Friend Affinity Finder

Team Members:

Our team consists of 4 members (all from IIIT Hyderabad):

Prajwal Krishna Maitin

Pratyush Kumar

Yash Chaurasia

Aradhya Tongia

Role of Team Members:

Prajwal Krishna Maitin: Data Collection and preprocessing

Pratyush Kumar: Basic ML

Yash Chaurasia: Data Collection

Aradhya Tongia: Website

Technologies/Platform/APIs used:

IBM Watson Personality Insights   Flask

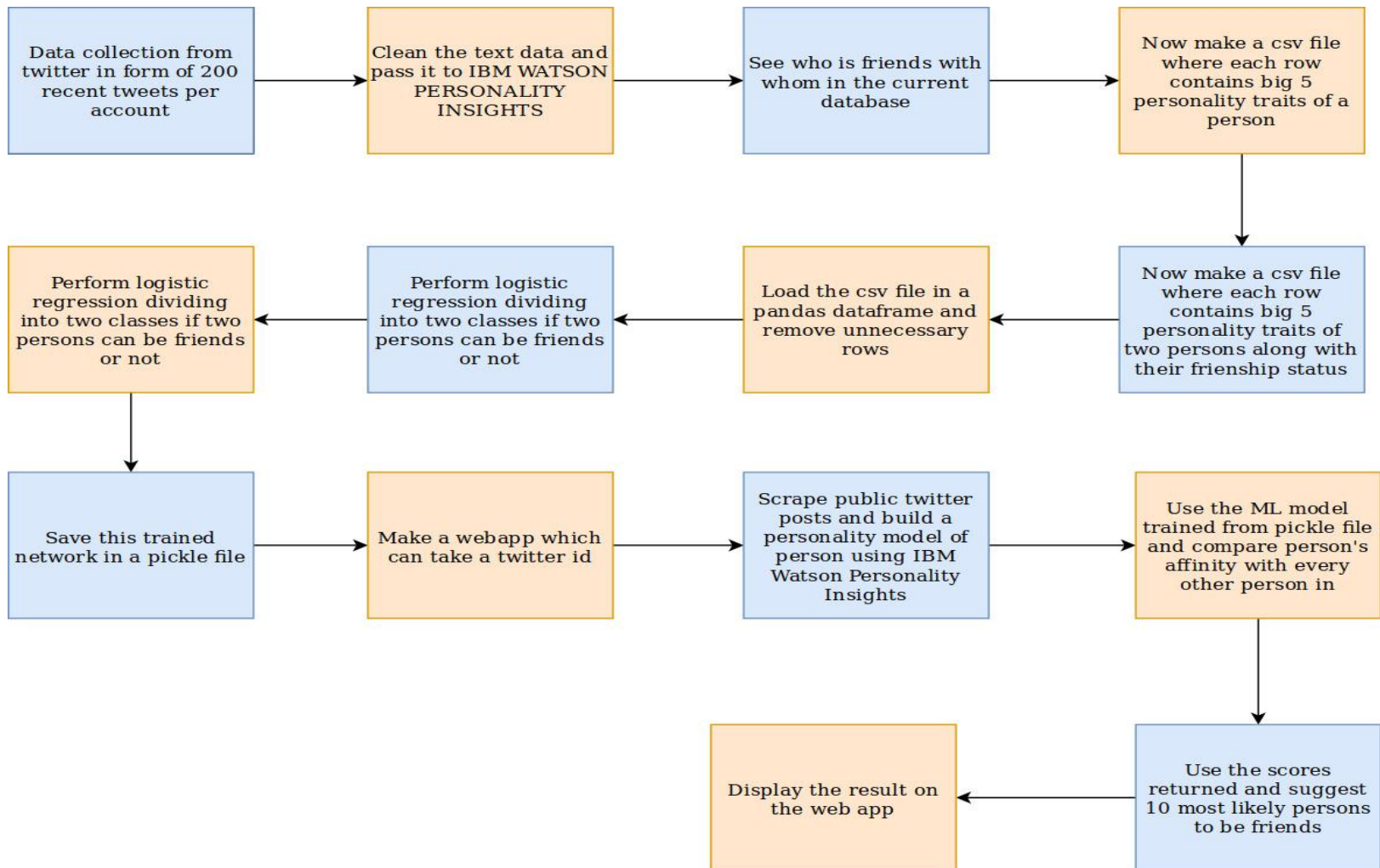IBM Cloud   Keras

Github   Numpy

Python3   Pandas

Tweepy   SKLearn

```
┌─────────────────┐      ┌─────────────────┐      ┌─────────────────┐      ┌─────────────────┐
│ Data collection │      │ Clean the text  │      │ See who is      │      │ Now make a csv  │
│ from twitter in │─────▶│ data and pass   │─────▶│ friends with    │─────▶│ file where each │
│ form of 200     │      │ it to IBM WATSON│      │ whom in the     │      │ row contains    │
│ recent tweets   │      │ PERSONALITY     │      │ current         │      │ big 5           │
│ per account     │      │ INSIGHTS        │      │ database        │      │ personality     │
└─────────────────┘      └─────────────────┘      └─────────────────┘      │ traits of a     │
                                                                           │ person          │
                                                                           └─────────────────┘
                                                                                    │
                                                                                    ▼
┌─────────────────┐      ┌─────────────────┐      ┌─────────────────┐      ┌─────────────────┐
│ Perform logistic│      │ Perform logistic│      │ Load the csv    │      │ Now make a csv  │
│ regression      │      │ regression      │      │ file in a       │      │ file where each │
│ dividing into   │◀─────│ dividing into   │◀─────│ pandas          │◀─────│ row contains    │
│ two classes if  │      │ two classes if  │      │ dataframe and   │      │ big 5           │
│ two persons can │      │ two persons can │      │ remove          │      │ personality     │
│ be friends or   │      │ be friends or   │      │ unnecessary     │      │ traits of two   │
│ not             │      │ not             │      │ rows            │      │ persons along   │
└─────────────────┘      └─────────────────┘      └─────────────────┘      │ with their      │
        │                                                                  │ frienship status│
        ▼                                                                  └─────────────────┘
┌─────────────────┐      ┌─────────────────┐      ┌─────────────────┐      ┌─────────────────┐
│ Save this       │      │ Make a webapp   │      │ Scrape public   │      │ Use the ML model│
│ trained network │─────▶│ which can take  │─────▶│ twitter posts   │─────▶│ trained from    │
│ in a pickle file│      │ a twitter id    │      │ and build a     │      │ pickle file and │
│                 │      │                 │      │ personality     │      │ compare person's│
│                 │      │                 │      │ model of person │      │ affinity with   │
│                 │      │                 │      │ using IBM Watson│      │ every other     │
│                 │      │                 │      │ Personality     │      │ person in       │
│                 │      │                 │      │ Insights        │      │                 │
└─────────────────┘      └─────────────────┘      └─────────────────┘      └─────────────────┘
                                                                                    │
                         ┌─────────────────┐      ┌─────────────────┐              ▼
                         │ Display the     │      │ Use the scores  │
                         │ result on the   │◀─────│ returned and    │
                         │ web app         │      │ suggest 10 most │
                         │                 │      │ likely persons  │
                         │                 │      │ to be friends   │
                         └─────────────────┘      └─────────────────┘
```

Frameworks/Tools:

- Python3 - Used Python3 for web development.
- Flask - Used as a framework for the development of backend of our website.
- Bootstrap - Most of the CSS features are taken from Bootstrap framework used for the developement of front-end.

The Prediction System:

We generated twitter account data of multiple users by scraping using twitter apis.

Then we performed sentiment analysis on them using IBM Watson Personality Insight and stored the features of each account.

Next we made NC2 pairs of the accounts and checked if they follow each other and accordingly assigned a boolean value to the pair and stored this (pair, boolean value)in a csv file.

Then we a logistic regression model on this training set and stored this in