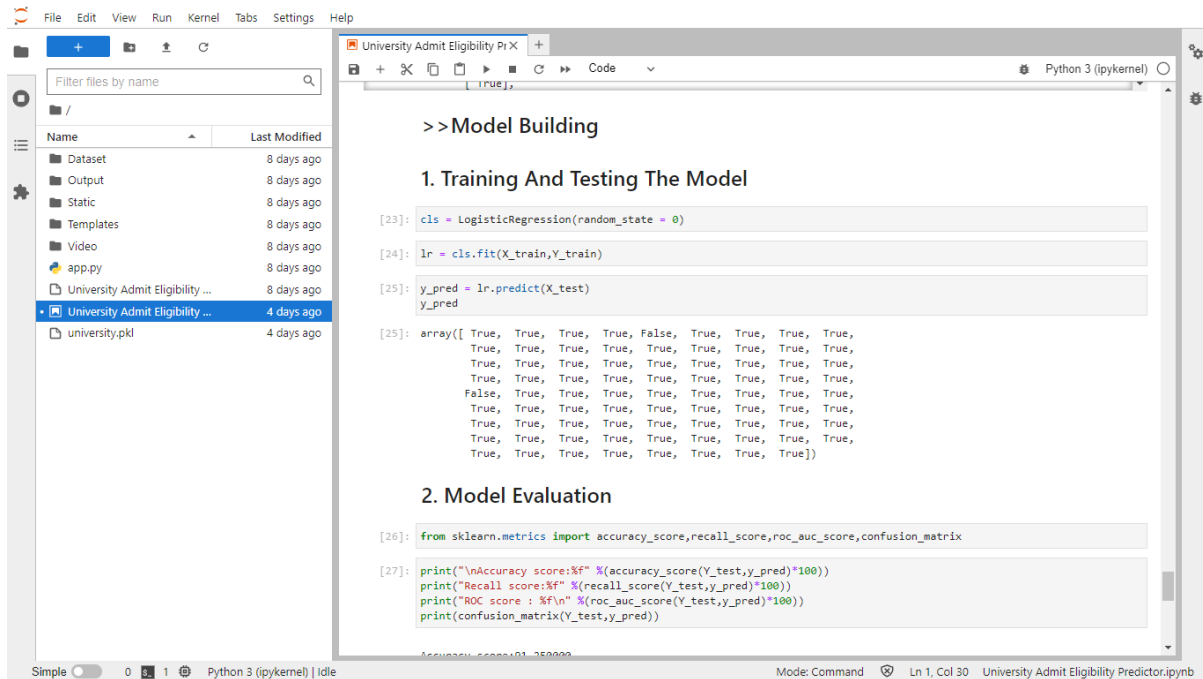
- After splitting the data into train and test, the data should be fed to an algorithm to build a model.
- There are several Machine learning algorithms to be used depending on the data you are going to process such as images, sound, text, and numerical values.
- Classification algorithms are Regression algorithms, etc. Here the Regression algorithm used.
- There are 4 types of methods such as
 - ❖ Logistic Regression
 - ❖ Decision Tree Classifier
 - ❖ Random Forest Classifier
 - ❖ KNN

Logistic Regression:

- ❖ Logistic Regression is used when the dependent variable (target) is categorical. For example,
 - ❖ To predict whether a website is an untrusted SSL certificate (1) or (0)
 - ❖ Whether the tumour is malignant (1) or not (0)
- ❖ Out of all the algorithms Logistic Regression got the highest accuracy, so let's build a model with Logistic regression.
- ❖ By using X_train and Y_train obtained above in the train_test_split section to train our decision tree regression model. Then using the fit method and passing the parameters as input.
- ❖ The algorithm from **Scikit** learn library to build the model for given input from splitted data.
- ❖ Once the model is trained, it's ready to make predictions. Then, using the predict method on the model and pass x_test as a parameter to get the output as y_pred.
- ❖ Notice that the prediction output is an array of real numbers corresponding to the input array.

Step-1:

- After splitting the data as X_train, X_test, Y_train, Y_test.
- Then, we need to train the data for further prediction.
- Import **sklearn.linear_model.logistic** package.
- Then use **fit ()** method to train the model.
- And then use **Predict ()** method to test created model.



The screenshot displays a Jupyter Notebook titled "University Admit Eligibility Predictor.ipynb". The left sidebar shows a file explorer with a list of files and folders, including "Dataset", "Output", "Static", "Templates", "Video", "app.py", "University Admit Eligibility ...", and "university.pkl". The main area of the notebook is divided into two sections: "1. Training And Testing The Model" and "2. Model Evaluation".

1. Training And Testing The Model

```
[23]: cls = LogisticRegression(random_state = 0)
[24]: lr = cls.fit(X_train,Y_train)
[25]: y_pred = lr.predict(X_test)
y_pred
```

The output of the prediction is a 1D array of boolean values:

```
array([ True,  True,  True,  True, False,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True,
        False, True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True])
```

2. Model Evaluation

```
[26]: from sklearn.metrics import accuracy_score,recall_score,roc_auc_score,confusion_matrix
[27]: print("\nAccuracy score:%f" %(accuracy_score(Y_test,y_pred)*100))
      print("Recall score:%f" %(recall_score(Y_test,y_pred)*100))
      print("ROC score : %f\n" %(roc_auc_score(Y_test,y_pred)*100))
      print(confusion_matrix(Y_test,y_pred))
```

The bottom status bar indicates the notebook is running on "Python 3 (ipykernel)" and is in "Simple" mode.