Working on soccer dataset

By

Himadri Bandyopadhyay
(173050004)

Arijit Mukherjee
(17305t002)

Shiva Kulshreshtha
(173050023)

Karanveer Singh
(173050027)

Linear SVM Classifier - 173050023

- Applied Linear SVM classifier.
- Achieved cross_val_score of 51%.
- Applied cross validation to fine tune the slack parameter C.
- Found poor accuracy for prediction of ‘draws’
- Results in line with the research papers studied.

SVM and SGD Classifier - 173050027

- Applied SVM classifier using 2 kernels namely RBF and Sigmoid
- Achieved average test accuracy of 45% and train accuracy of 55% on a train data of 15000 examples and test data of 7000 examples.
- Applied GridSearch to fine tune the parameters C and Gamma.
- SGDClassifier also gave comparable results with test accuracy of 42-43% and train accuracy of 49%.
- Results in line with the research papers studied.

Adaptive Boosting with decision tree as base model 173050004

- Adaboost was implemented with the base model as decision tree
- The number of weak learners were varied from 1 to 41 with depth of the tree limited to 12.
- Entropy was selected as the impurity measure, and the split feature was selected to be the best feature among randomly selected set of features.
- Test accuracy obtained was 54%.

Neural Networks 17305T002

- Fully Connected Neural Network with dropout using Keras .
- Features used were team attributes ,betting odd and last 5 match win rate . achieved 52%.
- Solve the problem from a different perspective , CNN.
- Encoding the formation , player’s skill information and betting odds for all the matches into images of size 200x300.
- Multiple such skills were stacked along the depth for each match , for each player .
- Used VGG 16 model to train with the above said features and result label .
- Achieved an accuracy of 45.6% in the test set.