

## Problem Statement - Friend Affinity Finder

As described in the problem statement, finding commonalities among friends breeds healthy relationships. In a social age, a person's social media data is comprehensive enough to gain apt knowledge about his behaviour and interests. This can be used to better understand the person and shape one's responses towards him/her.

**Novelty** - This project allows users to get an idea of how similar he is to his friends by providing their social media handles. Though the approach to predict affinity between friends in this approach is quite mathematical, it gives a fair idea about it.

The web application designed in this approach takes data from social media websites and using IBM Watson service to analyse this data. The 'Personality Insights' and 'Tone Analyzer' services are used to analyze data from Twitter and Reddit to get scores for the Big Five personality characteristics and the tone of the data respectively for both the user and his friends. The affinity is then calculated using the following formula :-

*Let  $t_{i,f}$  = Score of friend on  $i^{th}$  trait,  $t_{i,u}$  = Score of user on  $i^{th}$  trait*

$$\text{Then, Affinity} = \frac{(\sum (1 - |t_{i,f} - t_{i,u}|)^2)}{(\text{Total number of traits})}$$

There are five personality traits and seven tones. Therefore the number of traits are 12 in total.

Using the 'Natural Language Understanding' service and data from StackOverflow, common interest of the user and the friend are predict. To include this in the affinity, a separate formula was used. The formula used was based on the principle of giving weightage to common interest and punish mildly for uncommon ones. The following formula was used:-

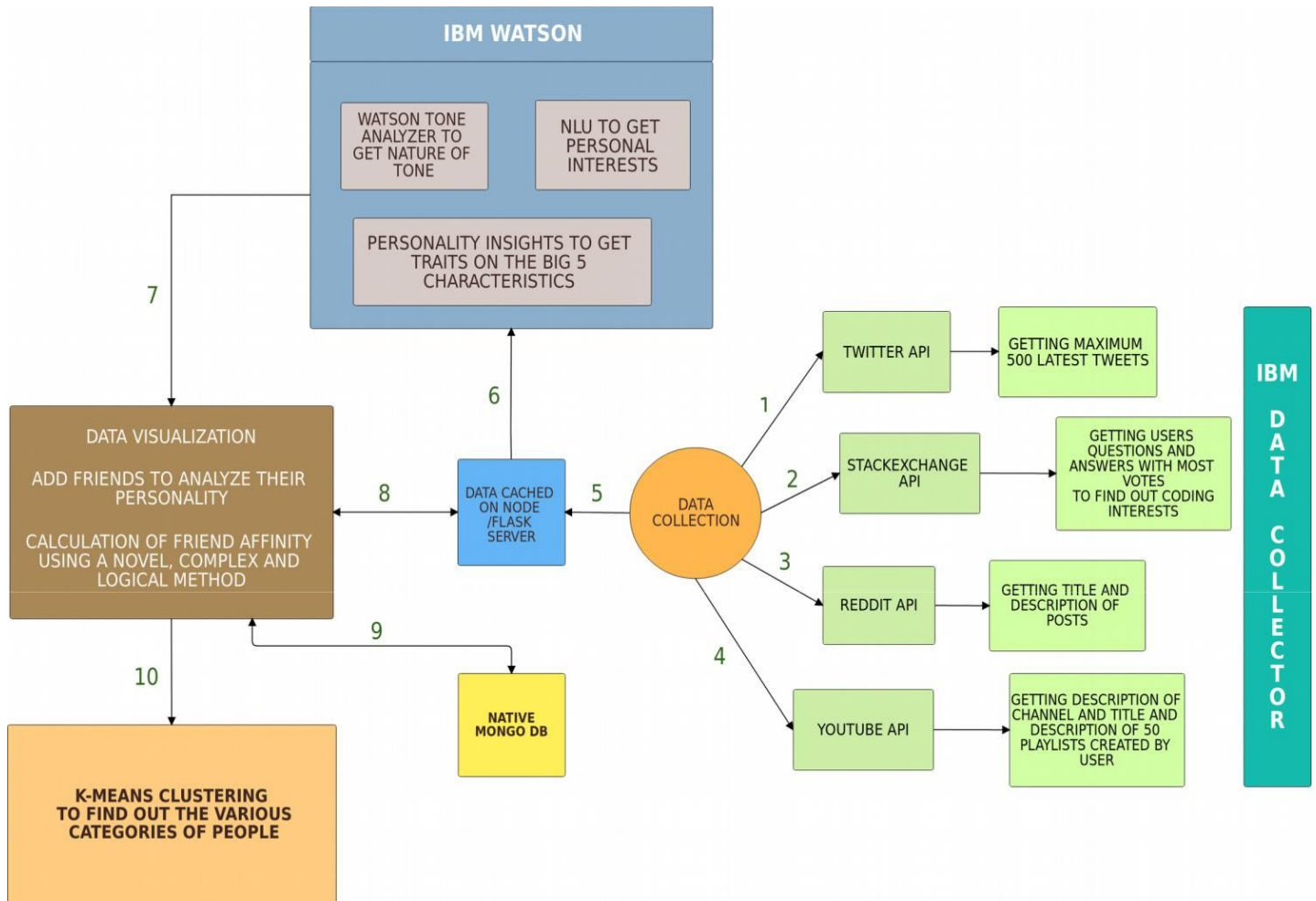
*Let  $y = 2 * (\text{Number of common interests}) - (\text{Number of uncommon interests})$*

$$\text{Then, Affinity} = \frac{(e^{\frac{y}{(\text{Total number of interests})}} - e^{-1})}{(e^2 - e^{-1})}$$

The constants are used to make the range of both the above affinities the same. The two affinities are combined by simply taking an average with hopes of giving same weightage to both.

The affinities of all provided friends of the user are calculated and run through a clustering algorithm to classify all the friends based on their affinities.

## Architectural flow -



## Technologies Used -

- Programming Languages - Node(Express framework) , Python(flask framework)
- Database - MongoDB
- Third-party APIs - Twitter API, Reddit API, StackOverflow API, Youtube API, Google Charts API, IBM Watson Natural Language Understanding, Tone Analyzer and Personality Insight

**Scope of Work -** The proposed solution would require development of the following modules :-

- **Data Collector -**  
It will be a server which collects data from different social media websites like Twitter, Reddit, StackOverflow and so on. It will also cache the data for further analysis.
- **Data Analyzer -**  
It will use the data aggregated by the collector and use IBM Watson services like Natural Language Understanding, Tone Analyzer and Personality Insight to analyze it. This module will also cluster the friends according to their affinities.
- **Frontend UI -**  
This module will be the one which interacts with the user. It will forward the data submitted by the user to the collector and display the data it receives from the analyzer.

**Impact of Solution on Business -** The above approach is very easily deployable and expandable. Adding new social media websites as data sources is a simple task. It can be used by people to find commonalities with their friends. Further, it can be used by businesses to find commonalities among potential employees and an ideal one. The concept can be extended to analyse any person's behaviour using their social media handles. Thus the solution has the potential to better relations between friends, help businesses to find the right employees and help people better understand each other in working and informal environments.