# Interest and Sentiment Analysis with Recommendation System

**Ayush Gupta . Debajyoti Halder . Pragyanshu Khare . Rishabh Singh**

## Statement of Purpose

**The Goal !**

Recommendation Systems of products of our daily use have always been black boxes for us. However everyone has a faint idea that recommendation systems are fueled by our interests, sentiments, “habits” of using a particular product. These three entities or elements can be referred to as Data or in basic programming terminology “input” for the program of the recommendation system. Through this project we tried to analyze the “Data” (except the habits), generate various visualizations and build a basic recommendation system. Our goal was to develop one tool that does it all - data collection, cleaning, analysis, visualizations and also recommend media.

## Data and Analysis

**The Input and the Program !**

We have used Twitter as our source of Data. Tweets, retweets contain various hints of what a user might be interested in (interest), what he/she might be thinking (sentiment).

We made various tweet searches - user based, topic (hashtag) based. A location filter helped us make various different analysis on a particular in a geographical region.

Links to our example datasets:

- Leonardo DiCaprio tweets
- #spacex tag tweets
- #TrumpvsBiden tag tweets

- Elon Musk tweets
- Airlines tweets

Every dataset has been downloaded and cleaned and processed through one tool. In every dataset there's a column of sentiment which signifies the sentiment of a particular tweet. The airlines dataset has been collected. It has columns of latitude and longitude which refers to the location from where the tweet was posted.

We've used textblob to analyse the tweets for sentiments, tf-idf to sort out the top most interests of a user and we've further used those topics to search videos using the YouTube API to suggest videos. Videos returned by the API was passed through a simple scoring function to make the suggestion more appropriate.

Normally analysis like this is done with static data - which has been collected for many days and processed to get results on sentiments or interests to be used later. However we wanted to make a tool which would use dynamic data and analyze them live. This way a tweet made even seconds ago can be used for analysis.

## Visualizations

**The Output !**

Our tool is made using Streamlit and hosted on Heroku. We have shown a number of visualization on user requested data fetched from Twitter using Twitter API. The attached Jupyter Notebook also has some visualizations specifically on trends of likes, dislikes, retweets of tweet topics or users.

Link for the tool : https://tweet-analyzer-alpha.herokuapp.com/
The Airlines Tweet analyzer tool - https://airline-tweet-stat-app.herokuapp.com/