DATA MINING PROJECT

EXPLORATORY DATA ANALYSIS OF DOTA 2

Requirements:

- IPython environment/editor like Jupyter Notebook, VS Code
- Python version 3.8 or 3.9
- **Python Libraries needed:**
  - Pandas 1.3.2
  - Numpy 1.21.1
  - Csv
  - Json
  - Seaborn
  - Matplotlib
  - Tensorflow
  - Keras
  - BeautifulSoup
  - Request

Submission Structure:

1) Analysis: Python scripts and IPython notebook which take tournament data as input and perform analysis and generate appropriate outputs stored in the output folder.
   a) Across Tournament Analysis.ipynb : Performs various analyses like Average Match Duration, Total Assists, Average First Blood Time and etc compared for each tournament. Displays results as various graphs.
   b) Hero Analysis Across Tournament.ipynb : Performs various analyses for each hero as to how many times is it picked and banned across tournaments. Also analyses no of wins per hero across tournaments. Displays analysis result as a bar chart for each year separately and then a stacked bar chart for all years.

c) Team Analysis Across Tournament.ipynb: Total kills, Unique Heroes Picked, Kill Death Ratio and etc for each team in a particular tournament. The analysis result is shown as bar graphs, horizontal bar graphs, and line charts.
d) Picks Bans.ipynb: Computes for each tournament individually which heroes have been either picked or banned most at some n^{th} position.
e) Hero-analysis.py: Take the input data of each tournament and perform various analyses like the total number of times the hero is picked, the total number of times hero is banned, etc, and output is stored in form of CSV in the output folder
f) Player-analysis.py: Take the input data of each tournament and perform various analyses like player's total kills, lowest death average, total assists, etc and output is stored in form of CSV file in output folder
g) tournament-analysis-2016.py: Take the input data of each tournament and perform various analysis like longest duration match, shortest duration match, unique heroes pick, etc and output is stored in form of a text file in the output folder. Similarly for 2017-2021 year
h) team-analysis.py: Take the input data of each tournament and perform various analyses like total matches played, total kill by the team, etc, and output is stored in form of csv in the output folder.
i) Win-Margin-Calc.py: Computes Net Win Margin of each time separately for each tournament. Net Win Margin is the sum of the difference of score between two sides divided by the total no of matches played in the tournament.
j) run.sh: Executes all the python scripts present in the current folder.

2) Prediction: Contains python scripts, csv and IPython Notebooks pertaining to Prediction Engine.
   a) prediction.ipynb: Builds two NN-based classification models which predict the winning chances of a team pre-game and mid-game. alldata5, alldata10, alldata14, data3, and data4 are taken as input.
   b) data3.csv and data4.csv- Contains data with features match id, hero ids of 10 players, and result.
   c) alldata5.csv, alldata10.csv, and alldata14.csv- Contains data with all the features.

3) Output: Contains CSV and text files generated as output from the analysis 2020 year file will not be generated as due to COVID there is no match data.
   a) hero-analysis-{2016-2021}.csv
   b) Team-analysis-{2016-2021}.csv
   c) NWM_{2016-2021}.csv
   d) Player-analysis-{2016-2021}.csv
   e) Tournament-Analysis-{2016-2021}.txt

4) Data: Contains Data extracted and gathered from various sources used to perform analysis and prediction engine. Also contains python scripts and IPython ustilised to gather this data

   a) Data Collection: Contains '.ipynb' files used to extract data from various sources.
      i) dotabuff-getmatch.ipynb: Uses Beautifulsoup to extract match ids
      ii) get_tournament_matchid.ipynb: Uses Beautifulsop to mine match ids of International Tournaments from year 2016 to 2021, except 2020.
      iii) hero_list.ipynb: Gives the mapping between hero id and hero name.
      iv) match_details.ipynb: Mines match details using Steam API for passed list of match ids.
      v) Pick Ban Generator.ipynb: Extracts Picks and Bans by each team in a match.
      vi) steam_getmatch.ipynb: Mines Match ids of recent matches being played using Steam API.
   b) Drafting: Contains output file generated by Pick Ban Generator.ipynb. Output files are in CSV format and contain for each tournament individually for each match of the tournament Picks and Bans by players in each match. Contains International_{2016-2021}_drafting.csv except for 2020
   c) Hero names: Contains a single CSV file which is the output of hero_list.ipynb and CSV contains a mapping of hero id to hero name.
   d) Match Details: Contains CSV files for each year from 2016-2021, except for 2020. Each CSV contains match details for each match id of a tournament.
   e) Rankings: Contains CSV files for each year from 2016-2021, except for 2020. Each CSV contains final team rankings for each tournament.

STEPS TO EXECUTE PROJECT:

1. For analysis, move into analysis folder, first execute run.sh file then for visualization run notebook from starting cell. Output will be stored in output folder
2. For prediction, move into prediction folder run the notebook from first cell.

NOTE:

1. **Do not delete 'Output' folder or else scripts will start generating error as all the csv files generated are stored in the 'Output' folder.**
2. **For files in analysis folder, first run shell script and then execute '.ipynb' files.**
3. **Data Gathering Code is just for reference.**