ACM MPSTME

NBA Prediction App Using ML

01

BY: SAUMYA JOSHI
SUMEDH PIMPALKHUTE

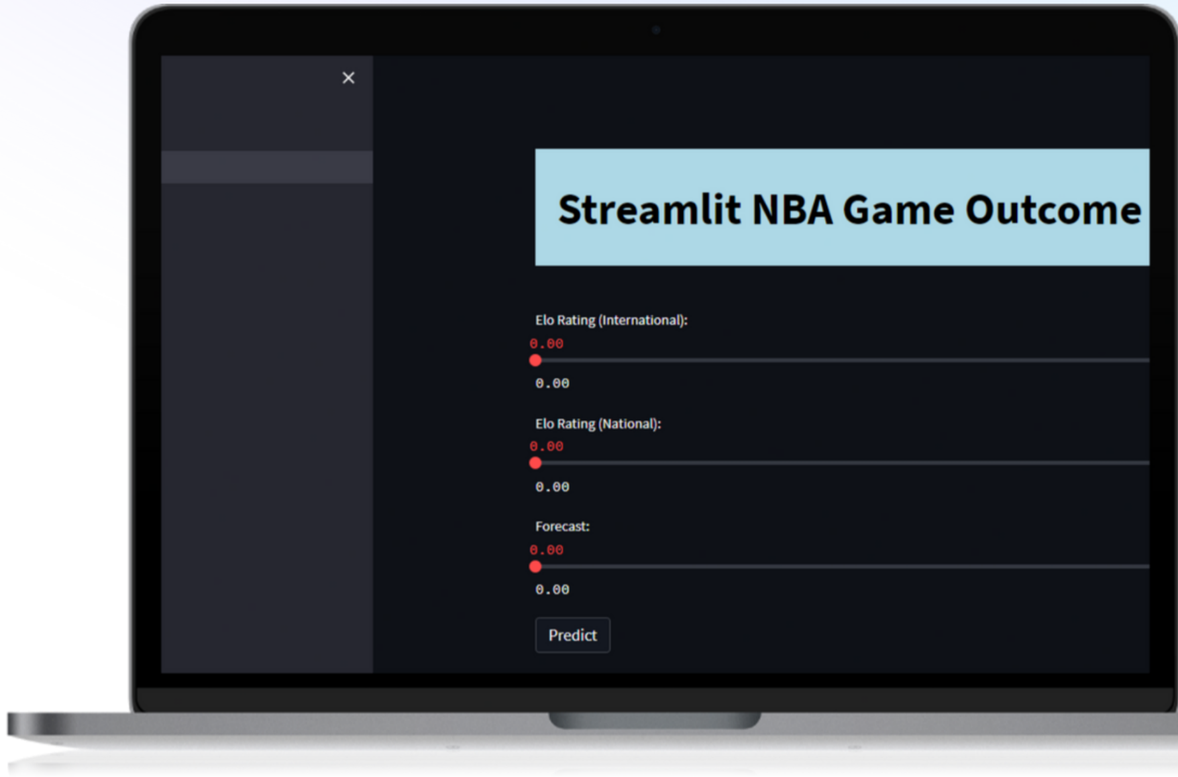
MENTOR: VIRAJ RAJENDRA SANAP

Problem Statement

The aim of the project is to curate a Machine Learning model to predict the outcome of NBA games and deploy it using Streamlit library.

Synopsis

Four different models were tested on the training set: Linear Regression, Logistic Regression, Decision Tree, and K Neighbors Classifier. The best performing model turned out to be **Logistic Regression**. After model testing, the need to interact with the model such that it predicts a win or loss arose. Hence, it was additionally upgraded with a web service application via Streamlit Cloud & GitHub.



MARKET SURVEY



- a tool for viewers to overview predictions before the match
- beneficial to the ever-growing betting industry
- understand and formulate strategies needed to win matches
- ML is one of the most promising technologies for prediction

Technologies Used

Numpy

NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), basic linear algebra, basic statistical operations, random simulation and much more.

Pandas

It is a python library which is used to analysis data.
The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" .Pandas allows us to analyse big data and make conclusions based on statistical theories. Pandas can clean messy data sets, and make them readable and relevant.

Streamlit

Streamlit is a free and open-source framework to rapidly build and share beautiful machine learning and data science web apps. It is a Python-based library specifically designed for machine learning engineers.

SKLearn

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python

Other technologies used

Matplotlib

Used for plotting 2d graphs

Seaborn

Used for plotting more complex graphs

Time

This module provides various time-functions.

Stasmodel.api

Used for gathering the stats needed in the project

mpl_toolkits.mplot3d

Used for making 3d graphs

dabl

Dabl offers ways of automating processes that otherwise take a lot of time and effort. Faster processing of data leads to faster model development and prototyping.

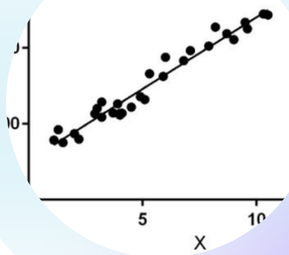
pickle

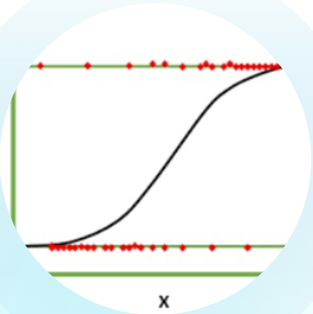
Used to store the model for the streamlit lib.

MODELS IMPLEMENTED

Linear Regression

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting





Logistic Regression

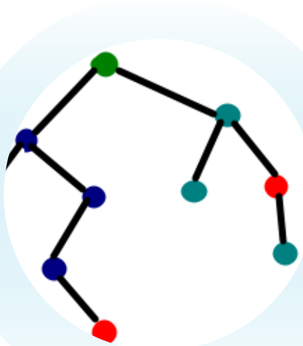
Logistic regression is used when one is trying to predict a dependent categorical variable. The two outcomes for a binary regression model are 1 and 0. Some things that could potentially be predicted are win or lose, spam or not spam, and so on.

08

MODELS IMPLEMENTED

KNN

K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. It is

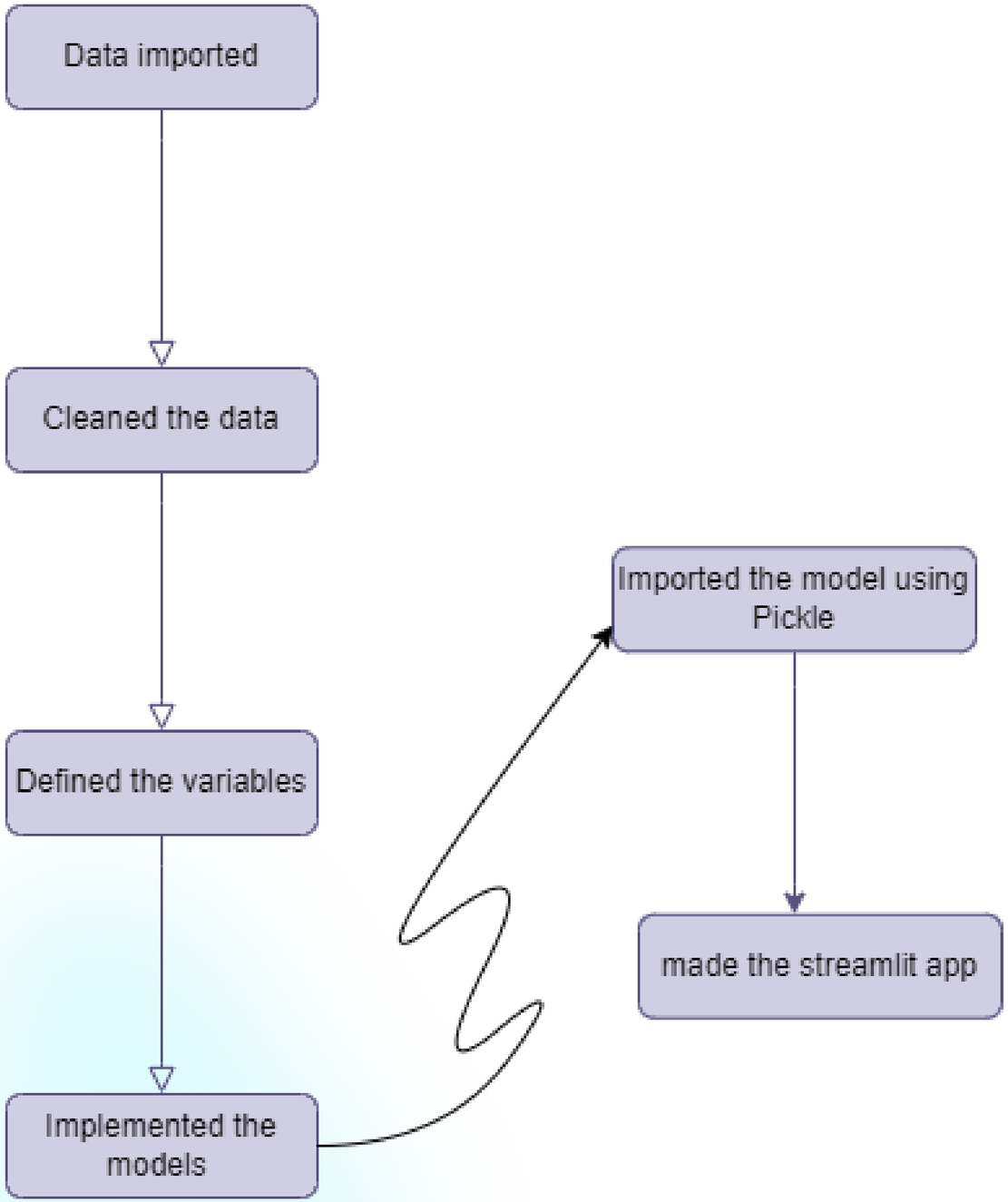


Decision Tree

Decision tree uses the tree representation to solve the problem in which each leaf node corresponds to a class label and attributes are represented on the internal node of the tree.

09

CODE IMPLENTATION



IMPLEMENTATION

The Code

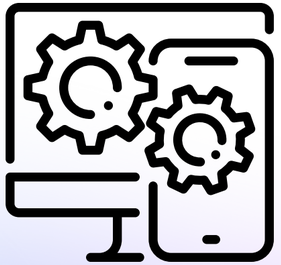
https://colab.research.google.com/drive/1iRm70F2EFnQj96F4H4S7dzsd6sjH-hLN?usp=sharing

The Dataset

https://www.kaggle.com/datasets/fivethirtyeight/fivethirtyeight-nba-elo-dataset

The App

https://bit.ly/3OJ6hwU



CONCLUSION

Best Performance Algorithm

After running all the different methods we concluded Logistic Regression gave the best accuracy.

Outcome Prediction App

After saving the model, we were able to make a Prediction App and work with Streamlit library, and use the latest upgrades of the python library.

Success of the Model

The best NBA game prediction models only accurately predict the winner about 75% of the time, so our Logistic Regression model is very close to the upper bound of predictions that currently exist.

Future Scope

Using an Updated Data

The dataset used in the model only dated till 2015 , hence working with a recent data set will be more relevant in future

Using Api

Instead of using aa static source like a dataset, we could add an api (nba_api package) so that we can get real time statistics

Adding more features to the Model

Adding more features like player prediction , chapionship odds, MVP odds etc.

A More Functional Application

Flask could've been used as the website framework which would in turn customize the application using Bootstrap.

REFERENCES

Streamlit Documentation

https://docs.streamlit.io/

Dataset handling

https://www.w3schools.com/python/pandas/default.asp

Model Building

https://www.geeksforgeeks.org/learning-model-building-scikit-learn-python-machine-learning-library/

ML Models

https://www.youtube.com/watch?v=QWYkQDvCo4Y