

## ABSTRACT

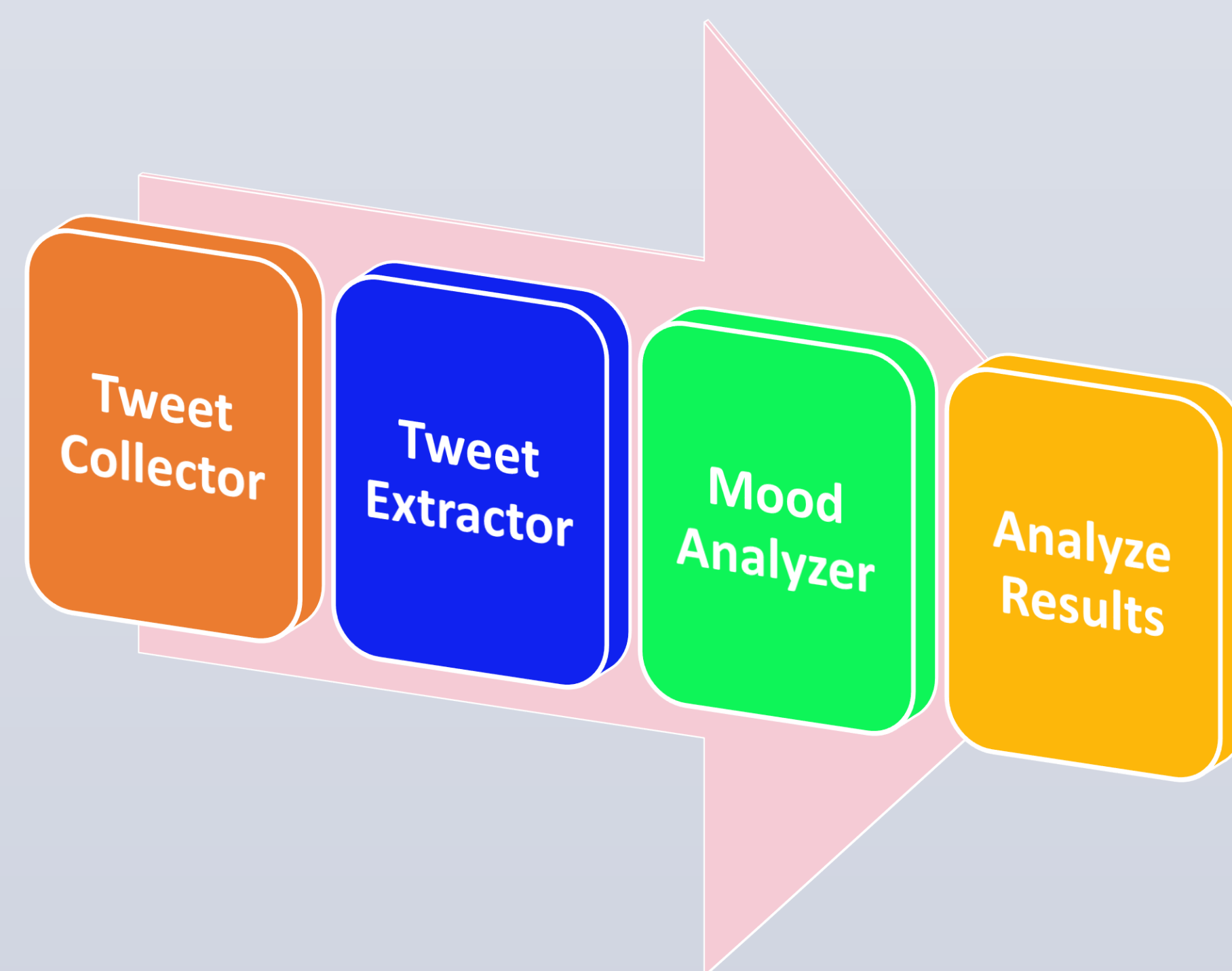
Since 2006 Twitter has been a platform that allows users to tweet no more than 140 characters. Tweets include daily conversations, information sharing, news critiques, and updates about a Twitter user's life. Thus Twitter promotes a wide array of emotional expression. In this research Twitter is queried for tweets and these tweets are used to identify users who might be in need of psychological help. To begin analysis, a collection of personal, expressive tweets will be gathered. These collected tweets will contain content where the Twitter user appears to be sincerely writing about their depression. Analysis will be done by using ANEW word list and other algorithms to compute Mood Quotient which can be used in user identification.

## MOTIVATION

A lot of papers have been published on how Twitter data could be used to capture human mood/sentiments. There are also quite a lot of interesting projects done driven by Twitter data. Some of them include sentiment analysis over a topic/event, behavioral patterns across regions, stock market predictions, etc. Most of these implementations capture the mood/sentiment of the crowd. We intend to concentrate on mood fluctuations of an individual rather than a crowd. Analyzing an individual's mood could be most valuable in early identification of mental health problems.

## SYSTEM DESIGN

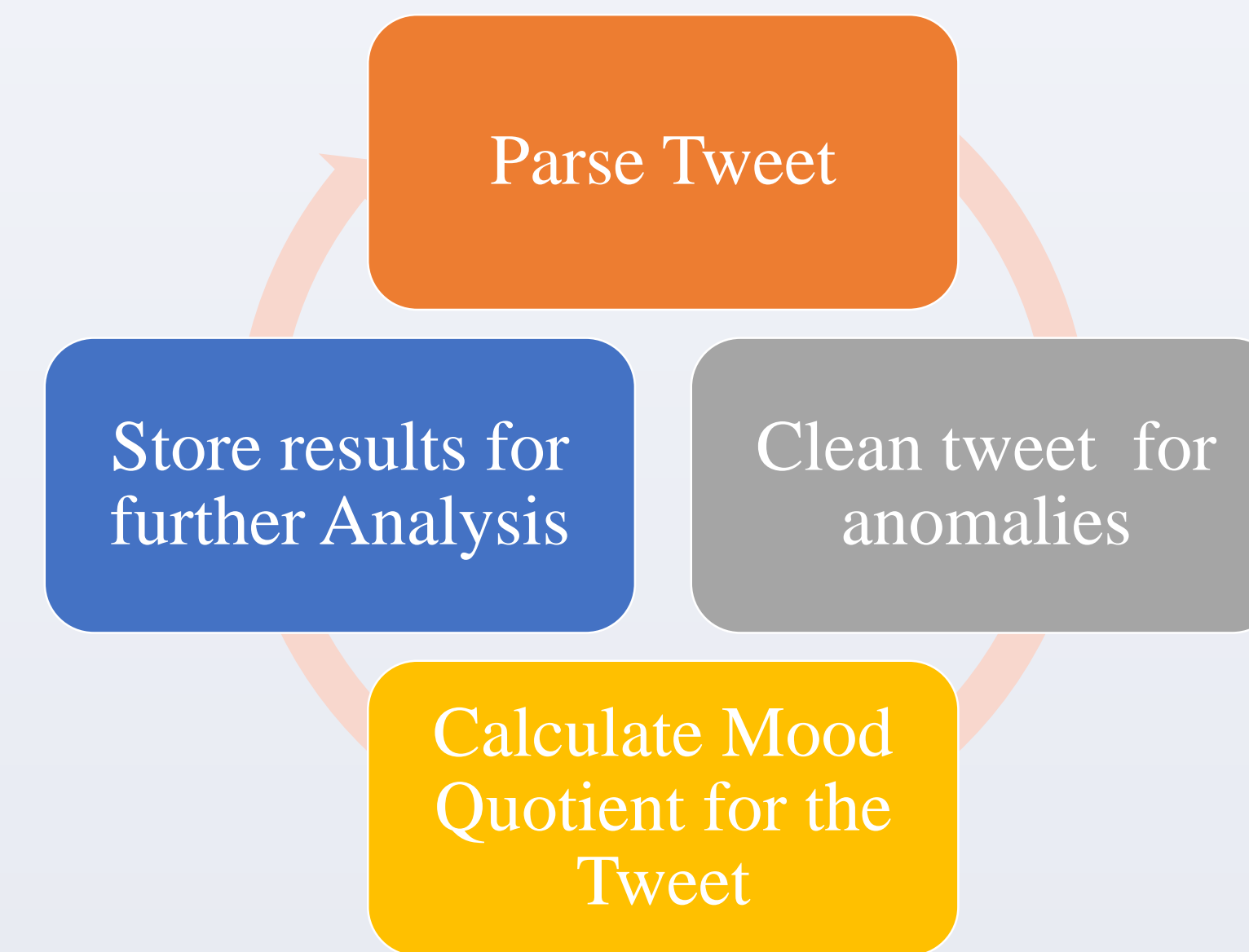
- ❖ **Tweet Collector:** Responsible for downloading tweets using Twitter API. Output is a huge set of twitter data in JSON format.
- ❖ **Tweet Extractor:** Responsible for parsing tweets and extracting relevant tweet data for analysis.
- ❖ **Mood Analyzer:** Key component of the whole design. Parses a tweet and calculates Mood Quotient using ANEW world list ratings for words in the tweet. Multiples tweets with lower Mood Quotient pertaining to a single user are identified.
- ❖ **Result Analyzer:** Provides analysis of the twitter data set that is used as input.



The design includes data cleaning involving purging of tweets with irrelevant data like tweets in local languages, junk text, anomalies etc.

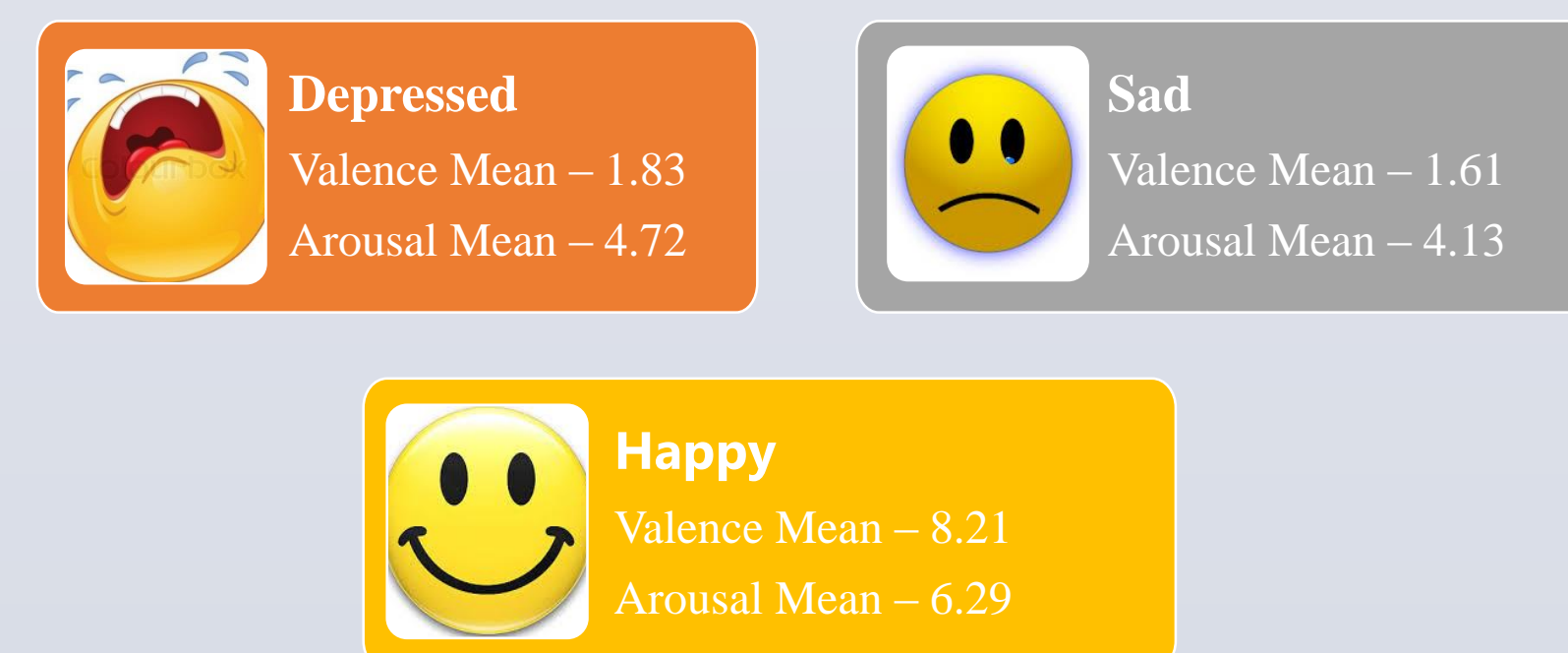
## ARCHITECTURE & IMPLEMENTATION

Tweet Life Cycle in the system:

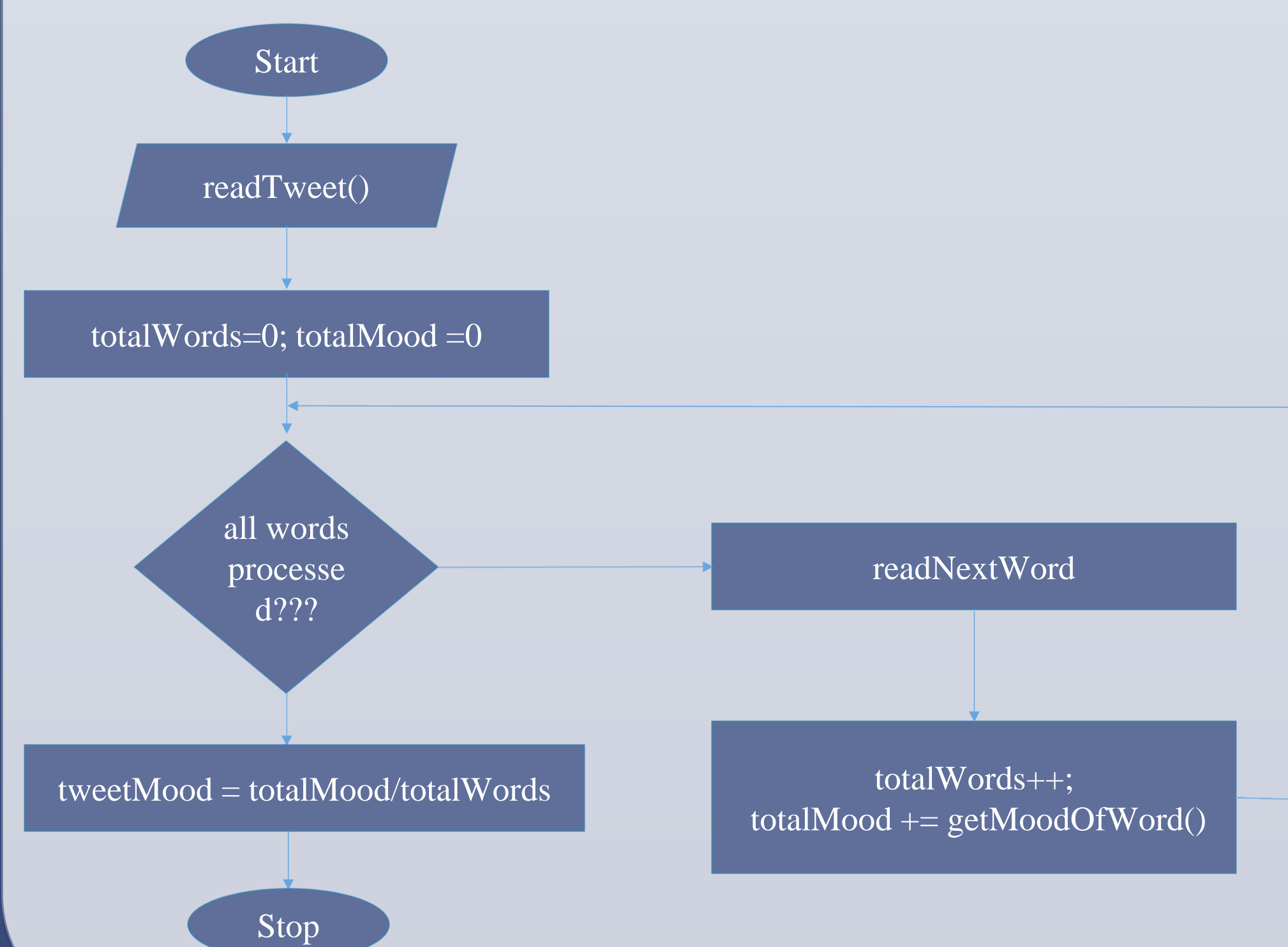


- ❖ Twitter data used to do analysis. Other social web data like Facebook etc. could be integrated as well.
- ❖ Analysis is done by collecting bulk tweets and rating them according to ANEW list which has a quotient of positivity or negativity related to each word based on Valence and Arousal Mean.
- ❖ The ANEW word list provides a set of emotional ratings for a large number of words in the English language. The words have been rated in terms of pleasure, arousal and dominance.
- ❖ Two values called Valence Mean and Arousal mean are used to represent the emotional quotient value for each word.
- ❖ The mean values are given on a scale of 1 to 9. These values are scaled to 100 for computational flexibility.

Some word rating samples:



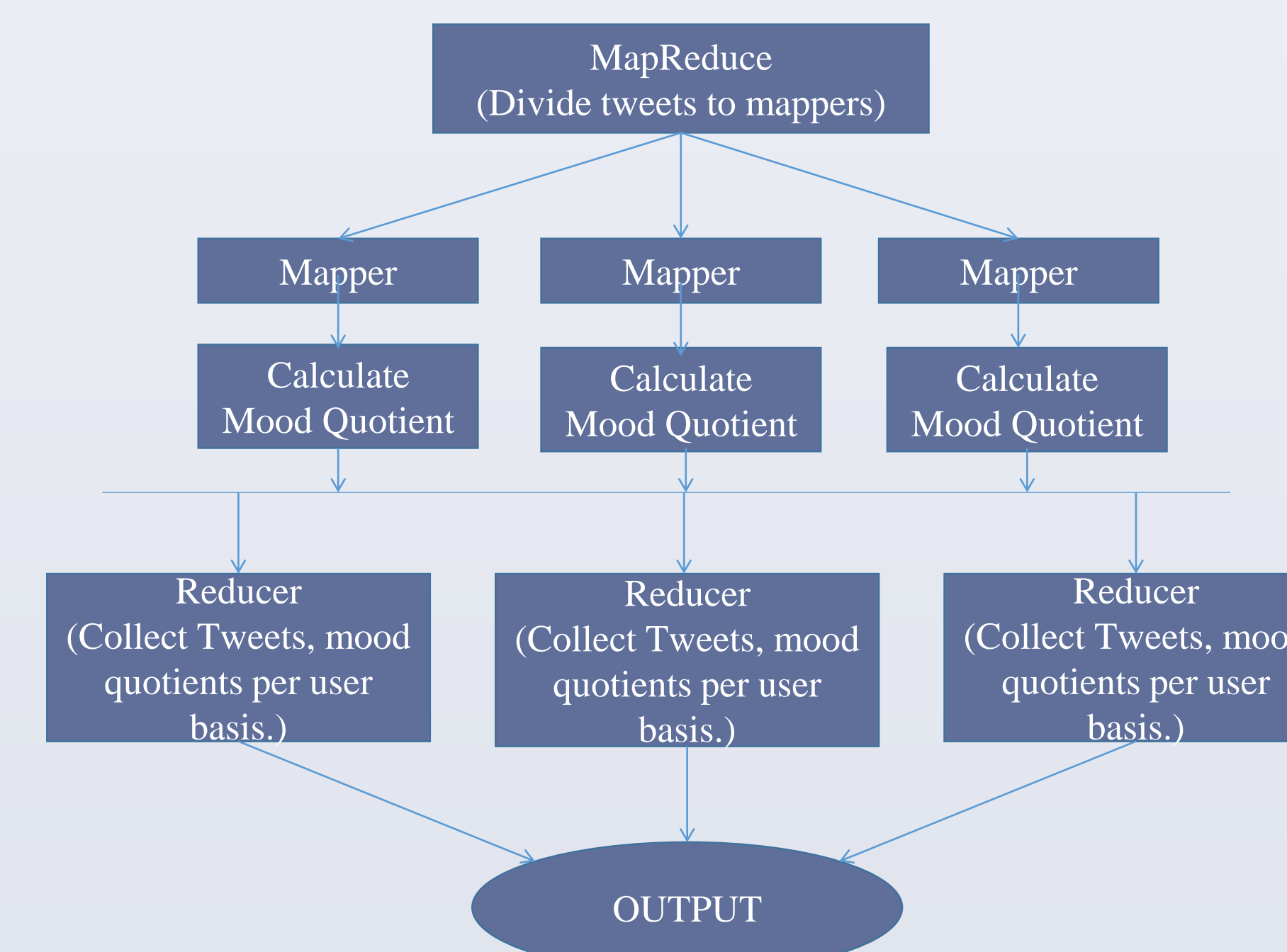
Mood Quotient calculation for a Tweet:



## HADOOP MAP-REDUCE

- ❖ The main MapReduce framework of the project resides in this module of Mood Analyzer .
- ❖ The tweets are divided among the mappers available and the tweets are processed individually at each mapper.
- ❖ The Hadoop's MapReduce framework helps in this division of tweets to each mapper.
- ❖ Tweets per user basis will be collected at Reducer and gives the output.

Map Reduce sample diagram



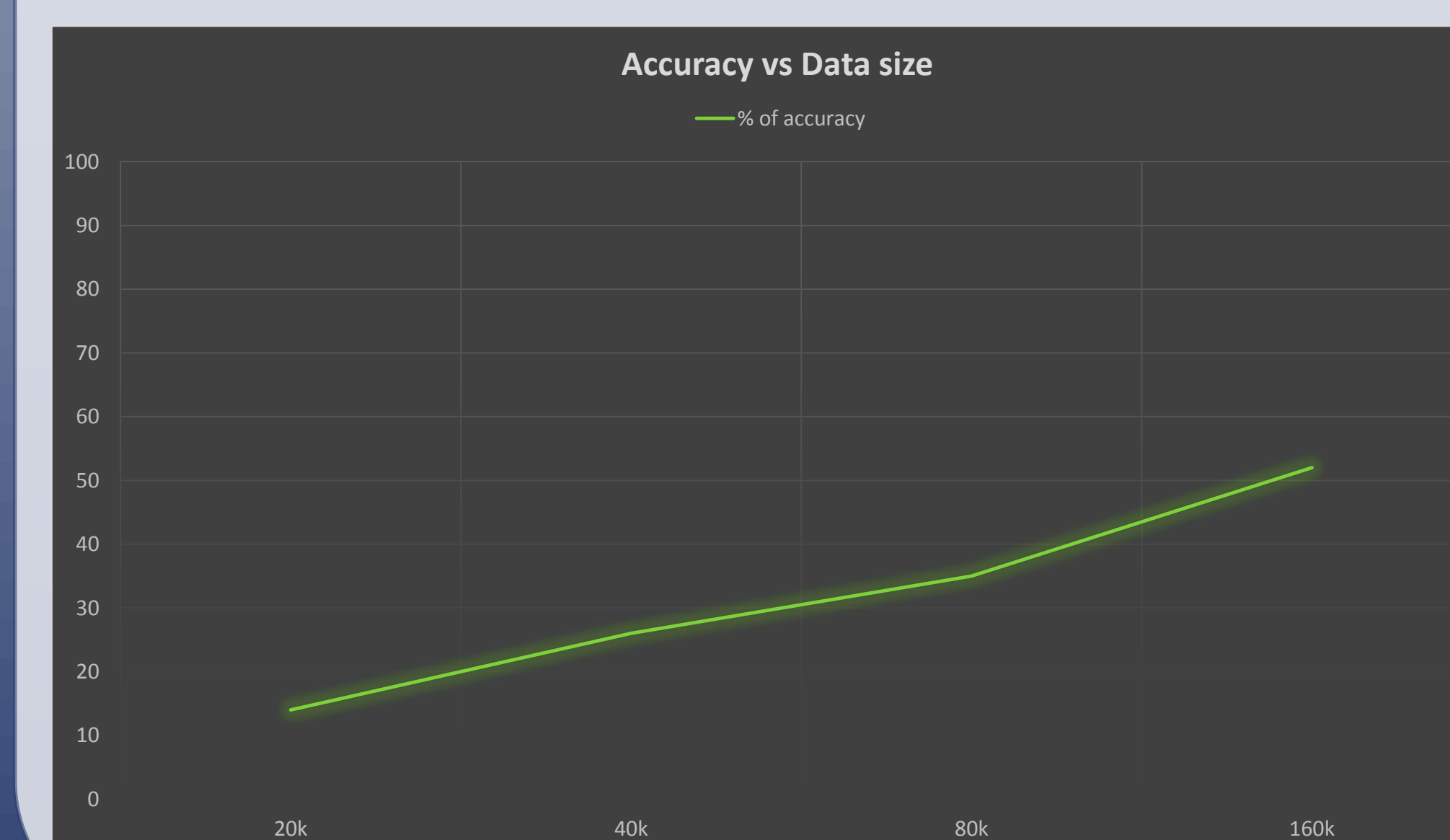
## RESULTS

Sample tweet mood quotient calculation

Tweeted text: RT @nicolehasedge: i feel really worthless. sad. unimportant. insignificant. low. depressed. hurt. used. shitty. scummy. useless.  
Time of creation: Tue Apr 01 06:57:07 +0000 2013  
User: marijamari\_  
Mood Quotient = 17

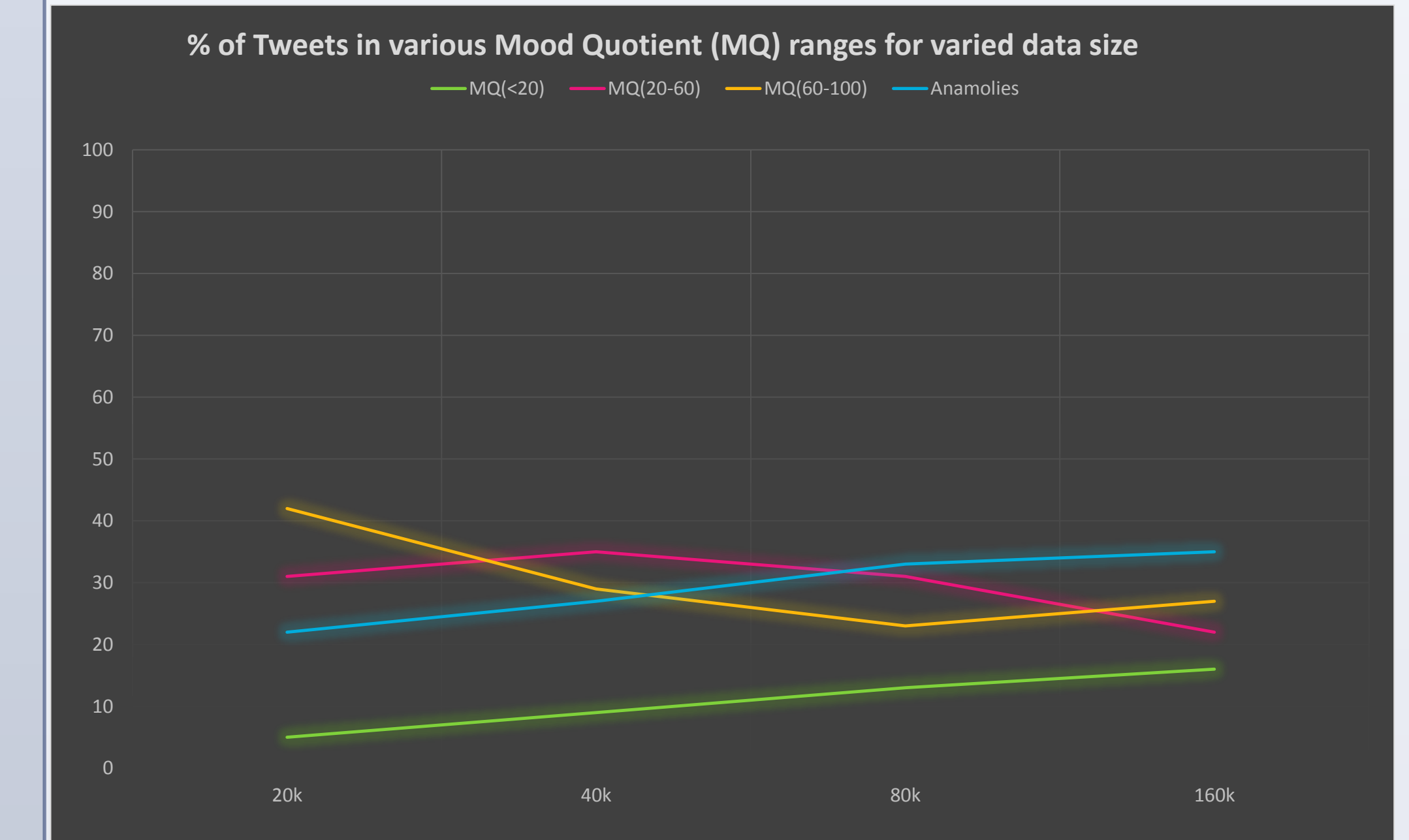
Tweeted text: sad mad sad mad  
Time of creation: Tue Apr 02 07:10:42 +0000 2013  
User: mayabrebs  
Mood Quotient = 19

- ❖ Precision variation with data size used



## RESULTS Contd...

- ❖ Mood Quotient ranges with varying data set sizes



## CONCLUSIONS & FUTURE WORK

- ❖ Twitter data could be effectively used to predict early mental health problems with a certain degree of precision.
- ❖ Understanding an individual's mood can be of most value when studied over a period of time.
- ❖ More sophisticated Lexicon construction will lead to improved precision in the prediction process.
- ❖ Implementation of a scalable system that can process huge twitter feed spanning across several month to years could help analyze user patterns and behavior.
- ❖ Crowdsourcing techniques along with automated algorithms could produce more accurate results.

## REFERENCES

- Motivation  
<http://www.cs.uccs.edu/~jkalita/work/reu/REU2011/FinalPapers/Billot.pdf>  
 ANEW Word List  
<http://arxiv.org/pdf/1103.2903v1.pdf>  
<http://csea.php.ufl.edu/media/anevmessage.html>  
 Hadoop-MapReduce Framework  
<http://hadoop.apache.org/>  
 Amazon Web Services  
<http://aws.amazon.com/ec2/>  
<http://aws.amazon.com/s3/>

## ACKNOWLEDGMENTS

Dr. Daisy Zhe Wang  
Computer and Information Science and Engineering,  
University of Florida

Michael Billot  
University of Colorado at Colorado Springs

## CONTACT

Harsha Vardhan Galla - [harshavgalla@ufl.edu](mailto:harshavgalla@ufl.edu)  
 Syam Sundar Rao Kolla - [syam@ufl.edu](mailto:syam@ufl.edu)