**PEARL -  An Interactive Visual Analytic Tool for Understanding Personal Emotion Style Derived from Social Media**

CSE 578 Project Report

Sanika Shah  
 Arizona State University  
 Tempe AZ USA  
 [ssshah33@asu.edu](mailto:ssshah33@asu.edu)

Mugdha Kolhe  
 Arizona State University  
 Tempe AZ USA  
 [mkolhe@asu.edu](mailto:mkolhe@asu.edu)

Divya Prakash Sivakumar  
 Arizona State University  
 Tempe AZ USA  
 [dsivaku2@asu.edu](mailto:mkolhe@asu.edu)

Tanmay Manolkar  
 Arizona State University  
 Tempe AZ USA  
tmanolka@asu.edu

1. **Introduction**

The digital footprint that a social media user leaves online can be utilised in many ways to gain insight into the way he/she thinks and behaves. This information can be used by several digital marketing industries to be more successful in the handling of customers and improving customer experience by understanding what triggers mood changes in customers.

The system we have implemented leverages the same and uses a twitter users past tweets to determine his personal emotional style. The PEARL system[1] is an interactive timeline based visual analytic tool which was implemented by Jian Zhao et al, which we implemented and extended.

For our implementation, we have used the tweets from popular twitter users and nation leaders, Narendra Modi, Donald Trump, and Barack Obama from June to September, 2018.

**2. Motivation**

The Pearl system seemed interesting to us because of its impressive, easy to understand visualisation and the thorough analysis done by Jian Zhao and his team on the twitter data.

There are numerous applications based on twitter data since it is so vast and contains the candid views of a social media user. The Pearl system is one of the most efficient

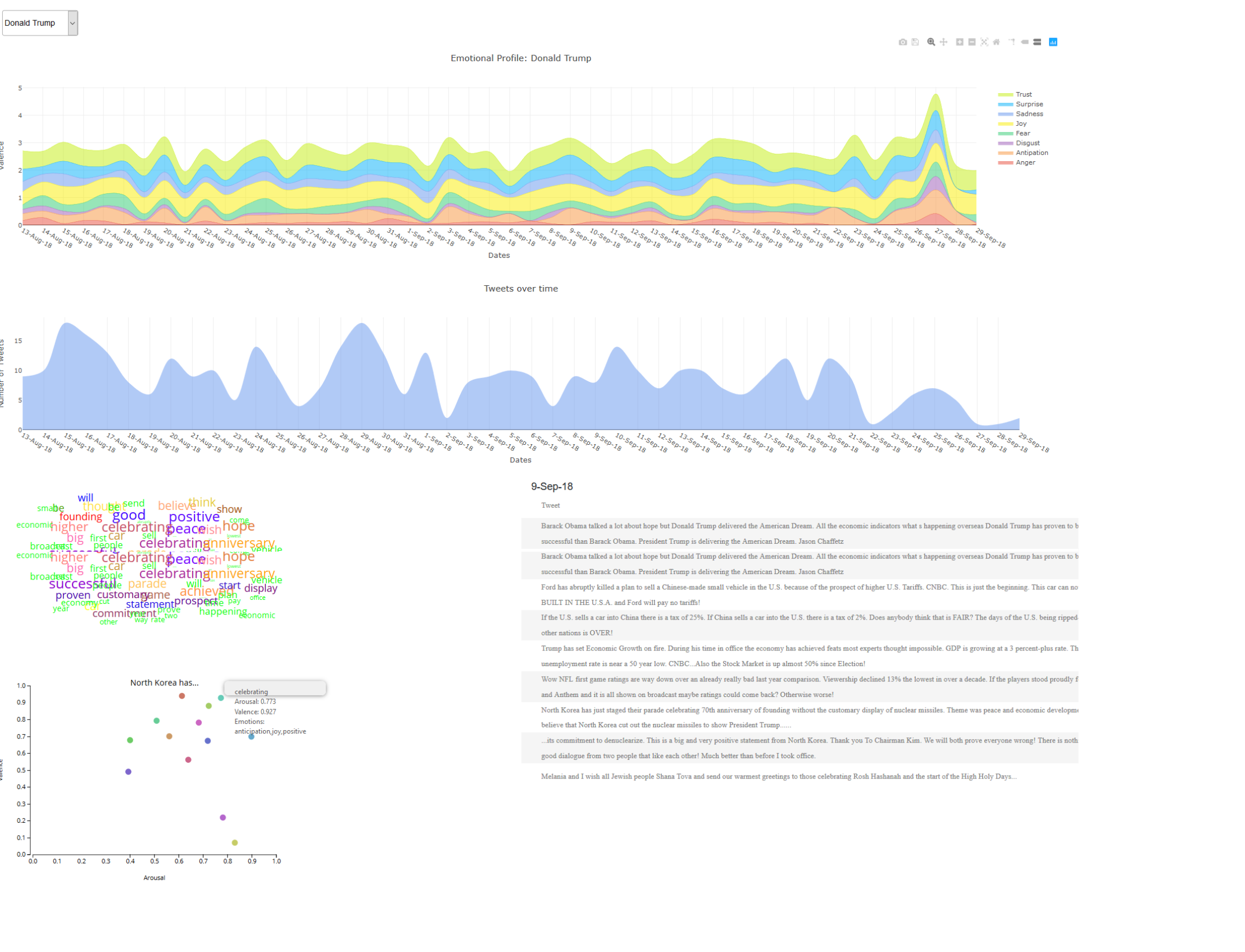
analysis we had seen and hence, chose it as the base paper for out system.

As mentioned above, several product developers can benefit from this kind of system that provides an insight into the psyche of the social media user. Customer relation management such as chat bots, customer service and developing a product that will catch the ned users fancy are all applications of this system and that was our motivation to develop this system.

**3. Working**

Pearl system is used to view the emotional status of the tweets for users. Three datasets are used in the system which are tweets from Narendra Modi, Barack Obama and Donald Trump ranging from June 2018 to September 2018. We used two different model for emotion detection, Plutchik[2] and VAD Model[3]. Plutchik Model provides the tweets based on 8 different emotions like Joy, Fear, Surprise, Trust, Sadness, Disgust, Anticipation and Anger. VAD model provides the Valence, Arousal and Dominance scores for all words of tweets.

We used NRC lexicon[4] based approach for finding the word wise VAD scores and emotion categories for tweets. NRC lexicon dataset is used to find the word wise VAD scores and emotions. Tweet preprocessing includes calculation of all VAD values and emotion categories for words and then averaging them out for tweet and days.

****

**Figure 1: Pearl System**

To plot the stacked area chart, we averaged out the valences scores with the emotions for all tweets posted on a particular date. This chart uses date wise averaged valence scores for each emotion.

Area chart that is shown in blue color above displays the number of tweets posted by user over time. The peaks in the chart can be viewed as dates of significant events. Clicking on the area chart provides the raw tweets view and the word cloud for the particular day. Raw tweets view displays all the tweets in a table for that date. Word cloud shows the significant emotional words of the tweets that contributed for categorization of the tweet. It’s done using D3[5] cloud layout library.

Clicking a tweet in raw tweets view displays the scatter plot which plots the words of that tweet with their Valence, Arousal scores and emotions. This plot is helpful to view the important words of the tweet with its VAD scores.

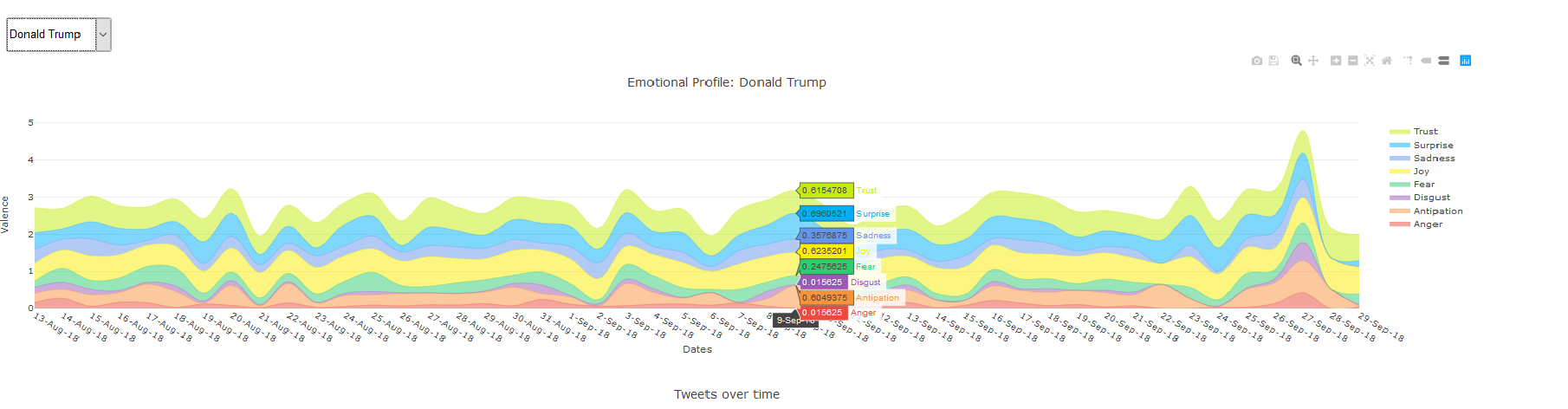
Extension of our project is a RNN based approach for Multi line chart. The paper mentions one of its limitation is the use of Lexicon Based Approach which doesn’t

consider negative words in the tweet(Example: Not Happy). Here Not and Happy are two different words

which gives a different emotion categorization of the tweet. To solve this limitation, we used a trained RNN based Plutchik model[2] using keras to get the probability values for 8 emotions for every tweet. This considers the entire sentence for emotion detection rather than word wise unlike Lexicon based approach. Multi line chart shows week wise count of tweets for 8 emotions. RNN based approach improved results of the lexicon based approach.

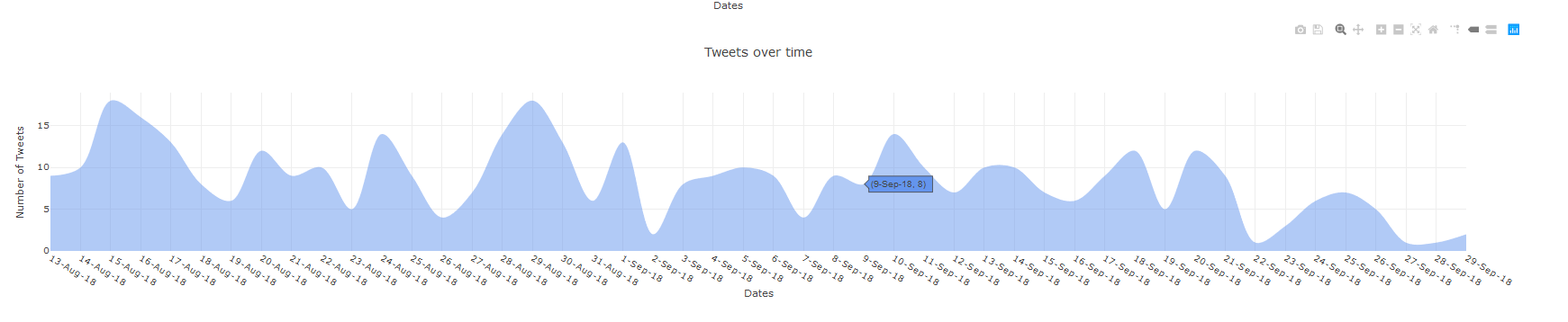
**4. Visualization Design**

We have a total of six components in our system. These visualisations were all implemented using D3[5] and javascript.



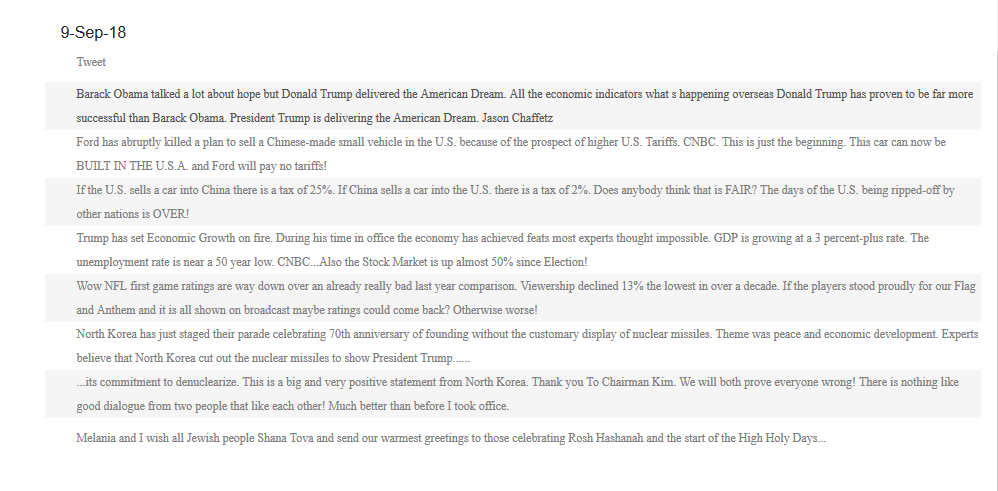
**FIgure 2: Stacked Area Graph shows the Donald Trump emotional profile.**

Stacked area graph ( Figure 2 ) shows the overview of a user’s emotional profile over four months. Each band represents an emotion as shown on the legend based on Plutchik’s emotion wheel[2].



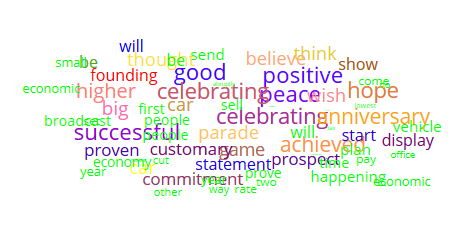
**Figure 3: Area chart shows number of Tweets over time**

Area graph ( Figure 3 ) shows the number of tweets over time. Clicking the date displays the raw tweet view and word cloud.



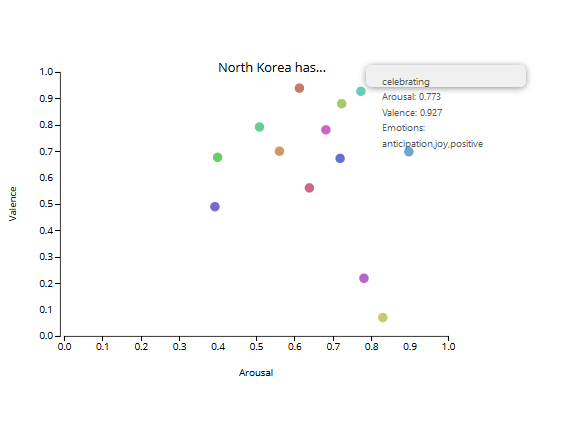
**Figure 4: Raw Tweet View shows the list of tweets**

Raw tweet view ( Figure 4 ) displays all the tweets of the user for the given date (that we get on clicking area chart). Clicking a tweet in this view displays a scatterplot of words for that tweet.



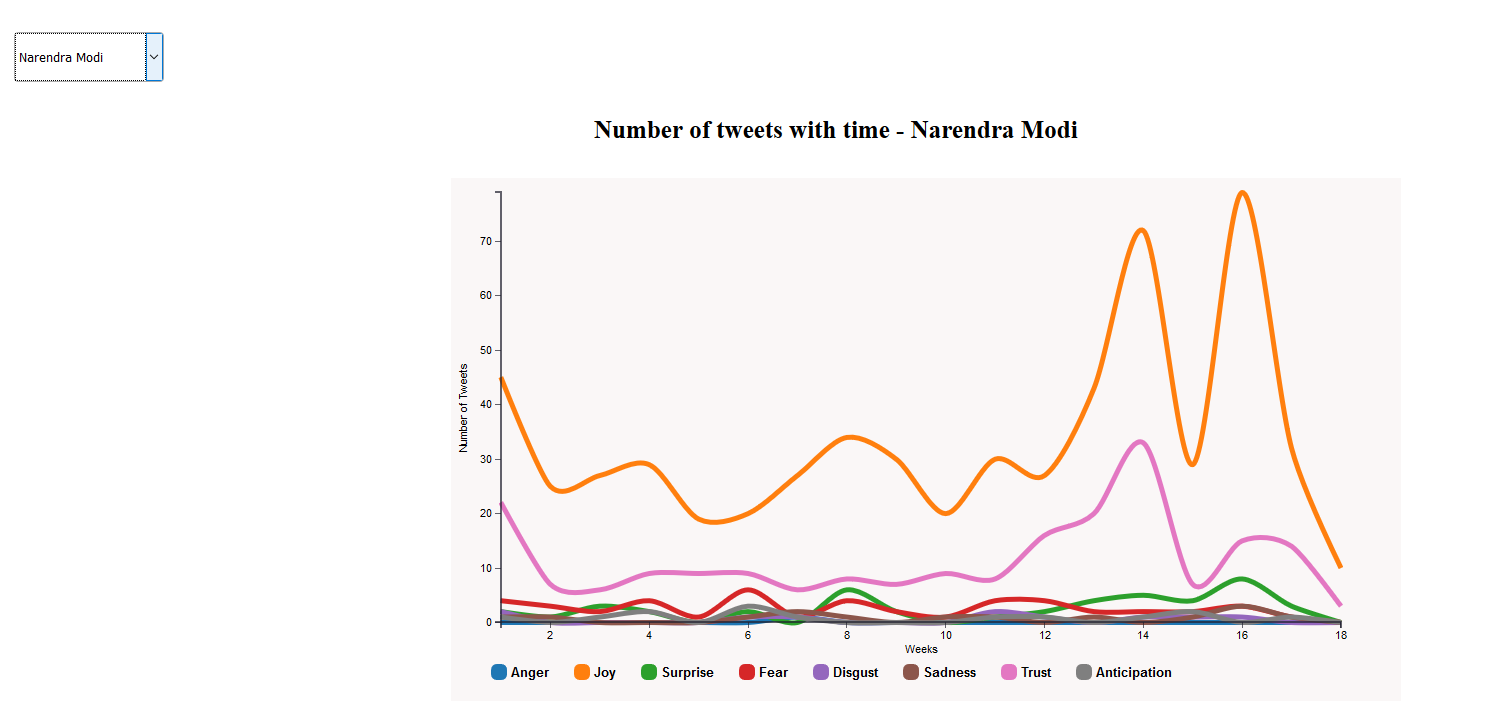
**Figure 5: Word Cloud shows the significant words of the tweets.**

Word Cloud ( Figure 5 ) highlights significant words of the tweet that are important for the categorization of the tweet. It displays words from the tweets posted on particular day.



**Figure 6: Scatter plot shows the VAD scores and word emotions for each word of the tweet.**

Scatter plot ( Figure 6 ) is based on VAD scores[3], clicking a tweet in Raw Tweet View displays this charts with circles representing words and X axis and Y axis representing Arousal and Valence values respectively. Hovering a point gives VAD scores and emotions of that word.



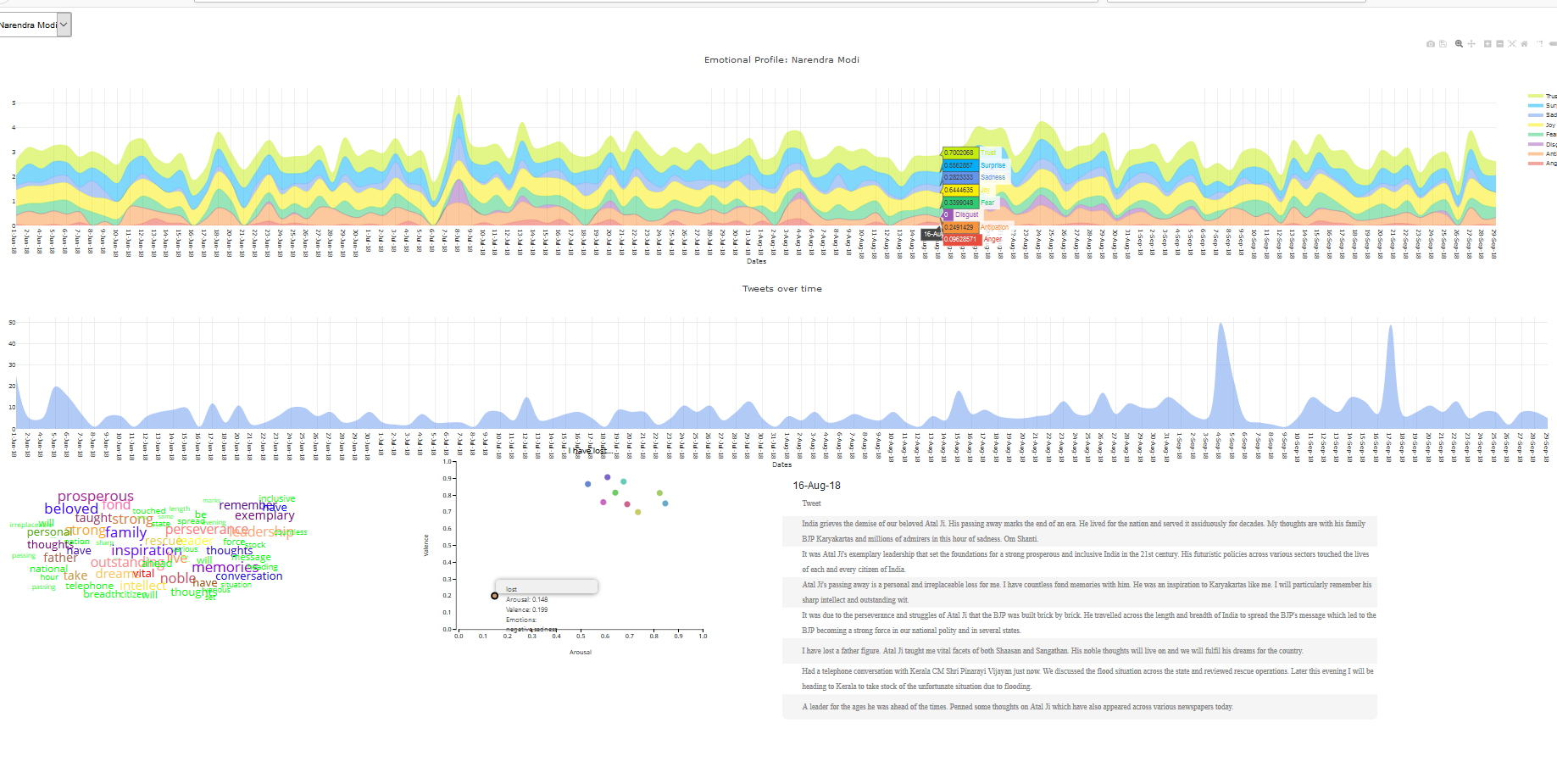
**Figure 7: Multi Line Chart shows the number of tweets for each emotion over weeks**

For the Extension Graph ( Figure 7 ), we used RNN based approach to find the emotions of the tweet. We plotted a multi line chart which shows number of tweets for every emotion over time. Mouse hover the line will display the number of tweets of the particular week.

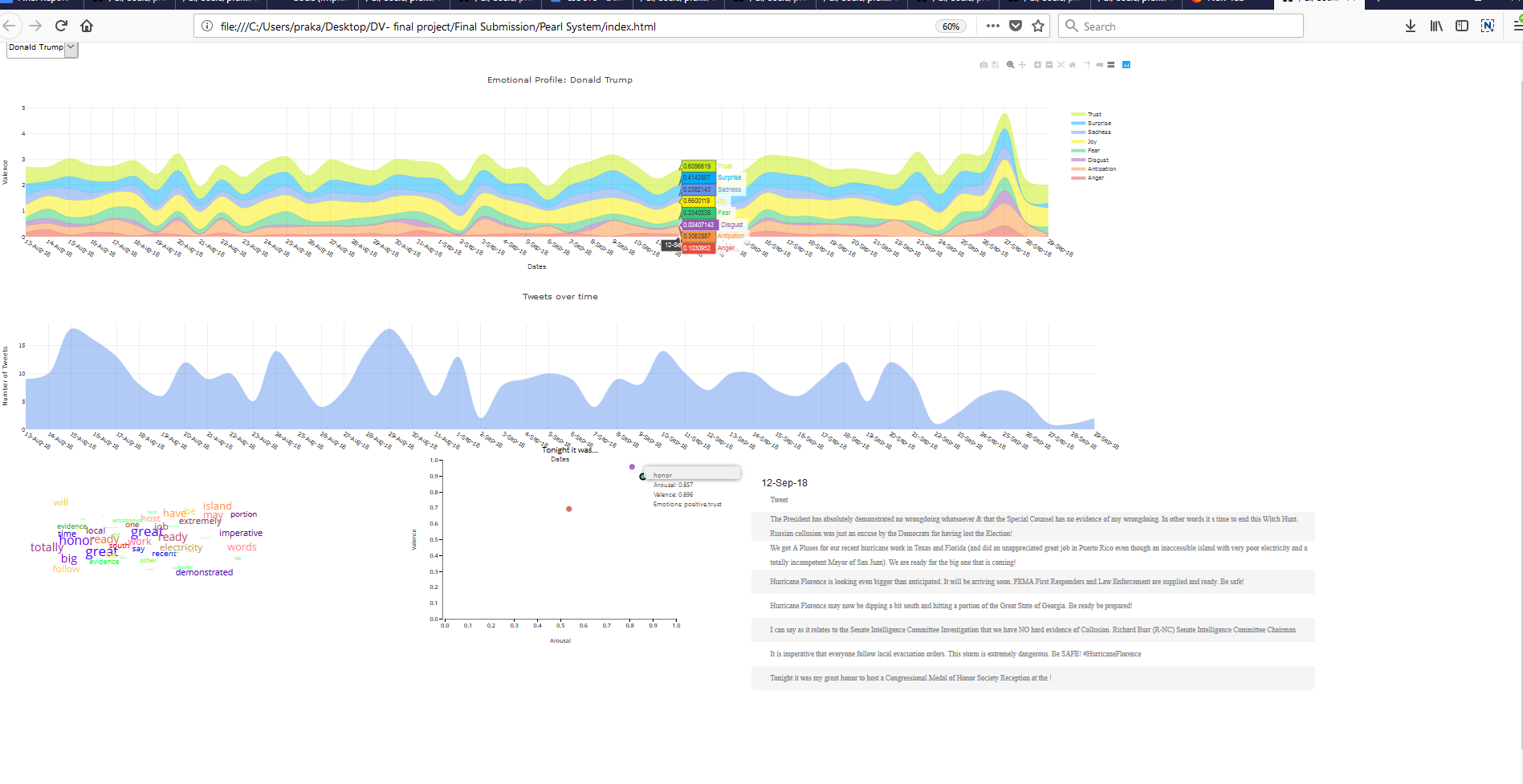
**5. Case Studies**

**5.1. Narendra Modi:**

For the first case study, we will be showcasing the emotion profile of Narendra Modi on 16th August, 2018. Figure 8 shows the system’s status on that day for the tweet: “I have lost a father figure. Atal Ji taught me vital facets of both Shaasan and Sangathan. His noble thoughts will live on and we will fulfil his dreams for the country.” On this day, Atal Bihari Vajpayee, an Indian politician and former prime minister passed away. The effect of this sadness can be seen in the graphs. The word lost represent negativity and sadness in scatterplot.

****

**Figure 8: Narendra Modi’s emotional profile on 16th August, 2018.**

****

**Figure 9: Donald Trump’s emotional profile on 12th September, 2018.**

**5.2. Donald Trump:**

For the second case study, we will be showcasing the emotion profile of Donald trump on 12th September, 2018. Figure 9 shows the system’s status on that day for the tweet: “Tonight it was my great honor to host a Congressional Medal of Honor Society Reception at the !”, which was a joyous occasion for him. The effect of this joy occurrence can be seen in the first graph as well. The word honor represents trust and positivity in scatter plot.

**6. Discussions and Future Work**

**6.1 Discussions**

1. This system helps to know about the emotional profile of user and the tweets posted daywise. It helps to see the important words of the tweets through word cloud. Scatter plot helps to see the word of each tweet with its VAD scores and emotions.
2. Lexicon based approach has few disadvantages where emotion of the tweet is categorized word wise. This has a disadvantage since it doesn’t take into account negative phrases like not happy.

**6.2 Future Work**

1. Real time analysis of tweets could be an extension. Once the user has tweeted we could use the tweet and update the emotional analysis of the system runtime.
2. More users data can be used for analysis which provides different emotions for different users. It helps us to see how their emotions vary for some given common event in time.
3. We could run a emotion analysis on a trending topics using the tweets from particular hashtag and view the emotions and the way it affects the users.

**7. References**

# [1] Jian Zhao, Liang Gou, Fei Wang, and Michelle Zhou. PEARL: An Interactive Visual Analytic Tool for Understanding Personal Emotion Style Derived from Social Media. IEEE Conference on Visual Analytics Science and Technology (VAST), 2014.

# [2] R. Plutchik. The nature of emotions. American Scientist, 89(4):344, 2001.

# [3] A. Mehrabian. Basic Dimensions for a General Psychological Theory. Oelgeschlager,Gunn & Hain Inc., 1980.

# [4] S. M. Mohammad and P. D. Turney. Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon. In Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, CAAGET’10, pages 26–34, 2010.

# [5] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-driven documents. IEEE Transactions on Visualization and Computer Graphics, 17(12):2301–2309, 2011.