# **Tasks:**

# **Classification Model:**

The classification model used is Multinomial Logistic Regression.

However, the data has been preprocessed before splitting and feeding into the model.

**Basic Idea:**

The categorical columns ‘Player and Tm’ have been dropped. Since, the dataset has the class label column ‘Pos’ which has been dropped too.

Then I had used **Random Forest** for feature selection where the model had assigned the ranking for each feature and the top features that affect the accuracy most have been selected. I have selected top 11 features because the accuracy was same even when I have included features more than 11.

The features that contributed for better accuracy are:

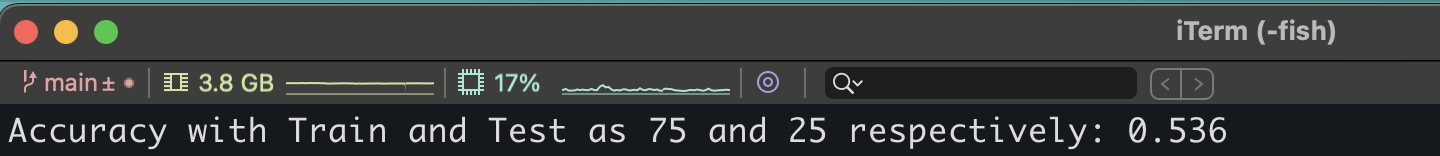
**top\_features['AST','ORB','FG%','TRB','3PA','BLK','DRB','2P%','eFG%','3P','3P%’]**

Then data is split into 75% training and 25% testing and fed to the Multinomial Logistic Regression model with parameter (multi\_class=’multinomial’)

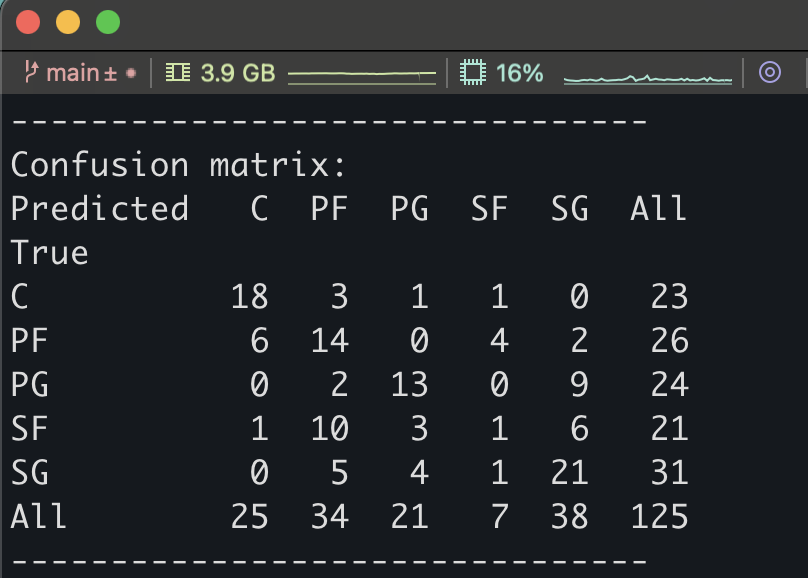
# **Accuracy of Selected model:**

The accuracy of the model is: 53.6%

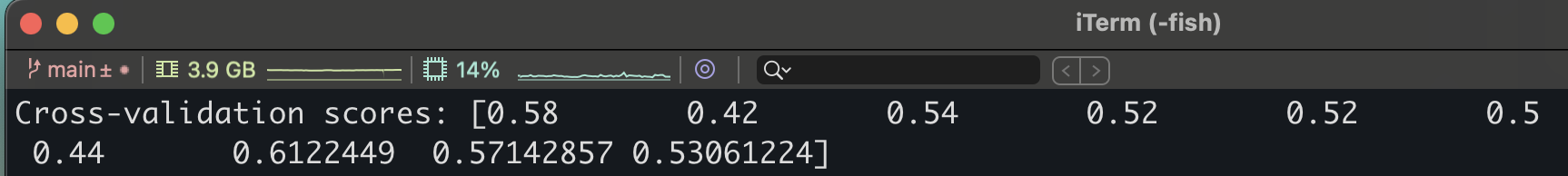
Output generated in Terminal:



# **Confusion Matrix:**

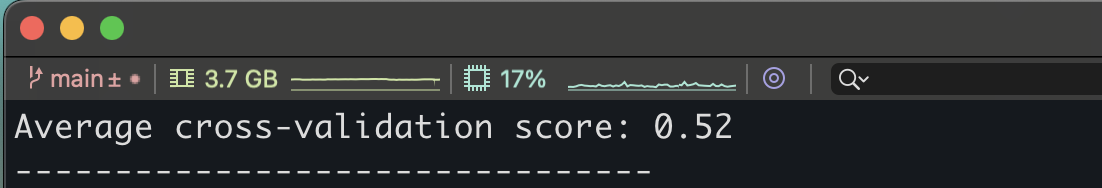


# **Cross-Validation:**



# **Accuracy:**

The average accuracy for cross validation.



1. **Observations:**

When the selected model has been run continuously the accuracy was noted >60%.

**Further Observations:**

Although, the accuracy was better in the selected process. I have implemented other feature selection methods as Chi-squared, univariate but the features selected by these methods did not contribute for better classification.