AMERICAN UNIVERSITY OF BEIRUT

AN AFFECTIVE DATA SCIENCE APPROACH FOR SPORTS RELATED TWEETS

by
MOHAMAD YAHYA KAMAREDDINE

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Science
to Computational Science Program
of the Faculty of Arts and Sciences
at the American University of Beirut

Beirut, Lebanon
March 2017

AMERICAN UNIVERSITY OF BEIRUT

AN AFFECTIVE DATA SCIENCE APPROACH FOR SPORTS RELATED TWEETS

by
MOHAMAD YAHYA KAMAREDDINE

Approved by:

Dr. Mariette Awad, Associate Professor Department of Electrical & Computer Engineering	Advisor
Dr. Nabil Nassif, Professor Department of Mathematics	Member of Committee
Dr. Mazen Al-Ghoul, Professor Department of Chemistry	Member of Committee

Date of thesis defense: March 30, 2017

AMERICAN UNIVERSITY OF BEIRUT

THESIS, DISSERTATION, PROJECT RELEASE FORM

Student Name: Kamareddine Mohamad Yahya
Last First Middle

✓ Master's Thesis ◯ Master's Project ◯ Doctoral Dissertation

☑ I authorize the American University of Beirut to: (a) reproduce hard or electronic copies of my thesis, dissertation, or project; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes.

☐ I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of it; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes
after : **One ---- year from the date of submission of my thesis, dissertation, or project.**
Two ---- years from the date of submission of my thesis, dissertation, or project.
Three ---- years from the date of submission of my thesis, dissertation, or project.

[signature] 4/4/2017

Signature Date

ACKNOWLEDGMENTS

Special thanks are for Dr. Mariette Awad, Prof. Nabil Nassif, and Prof. Mazen Al-Ghoul for their great help. In addition, I would like to thank Yara Rizk for her support and contribution.

My recognition and gratitude are addressed to the American University of Beirut for its financial support.

AN ABSTRACT OF THE THESIS OF

Mohamad Yahya Kamareddine for Master of Science
Major: Computational Science

Title: An Affective Data Science Approach for Sports Related Tweets

With the richness of data and information, particularly for sports in social media, especially Twitter and Facebook, natural language processing could identify subjectivity and objectivity of phrases in one hand, and could extend analysis to pinpoint to the existence of consensus for these phrases on the other hand. However, finding the ground truth of a phrase or a sentence being subjective or objective, is complicated due to the underlying English context for some sentences that might embed different forms of subjectivity. Indeed, objectivity and subjectivity in phrases could be directly or indirectly established in a sentence. Moreover, humans could agree or disagree on different topics, in which each individual could express his/her opinions in a manner that might be with or against the considered topic.

Motivated to apply artificial intelligence to sports phrases classification, this work presents an innovative *sports related Twitter data analytics framework* wrapped in a graphical user interface (GUI). This framework classifies sport phrases as subjective or objective, taking an additional step to introduce a new consensus label for subjective phrases according to Twitter bloggers. Our proposed workflow preprocesses and analyzes sport phrases, before generating decision trees based on tweets to identify objective and consensus-subjective polarities. Experimental results on a homemade corpus of 1007 phrases, reached an accuracy of 88% on the test set for objectivity and subjectivity of phrases using our proposed set of syntactic and semantic features. The syntactic and semantic features were tested on five classifiers (KNN, SVM, NN, NB, and AB) and results were validated using various statistical measures. Moreover, applying the decision tree on the corpus recorded that 57% of the subjective phrases are with consensus, while 43% of them are without consensus. Furthermore, our framework extends analysis to identify that herding behaviors of bloggers lead to consensus on subjective phrases rather than objective ones. Additionally, the proposed approach recorded an accuracy of 72.7% using our set of semantic features on the Stanford dataset used by a similar existing tool (Tweenator), which recorded an accuracy of 86.3% using its sematic topic features.

CONTENTS

ACKNOWLEDGMENTS v

ABSTRACT vi

LIST OF ILLUSTRATIONS x

LIST OF TABLES xii

Chapter

1. INTRODUCTION 1

2. LITERATURE REVIEW 3

2.1 Twitter Related Sentiment Analysis 3

2.2 Positive,Negative, and Neutral Classifications 4

2.3 Objective and Sbjective Classifications 5

2.4 Summary of Existing Tools 6

3. METHODOLOGY 12

3.1 Implementations 14

3.1.1 Phrase Input 15

3.1.2 POS Tagger 15

3.1.3 Keyword Extraction 16

3.1.4 WordNet 18
3.1.5 Twitter Search API 20
3.2 Computations 25
3.2.1 Pre-Processing Tweets 25
3.2.2 Tweet Features 29
3.2.3 Objective vs. Subjective Phrase Classification. 35
4. Results. 38
4.1 Clustering Results. 38
4.2 Phrase Classification (Objective vs. Subjective). 46
4.3 Herding Effect on Consensus. 50
4.4 Comapring Results with Tweenator. 56
5. Application 58
5.1 User Login. 58
5.2 Graphical User Interface (GUI). 60
5.3 Clustering Pie Chart 62
5.4 Graphical Plot 64
5.5 Tweets Visualization. 64
6. Conclusion 66

Appendix

1. SOURCE CODE 68

2. CORPUS 72

3. CLASSIFICATION FAILURES 118

BIBLIOGRAGHY 130

ILLUSTRATIONS

Figure		Page
1.	Subjective/Objective Classification and Tweet Herding	14
2.	Extracting Keywords Process	17
3.	Pseudo Code for Tag Count	19
4.	Query Handling Limit Length Exception for the Word Hold	19
5.	Nicknames Relations	20
6.	Protocol Workflow	21
7.	Twitter Search API Query Example	22
8.	Application Run Time	23
9.	Collect Tweets (5 Successful Queries).	24
10.	Collect Tweets (2 Successful Queries)	24
11.	JSON Instance	26
12.	Parsed Data	26
13.	Pre-Processing	28
14.	Subjective Phrases with/out Consensus	36
15.	Workflow of the Proposed Framework	37
16.	Herding Scheme	50

17. # of Followers and Retweets for Subjective and Objective Phrases 51

18. Connection Status 58

19. URL for Authorization 59

20. User Account Access 59

21. User PIN 60

22. Application Overview 60

23. Methodology(I), Methodology(II) 61-62

24. Clustering Pie Chart 63

25. Tweets Orientation 64

26. Raw Tweets 65

TABLES

Table Page

1. Existing Tools 6
2. Input Declaration 15
3. Stanford Tagger POS Tags 15
4. POS Tagger Examples 16
5. Tags Used for Keyword Extraction 17
6. WordNet Synonyms and Antonyms 18
7. Synonyms for Nouns from WordNet 20
8. Twitter Search API Parameters 22
9. 2-D and 3-D Emoticons 27
10. Samples of Slangs Lookup Table 28
11. Stemming Words 30
12. 10 Semantic Features 30
13. Senti-WordNet Scores 31
14. Semantic Feature Vector 32
15. 55 Syntactic Features 33
16. Objective and Subjective Phrases 36

17. Clustering Evaluation Measures 38

18. Datasets 39

19. Baseline Measures 40

20. Clustering with Supervised Datasets 40

21. Clustering Measures for Libras Dataset 41

22. Clustering Results for Dataset 1 (DTS1) 43

23. Clustering Results for Dataset 2 (DTS2) 44

24. Clustering Results for Dataset 3 (DTS3) 45

25. Classification Measures 46

26. Classifiers Parameters 47

27. Testing Results for Different Features on Several Classifiers 47

28. Some Misclassified Data (Syntactic + Semantic Features) 48

29. Herding in Subjective Phrases 51

30. Extracted Keywords for each Classified Subjective Phrase 52

31. Extracted Keywords for each Classified Objective Phrase 54

32. Naïve Bayes Classification Results on the Stanford Dataset 56

33. Miscalssified Data on Stanford Dataset 57

CHAPTER 1

INTRODUCTION

According to a recent study in 2016 [1], out of 3419 billion internet users, 2307 billion of them are considered to be active social media members. Internet supremacy and the invasion of social media (Twitter, Facebook, Instagram, etc.) has played a vital role in sentiment analysis. In accordance, different fields migrated to use such data (Social Media) and generated various categories of artificial intelligent agents and frameworks for data analysis purposes. For instance, [2] proposed a "filtering based model" to predict Twitter sentiments, while [3] introduced a "Twitter Opinion-Mining Framework" (TOM) for sentiment analysis and classifications. Furthermore, thousands of people use social media on daily basis to post their sport blogs in which they could insinuate the sense of consensus or not to other blogger posts regarding matches, players, sport events, etc. Yet, to authenticate the agreement or disagreement of bloggers on the sought sport topics, and to identify the concord of users to sport phrases, some statistical analysis should be conducted.

Focusing on sports data relies on the fact that billions of dollars are spent on sports gambling and betting. According to a recent statistical study of sports betting [4], the value of the sports gambling commerce is between 700 and 1000 billion U.S dollars. Moreover, gross gaming yield reached 452 billion U.S dollars in 2014, and forecasted to reach 511 billion U.S dollars in 2019 [5].

Motivated to design agents that perform syntactic and semantic sentiment analysis for identifying sport phrases as subjective with/out consensus or objective, which to our knowledge is still not investigated, we propose in this thesis a new phrase classification procedure for this task. Hereafter, we will use interchangeably the terms: consensus, agreement, and social media herding (leading to consensus).

This work targets only English phrases, specifically sport phrases, in which it preprocesses the provided input expressions to extract keywords based on Stanford POS Tagger [6]. Synonyms and antonyms are then specified using WordNet [7] to build a twitter search query to retrieve tweets using the Twitter search API [8]. Additionally, to enhance the search procedure, and to ensure a complete retrieval of related tweets, we look for some alternatives of players and team original names (nicknames) by [9] adding an "OR" combinational relation between them. Accordingly, tweets are clustered to agree or disagree with the specified phrase. Moreover, a set of syntactic and semantic features are used to classify phrases as objective or subjective, and a decision tree is structured to identify consensus in subjective expressions which could be indicative of herding effect in the social media as it is related to sports. To test our proposed phrase classification methodology, a dataset of 1007 phrases is manually constructed and annotated, and around 60,000 tweets related to these phrases are collected and filtered in about a month time (Sept, 2016). Grouping syntactic and semantic features together improved our accuracy in the objectivity and subjectivity analysis.

CHAPTER 2

LITERATURE REVIEW

This section mainly summarizes some existing approaches and tools on sentiment analysis for twitter data that are most related to our work. Table.1 summarizes other online tools regarding sentiment analysis.

2.1 Twitter Related Sentiment Analysis

Reference [10] proposed a web based tool (Sentiment 140) that classifies online streaming tweets as positive or negative. Training data was collected by a “scraper” while test data was collected by the twitter Search API. The highest reported accuracy was 83.0 **%** when using unigram and bigram feature. Additionally, [11] introduced an online/offline sentiment analysis tool (Senti-Strength) that is related to the “linguistic and word count program” (LIWC). Using this lexicon, it performed a “Kleene star stemming” reporting a highest accuracy of 63.7% and 67.8% for true positives and true negatives under supervised training respectively. Furthermore, [12] also provided an online sentiment tool that classifies tweets as positive or negative (Tweenator). The highest accuracy reported was 86.3% using the “Stanford twitter sentiment test set” based on the Naïve Bayes classifier.

However, we consider Tweenator as a comparable tool for our framework. Both tools focus on twitter, yet, Tweenator deals with products, companies, etc. as our framework is specified mainly for sports. Furthermore, it assigns positive and negative polarities to tweets and provides a sentiment annotator interface for tweet labeling. However, Twitter Sports Data

Analytics Framework specifies objective and subjective polarities to phrases taking an additional step to insinuate consensus in subjective expressions.

Furthermore, [13] also proposed a web based sentiment analysis tool (Sentimentor) that categorizes online-streaming tweets as positive, negative, or objective based on the Naïve Bayes classifier. It uses the Twitter API for Data extraction. The highest false positive accuracy obtained was 52.31% using bigrams.

2.2 Positive, Negative, and Neutral Classifications

[14] Classified tweets as positive, negative, or neutral. It assigned polarity scores to tweets using a dictionary of effect in language (DAL) and WordNet. Unigrams and Senti-features in a 2-way classification obtained the highest average accuracy of 75.39%. Furthermore, [15], classified tweets restricted to sports as positive or negative using the streaming twitter API. In the classification phase, it focused on "Profiles of mood states" (POM), negation words, and special conditions as the existing of the "but" word in a tweet. Results reported that positive polarities dominated tweet sentiments. Additionally, [16] identified tweet polarities as positive, negative, or neutral. After preprocessing tweets, it extracted n-gram, lexicon, POS, and microblogging features. Using the Adaboost classifier, and with the Hash set and Emoticon dataset the highest accuracy obtained was 75% by including the n-gram, lexicon, and microblogging features. On the other hand, [17] used a hybrid approach of symbolic (Knowledge-based approach) and machine learning techniques on the "Sanders analytics" data set to classify tweets as positive, negative, or neutral. The achieved accuracy was 100%. Furthermore, [18] restricted its sentiment analysis to only

subjective tweets relating them to climate change. It obtained around 76% of accuracy for number of features varying from 100 to 600, and an average of around 80% of tweets related to subjective-positive polarities that are certain of climate change. Moreover, [19] predicted positive, negative, and neutral polarities of tweets conveying to certain topics. It also assigned their distribution on a five-class scale point from highly negative to highly positive. The highest achieved accuracy was 64.6% for polarity predictions. Additionally, [20], computed the weighted average scores of sentiments in texts to determine their positive, negative, and neutral polarities. The highest obtained accuracy was 80.47%. Finally, [21], introduced a new senti-circles lexicon methodology based on occurrences of patterns of words in tweets to assign sentiment polarities. The highest recorded F1-measure was 85.45% by senti-median approach.

2.3 Objective and Subjective Classifications

Reference [22] proposed a genetic algorithm (GA) agent that classifies sports articles as subjective or objective based on syntactical features of texts. Based on a corpus of 300 sport article, training and testing accuracy reached 96.2% and 94.5% respectively with a 3-fold cross validation. Finally, [23] proposed sentence based classifiers for objective and subjective sentences for unannotated texts. It achieved an accuracy with pattern extraction of 69.7% for a subjective rule based classifier and 66.7% for an objective one. The Naive Bayes recorded an accuracy of 73.8%. Generating a new training set using the classifier itself, recorded an increase in accuracy with pattern extraction of 71.0% and 67.3% for subjective and objective rule based classifiers respectively.

2.4. Summary of Existing Tools

Table.1: Existing Tools

Tools	**Features**			
	Web Based	**Twitter Focus**	**Supported SDKs**	**Capabilities**
Data analytics Framework for sports related tweets	Yes	Yes	Java	• Provides sentiment analysis based on: 1.Pos Tagging 2. WordNet 3.Nicknames. 4. Tweets. • Classifies sport phrases as objective or subjective with/out consensus. • Clusters tweets to agree or disagree to the phrase. • Provides a decision tree for subjective phrases. • Figures agreement and disagreement of tweets over time. • Identifies herding behaviors of subjective phrases • Permits surfing tweets.
Datum-Box (API) [24]	Yes	No	Android, Java, PHP, python, C#, Ruby, Objective-C	• Supports Sentiment Analysis (including Twitter), Subjectivity Analysis, Topic Classification, Spam Detection, etc. • Supports some algorithms as: LDA, Max Entropy, Naïve Bayes, SVM, etc.

Alchemy (API) [25]	Yes	No	Java, Python, PHP, Node.js, Ruby, Android, C, C++, Perl, C#	• Classifies documents, texts, and sentences for user specified texts and webpages as (+, -). • Supports other features as keywords extraction, Taxonomy, etc.
Twitty City [26]	Yes	Yes	-----	• Launches a sentiment analysis comparison for tweets of five different cities.
Stream Crab [27]	Yes	Yes	-----	• Specified for real time tweets sentiment analysis. • Input: Keyword • Output: 1) Percentage of tweet counts. 2) "Polarity Sums". 3) "Polarity trend overtime" 4) "Polarity distribution overtime". • Classifiers: MaxtEnt (Default) and NB.
Mood Map [28]	Yes	Yes	-----	• Detects the mood of the world by analyzing public stream twitter data. • Data is extracted randomly every minute based on existing emoticons' polarities (+, -). • The mood of each tweet is calculated and geocoded to be located on the map.

Mappiness (Mobile App) [29]	Yes	No	-----	• Analyzes the effect of local surroundings to people's happiness as: Pollution, friends, etc. • It is based on asking you through phone beep(s) on daily bases on your mood, place, people you are with, etc.
Lymbix (API) [30]	Yes	No	Ruby, Python, PHP, JavaScript, Java, Objective-C, C#	• Based on detecting emotions existing in a text specified by a user as an input. • Users can select what fields to detect for their output (Anger, Fear, etc.) • Results display scores with respect to the chosen fields.
Sentiment Analyzer [31]	Yes	Yes	-----	• Input: A specified text by a user with a specified language. • Output: A measurement of the polarity of the given text as (+, -, neutral).
NLTK (API) [32]	Yes	No	Python	• Supports sentiment analysis, stemming, POS tagging, etc. • Supports some classifiers as: Maxent, and NB.

Repustate (API) [33]	Yes	No	Ruby, Python, PHP, JavaScript, Java, Objective-C, C#	• Provides sentiment analysis based on: 1) POS tagging. 2) Lemmatization. 3) "Prior Tagging". 4) "Negations and Amplifiers". 5) "Document Sentiment". 6) "Scoped Sentiment" 7) "Customized Sentiment".
Opinion Crawl [34]	Yes	No	-----	• Uses "text mining", "semantic" and "sentiment analysis". • Determines sentiments not only based on calculating (+, -) sentiments, but also including major events that could affect our search. • Users choose categories for their search or could enter a topic. • Output: 1) Pie chart. 2) "Sentiment Trend". 3) (+ TO -) ratio. 4) "News headlines". 5) Thumbnail Images. 6) "Tag Cloud".
Twit Fight [35]	Yes	Yes	------	• Input: Two keywords. • Output: 1) Real time tweets. Polarity for each input specified as (+, -, neutral

Social Mention (API) [36]	Yes	No	SDKs that can issue HTTP requests.	• Provides a real-time search data of different Medias. • Output: 1) "Title of search". 2) "Timestamp of the search". 3) "Number of items". 4) Array of items including ID, timestamp, etc. • For Twitter: 1) (+, -) sentiment scores. 2) Top hashtags. 3) Retweets. 4) Top keywords. 5) "URLS-cited". 6) Hashtags. 7) Number of "@references" in mention. 8) Top active users.
Mozvo [37]	Yes	Yes	-----	• Specified for movies only. • Analyzes movie reviews from different public sources as: Twitter, Facebook, etc. • Assesses the sentiment for movies and assign scores. Some features: 1) Check friend's recommendations. 2) Check friend's feeds of movies (+, -). 3) Check celebrity's recommendations for movies. 4) Surf tweets and watch latest trailers.

Semantria (API) [38]	Yes	No	C, C++, Java, PHP, .Net, Python, Ruby, JavaScript	• Supports: 1) Summarization. 2) Pos Tagging. 3) Sentiment. 4) Boolean Queries. 5) Categorizations. 6) Intentions. 7) Entities. 8) Themes.
TweetTronics [39]	Yes	Yes	-----	• Extracts real time data. • Detects: 1) Sentiments. 2) Participants and reasons they participate. 3) Persona and effects. • Measures the efficiency of conversations. • Outputs: 1) How roughly the brand is being conversed? 2) A (+ or -) buzz. 3) Key expressions used. 4) Most influenced people. 5) Main web pages' people speaking about.
Sentiment Viz [40]	Yes	Yes	-----	• Input: keywords. • Output: 1) Recent tweets with keywords. 2) Sentiments of tweets as (serene, calm, relaxed, etc.) 3) Topics. 4) Heat map. 5) Tag Cloud. 6) Timeline. 7) Location of tweets (Map). 8) Affinity of tweeters. • Uses "sentiment dictionary" for sentiment analysis.

CHAPTER 3

METHODOLOGY

This section discusses the proposed approach. The methodology goes through three main stages. In the first phase and after tagging phrases, all keywords are extracted using the POS Tagger, pulling their synonyms and antonyms from Word-Net. Subsequently, nicknames are added, when available, to the twitter search query for a complete retrieval of tweets by the Twitter Search API. In the next level, features are extracted from tweets to create two clusters that agree or disagree to the input phrase. Furthermore, phrases are classified as objective or subjective, were tweets that agree to the specified phrase distinguishes between consensual subjective phrases by our predefined decision tree.

The decision tree is composed of 5 levels. The first level is the base state level of the tree; in which it checks the availability of sufficient tweets to proceed to the next level. At this point we impose a threshold of 30 tweet as a minimum number of required data for the cluster that the phrase belongs too. This threshold is computed according to (1).

$$\frac{\sum_{i=1}^{n} \#\text{of tweets}}{n} - c \qquad (1)$$

"n" is the total number of phrases in our corpus (1007 phrases) and "c" is a heuristic number to ensure that tweets belonging to the same phrase cluster are at least more than half the number of average tweets per phrase of the corresponding used corpus. Taking into consideration that tweets are distributed between 2 clusters. Consequently, to predict consensus or social media herding. In our case, and according to our corpus, we recorded an average of 55 tweets per phrase for “c” to be “25”.

Level 2 computes the intra cluster distances between all points in the cluster that the phrase belongs to (2), in which if this distance is smaller or equal to the maximum distance between the farthest 2 points in the same cluster, we proceed to the next level of the tree.

$$\frac{\sum_{s \in \text{cluster}} \sum \frac{d(i)d(j)}{\text{size} - 1}}{\text{size}} \qquad (2)$$

“s” is the phrase belonging to the cluster, “d” is the distance between the cluster points, and “size” is the number of cluster points.

However, if the intra cluster distance is not larger than the distance between the farthest two points, no conclusion could be made regarding consensus.

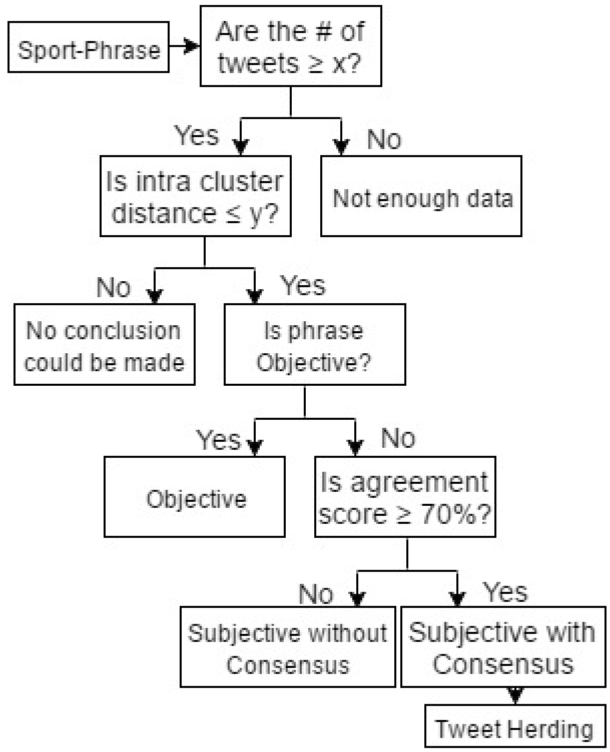
In the third level, we check the objective or subjective polarity of a phrase. If the phrase is objective, the decision tree is terminated, however, in case subjectivity is encountered, the fourth level of the tree computes an agreement score (3).

$$\frac{\sum_{s \in \text{agree cluster}} \#\text{ of times } 's' \text{ is retweeted}}{\sum_{s \in \text{agree cluster}} \#\text{of times } 's' \text{ is retweeted} + \sum_{s \in \text{disagree cluster}} \#\text{ of times } 's' \text{ is retweeted}} \qquad (3)$$

Yet, if this score is greater or equal to 70%, this means that the phrase has consensus, otherwise, there is no consensus as the last level of the decision tree insinuates.

Finally, herding behaviors are identified in subjective tweets. Figure.1 shows the workflow of the subjective and objective classification with tweet herding.

Figure.1: Subjective/Objective Classification and Tweet Herding



3.1 Implementations

This part focuses on the implementation level of phrase classification, in which it introduces the first five phases of the proposed methodology, starting from sport phrase inputs by users, proceeding with the POS Tagger, Keywords extraction, WordNet, reaching the Twitter Search API.

3.1.1 Phrase Input

Acceptable phrases are restricted to English language with a minimum of **15** characters as shown in Table.2.

Table.2: Input Declaration

Accepted phrase	Barcelona played well in their match today.	43>15 characters (spaces included)
Un-Accepted Phrase: (less than 15 character)	Ronaldo	7<15 characters (spaces included)
Un-Accepted Phrase: (Language not supported)	ميسي افضل لاعب في برشلونة	24>15 characters (spaces included)

3.1.2 POS Tagger

The process of tagging phrases with their part of speech is done by the **P**art **of S**peech tagger supported by Stanford University [6]. This tagger tags each word, special characters, and other symbols by their specific part of speech (verbs, nouns, etc.). Table.3 summarizes all the supported tags.

Table.3: Stanford Tagger POS Tags

Description	**Tags**		**Description**	**Tags**		**Description**	**Tags**
Coordinating conjunction	CC		Adjective	JJ		Noun, plural	NNS
Cardinal number	CD		Adjective, Comparative	JJR		Proper noun, singular	NNP
Determiner	DT		Adjective, superlative	JJS		Proper noun, plural	NNPS

Existential there	EX		List item marker	LS		Predeterminer	PDT
Foreign Word	FW		Modal	MD		Possessive ending	POS
Preposition or subordinating conjunction	IN		Noun, singular, or mass	NN		Personal pronoun	PRP
Possessive pronoun	PRP$		To	TO		Verb, non 3^{rd} person singular present	VBP
Adverb	RB		Interjection	UH		Verb, 3^{rd} person singular present	VBZ
Adverb, comparative	RBR		Verb, base form	VB		Wh-determiner	WDT
Adverb, superlative	RBS		Verb, past tense	VBD		Wh-pronoun	WP
Particle	RP		Verb, gerund or present participle	VBG		Possessive wh-pronoun	WP$
Symbol	SYM		Verb, past participle	VBN		Wh-adverb	WRB

These are some examples on how the POS tagger tags phrases:

Table.4: POS Tagger Examples.

Before Tagging	**Messi is the best player.**
After Tagging	Messi/NNP is/VBZ the/DT best/JJS player/NN./.
Before Tagging	**Stephen Curry is the MVP.**
After Tagging	Stephen/NNP Curry/NNP is/VBZ the/DT MVP/NN./.
Before Tagging	**Barca won 2-1 on Real Madrid.**
After Tagging	Barca/NNP won/VBD 2-1/CD on/IN Real/NNP Madrid/NNP./.

3.1.3 Keyword Extraction

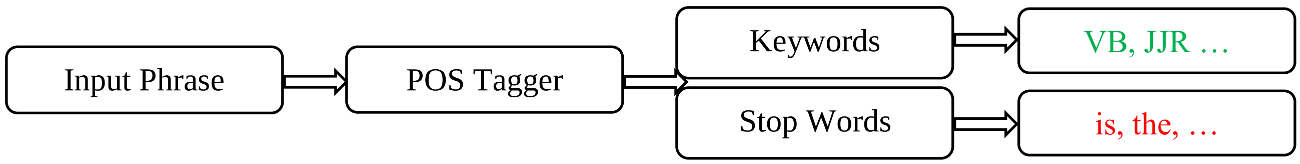
Though every word in the phrase is tagged, keywords are considered to hold specific tags corresponding to verbs, adjectives, nouns, pronouns, numbers, and adverbs according to Table.5.

Table.5: Tags Used for Keyword Extraction

Description	Tags
Cardinal number	CD
Adjective	JJ
Adjective, Comparative	JJR
Adjective, Superlative	JJS
Noun, singular, or mass	NN
Noun, plural	NNS
Proper noun, singular	NNP
Proper noun, plural	NNPS
Adverb, comparative	RBR
Adverb, superlative	RBS
Adverb	RB
Verb, base form	VB
Verb, past tense	VBD
Verb, gerund or present participle	VBG
Verb, past participle	VBN
Verb, non 3rd person singular present	VBP
Verb, 3rd person singular present	VBZ

However, as Figure.2 shows, some words could be tagged as one of the tags specified in Table.5, but they would not be considered as keywords. A bag of words, stop words, of around 487 tokens based on Princeton University [41] that provides a list of tokens to be considered as stop words is used to handle such cases. Additionally, some special characters as '*, %, #, &, +. -, /, \ and @' are also filtered as they are not part of our features. However, for future use they would be maintained as attributes in the feature vector, especially in subjective cases.

Figure.2: Extracting Keywords Process



3.1.4 WordNet

WordNet [7] is a lexical database for the English language. It is used for grouping words with their synonyms to form different Synsets. A Synset is the semantic and lexical relation between words. In phrase classification, WordNet is used to extract synonyms and antonyms only of keywords tagged as Verbs, Adverbs or Adjectives. Table.6 illustrates some examples:

Table.6: WordNet Synonyms and Antonyms

Word	**Tag**	**Synonyms**	**Antonyms**
Best	Adjective	*Good*	*worst*
Defend	Verb	*support, hold, guard, fight, oppose, champion, represent, maintain*	*attack, assail, prosecute*
Surprisingly	Adverb	*astonishingly, amazingly*	*unsurprising*

Moreover, these keywords are passed to the Twitter Search API in-order to retrieve tweets containing

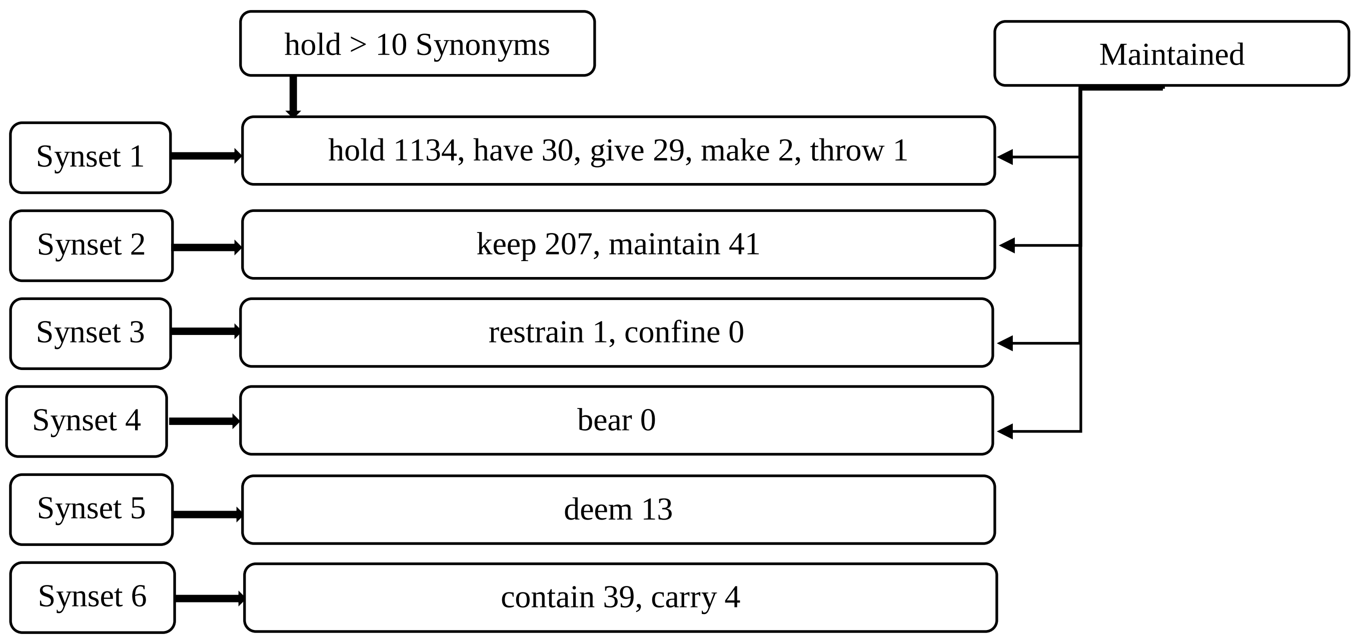
1- (Extracted Keywords) OR
2- Extracted Keywords' synonyms including negations (not good, etc.)
3- OR antonyms including negations (not worst, etc.)
4- OR all combinations if possible.

Some phrases might contain multiple keywords with multiple synonyms and antonyms in which might exceed the 500-character limit of a query. According to the Twitter Search API documentation: the UTF-8, URL-encoded search query is limited to 500-characters as a maximum length of a query including operators (OR in our case). To solve this problem, we limited the number of synonyms. To select the most reliable synonyms from all Synsets, we used the Tag-Count parameter provided by WordNet to count the semantic frequency for each synonym in texts of WordNet (Figure.3). Synonyms are then sorted according to their Tag Counts and Synsets. Figure.4 illustrates the idea.

Figure.3 Pseudo Code for Tag Count.

```
1.Initialize input word "w"
2.For each Synset
3.Loop over each synonym "i"
4.Count semantic frequency for "i" in texts
5.End
```

Figure.4: Query Handling Limit Length Exception for the Word Hold



As mentioned earlier, WordNet is only used for adjectives, adverbs, and verbs, however, nouns are excluded, for WordNet might provide synonyms which may not be related to our context. Such a limitation, detected by manually screening of tweets, accounted by retrieving irrelevant tweets according to unrelated nouns. Table.7 shows a detailed example.

Table.7: Synonyms for Nouns from WordNet

Phrase	Word (Noun)	Synonyms	Sample of Retrieved Tweets	Relevant
The first **quarter** is the best	quarter	one-fourth, fourth, quartern, stern, poop, tail…	It's a difficult time when you're holding in your **poop** to clock in first. Gotta get paid for what I do best! First Trollcraft episode is out tonight! I don't know which emoji best symbolises Trollcraft so here is a smiley **poop**	×
The first **quarter** is the best	quarter	Quarter	that's the best first **quarter** team 😊😊	✓

Furthermore, proper nouns are treated differently, a pre-gathered corpus from a set of nicknames [9] for famous teams and players is added to our search query, often players and teams are not just known by their original names, some are known for their nicknames. To ensure a complete retrieval of related tweets, we looked for some alternatives for players and team original names and add an “OR” combinational relation between them. Figure.5 illustrates the idea.

Figure.5: Nicknames Relations

Jerry West → logo; Mr clutch; Mr outside

Shaquille O'Neal → Shaq; Shaq Daddy; Shaq Fu; Diesel; Big Aristotle; Superman

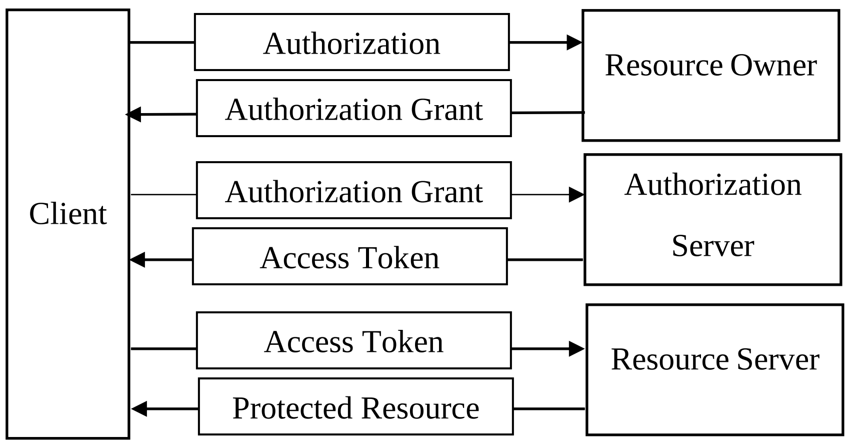
Kobe Bryant → Black mamba

3.1.5 Twitter Search API

Building our search engine relies on “twitter4j” API that permits developers to contact twitter servers and retrieve tweets according to the following Application settings:

- Consumer Key (API Key) for API calls.
- Consumer Secret (API Secret) serving as the client password for application authentication by twitter servers.
- Access Token serving as the secret identity of the processes of threads.
- Access Token Secret serving as a password and sent with the Access Token to issue queries.

Figure.6: Protocol Workflow



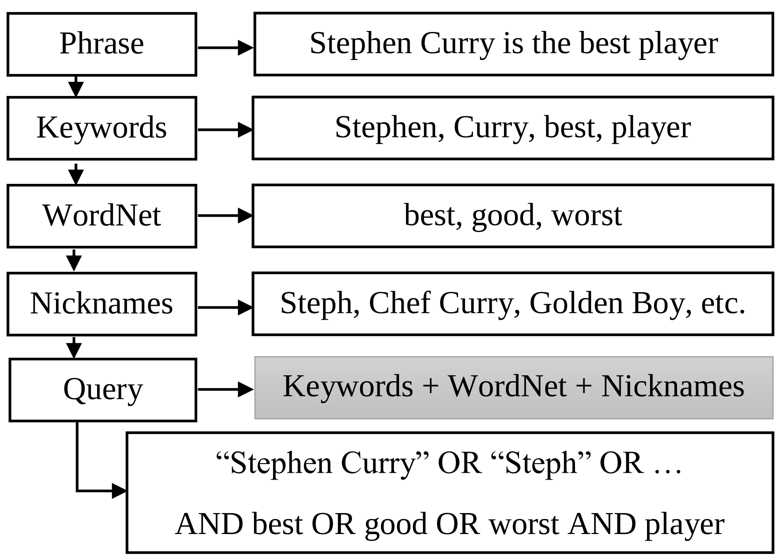
We use the twitter search API to extract the most recent, and popular tweets available within a 7-day period. However, Twitter also provides a stream API in which all retrieved tweets are current (live). In-order to increase the probability for retrieving relevant tweets, we rely on the Search API. Table.8 specifies the parameters used to retrieve tweets. **Each issued query - no matter the number of tweets it retrieves, decreases the token count by one**; tokens specify the number of queries a user can issue.

Table.8: Twitter Search API Parameters

Max number of queries	Tweets per Query	Language	Query Type	Authentication Type	Number of tokens/user
5	**Max** of 100	English	Mixed (popular + recent)	Oauth- user Authentication	180

For instance, if a user enters a phrase as: “Stephen Curry is the best player”, our tool will issue a request to retrieve tweets for this phrase after extracting its keywords as discussed in the previous steps. Figure.7 shows the details for such an example.

Figure.7: Twitter Search API Query Example

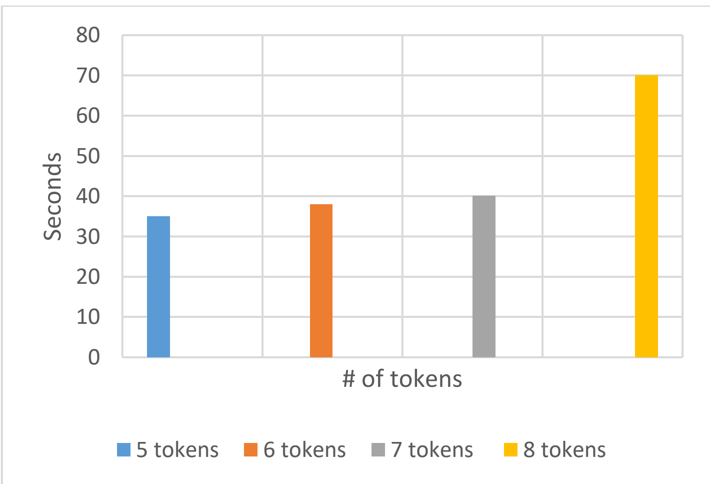


However, our application is modeled to request a total of 5 queries in an iterative manner. If the first query could retrieve the maximum number of tweets (100) or less per token, out of 180 tokens, the subsequent one will be issued to retrieve the next 100 (or less) tweets until we reach the maximum threshold of 5 queries (500 tweet at a maximum). Furthermore, if a query could not receive any tweets anymore which means all possible tweets including these keywords are retrieved, our tool shuts the connections with the Twitter API but maintains the retrieved tweets.

We chose the threshold for the total numbers of queries per user request to be “5” after trial and error for the following reasons:

- The run time to process a maximum of 500 tweet is efficient.
- The run time to retrieve tweets from 5 queries at a max gives better performance for our tool, the more queries, the more time is required to retrieve and process tweets. (Figure.8).
- Users can issue a resourceful number of requests (36 or more), depending on the phrase.
- Issuing 36 requests or more, corresponding to 180 tokens, and according to the run time of our application to process each request, a user might not have to wait much or even not to wait at all to refill the 180 tokens after the twitter time limit exceptions for requesting queries.

Figure.8: Application Run Time



Moreover, our model restricts 180 tokens per user, so each user has around 36 requests (if 5 queries are issued per user request). However, each user is obliged to authorize our application to use our tool, this is known as: Oauth- user Authentication, in which all users should sign with their twitter accounts and authenticate our app. At this point, each end-user is assigned an authorized token number covering those 180 tokens. In case users have consumed and exceeded their limit, an automated time counter will notify them with the waiting time in-order to process their queries again. In that way, all users would have equal opportunities to use our tool with a private token number for each.

Figure.9: Collect Tweets (5 Successful Queries).

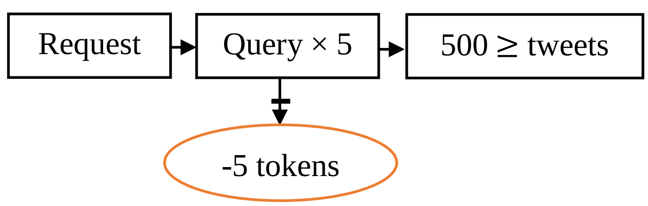
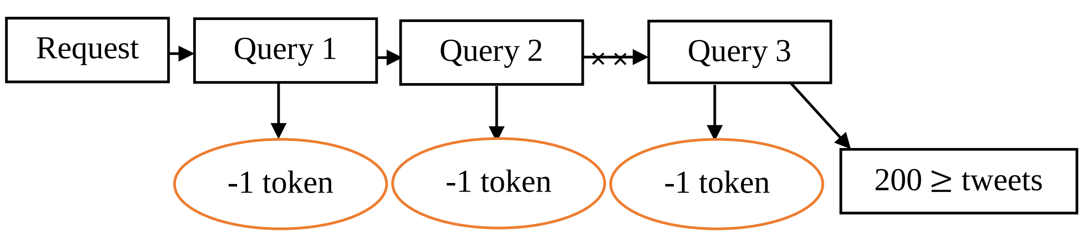


Figure.10: Collect Tweets (2 Successful Queries).



3.2 Computations

In this section, we discuss the computational part of phrase classification, going through pre-processing tweets, extracting tweet features, clustering tweets and phrases, and heading to Objectivity and Subjectivity classification to construct the decision tree for subjective-consensus conclusions.

3.2.1 Pre-Processing Tweets

After contacting the “twitter4j API”, we receive data displayed in “JSON” format including all components related to a tweet and all information about its blogger. Subsequently, the received data are parsed and filtered out to extract the needed entities including tweet texts. Moreover, we filter redundant tweets and insure that all retrieved tweets include the extracted keywords (with synonyms and antonyms). Figure.11 is a sample of a JSON instance and Fig.12 is a sample of parsed data.

Figure.11: JSON Instance

{"in_reply_to_status_id_str":null,"in_reply_to_status_id":null,"created_at":"Wed Aug 12 13:02:43 +0000 2015","in_reply_to_user_id_str":null,"source":"<a href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client<\/a>","retweet_count":0,"retweeted":false,"geo":null,"filter_level":"low","in_reply_to_screen_name":null,"id_str":"631450744164585472","in_reply_to_user_id":null,"favorite_count":0,"id":631450744164585472,"text":"NFL Schedule http://t.co/N58JS1aubO","place":null,"lang":"en","favorited":false,"possibly_sensitive":false,"coordinates":null,"truncated":false,"timestamp_ms":"1439384563853","entities":{"urls":[{"display_url":"espn.go.com/nfl/schedule","indices":[13,35],"expanded_url":"http://espn.go.com/nfl/schedule","url":"http://t.co/N58JS1aubO"}],"hashtags":[],"user_mentions":[],"trends":[],"symbols":[]},"contributors":null,"user":{"utc_offset":null,"friends_count":59,"profile_image_url_https":"https://pbs.twimg.com/profile_images/629384559885287424/xi9XMt2z_normal.png","listed_count":0,"profile_background_image_url":"http://abs.twimg.com/images/themes/theme1/bg.png","default_profile_image":false,"favourites_count":0,"description":null,"created_at":"Thu Aug 06 20:06:09 +0000 2015","is_translator":false,"profile_background_image_url_https":"https://abs.twimg.com/images/themes/theme1/bg.png","protected":false,"screen_name":"billgiles3810","id_str":"3308031524","profile_link_color":"0084B4","id":3308031524,"geo_enabled":false,"profile_background_color":"C0DEED","lang":"en","profile_sidebar_border_color":"C0DEED","profile_text_color":"333333","verified":false,"profile_image_url":"http://pbs.twimg.com/profile_images/629384559885287424/xi9XMt2z_normal.png","time_zone":null,"url":null,"contributors_enabled":false,"profile_background_tile":false,"statuses_count":4,"follow_request_sent":null,"followers_count":5,"profile_use_background_image":true,"default_profile":true,"following":null,"name":"Bill Giles","location":"","profile_sidebar_fill_color":"DDEEF6","notifications":null}}

Figure.12: Parsed Data

550830572, siiiiiickladlwt,Wed Aug 12 13:02:42 +0000 2015,Wed Aug 12 16:02:42 EEST 2015,⚽🏁⚽🏁⚽\nYou run after the ball \nI run after you\n\nYou reach your goals, help me reach mine? \n⚽🏁⚽🏁⚽\n\nFollow **me**? ☒˙ @Louis Tomlinson\n—121.953

Following Retrieving tweets, based on the previous steps, data is preprocessed, were all emoticons and emoji are temporary eliminated as the POS tagger does not recognize them. Emoticons are divided into 2 datasets, the 2-D and the 3-D emoticons, in which we labeled them as -1 for negative emoticons, +1 for positive emoticons, and 0 for neutral ones (Table.9). All URLs are also replaced by the string “URL”, some special characters are also removed, some abbreviations are replaced with their original words, and some slang notations are replaced by their corresponding proper English representations using our slangs look up dataset (Table.10) [11]. Tweets are now tagged using the “Stanford POS tagger” and post processed to generate well-defined and tagged tweets for feature extraction. The post-process phase is eventually for cleaning up some tagged symbols of the POS Tagger. Figure.13 illustrates this phase:

Table.9: 2-D and 3-D Emoticons

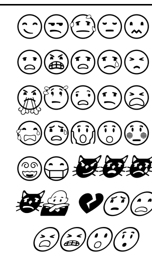
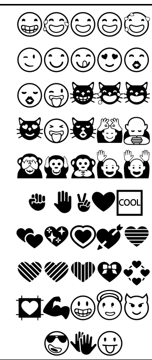
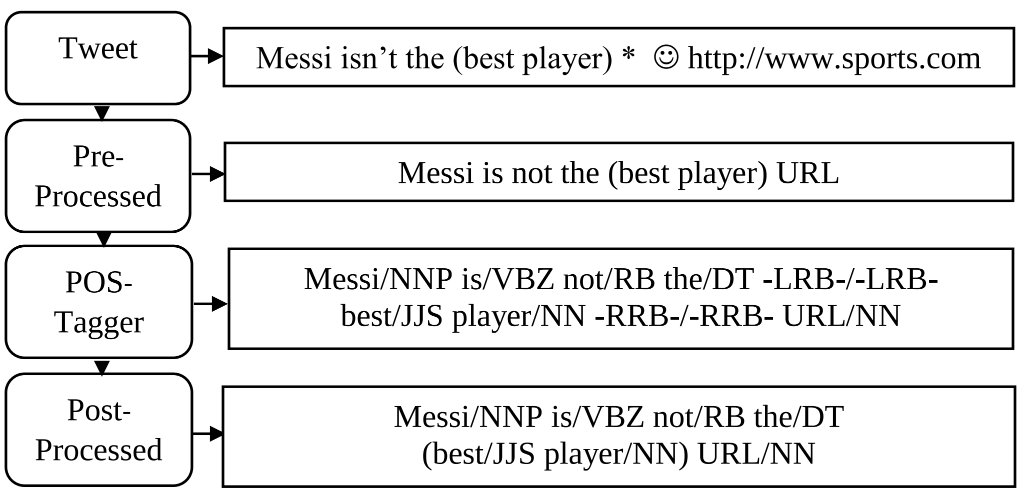
<table>
<tr><th>2D Emoticons</th><th>3D Emoticons</th><th>Labels</th></tr>
<tr><td>%-(,)-:,):,)o:, 38*, 8-0, 8/, 8\, 8c, :#, :'(,:'-(, :(, :*(, :,(, :-&, :-(, :-(o), :-/, :-S, :-\, :-|, :/, :E, :F, :O, :S, :[, :\, :_(, :o(, :s, :|, :...(, </3-1, <o<, =(, =[, >/, >:(, >:L, >:O, >[, >\, >o>, B(¸ Bc, D:, X(,X-(, ^o), xP, |8C, |8c</td><td></td><td>-1</td></tr>
<tr><td><:}, ;o), ;), :I¸ :0--<|:¸ :-O, 8-), *), (o;,</td><td></td><td>0</td></tr>
<tr><td>%-), (-:, (:,(^ ^), (^-^),(^.^), (^_^),(o:, *\o/*, --^--@, 0:), 8), :), :-), :-*, :-D, :-P, :-}, :3, :9, :D, :P, :X, :], :b), :o), :p, :Þ, ;^), <3, =), =], >:), >:D, >=D, @}->--, XD, XO, xo, XP, ^_^, x3?, xD, |D, }:)</td><td></td><td>1</td></tr>
</table>

Table.10: Samples of Slangs Lookup Table

Slangs	***Proper English***
121	*one to one*
a/s/l	*age, sex, location*
And	*any day now*
Afaik	*as far as I know*
Alol	*Actually laughing out loud*
b4	*Before*
bfn	*Bye for now*
Bg	*Big grin*
btw	*By the way*
Cp	*Chat post*

Figure.13: Pre-Processing



3.2.2 Tweet Features

A feature vector is an essential instance for classification. Therefore, we need a reliable feature vector to embed into our classifiers. Accordingly, semantic and syntactic features are used.

However, choosing features can differ according to the type of processed texts. For tweets, short texts or informal sentences, and particularly targeting a subjective-objective classification approach as one of our methodological steps, we suggest the following:

a) Semantic Features

For semantic features, we use online dictionaries and lexicons such as “Senti-WordNet” [42]. Senti-WordNet assigns positive, negative and neutral polarity scores to each Synset as it identifies a semantic score by normalizing the scores of the ternary classifiers that decide if a Synset is objective, positive or negative. Consequently, we compute the average of scores. However, not all words are available in the dictionary of Senti-WordNet, so some word-tokens need to be stemmed before finding their scores (Table.11). Furthermore, using our pre-labeled emoticon databases (2-D and 3-D emoticons) that assign polarities to emoticons could also insinuate the semantic orientation to tweets. For instance, a smiley face in a tweet (☺) could point out to a positive sentiment, similarly a sad face (☹) could theme a negative sentiment. Table.12 shows the semantic attributes.

Table.11: Stemming Words

Word	Tag	Score	Stemmed Word	Stemmed Score
Playing	*Verb*	------	*Play*	0.036409580058674305
Singing	*Verb*	------	*Sing*	0.01824817518248175

Table.12: 10 Semantic Features

SEMANTIC FEATURES	
1	Average word score
2	# of + scores
3	# of – scores
4	# of neutral scores
5	# of + emoticons
6	# of – emoticons
7	# of neutral emoticons
8	# of words
9	# of emoticons
10	Average emoticon scores

We clarify the selected semantic features in the following list:

1. Average word score:

   Every word of the tweet tagged as a verb, adverb, adjective, or noun is passed to Senti-WordNet to assign a polarity for it. The intuition behind these polarities is that words that are positive or negative could point out to subjective phrases, whereas words assigned neutral polarities or scores close to zero most probably will theme objective phrases. As an overall feature, we calculate the average scores of all words for each tweet as a part of our feature vector. Table.13 shows some polarities assigned by Senti-WordNet to various word examples.

Table.13: Senti-WordNet Scores

Word	***Tag***	***Score***
Love	Verb	0.6100000000000001
Hate	*Verb*	-0.75
Nice	*Adjective*	0.708941605839416
Ugly	*Adjective*	-0.515
Luckily	*Adverb*	0.5
Honestly	*Adverb*	0.375

2. Number of positive scores:

   All words of positive polarities are counted in each tweet and the overall number of positive scores are saved as one of the semantic features.

3. Number of negative scores:

   All words of negative polarities are counted in each tweet and the overall number of negative scores are saved as one of the semantic features.

4. Number of neutral scores:

   All words of neutral polarities are counted in each tweet and the overall number of neutral scores are saved as one of the semantic features.

5. Number of Positive Emoticons:

   According to our predefined emoticons (2-D and 3-D) dataset, our tool is able to detect all existing emoticons in tweets. However, all these emoticons are replaced with their corresponding labels. For the Positive emoticons, +1 value is to represent their positive theme.

6. Number of Negative Emoticons:

   For the Negative emoticons, -1 value is to represent their negative theme.

7. Number of Neutral emoticons:

Similarly, For the Neutral emoticons, 0 value is to represent their neutral theme.

8. Number of words:

Words of each tweet are counted excluding all special characters and “URLs”.

9. Number of emoticons:

The overall number of emoticons is also saved as one of the attributes of our feature vector particularly our semantic part of it.

10. Average Emoticons score:

All numerical representations of emoticons in tweets are averaged to insinuate the polarity of tweets as a first step, which indeed indicated the flavor of Objectivity and Subjectivity.

Table.14 exemplifies a sample of the semantic part of our feature vector:

Table.14: Semantic Feature Vector

Tweet	@warriors You guys deserve to win. ❤❤❤❤ #Warriors #DubNation									
Semantic Features	*Avg. word score*	*Avg. Emot score*	*#of Emot*	*#of neutral Emot*	*#of neg. Emot*	*#of pos. Emot*	*#of neut. score*	*#of neg. scores*	*#of pos. scores*	*#of words*
Numerical Numbers	*0.256667*	*1*	*4*	*0*	*0*	*4*	*1*	*0*	*2*	*8*

b) <u>Syntactic Features</u>

Table.15: 55 Syntactic Features

SYNTACTIC FEATURES	
1	Question marks (?)
2	Exclamations marks (!)
3	Full stops (.)
4	Commas (,)
5	Colons (:)
6	Semi colons (;)
7	Triple dots (…)
8	Quotations (‘’ or “”)
9	If retweet or not
10	Pronouns (I, we, me, us, mine, our)
11	Pronouns (you, your)
12	Pronouns (he, his, she, her it, its, they, their, them)
13	Verbs (past)
14	Verbs (imperative)
15	Verbs (present3rd)
16	# of retweets
17	Verbs (present1st2nd)
18	Verbs (infinitive/base form)
19	comparative/superlative adj /adv
20	All POS tags **(which are 36 from Table. 3**) ("cc","cd","dt","ex","fw","in", "jj","jjr","jjs","ls","md","nn", "nnp","nnps","nns","pdt","pos","prp", "prp$","rb","rbr","rbs","rp","sym", "to","uh","vb","vbd","vbg","vbn", "vbp","vbz","wdt","wp","wp$","wrb")

Our syntactic features selection is based on the following [22]:

Punctuations: From Table.15, the first 8 features represent the punctuations that appear in tweets or phrases, as Exclamation and Question marks insinuate the appearance of subjectivity, in which the author or user could be emphasizing events or reports, or even expressing revelations. A user could also be questioning for information or knowledge.

Quotations, in turn, are indications of objectivity, as the author would narrate or state what others have said without interference for his/her emotions or opinions. However, additional features under this category are also recorded as full stops, triple dots, commas, colons, and semi-colons.

Retweet or not: While parsing the Jason object for each tweet, as discussed before, we check the availability of the original tweet for each Re-tweet, in which all retweets with the same original tweet are counted as a single tweet (Original); in such a way, we ensure eliminating all redundant tweets. However, Some Jason objects does not contain the original tweet for retweets, in such cases we mark these phrases by a "1" value in our feature vector to identify retweets and original ones. Furthermore, this feature would also insinuate the sense of consensus and social media herding.

Number of retweets: As mentioned earlier, all retweets are referred by their original tweet to reduce redundancy in tweets, still, for each original tweet, the number of retweets is maintained. Such a feature would refer to consensus and sense of herding.

Pronouns: First, second, and third person pronouns are also part of our feature vector, in which third person pronouns as: him, her, it, etc. would insinuate objective statements while first and second pronouns as: I, we, you indicate subjectivity.

Past tense verbs: Such verbs usually refer to objective phrases, in which an author would report or restate actions and incident that happened in the past, however, past tense verbs would somehow insinuate subjectivity with first and second pronouns.

Present tense verbs: These verbs combined with first and second pronouns would represent subjectivity in which they would show the opinion of the author, yet, present tense verbs using third pronouns in most cases indicate objectivity, though subjectivity might still occur in some cases.

Imperative verbs: Imperative verbs are verbs that create a sentence to give orders or inforce actions in a way that a user expresses his/her feelings to insinuate subjectivity.

Comparative and Superlative adjectives and adverbs: These part of speech tags represent in most cases subjectivity, the author would be expressing his/her opinions in a descriptive manner.

Numerical Data: Numbers in phrases most probably indicate objectivity, in which statistics could be reported. Numbers are maintained by the part of speech tag ("CD") referring to cardinal numbers.

POS Tags [6]: ALL 36 Part of Speech Tags in Table.3 are preserved beside the mentioned features.

3.2.3 Objective vs. Subjective Phrase Classification

Sport phrases that are inputted by users, are somehow treated as tweets, in which we extract the same syntactic and semantic features as listed and discussed above. However, to classify phrases as subjective or objective, we need a corpus to train our phrase features on.

A supervised corpus of 1007 phrase is built including pre-labeled 508 subjective phrase and 499 objective phrase. Table.16 shows a sample of these phrases.

Table.16: Objective and Subjective Phrases

Objective Phrases (Label =1)	**Subjective Phrases (Label=0)**
Lakers is an NBA team	Messi is the best player
Kyrie Irving scored 16 of his 30 in the first quarter	Atletico could not catch up on Real
Portugal finished third	Thunder must be more than just Big 3 to win
Cavs whip Warriors in Game 3	Barcelona would be better without Alves
Cavs took the lead 120-90	It didn't go right tonight for the Warriors

Figure.14 shows the number of phrases with/out consensus in our labeled dataset. Applying and experimenting the decision tree on our handmade corpus indicated that 57% of the subjective phrases are with consensus while 43% of them are consensus according to the corresponding tweets for each subjective phrase.

Figure.14: Subjective Phrases with/out Consensus

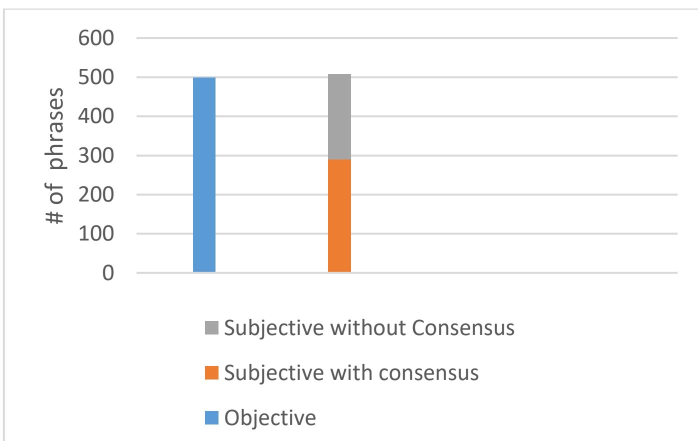


Figure.15 illustrates and summarizes our proposed Methodological framework.

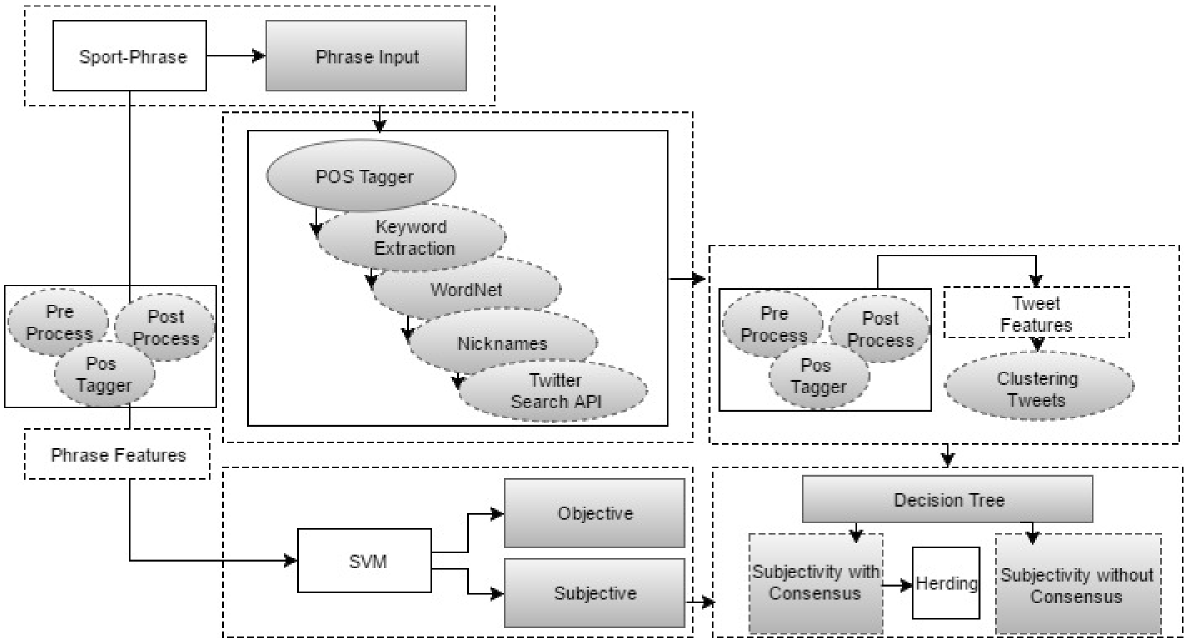


Figure.15: Workflow of the Proposed Framework

CHAPTER 4

RESULTS

This chapter shows our experiments and results of clustering and classification procedures for unsupervised and supervised datasets. Herding behaviors are also identified for subjective tweets. Moreover, we compare and analyze the performance of our tool with Tweenator.

4.1 Clustering Results

This sections shows the experimental results on different clustering algorithms to identify the best algorithm that fits our methodology. Table.17 lists the measures that were used for validation and evaluation. Table.18 shows a description of all datasets used.

Table.17: Clustering Evaluation Measures

Measure	Description
Accuracy (Acc %)	$100\times\left(1-\frac{number\ of\ misclassified\ data}{total\ number\ of data}\right)\ \%$
Silhouette Index (SI)	Measures the similarity of each point to its corresponding points in the same cluster, when matched to other data points in other clusters. This index is defined by: $$Si=\frac{bi-ai}{\max(bi,ai)}$$ Were "ai" is the average distance between the ith point with the corresponding data points in the equivalent cluster, and "bi" is the average distance between the ith point and other data points in different clusters, (Minimum distance to take). This index ranges from values of [-1 +1]; high values indicate a better matching for data points within the same cluster, and a poor matching with other data points in other clusters. Negative values might occur indicating insufficient or a lot of clustering solutions.

Dunn Index (DI)	Measures the maximum distance between the points in the same cluster and the minimum distance between the clusters, this is known as compactness and separation measurements respectively. The higher the value of this index is, a better partitioning of data we get.
Davies Bouldin Index (DBI)	This index is based on the following ratio: $$DBI = \frac{1}{K} \sum_{i=1}^{k} \max_{j \neq i} D\{i,j\}$$ $D\{i,j\} = (\frac{d_i + d_j}{d_{i,j}})$ is the distance between and within the clusters $i\ and\ j$ d_i: Is the average distance of every data point in cluster i and the centroid of this cluster, similarly for d_j $d_{i,j}$: Is the Euclidean distance between the centroids of clusters $i\ and\ j$ The minimum value of this index represents a better (within-between) ration.

Table.18: Datasets

Dataset Name	Dataset Type	# of attributes	# of instances	# of Classes
Arcene	Supervised	10000	200	2
Balance-Scale	Supervised	4	625	3
Forest Types	Supervised	27	523	4
100 leaves plant species	Supervised	64	240	5
Libras Movement	Unsupervised	91	360	--
DTS1 (Messi is the best player.)	Unsupervised	65	564	--
DTS2 (Champions League: Real vs. Atletico Madrid.)	Unsupervised	65	996	--
DTS3 (Cavaliers are taking the lead.)	Unsupervised	65	267	--

From Table.18, Arcene, Balance-Scale, Forest Types, 100 leaves plant species, and Libras Movement are datasets found in the UCI repository [43]. However, DTS1, DTS2, DTS3, and DTS4 are homemade corpuses from 4 different phrases from our corpus of 1007 phrase. For each phrase, we retrieved a different number of tweets, in which features (Table 12, Table 15) are extracted to create these datasets.

Table 19, identifies our baseline for the clustering measures, as it records the values of indexes on the true labels of each supervised dataset.

Table.19: Baseline Measures

Dataset (Labeled)	# of clusters	SI	DI	DBI
Arcene	2	0.0665	0.2005	4.4948
Balance-Scale	3	0.1587	0.1250	4.2187
Forest Types	4	0.2974	0.0280	1.4115
100 leaves plant species	5	-0.0683	0.0012	3.1143

Five clustering algorithms have been tested, and results are presented in Table.20. The cortical algorithm (CA) [44], a six-layered structure of cortical columns using a MapReduce parallel implementation, outperformed other algorithms on the Arcene and 100 leaves plant species datasets of 2 and 5 classes respectively. Additionally, it recorded better values than our baseline on all measures (Table.19). Meanwhile, on Balance-Scale dataset, the CA and kmeans algorithms came to record the top values. Though, the hierarchical algorithm scored the best DI and DBI indexes for the Forest type dataset, the CA clustering algorithm recorded the highest accuracy of 96.77%.

Table.20: Clustering with Supervised Datasets

Dataset	**Arcene**			
Clustering Algorithm	Acc (%)	SI	DI	DBI
Maximin	66	0.5122	0.6772	1.2929
K-means	59	0.5318	0.6936	**1.0762**
Hierarchical	59	0.5318	0.6936	**1.0762**
K-medoids	68.5	0.4439	0.2435	**1.0762**
CA	**73**	**0.6488**	**0.7893**	1.0948
Dataset	**Balance-Scale**			
Clustering Algorithm	Acc (%)	SI	DI	DBI
Maximin	53.92	0.2424	0.1250	1.9758
K-means	56.32	**0.2814**	**0.1429**	1.6996
Hierarchical	46.4	-0.2223	0.1250	1.1428
K-medoids	51.36	0.2755	**0.1429**	1.1428
CA	**83.64**	0.2587	0.1250	**1.1279**

Dataset	**Forest Types**			
Clustering Algorithm	Acc (%)	SI	DI	DBI
Maximin	60.4207	0.1806	0.0619	2.6339
K-means	77.8203	**0.4803**	0.0234	1.1431
Hierarchical	37.6673	0.3926	**0.2625**	**0.5923**
K-medoids	74.5698	0.4496	0.0254	**0.5923**
CA	**96.77**	0.2974	0.0480	1.1457
Dataset	**100 leaves plant species**			
Clustering Algorithm	Acc (%)	SI	DI	DBI
Maximin	40	0.0106	6.0399e-04	2.4949
K-means	40	**0.5168**	0.2429	1.2330
Hierarchical	33.3333	0.4684	**0.3539**	0.8668
K-medoids	35.4167	0.4349	0.1182	0.8668
CA	**79.909**	0.0683	0.1012	**0.8637**

Table.21 below shows the measurement values for the Libras unsupervised dataset. The CA algorithm proved to form the best clusters on 3 and 5 cluster numbers. On 2 clusters, both The CA and kmeans algorithm gave the best values. However, on 4 clusters, k-medoids gave the best results.

Table.21: Clustering Measures for Libras Dataset

# of clusters	**2**		
Clustering Algorithm	SI	DI	DBI
Maximin	0.7187	0.0960	0.6897
K-means	0.7199	**0.1546**	**0.6793**
Hierarchical	0.2032	0.1542	0.7862
K-medoids	**0.7301**	0.1411	0.7862
CA	**0.7301**	0.1411	**0.6793**
# of clusters	**3**		
Clustering Algorithm	SI	DI	DBI
Maximin	0.5113	0.0178	0.8311
K-means	**0.6301**	**0.2582**	0.8197
Hierarchical	0.1096	0.1499	1.0938
K-medoids	**0.6301**	**0.2582**	1.0938
CA	**0.6301**	**0.2582**	**0.8176**

# of clusters	**4**		
Clustering Algorithm	SI	DI	DBI
Maximin	0.3024	0.0528	1.0959
K-means	0.5449	0.1753	0.9790
Hierarchical	0.1217	0.1451	**0.7570**
K-medoids	**0.5450**	0.0839	**0.7570**
CA	0.5353	**0 2168**	0.9865
# of clusters	**5**		
Clustering Algorithm	SI	DI	DBI
Maximin	0.2217	0.0267	2.0446
K-means	0.4950	0.2338	1.1043
Hierarchical	0.1167	0.1446	**0.6408**
K-medoids	0.5040	0.2077	**0.6408**
CA	**0.5203**	**0.2617**	1.0424

Furthermore, Tables.22, 23 and 24 below form our homemade test corpuses for unsupervised clustering. On 2 clusters, SI, DI, and DBI measures recorded better values than 3, 4, and 5 which indicates that our procedure requires only 2 clusters. Moreover, the four datasets show consistency in results regardless their sizes (number of retrieved tweets).

On DTS1, specifically on 2 clusters, the Maximin, and hierarchical algorithms recorded best SI, DI, and DBI values (highest SI and DI, and lowest DBI) according to Table.17. Additionally, on DTS2 and DTS3, also on 2 clusters, the Maximin algorithm showed identical results to K-means, Hierarchical, and K-medoids on all measures. However, applying 3, 4, and 5 clusters to our datasets, Maximin detected poor cluster formulations (very low SI and DI values, and high DBI indications). The fact that applying our features on short phrases, tweets, is generating relatively close tweet features (short Euclidean distances), besides that Maximin starts by identifying the farthest "n" points as cluster centers, cluster measurements would show poor cluster formulations. Indeed, Maximin algorithm proved to fit our methodology especially requiring 2 clusters under unsupervised learning, while the

CA algorithm proved to record the best performance on supervised datasets and Libras unsupervised dataset.

Table.22: Clustering Results for Dataset 1 (DTS1)

# of clusters	**2**		
Clustering Algorithm	SI	DI	DBI
Maximin	**0.9945**	**1.0112**	**0.1237**
K-means	0.9886	0.1052	0.4705
Hierarchical	**0.9945**	**1.0112**	**0.1237**
K-medoids	0.9865	0.0207	**0.1237**
CA	0.9896	0.0730	0.4487
# of clusters	**3**		
Clustering Algorithm	SI	DI	DBI
Maximin	-0.4239	0.0020	1.2801
K-means	0.9864	0.0721	0.3424
Hierarchical	0.9923	0.4951	0.0139
K-medoids	0.9875	0.0922	0.0139
CA	0.9864	0.0721	0.3292
# of clusters	**4**		
Clustering Algorithm	SI	DI	DBI
Maximin	-0.0085	9.8545e-04	1.2224
K-means	0.9857	0.0951	0.3205
Hierarchical	0.9863	0.3438	0.1035
K-medoids	0.9857	0.0951	0.1035
CA	0.9857	0.0951	0.3115
# of clusters	**5**		
Clustering Algorithm	SI	DI	DBI
Maximin	0.0400	9.8547e-04	1.6565
K-means	0.9862	0.1294	0.3900
Hierarchical	0.9804	0.2237	0.1896
K-medoids	0.9862	0.1294	0.1896
CA	0.9651	0.0192	0.3902

Table.23: Clustering Results for Dataset 2 (DTS2)

# of clusters	**2**		
Clustering Algorithm	SI	DI	DBI
Maximin	**0.9988**	**1.7061**	**0.0061**
K-means	**0.9988**	**1.7061**	**0.0061**
Hierarchical	**0.9988**	**1.7061**	**0.0061**
K-medoids	**0.9988**	**1.7061**	**0.0061**
CA	**0.9988**	**1.7061**	0.5274
# of clusters	**3**		
Clustering Algorithm	SI	DI	DBI
Maximin	-0.4111	0.0038	1.7034
K-means	0.9955	0.1706	0.2085
Hierarchical	0.9910	0.4007	0.0119
K-medoids	0.9955	0.1706	0.0119
CA	0.9955	0.1706	0.3115
# of clusters	**4**		
Clustering Algorithm	SI	DI	DBI
Maximin	-0.0859	0.0038	1.5987
K-means	0.9918	0.1066	0.3331
Hierarchical	0.9912	0.2211	0.0335
K-medoids	0.9911	0.0911	0.0335
CA	0.0842	0.0038	0.8489
# of clusters	**5**		
Clustering Algorithm	SI	DI	DBI
Maximin	-0.0905	0.0038	1.9186
K-means	0.9923	0.2082	0.2480
Hierarchical	0.9921	0.3210	0.1550
K-medoids	0.9923	0.2082	0.1550
CA	0.3243	0.0047	0.6338

Table.24: Clustering Results for Dataset 3 (DTS3)

# of clusters	**2**		
Clustering Algorithm	SI	DI	DBI
Maximin	**0.9992**	**2.6858**	**0.0032**
K-means	**0.9992**	**2.6858**	**0.0032**
Hierarchical	**0.9992**	**2.6858**	**0.0032**
K-medoids	**0.9992**	**2.6858**	**0.0032**
CA	**0.9992**	**2.6858**	1.2028
# of clusters	**3**		
Clustering Algorithm	SI	DI	DBI
Maximin	0.0027	0.0017	1.3556
K-means	0.9996	4.8873	0.0043
Hierarchical	0.9996	4.8873	0.0043
K-medoids	0.9996	4.8873	0.0043
CA	0.9996	4.8873	0.7032
# of clusters	**4**		
Clustering Algorithm	SI	DI	DBI
Maximin	0.1202	0.0017	1.3138
K-means	0.9912	0.4853	0.1851
Hierarchical	0.9871	0.7247	0.0160
K-medoids	0.9912	0.4853	0.0160
CA	0.9912	0.4853	0.4102
# of clusters	**5**		
Clustering Algorithm	SI	DI	DBI
Maximin	0.1856	0.0017	1.3250
K-means	0.3201	0.0205	0.2993
Hierarchical	0.9899	1.1090	0.0855
K-medoids	0.9899	1.1090	0.0855
CA	0.3201	0.0205	0.7441

4.2 Phrase Classification (Objective vs. Subjective)

Table.25 shows different classification measures to validate our results based on True positives (TP), True negatives (TN), False positives (FP), and False negatives (FN).

Table.25: Classification Measures

Measure	Description
Sensitivity	TP / (TP+FN)
Specificity	TN / (TN+FP)
Accuracy	(TP+TN) / (TP+TN+FP+FN)
Precision	TP / (TP+FP)
Recall	TP / (TP+FN)
F-measure	2*((precision*recall) / (precision + recall))

We performed some experiments on 5 different classifiers (Table.26) with various features on our homemade corpus of 1007 pre-relabeled phrases as objective or subjective. 499 phrase were labeled as objective and 508 as subjective in which 1/3 of data are randomly selected for testing and the other 2/3 of data for training. With a 1-fold cross validation Table.27 recorded the obtained results.

Table.26: Classifiers Parameters

Classifiers	Parameters
K-nearest neighbor (KNN)	K=5
Support vector Machine (SVM)	Linear SVM
Neural Networks (NN)	1 Layer with 20 neurons
Naïve Bayes (NB)	Kernel Distribution with empirical Prior
Adaboost (AB)	15 ensemble learning cycles with (Tree) as a weak learner

Table.27: Testing Results for Different Features on Several Classifiers

Features	**Syntactic**					
Classifier	Sensitivity	Specificity	Accuracy	Precision	Recall	F-measure
KNN	0.7168	0.9018	0.8065	0.8857	0.7168	0.7923
SVM	**0.7977**	**0.9387**	**0.8661**	**0.9324**	**0.7977**	**0.8598**
NN	0.7861	0.9325	0.8571	0.9252	0.7861	0.8500
NB	0.7399	0.9018	0.8185	0.8889	0.7399	0.8076
AB	0.7457	0.9080	0.8244	0.8958	0.7457	0.8139
Features	**Semantic**					
Classifier	Sensitivity	Specificity	Accuracy	Precision	Recall	F-measure
KNN	**0.6901**	0.8242	0.7560	0.8027	0.6901	0.7421
SVM	0.6316	0.8121	0.7202	0.7770	0.6316	0.6968
NN	0.6667	**0.8848**	**0.7738**	**0.8571**	0.6667	**0.7500**
NB	0.6784	0.8606	0.7679	0.8345	**0.6784**	0.7484
AB	0.6784	0.8061	0.7411	0.7838	**0.6784**	0.7273
Features	**Syntactic + Semantic**					
Classifier	Sensitivity	Specificity	Accuracy	Precision	Recall	F-measure
KNN	0.7529	0.9212	0.8358	0.9078	0.7529	0.8232
SVM	**0.8118**	0.9394	0.8746	0.9324	**0.8118**	0.8679
NN	**0.8118**	**0.9515**	**0.8806**	**0.9452**	**0.8118**	**0.8734**
NB	0.7882	0.9333	0.8597	0.9241	0.7882	0.8508
AB	0.7941	0.8909	0.8418	0.8824	0.7941	0.8359

Table.27 shows that the combination of syntactic and semantic features over performed the semantic and syntactic features alone respectively on the different classifier used. SVM classifier recorded the highest accuracy of 86.61% on syntactic features, it dropped to reach 77.38% using Neural Networks on semantic features. However, a combination of both semantic and syntactic features recorded the highest accuracy of 88.06% using Neural Networks. Table.28 indicates some common failures for our classifiers differentiated in 4 categories on the combination of syntactic and semantic features. The whole set is available in the appendix.

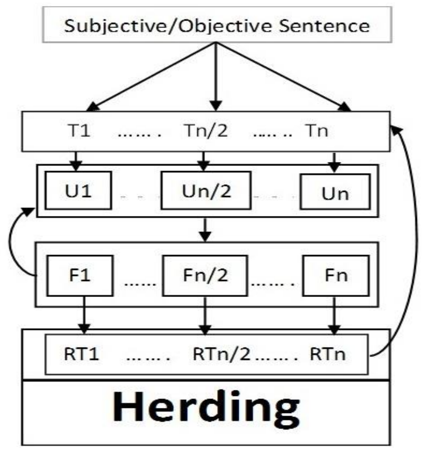
Table.28: Some Misclassified Data (Syntactic + Semantic Features)

Misclassified Data	Actual Label	Predicted Label	**Categorized failures**
• Love was hit by Barnes elbow • Man United are the Red Devils. • Warriors miss green in game 5'	Objective	Subjective	Weakness in Nicknames Database
• Warriors have the best record in the NBA • Cavs win first NBA title • Bartolo Colon was injured • Murray to win the Queen's title • Stephen Curry will not play in the 2016 Summer Olympics	Objective	Subjective	Objective Phrases with subjective word scores

• Warriors in control • Game 5: LeBron, Kyrie keep Cavs alive • Tough match, Portugal vs. Croatia • France to round 16 in Euro 2016 **(tweets for this phrase were retrieved before knowing if France will be qualified to round 16)**	Subjective	Objective	Phrases not grammatically correct
• Thierry Henry is a legend • Baseball is in the blood • Warriors are nothing without Steph Curry • Curry is the face of NBA • Wolves plan to target Joakim Noah • Rumours: Vermaelen to Liverpool	Subjective	Objective	Subjective Phrases without adjectives

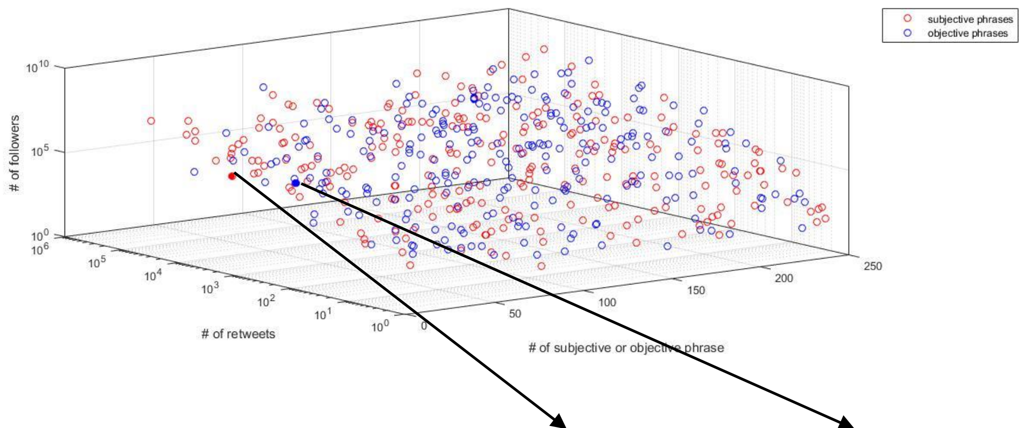
4.3 Herding Effect on Consensus

Figure.16: Herding Scheme



Despite that our framework introduces the concept of subjective tweets with/out consensus, it extends analysis for tweet herding. According to [45], herd behaviors occur when individuals witness other behaviors and take actions affiliated with them. Experimenting this concept in our work, we identified that herding may lead to consensus on subjective phrases rather than objective ones using figure.16, were “T” stands for tweet, “U” for user, “F” for follower, and “RT” for retweet. In this scheme, subjective and objective tweets belonging to the same cluster of the subjective/objective phrases are extracted. However, each tweet has a user, which in return has a certain number of followers that could witness his/her blogs. Consequently, these followers could themselves retweet the original tweet of the user indicating an agreement. Furthermore, followers are also considered as users that have their own groups that also could seek their retweets. In such a structure, results are recorded in figure.17.

Figure.17: # of Followers and Retweets of Subjective and Objective Phrases



Subjective Phrase / Objective Phrase	Football is a boring game	D'Angelo Russell plays with Lakers
#of followers	304507	690216
#of retweets	2347 > 197 (Herding)	197

Table.29: Herding in Subjective Phrases

Around 250 subjective phrases and 250 objective ones were manually collected, in which tweets are extracted for each classified phrase according to its keywords (Table.30 and Table.31) in the same process discussed before. Consequently, the number of followers and retweets were recorded. Indeed, herding behaviors appeared in subjective phrases. Additionally, the number of retweets of the subjective tweets may exceed the number of retweets of the objective ones, though the number of followers of the latter may exceed the number of followers of the subjective phrases. However, herding may lead to a subjective consensus of tweets. For instance, Figure.18 highlights a case (red, and blue dotted spots), in which the subjective classified phrase ("football is a boring game.") retrieved 2347 retweet from 304507 followers and the objective classified phrase ("D'Angelo Russell plays with Lakers.") retrieved 197 retweet from 690216 followers. In such a case, the herding effect lead to an agreement on the subjective phrase rather than that of the objective one.

Table.30: Extracted Keywords for each Classified Subjective Phrase

# of subjective phrase	Keywords	# of subjective phrase	Keywords	# of subjective phrase	Keywords	# of subjective phrase	Keywords	# of subjective phrase	Keywords	# of subjective phrase	Keywords
1	Lakers eliminated	21	Football best game	41	Scottish football bad	61	Xavi best player	81	Real Madrid better Man City	101	One win good
2	Warriors best record NBA	22	Neymar better Messi	42	Bayern Munich better Real Madrid	62	Ibrahimovic best player	82	Bayern UEFA champions League	102	game lot better
3	Ramos Spanish player	23	Neymar better Ronaldo	43	Michael Jordan best player	63	Neymar best player	83	Ronaldo scored best goals	103	Celtics chance win
4	NYY called Yankees	24	Liverpool stronger Chelsea	44	Michael Jordan worst player	64	Mesut Ozil best player	84	motivated want win	104	play Liverpool
5	Two games series	25	Manchester City better United	45	Kobe Bryant best player	65	David Silva best player	85	Manchester City play well	105	expect defending champs
6	Warriors 2 0 Cavs	26	Basketball best game	46	LeBron James best player	66	Sergio Ramos best player	86	Cavaliers dominant	106	Bayern strong team
7	Bayern Munich better Barcelona	27	Messi scores best goals	47	Larry Bird best player	67	Willian best player	87	crowd great week	107	Atletico better Bayern
8	Chelsea better Manchester City	28	Ray Allen best 3 point shooter	48	Wilt Chamberlain best player	68	Willian worst player	88	Ronaldo fans Messi	108	Bayern play well
9	basketball better football	29	football exciting basketball	49	Kevin Durant best player	69	Wayne Rooney best player	89	Zidane blamed	109	Today match excited
10	football boring game	30	Guardiola better Mourinho	50	Kevin Durant best scorer	70	Juan Mata best player	90	good European game	110	Game 2 better 1
11	Basketball boring game	31	Guardiola best coach	51	Chris Paul best player	71	Thomas Muller best player	91	Real win champions league	111	Atletico push
12	Lakers best team	32	Manuel Neuer best goalkeeper	52	Vince Carter best dunker	72	Oscar best player	92	Pistons difficult beat	112	winner good today
13	Soccer popular sport world	33	Kobe Bryant better LeBron James	53	Bradley Beal best player	73	Oscar worst player	93	PSG great team	113	Pacers play better
14	Juventus better Chelsea	34	Mourinho best coach	54	Dwyane Wade best player	74	Fernandinho best player	94	Blazers keep	114	Celtics perform well
15	Champions League best	35	Ronaldo plays better Messi	55	Harrison Barnes best player	75	Gareth Bale best player	95	good Clippers	115	Cavaliers leave Lebron
16	Benzema best player	36	Real Madrid best team	56	Tim Duncan best player	76	Dani Alves best player	96	Defense similar	116	Lebron take court
17	Buffon best goalkeeper	37	Barcelona best team	57	Ryan Anderson best player	77	Petr Cech best player	97	Raptors better Pacers	117	Things Warriors
18	Benzema good player	38	Messi better Ronaldo	58	Jamal Crawford best player	78	Suarez better Messi	98	Blazers beat Clippers	118	Arsene Wenger work
19	Xavi better Iniesta	39	Spanish league worst	59	Lionel Messi best player	79	Manchester City best team	99	pitch disgrace	119	Thunder solid
20	Gerrard best player	40	Premier League best league	60	Cristiano Ronaldo best player	80	Manchester City Champions	100	injuries legitimate	120	disappointing season Chelsea

# of subjective phrase	Keywords	# of subjective phrase	Keywords	# of subjective phrase	Keywords	# of subjective phrase	Keywords	# of subjective phrase	Keywords	# of subjective phrase	Keywords	# of subjective phrase	Keywords
121	Man City good Madrid	141	Pistons best shooters	161	Curry starting lineup	181	Barcelona better Alves	201	Cavs win game 5	221	Durant play Warriors	241	Ronaldo best player Portugal
122	Ronaldo Benzema play	142	bad season Lakers	162	Curry fully healthy	182	Lebron James better	202	Lebron win game	222	Italy win Euro	242	Portugal depend Ronaldo
123	Aguero rested	143	Celtics good season	163	horrible series	183	Atletico deserve win	203	Cavs game home	223	England one best teams	243	Portugal defensive
124	City rest weekend	144	Bayern playing team	164	Raptors horrible	184	Cavs gave end	204	Warriors defend title	224	Mets struggling score	244	Ronaldo better Euro
125	Atletico win UEFA	145	Conley Gasol	165	Curry best shooter time	185	Cavs chance win	205	Cavs great job	225	Jamie Vardy remain Leicester City	245	Italy great job
126	Spanish teams dominating	146	pressure Leicester	166	Lillard McCollum best	186	Ronaldo special player	206	Cavs Champs	226	Russia won Wales	246	Sanchez traded
127	Last game exciting	147	Manchester United fight	167	Russell played unbelievable	187	Hornacek change Knicks	207	dunk foul	227	Germany better Euro		
128	Liverpool achieve Klopp	148	Leicester fearless Champions League	168	foul Spurs	188	Klay Thompson left first quarter	208	James finished triple double	228	Lebron great job		
129	FA cup magic	149	Leicester impossible	169	Tonight emotional game	189	Barcelona UEFA finals	209	Cavaliers series win	229	Curry blamed loss		
130	Arsenal improve performance	150	Chelsea nothing play	170	Durant know	190	Warriors win championship	210	historic season Cavs	230	Wales great team		
131	Leicester slip	151	Messi better age	171	Durant stay Thunder	191	Cavs finally back game	211	Warriors greatest team	231	Lebron played game life		
132	Newcastle survive	152	Final favourite	172	Azarenka good	192	good follow Cavs	212	win Cavs expected	232	Ozil talented player		
133	More expected Arsenal	153	Real Madrid lucky	173	Atletico catch Real	193	Warriors ready play	213	GSW won	233	better control Moore		
134	Hornets playing better	154	Real dominated game	174	Warriors deserve win	194	tonight Warriors	214	England knocked Euro	234	Djokovic beat Murray		
135	Hornets worst team	155	City win fans	175	Curry Warriors win	195	playoffs hard	215	Lukaku stay Everton summer	235	NBA game 7 easy		
136	Rockets push	156	Curry valuable player	176	Real played better Mourinho	196	Warriors tough	216	Warriors	236	Wales make finals		
137	Bulls play better	157	Curry shooting passing	177	Spurs make major	197	Warriors struggle road	217	Warriors take Durant	237	Germany tough team		
138	Stanley Johnson better	158	Curry best shooter NBA	178	Spurs push	198	Cavs trust	218	Cavs trade Love	238	Warriors win year		
139	good season Cavs	159	Curry led Warriors win	179	Thunder defensive	199	Warriors better Cavs	219	best NBA finals	239	Round 16 exciting round		
140	Walton Kings	160	Curry best player Warriors	180	Thunder Big 3 win	200	Cavs prove	220	Ronaldo star real Madrid	240	James resign Cavs		

Table.31: Extracted Keywords for each Classified Objective Phrase

#of objective phrase	Keywords	#of objective phrase	Keywords	#of objective phrase	Keywords	#of objective phrase	Keywords	#of objective phrase	Keywords
1	Derrick basketball player	26	Kobe Bryant basketball player	51	Miami Heat basketball team	76	Aston Villa football team	101	Kei Nishikori tennis player
2	Zidane football player	27	LeBron James basketball player	52	Miami Heat NBA team	77	Chelsea football team	102	Serena williams tennis player
3	Tottenham football team	28	Larry Bird basketball player	53	Milwaukee Bucks NBA team	78	Everton football team	103	simona halep tennis player
4	Iniesta plays Barcelona	29	Shaquille O'Neal basketball player	54	Minnesota Timberwolves NBA team	79	Leicester City football team	104	Spain won world cup 2010
5	Arda Turan plays Barcelona	30	Kevin Durant basketball player	55	New York Knicks NBA team	80	Liverpool football team	105	Henderson Captain Liverpool
6	Lionel Messi plays Barcelona	31	Steve Nash basketball player	56	Oklahoma City Thunder basketball team	81	Manchester City football team	106	Man United Red Devils
7	Neymar plays Barcelona	32	Dirk Nowitzki basketball player	57	Oklahoma City Thunder NBA team	82	Manchester United football team	107	Barcelona Spanish team
8	LeBron James plays Cavaliers	33	Tim Duncan basketball player	58	Phoenix Suns basketball team	83	Newcastle United football team	108	Barcelona played Real Madrid
9	Kyrie Irving plays Cavaliers	34	Atlanta Hawks NBA team	59	Phoenix Suns NBA team	84	Norwich City football team	109	Germany won world cup 2014
10	Nick Young plays Lakers	35	Boston Celtics basketball team	60	Portland Trail Blazers NBA team	85	Southampton football team	110	John Terry captain Chelsea
11	Jordan Clarkson plays Lakers	36	Charlotte Bobcats NBA team	61	Sacramento Kings NBA team	86	Stoke City football team	111	Beckham played England
12	D'Angelo Russell plays Lakers	37	Chicago Bulls basketball team	62	Toronto Raptors NBA team	87	Sunderland football team	112	Ozil plays Arsenal
13	Julius Randle plays Lakers	38	Chicago Bulls NBA team	63	Utah Jazz NBA team	88	Tottenham Hotspur football team	113	Arsenal play Premier League
14	Keylor Navas plays Real Madrid	39	Cleveland Cavaliers NBA team	64	Washington Wizards NBA team	89	Watford football team	114	Man United English team
15	Gareth Bale plays Real Madrid	40	Cleveland Cavaliers NBA team	65	Lionel Messi football player	90	West Ham United football team	115	Real Madrid Spanish team
16	Casemiro plays Real Madrid	41	Denver Nuggets basketball team	66	Cristiano Ronaldo football player	91	Juventus football team	116	Brad Stevens coach Boston Celtics
17	Toni Kroos plays Real Madrid	42	Denver Nuggets NBA team	67	Neymar football player	92	Roma football team	117	Suarez plays Uruguay
18	Cristiano Ronaldo plays Real Madrid	43	Detroit Pistons basketball team	68	Wayne Rooney football player	93	Palermo football team	118	Pique defender Barcelona
19	LeBron James NBA player	44	Detroit Pistons NBA team	69	Oscar football player	94	Bayern Munich football team	119	Iniesta midfielder Barcelona
20	Mo Williams NBA player	45	Golden State Warriors basketball team	70	Gareth Bale football player	95	Dortmund football team	120	Navas goalkeeper Real Madrid
21	Kyrie Irving NBA player	46	Golden State Warriors NBA team	71	Eden Hazard football player	96	Djokovic tennis player	121	Premier League football
22	D'Angelo Russell NBA player	47	Houston Rockets basketball team	72	Atletico Madrid football team	97	Andy Murray tennis player	122	Champions League football
23	Kobe Bryant NBA player	48	Houston Rockets NBA team	73	Barcelona football team	98	Roger Federer tennis player	123	World cup 2014 Brazil
24	Michael Jordan basketball player	49	LA Lakers NBA team	74	Real Madrid football team	99	Stan Wawrinka tennis player	124	Juventus Italian club
25	Magic Johnson basketball player	50	Memphis Grizzlies NBA team	75	Arsenal football team	100	Rafael Nadal tennis player	125	Ronaldo Portuguese player

#of objective phrase	Keywords	#of objective phrase	Keywords	#of objective Phrase	Keywords	#of objective phrase	Keywords	#of objective phrase	Keywords
126	Buffon goalkeeper Juventus	151	Mets baseball team	176	Astros baseball team	201	World cup boring	226	Cavs won game 5
127	Neymar Brazilian player	152	Braves baseball team	177	3 assists Pistons	202	World cup exciting	227	Cleveland Cavaliers beat Golden State Warriors
128	Chelsea play Champions League	153	Blue Jays baseball team	178	Noah plays Bulls	203	Baseball blood	228	NBA finals Game 5
129	Sturridge plays Liverpool	154	Royals baseball team	179	Zinedine Zidane Real Madrid coach	204	Hart hero	229	Italy 2 Belgium 0
130	NBA basketball association	155	Vikings Titans	180	Starlin plays Yankees	205	Problems Barcelona	230	Comeback Cavs
131	FIFA football association	156	Red Sox 7 Astros 5	181	Girardi manager Yankees	206	Kevin Durant led	231	Arsenal Liverpool Premier League 2016
132	Liverpool played Arsenal	157	Mets baseball team	182	Draw Arsenal Sunderland	207	Guardiola genius	232	Cavaliers win NBA championship
133	Cech goalkeeper	158	Braves baseball team	183	UEFA Champions League final Milan	208	Ronaldo leads Messi	233	4 3 lead Cavs
134	Welbeck plays Arsenal	159	Blue Jays baseball team	184	Isaiah plays Celtics	209	Real president Florentino Perez	234	Cavaliers Warriors Game 7
135	Walcott plays Arsenal	160	Royals baseball team	185	Curry MVP	210	Thompson Leads Warriors	235	Warriors series loss
136	Hazard midfielder Chelsea	161	Padres Cardinals	186	Curry second MVP	211	Dwyane Wade takes control	236	Lebron James back reason
137	Paul plays Clippers	162	Cavaliers beat Pistons	187	Real Madrid won Champions League	212	Steph Curry trending	237	James plays Cleveland
138	Messi 28 years	163	Football referee	188	Atletico lost Real Madrid	213	Curry changed game	238	France 0 Switzerland
139	Rivers plays Clippers	164	March Madness basketball	189	Warriors 2 0 lead Cavs	214	Warriors control	239	NBA rigged
140	Bournemouth Man City	165	Real Madrid Manchester City	190	LeBron James 19 points	215	Warriors Finals	240	NBA finals rigged
141	Norwich Newcastle	166	Manchester United Barcelona	191	Durant playoffs	216	Cavs finals	241	NBA fixed
142	Stoke Swansea	167	Coventry City football team	192	Warriors lead series 2 0	217	Warriors control	242	Cavs made history
143	Sky Bet League football league	168	Real Madrid semi finals	193	Game 2 Warriors	218	Cavs missed Love	243	Warriors disappointed
144	Charlotte Atlanta	169	Barcelona won Real	194	Warriors finals	219	Cavs get Game 3	244	Hodgson fear
145	Atletico Barcelona 2 0	170	Olivier plays Arsenal	195	Game 1 Warriors won Stephen Curry	220	Curry face NBA	245	Rio Olympics Brazil
146	Hawks Celtics	171	Arsenal lost Barcelona	196	Real Madrid 1 Atletico	221	Game 3 Wrap	246	Durant Free Agency
147	Serie A Italy League	172	Roma football team	197	Real Madrid president Florentino Perez	222	Warriors step road	247	Rosberg wins
148	Chargers Chiefs	173	Santi plays Arsenal	198	Vardy Arsenal Leicester	223	Golden State leads series 3 1	248	Johnson win Open
149	Vikings Titans	174	Hector plays Arsenal	199	Roger Federer champion	224	Warriors week Green	249	England round 16
150	Red Sox 7 Astros 5	175	Ramos plays Spain	200	Thierry Henry legend	225	Warriors 3 2 lead Cavs	250	Wayne Rooney striker

4.4 Comparing Results with Tweenator

To compare our proposed method to Tweenator, we also conducted experiments on the Stanford dataset [46] which consist of 177 negative and182 positive tweets as testing data. Alike Tweenator, a subset of 60,000 tweet of the original Stanford training data was randomly selected. Using the Naïve Bayes classifier, Table.32 shows the results.

Table.32: Naïve Bayes Classification Results on the Stanford Dataset

Tool	Method	Accuracy
Tweenator	Unigrams	81.0%
	Semantic replacement	77.3%
	Semantic augmentation	80.45%
	Semantic interpolation	84.1%
	Sentiment-topic features	**86.3%**
Data Analytics Framework for Sport Related Tweets	Syntactic features	54.87%
	Semantic features	**72.7%**
	Syntactic + Semantic	60.72%

The proposed set of semantic features recorded the highest accuracy of 72.7% than our syntactic features alone and a combination of the latter with the semantic features of accuracies 54.87% and 60.72% respectively. However, tweets in the training and testing data are not related to sports, as our framework is targeted specifically for sports related phrases in which our corpus is trained on. Moreover, this data eliminates most emoticons from tweets as they are part of our semantic feature vector, considering also that tweets are not complete and correct English sentences affecting our tagging process by the POS tagger. An additional experiment is conducted on the same 60,000 tweets randomly selected from the same training set combined with the provided test set, using the leave one out (LOO) technique as a worst scenario. The following accuracies were recorded: 60.20%, 63.73%, and 64.41% for syntactic, semantic, and a combination of syntactic and semantic features respectively.

Table.33 shows some common speculated fails. Others are provided in the appendix.

Table.33: Misclassified Data on Stanford Dataset

<table>
<tr><th>Misclassified Data</th><th>Categorized failures</th></tr>
<tr><td>I hate the effing dentist.
my cluster is back and I'm a happy man
@ArunBasilLal I love Google Translator too
reading on my new Kindle2
Stanford Charity Fashion Show a top draw</td><td>Phrases not related to sports</td></tr>
<tr><td>curses the Twitter API limit
Time Warner's HD line up is crap.
Lyx is cool
yahoo answers can be a butt sometimes</td><td>Eliminated Emoticons</td></tr>
<tr><td>zomg!!! I have a G2!!!!!!!
Blink by malcolm gladwell amazing book and
The tipping point!
GM files Bankruptcy, not a good sign…
sad day...bankrupt GM
Comcast sucks.
@SCOOBY_GRITBOYS</td><td>Phrases not grammatically correct</td></tr>
</table>

CHAPTER 5

APPLICATION

This section demonstrates the proposed Application. We show screen-shots of all steps from login phase until classifying the input phrases.

5.1 User Login

In the first step, as the user launches the application (APP), the tool checks for internet connectivity and informs the user about its status as Figures.18 (a) and (b) illustrates.

Figure.18(a): Internet Connection exist

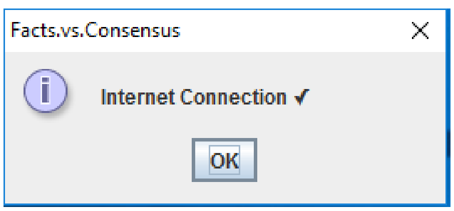
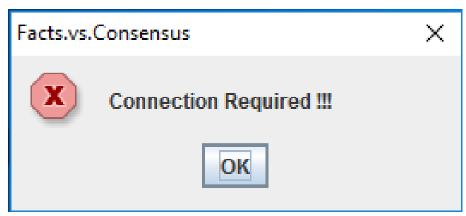


Figure.18(b): Connection Required



However, all users intending to use the application must authorize our APP, in a sense they must login with their Twitter credentials. Indeed, as the application is launched, it provides the user with a URL for this process (Figure.19, Figure.20).

Figure.19: URL for Authorization

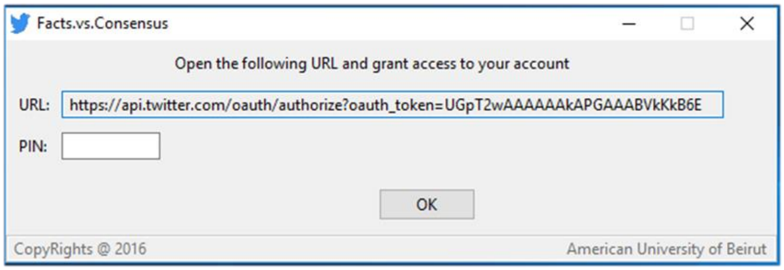
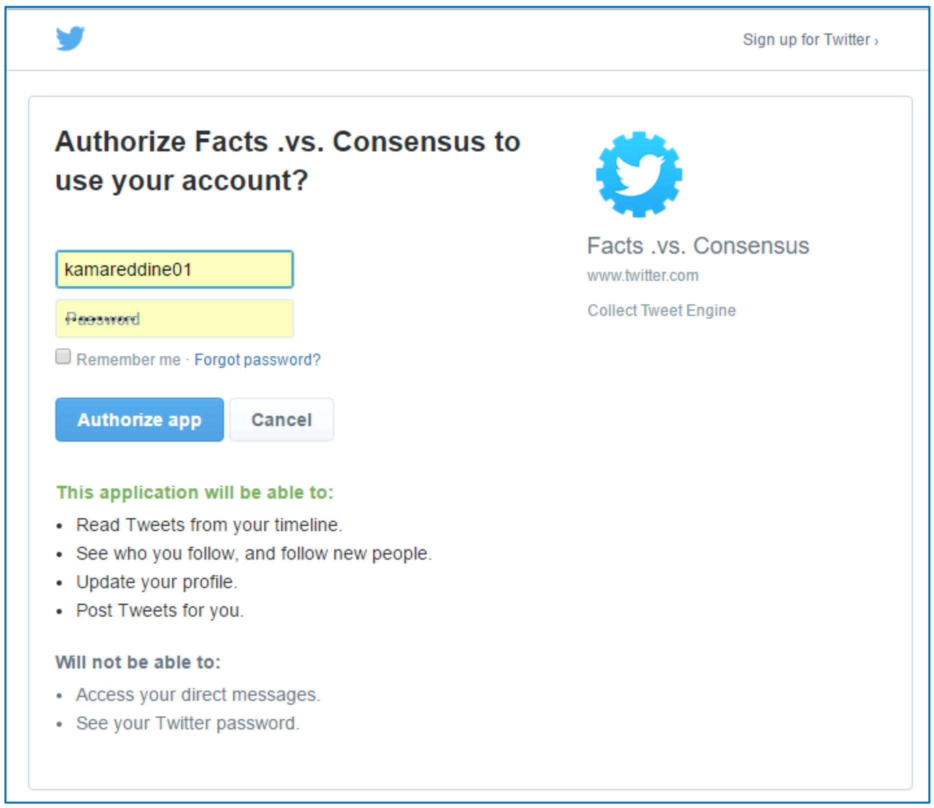
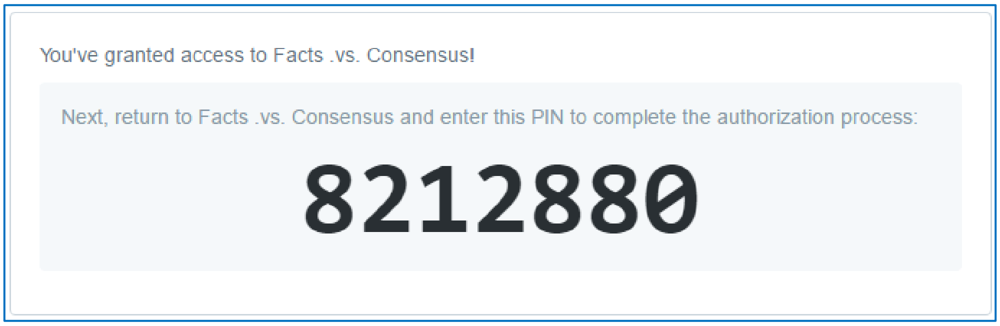


Figure.20: User Account Access



Subsequently, the user is provided with his/her PIN by twitter servers to complete the Authorization Process(Figure.21) and grant access to our tool. At this point, the tokens for each end user is saved as previously discussed above.

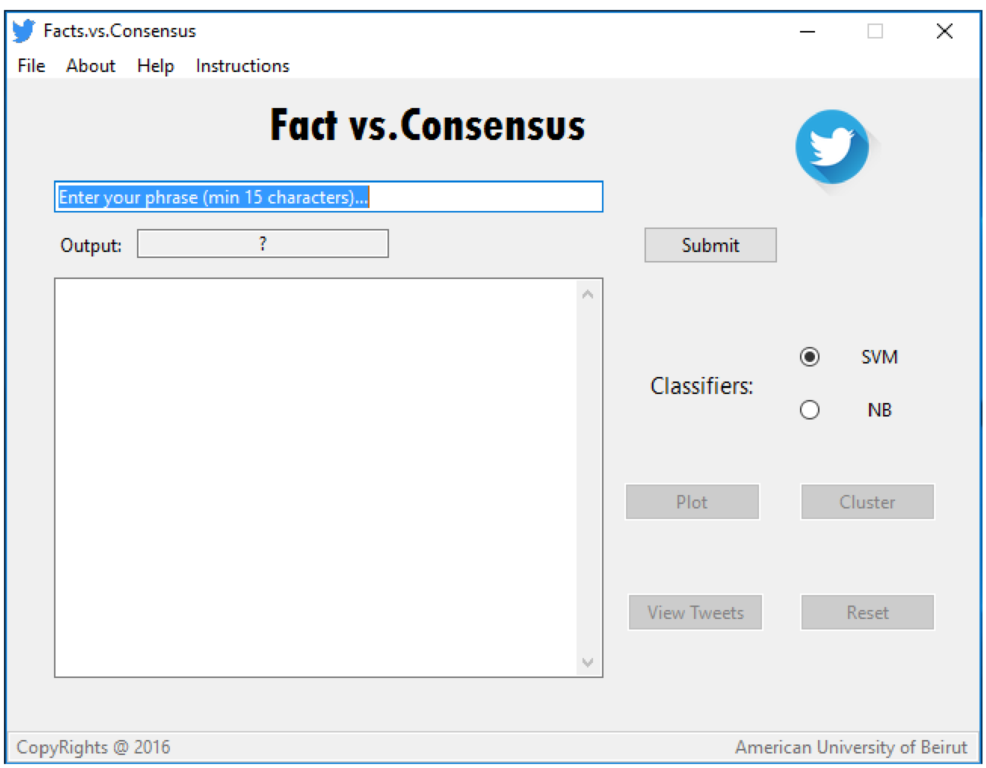
Figure.21: User PIN



5.2 Graphical User Interface

After the authorization process is completed, the tool would be displayed as Figure.22 demonstrates.

Figure.22: Application Overview



The app constitutes of a single text bar in which the user enters his/her Sport-phrase in, once the user chooses his/her classification method (SVM, NB, etc.) and after clicking the submit button, the tool starts executing the proposed methodology. Figures.23(a) and (b) shows the process.

Figure.23(a): Methodology(I)

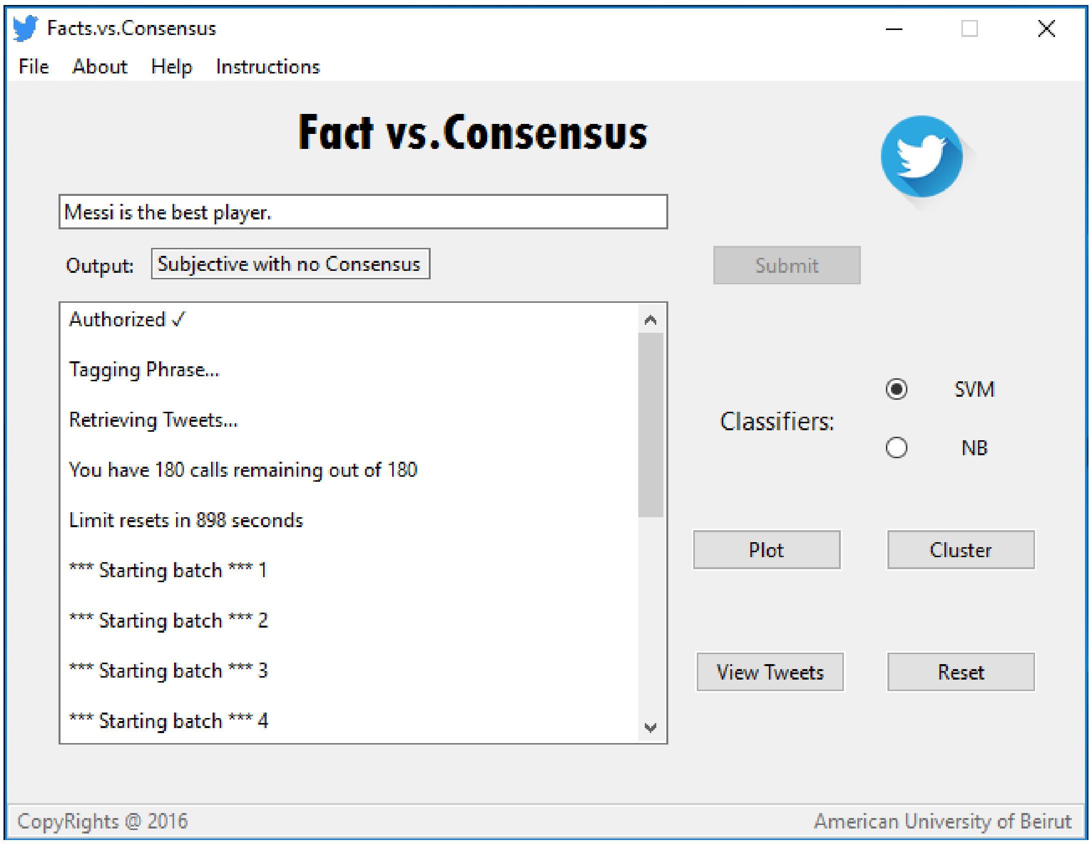
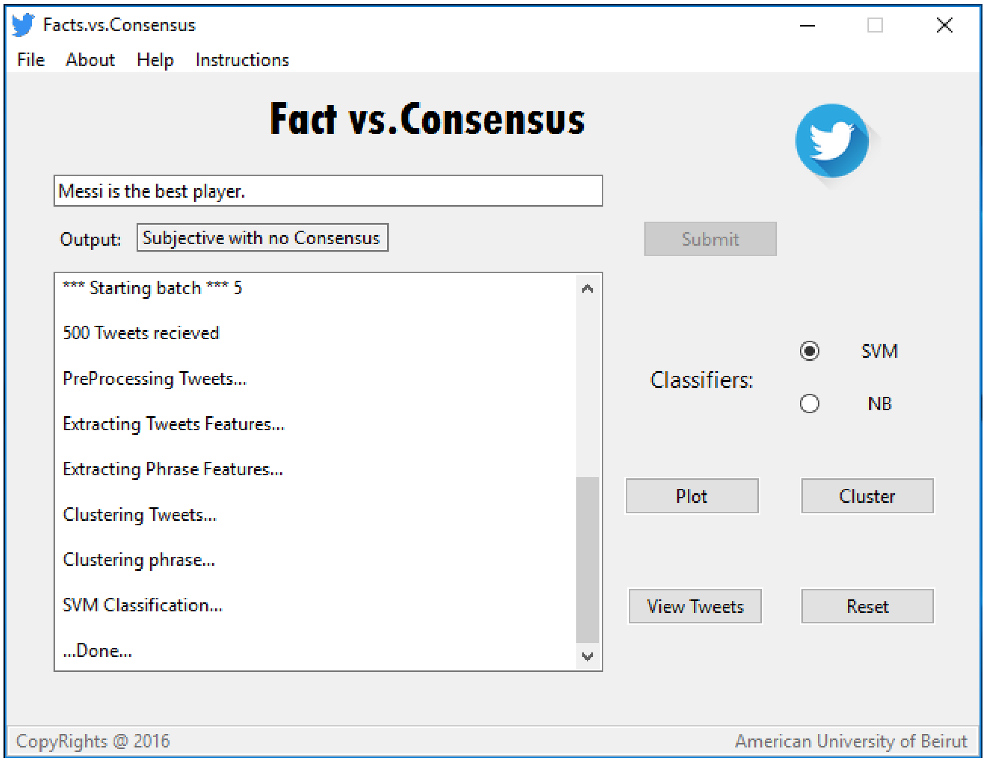


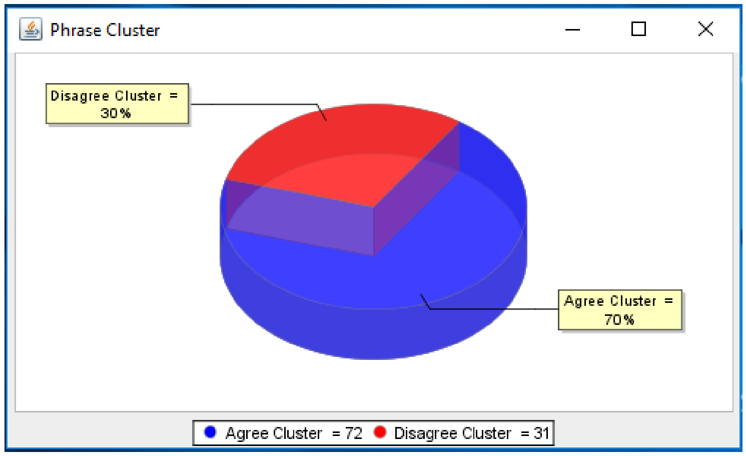
Figure.23(b): Methodology(II)



5.3 Clustering Pie Chart

Figure.24, shows a statistical analysis sample done on clustering method, in which it displays the percentage of splits between tweets as Agree or Disagree to the phrase entered by a user. The “Cluster button” in our tool is responsible for this task.

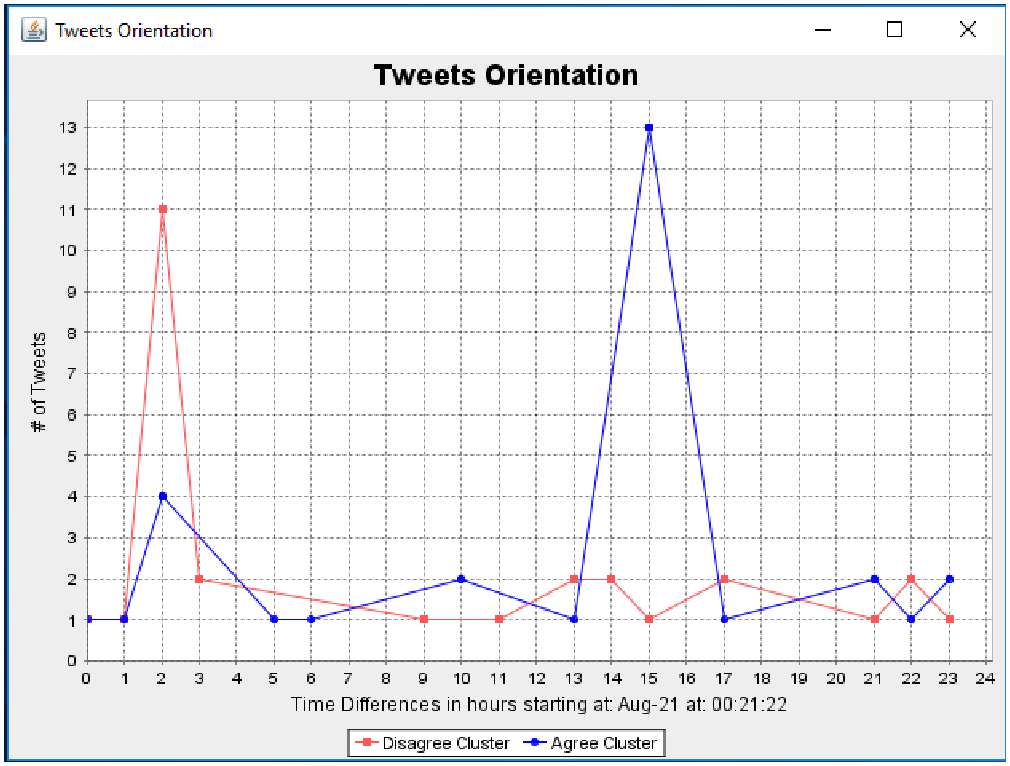
Figure.24: Clustering Pie Chart



In the above example, 30% of tweets are positioned in the disagree cluster, while 70% of tweets are positioned in the agree cluster. We color the section of the Pie Chart in which the phrase belongs too as blue indicating a positive relation between the input phrase and the clustered tweets. In the bottom of the above figure a user can also visualize the number of tweets in each cluster.

5.4 Graphical Plot

Figure.25: Tweets Orientation

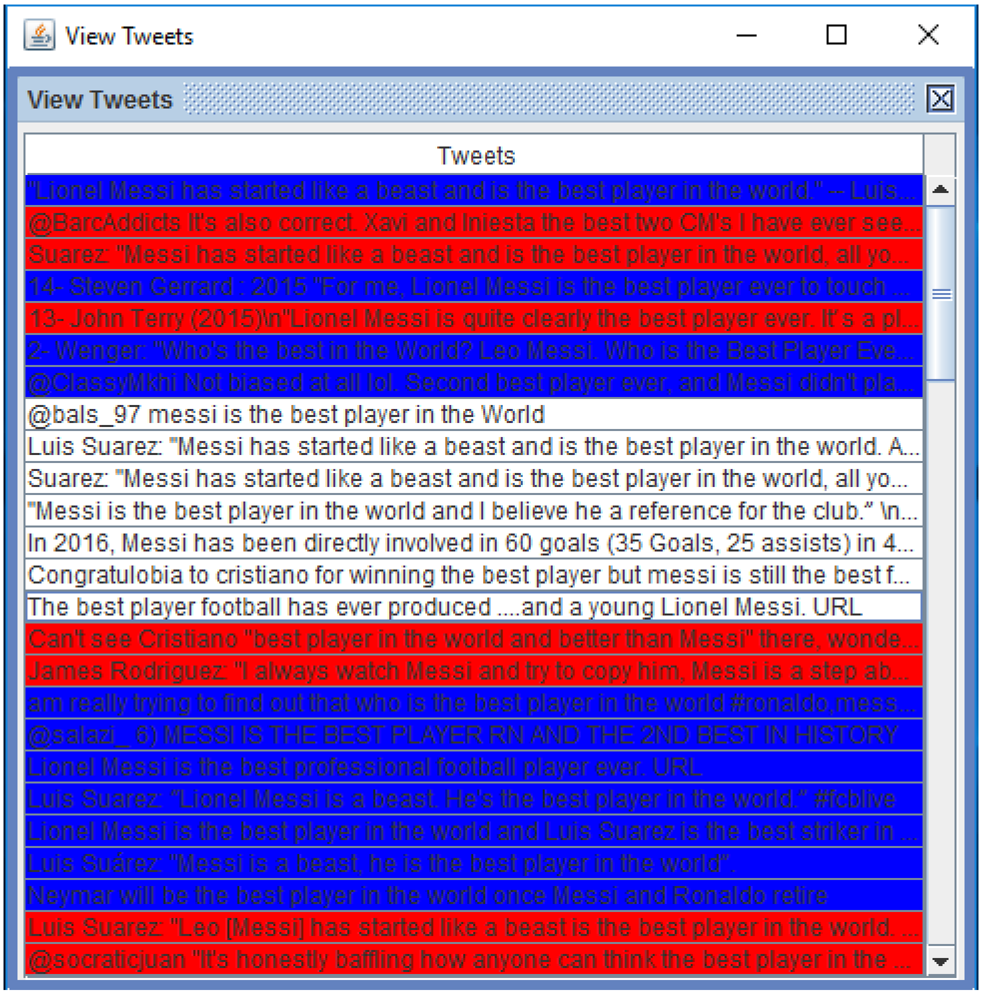


The Twitter Sport-Phrase framework supports graphical plots ("Plot button"), in which a user can visualize the time differences of tweets as a function of their number. Figure.25.

5.5 Tweets Visualization

The "View Tweets" option permits access to some of the retrieved tweets, as tweets are highlighted in red and blue for (Objectivity, Subjectivity) disagreement and agreement respectively. Blue tweets correspond to posts that belong to the same cluster of the phrase while Red ones belong to the other cluster. Figure.26 below. Additionally, users can select tweets to highlight them for better reading presentations.

Figure.26: Raw Tweets



All users could use the reset button to clear his/her current query with all its corresponding visualizations.

CHAPTER 6

CONCLUSION

We proposed in this work a new sentiment analysis phrase classification framework that classifies sport phrases as objective or subjective with/out consensus and extends analysis for tweet herding. Experimental results were conducted on different classifiers and showed that the combination of syntactic and semantic features on a handmade supervised corpus of 1007 phrase recorded the highest accuracies using the Neural network and SVM classifiers. The model was designed to target only English language. Phrase classification is based on the POS tagger for keyword extraction, additionally WordNet was used to extract synonyms and antonyms. The procedure handled the twitter Search API to retrieve tweets. Syntactic and semantic features were used after preprocessing tweets and phrases. Moreover, we introduced a decision tree that created new consensus labels for phrase subjectivity as tweets are clustered into two clusters that agree or disagree with phrases.

Experiments on different clustering algorithms showed that the CA algorithm outperformed other clustering procedures on supervised datasets and Libras unsupervised dataset. However, the Maximin algorithm proved to fit the proposed approach, especially on our proposed datasets particularly requiring 2 clusters.

Experimenting with Stanford dataset, the proposed semantic features recorded a highest accuracy of 72.7% than our syntactic features alone and a combination of the latter with the semantic features of accuracies 54.87% and 60.72% respectively. However, the nature of the provided tweets in the training and testing datasets affected our accuracies especially that most emoticons are ignored as they are part of our semantic features. Furthermore, our framework and set of features are specified for complete sports phrases.

A Sports Related Twitter News Data Analytics application was also presented that supports the phrase classification methodology. It provides some visualization features as clustering pie charts, graphical plots, and tweets visualizations in which a user could conduct some analysis on sport-phrases.

We proposed a novel application with a new set of features that achieved high objective, and subjective classification accuracies and defining a new consensus label for subjective expressions.

For future work, we intend to extend our work to sustain other languages. New labels could also be introduced for objectivity to distinguish factual and non-factual objective phrases. However, more experiments are to examine additional features.

APPENDIX 1

SOURCE CODE

```
package FeatureExtraction;
import edu.stanford.nlp.ling.HasWord;
import edu.stanford.nlp.ling.Sentence;
import edu.stanford.nlp.ling.TaggedWord;
import edu.stanford.nlp.parser.Parser;
import edu.stanford.nlp.tagger.maxent.MaxentTagger;
import java.io.FileNotFoundException;
import java.io.FileOutputStream;
import java.io.FileReader;
import java.io.IOException;
import java.io.OutputStreamWriter;
import java.util.List;
import java.util.logging.Level;
import java.util.logging.Logger;
public class POSTagger {
static OutputStreamWriter POS;

public static void main (String [] args) throws ClassNotFoundException,
SecurityException, FileNotFoundException, IOException {
Logger logger = Logger.getLogger(POSTagger.class.getName());
logger.log(Level.INFO, "Msg");
RedwoodConfiguration.empty(). capture(System.err). apply ();
tagFile ();}
public static void tagFile () throws ClassNotFoundException, IOException {
MaxentTagger tagger = new MaxentTagger ("english-bidirectional-distsim.tagger ");
try {POS = new OutputStreamWriter (new FileOutputStream ("……"), "UTF-8");
List<List<HasWord>> sentences = MaxentTagger.tokenizeText(new
FileReader("words.txt"));
for (List<HasWord> sentence: sentences)
{List<TaggedWord> tSentence = tagger.tagSentence(sentence);
POS.append(Sentence.listToString(tSentence, false) +"\n");}
POS.close();}
catch (IOException ex) {
Logger.getLogger(Parser.class.getName()).log(Level.SEVERE, null, ex);}}}
```

```
import java.io.File;
import java.util.ArrayList;
import java.util.Arrays;
import java.util.HashSet;
import java.util.List;
import edu.smu.tspell.wordnet.Synset;
import edu.smu.tspell.wordnet.SynsetType;
import edu.smu.tspell.wordnet.WordNetDatabase;
import edu.smu.tspell.wordnet.WordSense;
public class TestJAWS {
private static WordNetDatabase database;

public static void main (String [] args) {
File f=new File("dict");
System.setProperty("wordnet.database.dir", f.toString());
database = WordNetDatabase.getFileInstance();
String wordForm = "……";
Synset [] synsets = database.getSynsets(wordForm);//,SynsetType.ADJECTIVE);

if (synsets.length > 0){
ArrayList<String> al = new ArrayList<String> ();
HashSet<String> hs = new HashSet<String> ();
for (int i = 0; i < synsets.length; i++){
String [] wordForms = synsets[i]. getWordForms ();
for (int j = 0; j < wordForms.length; j++) {
al.add(wordForms[j]);}
hs.addAll(al);
al.clear();
al.addAll(hs);}
for (int j = 0; j < al.size(); j++) {
String [] spl=al.get(j). split (" ");
if(spl.length==1){
System.out.print(al.get(j)+",");}}}
else {
System.err.println("No synsets exist that contain the word form '" + wordForm + "'");}}}
```

```
import java.io.FileOutputStream;
import java.io.IOException;
import java.io.OutputStreamWriter;
import java.util.ArrayList;
import java.util.Map;
import twitter4j.Paging;
import twitter4j.Query;
import twitter4j.QueryResult;
import twitter4j.RateLimitStatus;
import twitter4j.Status;
import twitter4j.Twitter;
import twitter4j.TwitterException;
import twitter4j.TwitterFactory;
import twitter4j.conf. ConfigurationBuilder;
import twitter4j.json. DataObjectFactory;
public class SearchAPI {
static Twitter twitter;
private static OutputStreamWriter tweetContentFile;

public static void main (String [] args) throws IOException, TwitterException {
 tweetContentFile = new OutputStreamWriter (new FileOutputStream ("……", true),
"UTF-8");
ConfigurationBuilder cb = new ConfigurationBuilder ();
cb.setJSONStoreEnabled(true);
cb.setDebugEnabled(true)
. setOAuthConsumerKey ("……")
. setOAuthConsumerSecret ("……")
. setOAuthAccessToken ("……")
. setOAuthAccessTokenSecret ("……");
TwitterFactory tf = new TwitterFactory(cb.build());
twitter = tf.getInstance();
test ();}
public static void test () throws IOException, TwitterException {
Map<String, RateLimitStatus> rateLimitStatus = twitter.getRateLimitStatus("search");
RateLimitStatus searchTweetsRateLimit = rateLimitStatus.get("/search/tweets");
System.out.printf("You have %d calls remaining out of %d, Limit resets in %d
seconds\n",
searchTweetsRateLimit.getRemaining(),
searchTweetsRateLimit.getLimit(),
searchTweetsRateLimit.getSecondsUntilReset());
int p=1;
```

```
Paging page = new Paging (p);
Query query = new Query("Messi"). lang("en");
query.resultType("mixed");
int numberOfTweets = 500;
long lastID = Long.MAX_VALUE;
ArrayList<Status> tweets = new ArrayList<Status> ();
while (tweets.size () < numberOfTweets) {
if (numberOfTweets - tweets.size() > 100)
query.setCount(100);
else
query.setCount(numberOfTweets - tweets.size());
try {
QueryResult result = twitter.search(query);
if (result.getTweets().size() == 0) { break; }
tweets.addAll(result.getTweets());
for (Status t: tweets) {
String jsonStr = DataObjectFactory.getRawJSON(t);
if (jsonStr! =null)
tweetContentFile.append(jsonStr+"\n");
if(t.getId() < lastID) {
lastID = t.getId();                    }}}
catch (TwitterException te) {
System.out.println("Couldn't connect: " + te);}
query.setMaxId(lastID-1);      }
 tweetContentFile.close();}}
```

APPENDIX 2

CORPUS

Phrase	Label
Derrick is a basketball player.	1
Butler is a basketball player.	1
Zidane was a football player.	1
Willian plays with Chelsea.	1
Tottenham is a football team.	1
Bartra plays with Barcelona.	1
Pique plays with Barcelona.	1
Busquets plays with Barcelona.	1
Iniesta plays with Barcelona.	1
Arda Turan plays with Barcelona.	1
Lionel Messi plays with Barcelona.	1
Neymar plays with Barcelona.	1
LeBron James plays with Cavaliers.	1
Mo Williams plays with Cavaliers.	1
Kyrie Irving plays with Cavaliers.	1
Metta plays with Lakers.	1
Nick Young plays with Lakers.	1

Jordan Clarkson plays with Lakers.	1
D'Angelo Russell plays with Lakers.	1
Julius Randle plays with Lakers.	1
Keylor Navas plays with Real Madrid.	1
Danilo plays with Real Madrid.	1
Gareth Bale plays with Real Madrid.	1
Casemiro plays with Real Madrid.	1
Isco plays with Real Madrid.	1
Toni Kroos plays with Real Madrid.	1
James Rodriguez plays with Real Madrid.	1
Cristiano Ronaldo plays with Real Madrid.	1
LeBron James is an NBA player.	1
Mo Williams is an NBA player.	1
Kyrie Irving is an NBA player.	1
Metta is an NBA player.	1
Nick Young is an NBA player.	1
D'Angelo Russell is an NBA player.	1
Kobe Bryant was an NBA player.	1
Michael Jordan was a basketball player.	1
Magic Johnson was a basketball player.	1
Kobe Bryant was a basketball player.	1
LeBron James is a basketball player.	1

Larry Bird was a basketball player.	1
Wilt Chamberlain was a basketball player.	1
Shaquille O'Neal was a basketball player.	1
Kevin Durant is a basketball player.	1
Steve Nash was a basketball player.	1
Dwight Howard is a basketball player.	1
Dwyane Wade is a basketball player.	1
Harrison Barnes is a basketball player.	1
Dirk Nowitzki is a basketball player.	1
Tim Duncan is a basketball player.	1
Atlanta Hawks is a basketball team.	1
Atlanta Hawks is an NBA team.	1
Boston Celtics is a basketball team.	1
Boston Celtics is an NBA team.	1
Charlotte Bobcats is an NBA team.	1
Chicago Bulls is a basketball team.	1
Chicago Bulls is an NBA team.	1
Cleveland Cavaliers is a basketball team.	1
Cleveland Cavaliers is an NBA team.	1
Dallas Mavericks is a basketball team.	1
Dallas Mavericks is an NBA team.	1
Denver Nuggets is a basketball team.	1

Denver Nuggets is an NBA team.	1
Detroit Pistons is a basketball team.	1
Detroit Pistons is an NBA team.	1
Golden State Warriors is a basketball team.	1
Golden State Warriors is an NBA team.	1
Houston Rockets is a basketball team.	1
Houston Rockets is an NBA team.	1
Indiana Pacers is a basketball team.	1
Indiana Pacers is an NBA team.	1
LA Clippers is a basketball team.	1
LA Clippers is an NBA team.	1
LA Lakers is a basketball team.	1
LA Lakers is an NBA team.	1
Memphis Grizzlies is an NBA team.	1
Miami Heat is a basketball team.	1
Miami Heat is an NBA team.	1
Milwaukee Bucks is a basketball team.	1
Milwaukee Bucks is an NBA team.	1
Minnesota Timberwolves is an NBA team.	1
New Jersey Nets is an NBA team.	1
New York Knicks is a basketball team.	1
New York Knicks is an NBA team.	1

Oklahoma City Thunder is a basketball team.	1
Oklahoma City Thunder is an NBA team.	1
Orlando Magic is a basketball team.	1
Orlando Magic is an NBA team.	1
Phoenix Suns is a basketball team.	1
Phoenix Suns is an NBA team.	1
Portland Trail Blazers is an NBA team.	1
Sacramento Kings is a basketball team.	1
Sacramento Kings is an NBA team.	1
San Antonio Spurs is a basketball team.	1
San Antonio Spurs is an NBA team.	1
Toronto Raptors is a basketball team.	1
Toronto Raptors is an NBA team.	1
Utah Jazz is a basketball team.	1
Utah Jazz is an NBA team.	1
Washington Wizards is a basketball team.	1
Washington Wizards is an NBA team.	1
Lionel Messi is a football player.	1
Cristiano Ronaldo is a football player.	1
Xavi is a football player.	1
Ibrahimovic is a football player.	1
Neymar is a football player.	1

Mesut Ozil is a football player.	1
Luis Suarez is a football player.	1
Marco Reus is a football player.	1
Wayne Rooney is a football player.	1
Oscar is a football player.	1
Gareth Bale is a football player.	1
Toni Kroos is a football player.	1
Eden Hazard is a football player.	1
Atletico Madrid is a football team.	1
Barcelona is a football team.	1
Granada is a football team.	1
Real Madrid is a football team.	1
Valencia is a football team.	1
Arsenal is a football team.	1
Aston Villa is a football team.	1
Chelsea is a football team.	1
Everton is a football team.	1
Leicester City is a football team.	1
Liverpool is a football team.	1
Manchester City is a football team.	1
Manchester United is a football team.	1
Newcastle United is a football team.	1

Norwich City is a football team.	1
Southampton is a football team.	1
Stoke City is a football team.	1
Sunderland is a football team.	1
Tottenham Hotspur is a football team.	1
Watford is a football team.	1
West Bromwich is a football team.	1
West Ham United is a football team.	1
Juventus is a football team.	1
Inter Milan is a football team.	1
Napoli is a football team.	1
Roma is a football team.	1
Fiorentina is a football team.	1
Palermo is a football team.	1
Bayern Munich is a football team.	1
Dortmund is a football team.	1
Djokovic is a tennis player.	1
Andy Murray is a tennis player.	1
Roger Federer is a tennis player.	1
Stan Wawrinka is a tennis player.	1
Rafael Nadal is a tennis player.	1
Kei Nishikori is a tennis player.	1

Gael Monfils is a tennis player.	1
Serena williams is a tennis player.	1
simona halep is a tennis player.	1
Maria sharapova is a tennis player.	1
Spain won the world cup in 2010.	1
Henderson is the Captain of Liverpool.	1
Man United are the Red Devils.	1
Barcelona is a Spanish team.	1
Barcelona played against Real Madrid.	1
Germany won the world cup 2014.	1
John Terry is the captain of Chelsea.	1
Alex Ferguson is Scottish.	1
Beckham played with England.	1
Ozil plays with Arsenal.	1
Arsenal play in Premier League.	1
De Gea is the goalkeeper of Man United.	1
Man United is an English team.	1
Real Madrid is a Spanish team.	1
Brad Stevens is the coach of Boston Celtics.	1
Suarez plays with Uruguay.	1
Pique is a defender in Barcelona.	1
Iniesta is a midfielder in Barcelona.	1

Navas is a goalkeeper in Real Madrid.	1
Premier League is a football League.	1
Champions League is a football League.	1
World cup 2014 was in Brazil.	1
Juventus is an Italian club.	1
Ronaldo is a Portuguese player.	1
Nadal is a Spanish player.	1
Guardiola is a Spanish coach.	1
Buffon is the goalkeeper of Juventus.	1
Pato plays in Chelsea.	1
Terry plays in Chelsea.	1
Neymar is a Brazilian player.	1
Chelsea play in Champions League.	1
Sturridge plays with Liverpool.	1
NBA is a basketball association.	1
FIFA is a football association.	1
Liverpool played against Arsenal.	1
Cech is a goalkeeper.	1
Welbeck plays with Arsenal.	1
Walcott plays in Arsenal.	1
Hazard is a midfielder in Chelsea.	1
Alaba is a defender in Bayern.	1

Paul plays with Clippers.	1
Messi is 28 years old.	1
Ronaldo is 31 years old.	1
Rivers plays with Clippers.	1
Neymar is 24 years old.	1
Lakers have been eliminated.	1
Champions League final on May 28	1
Bournemouth against Man City	1
Norwich versus Newcastle	1
Stoke versus Swansea	1
Sky Bet League is a football league	1
Charlotte against Atlanta	1
Toronto versus Detroit 2016	1
Warriors have the best record in the NBA	1
Atletico vs Barcelona 2-0	1
Game 4 thunder vs Mavericks	1
Byron Scott lakers head coach	1
Hawks vs Celtics	1
Pacers and Raptors May 1	1
Grizzlies lost to Spurs	1
Serie A is an Italy League	1
Chargers vs Chiefs	1

Vikings in NFL	1
Vikings against Titans	1
Red Sox 7 Astros 5	1
Mets is a baseball team	1
Braves is a baseball team	1
Phillies 10 Brewers 6	1
Reds lost to Cubs	1
Tigers lost to Indians	1
Blue Jays is a a baseball team	1
Royals is a baseball team	1
Orioles is a baseball team	1
Padres against Cardinals	1
Cavaliers beat Pistons	1
Messi is an attacker.	1
Benzema is an attacker.	1
Ronaldo scores against Barcelona	1
Football has a referee	1
Referee gives red cards in football	1
March Madness basketball	1
Real Madrid against Manchester City	1
Manchester United against Barcelona	1
Coventry City is a football team	1

Real Madrid in semi finals	1
Barcelona won against Real	1
Olivier plays with Arsenal	1
Arsenal lost to Barcelona	1
Roma is a football team	1
Welbeck plays football	1
Santi plays with Arsenal	1
Hector plays with Arsenal	1
Ramos plays with Spain	1
Ramos is a Spanish player	1
Astros is a baseball team	1
David Ortiz plays in Red Sox	1
David Ortiz is baseball player	1
3 assists against Pistons	1
Noah plays with the Bulls	1
Curry out at least 2 weeks with sprained right knee	1
Zinedine Zidane, Real Madrid coach	1
NYY is a baseball team	1
NYY are called the Yankees	1
White Sox win on odd out	1
ESPN: Entertainment and Sports Programming Network	1
Starlin plays with Yankees	1

Andrew Miller is a Pitcher	1
Alex is a hitter in Yankees	1
Girardi is the manager of Yankees	1
UEFA, The Union of European Football Associations	1
Draw between Arsenal and Sunderland	1
The UEFA Champions League final is in Milan	1
Isaiah plays with Celtics	1
Damian plays with Blazers	1
Final Score: Warriors 110 Cavs 77	1
Curry, the MVP.	1
Curry, his second MVP.	1
Game 1: Warriors 104 vs. Cavs 89	1
Warriors won on Thunder 96-88	1
Game 3: Warriors vs. Cavaliers	1
Real Madrid won the Champions League	1
Atletico lost to Real Madrid	1
Alves to leave Barcelona	1
Zidane has become the seventh man to win the European Cup as a player and a coach	1
Warriors 2-0 lead against the Cavs	1
Muhammad Ali is dead at 74	1
Love leaves Game 2 after blow to back of head	1
Draymond Green had 28 points	1

Jerry West defends LeBron	1
Barnes hit Love in the back of the head	1
LeBron James had 19 points	1
Durant out of the playoffs	1
Once again the cup is in Madrid	1
Cristiano Ronaldo takes scoring prize again	1
Warriors lead series 2-0	1
Two games into the new series	1
Warriors go up 2-0 on the Cavs.	1
Game 2 for the Warriors	1
Stephen Curry will not play in the 2016 Summer Olympics	1
Warriors to the finals	1
Ronaldo scoring the winning penalty	1
Jake Arrieta lost on Sunday	1
Love was hit by Barnes elbow.	1
In Game 1, the Warriors won without Stephen Curry	1
Jeff has authority to choose Knicks' assistant coaches	1
Kevin Love has been placed in the NBA Concussion Protocol.	1
Sharks Win their first Stanley Cup Game	1
Canadiens hire Kirk Muller as associate coach	1
Fan throws beer bottle at Ryan Howard in Philadelphia	1
Giants outfielder Hunter Pence will undergo surgery	1

Warriors rout Cleveland again	1
Game 2 of the NBA Finals, a 110-77 Warriors win	1
Real Madrid 1-1 Atletico Madrid	1
Real Madrid president Florentino Perez	1
5-3 on penalties, Real Madrid vs. Atletico.	1
Cristiano Ronaldo finished Champions League top scorer for a fourth season running	1
Mexico 3 Uruguay 1 in Copa America	1
Vardy to Arsenal or Leicester	1
NBA Finals: Golden State Warriors crush Cleveland Cavaliers to double lead	1
Murray lost to Djokovic on Sunday	1
Novak Djokovic beats Andy Murray to win first French Open title	1
Garbine defeated Serena in the French Open	1
Brazil will host Olympics 2016	1
Two more wins for the Warriors to become the champs	1
Curry pulls out of 2016 Olympics	1
Knick's new coach, Hornacek.	1
Juanfran had missed the penalty	1
Bayern Munich are better than Barcelona.	0
Djokovic is the best player.	0
Petr Cech is the best goalkeeper.	0
Chelsea are better than Manchester City.	0
basketball is better than football.	0

football is a boring game.	0
Basketball is a boring game.	0
Lakers are the best team.	0
Soccer is the most popular sport in the world.	0
Juventus are better than Chelsea.	0
Champions League is the best League.	0
Roger Federer is a champion.	0
Benzema is the best player.	0
Thierry Henry is a legend.	0
Buffon is the best goalkeeper.	0
Benzema is a good player.	0
Xavi is better than Iniesta.	0
Gerrard is the best player.	0
Football is the best game.	0
Neymar is better than Messi.	0
Neymar is better than Ronaldo.	0
Casillas is better than Buffon.	0
Liverpool are stronger than Chelsea.	0
Manchester City are better than Manchester United.	0
Basketball is the best game.	0
Messi scores best goals.	0
Ray Allen is the best 3 point shooter.	0

Tennis is a boring game.	0
World cup is boring.	0
football is more exciting than basketball.	0
Guardiola is better than Mourinho.	0
Guardiola is the best coach.	0
Manuel Neuer is the best goalkeeper.	0
Kobe Bryant is better than LeBron James.	0
World cup is exciting.	0
Mourinho is the best coach.	0
Luis Suarez should be the best player.	0
Messi is faster than Ronaldo.	0
Ronaldo plays better than Messi.	0
Real Madrid are the best team.	0
Barcelona are the best team.	0
Messi is better than Ronaldo.	0
Spanish league is the worst league.	0
Premier League is the best league.	0
Scottish football is so bad.	0
Djokovic is the best player.	0
Bayern Munich are better than Real Madrid.	0
Sacramento Kings is the best team.	0
Michael Jordan is the best player.	0

Michael Jordan is the best defender.	0
Michael Jordan is the worst player.	0
Magic Johnson is the best player.	0
Kobe Bryant is the best player.	0
Kobe Bryant is the best scorer.	0
LeBron James is the best player.	0
Larry Bird is the best player.	0
Wilt Chamberlain is the best player.	0
Kevin Durant is the best player.	0
Kevin Durant is the best scorer.	0
Chris Paul is the best player.	0
Chris Paul is the best defender.	0
Steve Nash is the best player.	0
Vince Carter is the best dunker.	0
Andre Drummond is the best player.	0
Pau Gasol is the best player.	0
Bradley Beal is the best player.	0
Dwight Howard is the best player.	0
Dwight Howard is the worst player.	0
Dwyane Wade is the best player.	0
Harrison Barnes is the best player.	0
Harrison Barnes is the worst player.	0

Chandler Parsons is the best player.	0
Dirk Nowitzki is the best player.	0
Tim Duncan is the best player.	0
Ryan Anderson is the best player.	0
Jamal Crawford is the best player.	0
Lionel Messi is the best player.	0
Cristiano Ronaldo is the best scorer.	0
Cristiano Ronaldo is the best player.	0
Xavi is the best player.	0
Andres Iniesta is the best player.	0
Ibrahimovic is the best player.	0
Neymar is the best player.	0
Mesut Ozil is the best player.	0
David Silva is the best player.	0
Luis Suarez is the best player.	0
Sergio Ramos is the best player.	0
Willian is the best player.	0
Willian is the worst player.	0
Wayne Rooney is the best player.	0
Juan Mata is the best player.	0
Thomas Muller is the best player.	0
Oscar is the best player.	0

Oscar is the worst player.	0
Fernandinho is the best player.	0
Gareth Bale is the best player.	0
Dani Alves is the best player.	0
Petr Cech is the best player.	0
Angel Di Maria is the best player.	0
Suarez is better than Messi	0
Manchester City are the best team.	0
Manchester City will be the Champions	0
Real Madrid are better than Man City	0
Bayern will be in UEFA champions League	0
Ronaldo scored best goals	0
Sanchez could be traded	0
City players are better than Real players	0
You are motivated by how much you want to win	0
Manchester City did not play well	0
Cavaliers weren't dominant	0
Lakers dominated Pistons	0
The crowd have been great this week	0
The audience did not cheer well	0
Ronaldo have more fans than Messi	0
Without Ronaldo Real have come for a Draw	0

Bale did not pass the ball	0
Zidane alone can hardly be blamed	0
We have not had a really good European game	0
Bale must have took a penalty	0
Yesterday game was a mess	0
Real should win the champions league	0
The Pistons have become very difficult to beat	0
Cavaliers didn't take care of the ball	0
Clippers should be more aggressive	0
PSG are not a great team	0
Blazers are not able to keep up	0
No good Clippers	0
Defense will have to be similar	0
Raptors are better than Pacers	0
Clippers will handle Blazers	0
Blazers will beat Clippers	0
Predictions say that Warriors will win	0
That pitch was a disgrace	0
All the injuries that he has gotten have been legitimate injuries	0
The best athletes were with Nike	0
One win isn't good enough	0
Djokovic and Murray are much better athletes	0

The game has gotten a lot better	0
Celtics had a chance to win	0
It is not enough to play for Liverpool	0
Madrid players still believe they can win the Champions League	0
We expect the defending champs to come out	0
Baseball is in the blood	0
Bayern are a strong team	0
Atletico are better than Bayern	0
Bayern did not play well	0
Atletico played better than Bayern	0
Today's match was not excited	0
Game 2 is better than Game 1	0
Bayern players are tough	0
Atletico should push further	0
The winner was very good today	0
Benzema is a good scorer	0
Pacers could play better	0
Walton is a good Head Coach of the Lakers	0
Celtics did not perform well	0
Bulls can perform better in their games	0
Lebron is the most talented player	0
Cavaliers should not leave Lebron	0

Thunder play better at home	0
Lebron can take over the court	0
Lebron can secure his ball	0
Lebron plays good in all positions	0
Warriors are hurt by injuries	0
Clippers are a weak team	0
Spurs are an aggressive team	0
Things are looking up for Warriors again	0
Luke Walton seems like the right choice	0
Luke Walton is smart	0
Lakers need a lot of help	0
Arsene Wenger still has work to do	0
Thunder looking solid	0
It is a disappointing season for Chelsea	0
Man City need not fear Chelsea	0
Hart is the hero	0
Man City not nearly as good as Madrid	0
Ronaldo and Benzema should play	0
Aguero can't be rested	0
City should rest everyone over the weekend	0
Atletico should not win UEFA	0
Atletico need to train well	0

Bayern have a good coach	0
Spanish teams are dominating	0
Last game was an exciting game	0
Liverpool can achieve something with Klopp	0
Ronaldo is not fit for UCL	0
FA cup still has magic	0
Arsenal did not improve their performance	0
Leicester will slip up	0
Problems in Barcelona	0
Newcastle will not survive	0
More is expected from Arsenal	0
Tyronn is the best coach	0
Hornets should be playing better	0
Hornets are the worst team	0
The Wolves have reached the end	0
Rockets are better than Pacers	0
Rockets should push further	0
Bulls could play better	0
Stanley Johnson is looking better and better	0
A good season for Cavs	0
Game 1 might have seemed like an outlier	0
Walton should consider Kings	0

Pistons have the best shooters	0
Cavs have the best guards	0
Celtics not good enough to beat Hawks	0
Cavs reached their peak	0
Isaiah Thomas is the best player in Celtics	0
This is a bad season for Lakers	0
Atletico are the team of the week	0
Celtics should target Lebron	0
The Celtics were so good this season	0
Warriors are nothing without Steph Curry	0
Bayern are not playing as a team	0
Thompson started out the game on fire	0
The Warriors can win without Curry but Green is really their MVP.	0
Kevin Durant led the way	0
The Thunder couldn't have won without Dion Waiters	0
Kevin Durant had different plans	0
Lebron was great in the series	0
Curry went down against Houston in these playoffs	0
Conley is much more likely than Gasol	0
Stephen Strasburg has actually been good for a while	0
Newcastle had too much anxiety	0
Kane and Vardy fit in for England	0

There's no more pressure on Leicester.	0
Manchester United still have something to fight for	0
Walcott is almost unstoppable	0
Leicester will be fearless in Champions League	0
Guardiola is a genius	0
Leicester done the impossible.	0
Chelsea had nothing to play for.	0
Ronaldo still leads Messi	0
Messi is only getting better with age	0
Cristiano Ronaldo maintains scoring	0
Real president Florentino Perez	0
Final has no favourite	0
Real Madrid were lucky	0
Henderson to be fit for Euro 2016	0
Real dominated for much of the game	0
Atletico can't defeat Real Madrid	0
Curry does not deserve the MVP award.	0
Curry is the best dribbler	0
City must win for fans	0
Thompson Leads the Warriors	0
Curry is a valuable player	0
Curry didn't just do it with his shooting. He did it with passing too	0

Lebron should take the MVP award, and not Curry	0
Curry is the best shooter in NBA	0
Curry is the most exciting player in the NBA.	0
Dwyane Wade takes control	0
Curry led Warriors to a win	0
Curry is the best player in Warriors	0
Steph Curry is still trending	0
The Warriors could not have survived this series without Curry	0
Curry should be in the starting lineup.	0
Curry looks fully healthy	0
This is still a horrible series	0
Raptors are horrible	0
Warriors have tough games	0
Curry is the best shooter of all time.	0
Lillard and McCollum did their best	0
Curry had to match Lillard	0
Russell played unbelievable	0
Curry is changing the game	0
There should have been a foul for Spurs	0
Tonight was an emotional game	0
Durant doesn't know what he will do	0
Durant should be a free agent this summer	0

Durant should stay with Thunder	0
Warriors must take Durant for the next season.	0
Warriors would be dead without Shaun Livingston.	0
Curry is a dream player	0
Curry changed the game	0
Azarenka is not good right now	0
Atletico could not catch up on Real	0
Warriors does not deserve to win	0
Without Curry, Warriors could not win	0
If Curry misses the game, Warriors will lose	0
Real played better with Mourinho	0
Spurs must make major changes.	0
Spurs should push further.	0
Thunder should be more defensive.	0
Thunder must take advantage of their win.	0
Thunder must be more than just Big 3 to win	0
West NBA teams are better than the East teams.	0
Warriors will take the championship.	0
Spurs can not play without David West	0
Barcelona would be better without Alves	0
Warriors in control	0
Cavs lose, Lebron is blamed	0

Warriors are safe with Green's lead.	0
With Green under control, Warriors are in control	0
Lebron James got to be better	0
Cavs need to block Curry	0
Atletico played better than Real	0
Atletico deserve to win	0
Cavs gave up at the end	0
Cavs don't have a chance to win	0
Lebron did nothing in game 2	0
Cristiano Ronaldo even more vital to Portugal than to Real Madrid	0
Irving should be quick in attacks	0
Wolves plan to target Joakim Noah	0
Diamondbacks are not ready To shift focus to 2017	0
Cubs Wouldn't Trade Kyle Schwarber For Andrew Miller	0
Steven Stamkos should stay with Lightning	0
Game 1 was a Bob Myers' win	0
The Heat should look to sign Noah.	0
Cavs must secure their home court	0
LeBron isn't good enough to beat these Warriors	0
The Warriors are too aggressive on defense	0
Warriors don't need their best to beat Cavaliers	0
Portugal has one of the best players in the world	0

Ronaldo can lead the success of Portugal to Euro 2016	0
Ronaldo is a very special player	0
Ronaldo's absence is a useful test for Portugal	0
Warriors are close to take the lead	0
Warriors are close to be the Champs	0
Warriors played better in the second quarter	0
Warriors are close to the title	0
Cavs should secure their court	0
Hornacek must do a change in Knicks.	0
Size does not matter in basketball	0
Warriors to the Finals	1
Cavs to the finals	1
Cavs took the lead 120-90	1
120-90 for the Cavs	1
Warriors lost to Cavs 120-90	1
Cavs whip Warriors in Game 3	1
James had 32 points in game 3	1
Harrison Barnes scored 18	1
Klay Thompson left in the first quarter	1
Warriors vs. Cavaliers: Game 3	1
Cavs pulled a 2-1 in the NBA Finals.	1
Kevin Love gets a concussion	1

Curry has 48 points in these three games	1
Former Bayer Leverkusen coach dies aged 44	1
Ronaldo now world's best paid athlete	1
Kyrie Irving scored 16 of his 30 in the first quarter	1
Barcelona should be in the UEFA finals	0
Warriors will win the championship	0
Cavs have been dominant all postseason	0
Cavs finally got back to their game	0
A good follow for Cavs	0
The Warriors didn't look a team that won a record of 73 games	0
Warriors did not play well in game 3	0
Cavs deserve to win in game 3	0
Warriors out of control	0
The Cavs hardly missed Love	0
The Cavs had to get Game 3	0
Kerr became emotional before the game	0
Curry didn't help his team win.	0
Curry is the face of NBA	0
Rumours: Vermaelen to Liverpool	0
Game 3: The Wrap	0
Warriors out of step again on road	0
Warriors weren't ready to play.	0

It didn't go right tonight for the Warriors	0
The playoffs are hard	0
Warriors weren't tough enough	0
cavs can't be soft in Game 4 if they want to win.	0
Warriors struggle on road again	0
Muhammad Ali gone but not forgotten by league.	0
Cavaliers expect more from Irving.	0
Warriors 108 Cavs 97 in game 4	1
Draymond Green is suspended for Game 5 of the NBA Finals	1
Warriors miss green in game 5	1
Warriors are one win away from title	1
Warriors' Green suspended for Game 5	1
Golden State leads the series 3-1	1
Green made a flagrant foul	1
Technical fouls for Lebron and Green	1
Russia played England in a Euro 2016 match	1
England 1-1 Euro 2016 draw with Russia	1
Euro 2016: Russia given disqualification warning	1
England and Russia have been warned by Uefa	1
Giaccherini scores for Italy at Euro 2016	1
Pjanic joins Juventus	1
Portugal vs Iceland, Euro 2016	1

UEFA Executive Committee issues warning after violence in Marseille	1
Penguins finish off Sharks to win Stanley Cup	1
Draymond Green was right to punch LeBron James	0
Green's suspension to bring out best in Warriors	0
Cavs will win game 5 with the absence of Green	0
The Cavs don't trust each other	0
Warriors are week without Green	0
Warriors seem better than Cavs	0
Russian fans are blamed for Euro violence	0
Russia should be charged for Euro 2016 violence	0
Spain made an unconvincing start in Euro 2016	0
England and Russia could be thrown out of Euro 2016	0
Green deserves to be suspended on game 5	0
Lebron is also blamed, not only Green	0
Cavs should at least prove themselves	0
Cavs will win game 5	0
Lebron alone can't win the game.	0
Cavaliers 112 vs. Warriors 97.	1
Warriors 3-2 lead over Cavs	1
Cavs won game 5	1
Warriors lead series 3-2	1
Kyrie Irving and LeBron James each scored 41 points	1

Game 5 of the NBA Finals at Oracle Arena in Oakland.	1
Warriors need one more win to be the champs	1
Two more wins for the Cavs for championship	1
The Cavs have another game at home	1
Game 6 is the last opportunity for the Cavs	1
The Cavaliers force the series back to Cleveland	1
The first time in Finals history that two teammates scored 40 or more points in the same game.	1
James had 41 points, 16 rebounds and seven assists	1
Cleveland Cavaliers beat the Golden State Warriors	1
NBA finals Game 5	1
Cavs force a Game 6 in the NBA finals	1
Italy 2 Belgium 0	1
A 0-0 game between Portugal and Iceland	1
Graziano Pelle scored vs Belgium	1
Italy began their Euro 2016 campaign with victory	1
Hungary won on Austria 2-0	1
Cavs played well in game 5.	0
Comeback for the Cavs.	0
Cavs proved themselves in game 5.	0
Cavs exploit Green's absence to stay alive in Game 5	0
Without Green, Cavs took their advantage	0
Defense helped Cavs Win Game 5	0

Lebron's dunk was awesome	0
The Warriors aren't the team facing elimination	0
Cavs should now take their home advantage	0
Curry did not play well in game 5.	0
Warriors missed their Golden opportunity	0
Warriors must defend their title	0
Cavs were aggressive in game 6	0
Cavs did a great job	0
Irving is playing as expected	0
Green's absence made a difference	0
Cavs will be the Champs	0
Warriors have a better chance than Cavs for the title.	0
Game 5: LeBron, Kyrie keep Cavs alive	0
Premier League starts on August 13.	1
Arsenal vs. Liverpool in Premier League 2016	1
Mourinho coaching Manchester United	1
Wade signs for Birmingham Bears	1
Fly-half Ford is to start for England	1
Andy Murray beats Nicolas Mahut in first round	1
Jordan Spieth to play at Rio Olympics	1
DeMar DeRozan Officially Opts Out Of Contract With Raptors	1
Grizzlies guard Adams getting cartilage transplant	1

Bogut out for rest of Finals with left knee injury	1
Game 6 is Thursday in Cleveland	1
A dunk and a foul.	1
Warriors' centers can't contain Cavs.	0
Without Green in Game 5, the Warriors had to play another way	0
Cavaliers win NBA championship	1
Cavs win first NBA title	1
Iguodala defends James	1
Cavs defeated Warriors 93-89	1
James finished with a triple double	1
James named Finals MVP	1
4-3 lead for Cavs	1
Cavaliers win series 4-3	1
LeBron James wins NBA Finals MVP for 3rd time	1
Cavaliers vs. Warriors: Game 7	1
Stephen Curry fouls out in game 6	1
Cavaliers celebrate title	1
The Cavaliers' win in Game 7 of The 2016 Finals	1
Cavaliers on series win	1
Warriors on series loss	1
Lebron James said: "I came back for a reason".	1
James win his third title and third Finals MVP	1

James plays for Cleveland	1
"I didn't do enough to help my team win," Curry said	1
The Golden State Warriors didn't repeat as champions	1
France 0 Switzerland 0	1
Albania won on Romania 1-0	1
Final Score: Wales 3 Russia 0	1
LeBron James had the game of his life.	0
Curry and LeBron legacies on the line	0
NBA is rigged	0
NBA finals are rigged	0
NBA is fixed	0
Curry didn't play efficient	0
Cavs played a great game in their finals	0
Warriors' MVP was useless	0
A historic season for the Cavs	0
Steve Kerr can take plenty of blame for Warriors' Game 7 loss	0
LeBron James inspired Cavaliers	0
Cavs made history	0
Warriors still to be the greatest team.	0
The Warriors' core is set	0
Warriors are disappointed	0
Warriors should have played better in game 7.	0

The playoffs proved to be a different beast for the Warriors	0
The win of Cavs was expected.	0
Cavs deserve to win.	0
Warriors did not give their best effort.	0
Lebron led the Cavs to a game 7 win	0
GSW still could have won	0
Blanc should leave PSG	0
Euro 2016: Wales and England through to knockout stages	1
Durant to play for team USA in Rio Olympics	1
Rockies 5 Marlins 3	1
Cardinals beat Cubs 3-2	1
Dodgers won on Nationals 4-1	1
NCAA baseball championship in Omaha	1
Orioles lose 4-3 in 1-game trip to Texas for makeup game	1
Hodgson: "We fear no one".	1
Higuain to stay at Napoli for the next five years	1
Kyle Edmund beats Lukas Rosol to reach second round	1
Murray to win the Queen's title	1
Zlatan to retire from international football after Euro 2016	1
Rio Olympics in Brazil	1
Durant on Free Agency	1
Versatile Layman to face NBA challenge	1

LeBron, Cavaliers bring trophy home to Cleveland	1
Davone Bess was arrested	1
Hingis and Mirza Qualify for Singapore	1
Mayer wins Halle title	1
Murray wins Queen's again	1
Rosberg wins	1
Johnson to win the US Open	1
Vincenzo Abbagnale banned for missing doping tests	1
Edmund loses to Dolgopolov in Nottingham	1
Leicester sign the defender Hernandez	1
Hodgson left Rooney on the bench	1
Stade Pierre Mauroy pitch to be replaced	1
Wales to round 16	1
England to round 16	1
Wayne Rooney is a striker	1
Rooney is 30 years old	1
Formula 1 2016	1
Ukraine 0 Poland 1	1
Germany beat Ireland 1-0	1
Ozil plays in Germany	1
Northern Ireland finish third in group C	1
Germany finished first in its group	1

Wales have 6 points	1
Albania scores its first ever EURO goal	1
Switzerland against Poland in round 16	1
Anderson wins Auckland title	1
Adrian was the defending darts champion	1
A draw between England and Slovakia	1
The Coach of Genzebe Dibaba is arrested in Spain	1
Panthers extend Rivera	1
Tyronn Lue is the youngest head coach	1
Steve Kerr, Warriors coach	1
Final score: Germany 1 Northern Ireland 0	1
NBA game 7	1
Minor League Baseball hires first female umpire since 2007	1
24 teams in the group stages at Euro 2016	1
Slovakia finished third in its group	1
Rain delays women's ODI series at Grace Road	1
ODI series at Grace Road	1
Johnson said: "Winning US Open is awesome".	1
Joachim, Germany Euro coach	1
Lebron kissed the trophy	1
Ibra to quit internationals	1
Wimbledon 2016	1

Portugal against Hungary, Euro 2016.	1
England will be knocked out in Euro	0
Stunning collapse in NBA Finals	0
A magical season for Cavs	0
Madison Bumgarner was good for Giants, but Pirates were better	0
Pirates ballgirl doesn't need a glove	0
Lukaku could stay at Everton this summer	0
Wales make history	0
Warriors are disappointed	0
Euro 2016: Belgium still need to find the right formula	0
There isn't a dominant team in Euro.	0
Murray Makes History	0
Warriors won't be the same	0
Lebron was brilliant against the Warriors	0
Warriors should take Durant	0
Cavs should trade Love	0
The best NBA finals ever	0
NBA is Corrupt	0
The Warriors were chasing history	0
A Pirates ballgirl might have had the best defensive play all day	0
Portugal need to be better at getting the ball into the box	0
Wales show they are more than Bales	0

A perfect preparation for Murray	0
Ronaldo is really the star of real Madrid	0
Ronaldo needs more support from his teammates	0
Portugal need not to rely on Ronaldo	0
Ronaldo is a selfish player	0
Durant should play with Warriors	0
Spain's tactical system has changed	0
Spain's tactical system matured for Euro 2016.	0
A good start for Italy in this Euro	0
Italy will win this Euro	0
Spain can not face Italy in this Euro	0
England one of the best teams	0
The Mets are struggling to score	0
Jamie Vardy must remain with Leicester City	0
Cavs won because the NBA is rigged	0
Russia could have won against Wales	0
Luck was with Wales against Russia	0
Spain vs. Croatia is an exciting game	0
Germany could be better this Euro	0
Spain is better than England in Euro 2016.	0
The Stephen Curry revolution is over	0
Lebron did really a great job	0

Curry to be blamed for the loss	0
Spain used to be a good team	0
Rooney should not be left on the bench	0
Injuries could affect this Euro	0
Wales are dominating	0
Wales seem to be a great team	0
LeBron James definitely wants to come back to the Cavs	0
Lebron guided Cavs to the win	0
Lebron played the game of his life	0
Formula 1 races are bad	0
Northern Ireland are the best fans of the Euro.	0
Ozil is a talented player	0
Spain never showed us a good game	0
Italy is the strongest football team	0
England is good in Euro, but Wales are better	0
Spain must survive this Euro	0
Murray will benefit from Lendl	0
Tyronn Lue is the best coach	0
Tyronn Lue is a rookie coach	0
Steve Kerr is still better than Tyronn Lue	0
NBA game 7 was rigged	0
We're seeing a better control from Moore	0

Murray is a legend	0
Djokovic can not beat Murray	0
It was never going to be easy in Euro games	0
NBA game 7 was not an easy game	0
Wales will not make it to the finals	0
Predictions: Spain might win this Euro.	0
Germany is a tough team	0
Germany have weaknesses	0
The Warriors will win next year	0
Spain to round 16	1
Italy vs. Spain in round 16	1
Dodgers extend Julio Urias' stay by one more start	1
England 2-1 Wales	1
Konta and Murray among seeds at Wimbledon	1
Wimbledon boosts security staffing for championships	1
Kyle Edmund loses to Alexandr Dolgopolov in Nottingham	1
Lizzie Armitstead wins stage three	1
Duncan Taylor leaves Japan tour after hamstring injury	1
Mayer beats Zverev to win Gerry Weber Open	1
Lionel Messi scoring on a free kick against USA	1
Bartolo Colon was injured	1
Morata to rejoin Real Madrid	1

Cheryshev joins Villarreal	1
Ronaldo missed a penalty against Austria	1
Portugal 3 Hungary 3	1
Ronaldo scored two goals against Hungary.	1
Round 16 will be an exciting round	0
Italy will knockout Spain in round 16	0
Nolan Arenado might hit 50 home runs	0
Euro 2016: England will not scare anyone	0
James must resign with Cavs	0
Ronaldo rescued Portugal	0
Hungary is a weak team	0
Tough match, Portugal vs. Croatia	0
Ronaldo is the best player in Portugal	0
Portugal should depend on Ronaldo	0
Italy is not an easy team to beat	0
Portugal should be more defensive	0
Portugal have good attackers	0
Ronaldo did his job as a captain of Portugal	0
Ronaldo can do better in this Euro	0
England second place in group B	1
Ronaldo throws reporter's microphone into lake.	1
Portugal 1-1 Iceland	1

A 1-1 draw between Iceland and Hungary	1
England to face Iceland in last 16	1
Portugal finished third	1
Iceland finished second after beating Austria.	1
Ramos lost a Penalty against Croatia	1
Ramos's penalty was saved	1
Full Time: Croatia 2-1 Spain	1
FT: Sweden 0-0 Belgium	1
Spain lost to Croatia	1
Croatia have 7 points	1
Italy is doing a great job	0
Cristiano Ronaldo was in a bad mood	0
Ireland don't need to fear Italy	0
Portugal did not play well against Hungary	0
Hungary played better than Portugal	0
Portugal should have topped their group	0
Italy vs. Spain is a tough match	0
It was tough for Portugal this Euro	0
France to round 16 in Euro 2016	0
France will be knocked out in round 16	0
Wales are expected to win their next match	0
England have good players in this Euro	0

APPENDIX 3

CLASSIFICATION FAILURES

Misclassified Data (1007 phrases)	Classifier	Actual Label	Predicted Label
'Man United are the Red Devils.'	KNN	1	0
'Lakers have been eliminated.'		1	0
'Warriors have the best record in the NBA'		1	0
'Ramos is a Spanish player'		1	0
'NYY are called the Yankees'		1	0
'Draw between Arsenal and Sunderland'		1	0
'Love leaves Game 2 after blow to back of head'		1	0
'Barnes hit Love in the back of the head'		1	0
'Stephen Curry will not play in the 2016 Summer Olympics'		1	0
'In Game 1, the Warriors won without Stephen Curry'		1	0
'Cristiano Ronaldo finished Champions League top scorer for a fourth season running'		1	0
'basketball is better than football.'		0	1
'Thierry Henry is a legend.'		0	1
'Spanish league is the worst league.'		0	1
'No good Clippers'		0	1
'Baseball is in the blood'		0	1
'Problems in Barcelona'		0	1
'Atletico are the team of the week'		0	1
'Warriors are nothing without Steph Curry'		0	1
'Bayern are not playing as a team'		0	1
'Kevin Durant led the way'		0	1
'Chelsea had nothing to play for.'		0	1
'Curry is a dream player'		0	1
'Warriors in control'		0	1
'Wolves plan to target Joakim Noah'		0	1
'Kyrie Irving scored 16 of his 30 in the first quarter'		1	0
'Curry is the face of NBA'		0	1
'Rumours: Vermaelen to Liverpool'		0	1
'Warriors miss green in game 5'		1	0

'Cavs won game 5'		1	0
'The Cavaliers force the series back to Cleveland'		1	0
'Cavs force a Game 6 in the NBA finals'		1	0
'Without Green, Cavs took their advantage'		0	1
'Game 5: LeBron, Kyrie keep Cavs alive'		0	1
'Bogut out for rest of Finals with left knee injury'		1	0
'Cavs win first NBA title'		1	0
'Iguodala defends James'		1	0
'"I didn't do enough to help my team win," Curry said'		1	0
'Higuain to stay at Napoli for the next five years'		1	0
'Murray to win the Queen's title'		1	0
'Zlatan to retire from international football after Euro 2016'		1	0
'The Coach of Genzebe Dibaba is arrested in Spain'		1	0
'The Warriors were chasing history'		0	1
'The Stephen Curry revolution is over'		0	1
'Bartolo Colon was injured'		1	0
'Tough match, Portugal vs. Croatia'		0	1
'Italy vs. Spain is a tough match'		0	1
'France to round 16 in Euro 2016'		1	0
'Juventus is an Italian club.'	SVM	1	0
'Ronaldo is a Portuguese player.'		1	0
'Nadal is a Spanish player.'		1	0
'Neymar is a Brazilian player.'		1	0
'Lakers have been eliminated.'		1	0
'Warriors have the best record in the NBA'		1	0
'Cavaliers beat Pistons'		1	0
'Ramos is a Spanish player'		1	0
'Curry out at least 2 weeks with sprained right knee'		1	0
'NYY are called the Yankees'		1	0
'Love leaves Game 2 after blow to back of head'		1	0
'Barnes hit Love in the back of the head'		1	0
'Stephen Curry will not play in the 2016 Summer Olympics'		1	0
'Love was hit by Barnes elbow.'		1	0
'Thierry Henry is a legend.'		0	1
'Baseball is in the blood'		0	1
'Problems in Barcelona'		0	1
'Warriors are nothing without Steph Curry'		0	1
'Kevin Durant led the way'		0	1
'Kane and Vardy fit in for England'		0	1
'Cristiano Ronaldo maintains scoring'		0	1

'Curry is a dream player'		0	1
'Warriors in control'		0	1
'Wolves plan to target Joakim Noah'		0	1
'Kyrie Irving scored 16 of his 30 in the first quarter'		0	1
'Curry is the face of NBA'		1	0
'Rumours: Vermaelen to Liverpool'		0	1
'Warriors miss green in game 5'		1	0
'The Cavaliers force the series back to Cleveland'		1	0
'Game 5: LeBron, Kyrie keep Cavs alive'		0	1
'Bogut out for rest of Finals with left knee injury'		1	0
'Cavs win first NBA title'		1	0
'The Cavaliers' win in Game 7 of The 2016 Finals'		1	0
'Cavaliers on series win'		1	0
'"I didn't do enough to help my team win," Curry said'		1	0
'LeBron James inspired Cavaliers'		0	1
'Higuain to stay at Napoli for the next five years'		1	0
'Murray to win the Queen's title'		1	0
'Zlatan to retire from international football after Euro 2016'		1	0
'Murray wins Queen's again'		1	0
'Albania scores its first ever EURO goal'		1	0
'The Coach of Genzebe Dibaba is arrested in Spain'		1	0
'The Stephen Curry revolution is over'		0	1
'Wales are dominating'		0	1
'Lionel Messi scoring on a free kick against USA'		1	0
'Bartolo Colon was injured'		1	0
'Tough match, Portugal vs. Croatia'		0	1
'France to round 16 in Euro 2016'		1	0
'Lakers have been eliminated.'	NN	1	0
'Warriors have the best record in the NBA'		1	0
'Ramos is a Spanish player'		1	0
'Curry out at least 2 weeks with sprained right knee'		1	0
'NYY are called the Yankees'		1	0
'Love leaves Game 2 after blow to back of head'		1	0
'Barnes hit Love in the back of the head'		1	0
'Stephen Curry will not play in the 2016 Summer Olympics'		1	0
'Love was hit by Barnes elbow.'		1	0
'Thierry Henry is a legend.'		0	1
'We expect the defending champs to come out'		0	1
'Baseball is in the blood'		0	1
'Things are looking up for Warriors again'		0	1

'Problems in Barcelona'		0	1
'Atletico are the team of the week'		0	1
'Warriors are nothing without Steph Curry'		0	1
'Kane and Vardy fit in for England'		0	1
'Cristiano Ronaldo maintains scoring'		0	1
'Warriors in control'		0	1
'Wolves plan to target Joakim Noah'		0	1
'Kevin Love gets a concussion'		1	0
'Kyrie Irving scored 16 of his 30 in the first quarter'		1	0
'Curry is the face of NBA'		0	1
'Rumours: Vermaelen to Liverpool'		0	1
'Warriors miss green in game 5'		1	0
'Cavs won game 5'		1	0
'Game 5: LeBron, Kyrie keep Cavs alive'		0	1
'Cavs win first NBA title'		1	0
'Cavaliers on series win'		1	0
'"I didn't do enough to help my team win," Curry said'		1	0
'LeBron James inspired Cavaliers'		0	1
'Higuain to stay at Napoli for the next five years'		1	0
'Murray to win the Queen's title'		1	0
'Germany finished first in its group'		1	0
'Albania scores its first ever EURO goal'		1	0
'The Warriors were chasing history'		0	1
'The Stephen Curry revolution is over'		0	1
'Wales are dominating'		0	1
'Lionel Messi scoring on a free kick against USA'		1	0
'Bartolo Colon was injured'		1	0
'Ronaldo missed a penalty against Austria'		1	0
'Tough match, Portugal vs. Croatia'		0	1
'France to round 16 in Euro 2016'		1	0
'Lakers have been eliminated.'	NB	1	0
'Warriors have the best record in the NBA'		1	0
'NYY are called the Yankees'		1	0
'Curry, his second MVP.'		1	0
'Stephen Curry will not play in the 2016 Summer Olympics'		1	0
'Love was hit by Barnes elbow.'		1	0
'Novak Djokovic beats Andy Murray to win first French Open title'		1	0
'Basketball is a boring game.'		0	1
'Soccer is the most popular sport in the world.'		0	1

'Thierry Henry is a legend.'		0	1
'Sanchez could be traded'		0	1
'Clippers will handle Blazers'		0	1
'Baseball is in the blood'		0	1
'Things are looking up for Warriors again'		0	1
'Liverpool can achieve something with Klopp'		0	1
'Problems in Barcelona'		0	1
'More is expected from Arsenal'		0	1
'Atletico are the team of the week'		0	1
'Celtics should target Lebron'		0	1
'Warriors are nothing without Steph Curry'		0	1
'Kevin Durant led the way'		0	1
'Cristiano Ronaldo maintains scoring'		0	1
'Warriors have tough games'		0	1
'Curry is a dream player'		0	1
'Warriors in control'		0	1
'Cavs lose, Lebron is blamed'		0	1
'Wolves plan to target Joakim Noah'		0	1
'The Warriors are too aggressive on defense'		0	1
'Kyrie Irving scored 16 of his 30 in the first quarter'		1	0
'Curry is the face of NBA'		0	1
'Rumours: Vermaelen to Liverpool'		0	1
'Cavs won game 5'		1	0
'Game 5: LeBron, Kyrie keep Cavs alive'		0	1
'Cavs win first NBA title'		1	0
'LeBron James wins NBA Finals MVP for 3rd time'		1	0
'Cavaliers on series win'		1	0
'"I didn't do enough to help my team win," Curry said'		1	0
'Hodgson: "We fear no one".'		1	0
'Murray to win the Queen's title'		1	0
'Zlatan to retire from international football after Euro 2016'		1	0
'Stade Pierre Mauroy pitch to be replaced'		1	0
'Germany finished first in its group'		1	0
'Albania scores its first ever EURO goal'		1	0
'The Warriors were chasing history'		0	1
'Spain's tactical system has changed'		0	1
'The Stephen Curry revolution is over'		0	1
'Wales are dominating'		0	1
'Spain must survive this Euro'		0	1
'Bartolo Colon was injured'		1	0

'Tough match, Portugal vs. Croatia'		0	1
'France to round 16 in Euro 2016'		1	0
'Chicago Bulls is a basketball team.'	AB	1	0
'Milwaukee Bucks is a basketball team.'		1	0
'Toronto Raptors is a basketball team.'		1	0
'Hazard is a midfielder in Chelsea.'		1	0
'Paul plays with Clippers.'		1	0
'Warriors have the best record in the NBA'		1	0
'Padres against Cardinals'		1	0
'Noah plays with the Bulls'		1	0
'Barnes hit Love in the back of the head'		1	0
'Stephen Curry will not play in the 2016 Summer Olympics'		1	0
'Love was hit by Barnes elbow.'		1	0
'Novak Djokovic beats Andy Murray to win first French Open title'		1	0
'Thierry Henry is a legend.'		0	1
'Liverpool are stronger than Chelsea.'		0	1
'Baseball is in the blood'		0	1
'Things are looking up for Warriors again'		0	1
'Luke Walton seems like the right choice'		0	1
'Ronaldo is not fit for UCL'		0	1
'More is expected from Arsenal'		0	1
'Atletico are the team of the week'		0	1
'Warriors are nothing without Steph Curry'		0	1
'Kevin Durant led the way'		0	1
'Cristiano Ronaldo maintains scoring'		0	1
'Curry is a dream player'		0	1
'Warriors in control'		0	1
'Wolves plan to target Joakim Noah'		0	1
'Ronaldo is a very special player'		0	1
'Curry is the face of NBA'		0	1
'Rumours: Vermaelen to Liverpool'		0	1
'Penguins finish off Sharks to win Stanley Cup'		1	0
'Lebron is also blamed, not only Green'		0	1
'Cavs won game 5'		1	0
'Game 5: LeBron, Kyrie keep Cavs alive'		0	1
'Bogut out for rest of Finals with left knee injury'		1	0
'Cavs win first NBA title'		1	0
'The Cavaliers' win in Game 7 of The 2016 Finals'		1	0
'Cavaliers on series win'		1	0

'"I didn't do enough to help my team win," Curry said'		1	0
'LeBron James inspired Cavaliers'		0	1
'The Warriors' core is set'		0	1
'Hodgson: "We fear no one".'		1	0
'Higuain to stay at Napoli for the next five years'		1	0
'Murray to win the Queen's title'		1	0
'Stade Pierre Mauroy pitch to be replaced'		1	0
'Albania scores its first ever EURO goal'		1	0
'The Coach of Genzebe Dibaba is arrested in Spain'		1	0
'The Warriors were chasing history'		0	1
'Wales show they are more than Bales'		0	1
'Spain's tactical system has changed'		0	1
'The Stephen Curry revolution is over'		0	1
'Wales are dominating'		0	1
'Bartolo Colon was injured'		1	0
'Tough match, Portugal vs. Croatia'		0	1
'Ronaldo throws reporter's microphone into lake.'		1	0
'Italy vs. Spain is a tough match'		0	1
'It was tough for Portugal this Euro'		0	1
'France to round 16 in Euro 2016'		1	0

Misclassified Data (Stanford Dataset)
'now im done with super stars, and all the tears on her guitar loves it!"'
'@DeniseHammock I like this & I like your site too "'
'@indiemoviemaker fantastic! "'
'We're moving to bigger faster servers today sorry for any service interruptions as a result."'
'@challyzatb Haha I suppose not!"'
'gonna go to sleep now. or maybe later.. hmm, now is good.. i'll be on friday "'
'hahahaha!!!! this tweffa got two chikn biscuits this morning because she's a G... ha! won't be eating the bis...'
'is finally fully recoverd from South Beach! "'
'@Cherylmcfly cause it will be a big change!   oohh thats a mcfly song :p, im sad lol x"'
'Just landed herself a job intervieww woopppp !!!!!!!!!!!!!"'
'@distinctgraphic Excellent. I will certainly keep you in mind for anything in that area "'
'Snap im going to be late for work, traffic stinks! Hope everyone has a great day today "'
'@alimony I'm sitting at your desk. i am going to steal your drawers. Derek has decimated your napkins. "'
'@cindy23cindy thanks!! "'
'@JF_Kennedy why, thank you you're interesting as well. that's why i'm following you of course."'
'@emvidaltx Buenos dias! Sounds like u r eatin d breakfast of champs? cereals, Wheaties? Enjoy yur day!"'
'but family comes tomorrow, so that'll be fun! "'
'@mfharrell @kellykendall. It is so good! Get season 3! Or just come visit and watch it with us. "'
'@joshthomas87 i must say, your win was pretty epic tonight "'
'Heya @surforama - diggin your Bondesque theme this mornin. I'm feelin the 80's synth bands so far this AM. ?...'
'happy birthday mother "'
'bhahahaha some of my new classmates already know that i like twilight and i am obsessed with David "'
'Off to work! "'
'getting ready to go walking with mya then hit the ground running. so much to do, no time for anymore breakdow...'
'watching the Covenant again w00t!!! Awesome movie "'
'5 #Evergreen jobs available: http://evergreen-ils.org/blog/?p=209 supporting a cool open source library syste...'
'@LorelieBrown Half of you is having a blue day, the other half a grey one "'
'@davidgregory aren't you supposed to be working on Today??? this is what you do during stories??   j/k - lo...'
'@mccmarianne Thanks. I haven't been around much either. Having 3 girls to take care of is exhausting..hugs ...'
'@maxmarkson hey max, didn't get to say thankyou before i left. so yes: thanks for tonight! i had heaps of fun...'
'@tennisqueen13 great here i am heading to bed and really wanting to know who was behind you?? Tell me... and...'
'@TracyeDukes It certainly is a good day "'
'@SOS_CaraBrook @tmcfeeley. :: I just lost... um like... 280 followers though! "'
'@IndywoodFILMS ...hello how goes it? ...things coming together for you?"'
'Away to watch drag me to hell "'
'@ManuelaFritschi btw, M5 now also has twitter although i think it's only jesse who updates @maroon5"'
'@springtree have fun! "'
'@NileyJirus I think I missed it to so plz when u find out tell what she said "'
'Wow, history isn't that hard! It's about the United States of America, and how it became independent! It's ac...'

'"can i buy you with my tickets?" "next week isnt good. the jonas brothers are in town." "'
'@aabradley82 When you're finished, will you come mow around our arena?? LOL Trainer comes 2day so we've got t...'
'visiting the old office always brings back waves of nostalgia. "'
'http://twitpic.com/7jdce - I love them"'
'wishes she could go see britney again. what an amzing show! xx"'
'getting ready for my rehearsal .. and listening to the kings "'
'@therealgoos The ultimate universal control "'
'is reminscing on Normanstock memories. "'
'@MollieK121 I use this basic recipe: http://bit.ly/KrZD8. I just change the sugar to 1/2 white & 1/2 br...'
'@Marievh des noms ! des noms ! "'
'1 hour and 35 mins til the new #Weeds episode downloads "'
'thanking god for all my wonderful blackfella friends and family-aint no one i trust more than them mob "'
'Vacation to Lake George! Will be back sometime on Friday. Text or call me if you need me. I'll miss you all ...'
'eating cake hihi.."'
'I'm back! It's even more than 10 minutes! "'
'Day 2 of elping the pre-k section of VBS is starting at nine...but I'm getting there early. Wish me luck. "'
'http://twitpic.com/7jdcq - roachedeggs @ coffeeshop homegrown fantasy, amsterdam. home grown "'
'Woohoo! Got my vanity url. Find me at facebook.com/chrismou "'
'@TheDarrenxshow Haha there was only one quick audience shot and you could hardly see me! "'
'@Chesneyh goodnight god bless sweet dreams "'
'amei jonas brothers e mcfly "'
'Anyone wanna go camping or something this summer? dead bored."'
'@Searock_ i know "'
'@RFLong Awesome Ruth! BTW, still have bookmarks to bring to Nationals. If you want me to bring anything else. "'
'@DeniceSy not really more like a advance class for the math challenged. haha but anyway it's still sumthin to...'
'@clarajonas i finished reading the whole series last year. at first,i kinda of hated EC while reading new moo...'
'@celsoportiolli bom diiiiiiiiiiia "'
'@taliabatalia we wish with kol "'
'saw HUNDREDS of ladyslippers in bloom in the upper slopes of the East Moose River gorge. too many bugs eating...'
'@ScruffyPanther hallo x"'
'Having fun today "'
'@ Lucas_Grabeel hellouu.... how r you "'
'going to see hangover again with colby "'
'Hope that's the 1st and last time I'm associated with the "n" word! "'
'@matthewkheafy wow, that Kent guy is a serious good drummer! awesome you let him drum Iron Maiden http://bi...'
'is gunna go shopping for sum food nd get her phne sorted out dropped it in the bath lol peace homies !!! "'
'@eliseeeherrera I hope for the same thing! lol IMY Elise "'
'Among all the people to see at the Groceries, I see you -__-. But I'm glad I did "'

'i didn't notice... it was rainin',, aha,, smiLe '''
'@katiapsyche Btw, your converting is AWESOME, but we use celsius in Canada. '''
'lapt n mtapos weeehhhh ngyn tapos na pla.. hehehe anyway.. tym 2 say smething sa pix q sa multiply'''
'@Cherylmcfly yeah i replied x'''
'Good Morning..at work already..hella early! Got my BK Mocha i should be good...i hope '''
'Off to work! '''
'@rashmibansal we just rent ours.. for Rs 500 a month! '''
'Kim Kardashian just told me she likes my voice. Hello ego! its going to be a good day. I'm in love '''
'@CaitDaviesUK Awesome.. You seem to be very busy these days! Can't be a bad thing '''
'http://tr.im/oEEo LOVE THIS! '''
'@jshe ah thanks! '''
'Things I learn last nite. Gay guys got some serious game. I wanna be a gay dude in nyc. I can't wait till Haw...'
'@davidlafuente That's true, that's true MUAH!'''
'@samerde thanks - guess my lazy nature was in control via http://twib.es/3ZQ'''
'@catarino I now first need to finish my THESIS Thanks for your reply - will get back to you in two weeks!'''
'@JeffWhite34 hey thanks for replying '''
'Sunrise. Vow #2 broken.'''
'going to quickly tidy my room, then off to town '''
'@theelfyone I'm going to sink two hours into "training updates" with her, I think '''
'gud mornin erryone..... just another day closer to payday '''
'@alexradsby yeah it's awesome '''
'@SusanScot I'm @ Carries, she's feeling abit better 2day I aint wearing that sling, it doesnt go with any my...'
'@othermuse Thanks, Lynn! It should be a good incentive for me to market my work '''
'yes! coffee was actually saved for me '''
'@njlitster one year at 10 months time flys?'''
'@goldigold oh got cha... so what's for breakfast this morning   I'll have live through someone else - too laz...'
'@Official_Leon - Leon!! your on the list as 'Mr Twitter 2009' we (fans on offish forums) are voting for you ...'
'@JustTwistie I AM IN BED LOSER =P haha and I will be '''
'Laundry day! buuuut, I am not feeling it, so I'll be naked today My door is locked...don't bother...'''
'installed windows 7 on laptop, so far its A LOT better than Vista   Think xp will finally have a successor'''
'@abduzeedo It's great! '''
'@Twyst thank you!!! '''
'good morning! enjoying a delish McCafe Iced Coffee Vanilla Yum!'''
'Although i have no money and no job i am happy happy happy and loving life right now yayyyyyya '''
'"Turbo C++ IDE" is a compiler for those people who have a fetish for blue screens after falling in ...'
'@renxin FOLLOWED LAAAA '''
'@vin495 @theroguegirl Don't you have to say it *before* the other person for them to be copying you??? ;-) Se...'
'@david_bosman oh que oui, on en redemande '''
'@GoDiegoGo12 at least you tried. '''
'Happy Tuesday '''

'@OfficialAkaye Hey, hope you & Nadine had a good b'day! Tour was the best yet, you & the girls were a...'
'@Alegrya oh if you ever need an excuse to wear one let me know and we can make one up togther '''
'@westlifepixie it is really warm here in Dublin. The sun is out.. '''
'@gavenoakley13 great movie '''
'@CDButler Thank you!! '''
': having Union BBQ today and we are expecting to have another beautiful day in Toronto. Listening to Viva La...'
'Father's Day....Aww I miss my dad. Happy Father's Day Dad!!! '''
'had an awful day but at least i there is dancing tomorrow '''
'@thomasrdotorg thanks thomas she's a bit of aright'''
'@colinkelly love the Manics version '''
'@davidrobinson11 I MISS YOU TOO! Horray for mutual missingness.'''
'Searching/ Studying twitter API... might have a project involving this. '''
'Just straightened my hair! So seeing how @HannahTYO baby is going on her date!!! She isn't updating me! Haha....'
'if there was a way to bring my uni friends and a local spar shop then awesome '''
'Those who know your name will trust N U, 4 U Lord have never 4saken those who seek U. Psm 9:10 Cherie Camp '''
'@scottross Great info from Forrester Scott, Thanks for sharing that with us '''
'@challyzatb ooh, cool. I think I saw a clip from it a while ago, now I come to think. Also I am sneezing. In...'
'Well let's hope today is better than yesterday '''
'@DotCom9000 Having an awesome day? You should... I expect as much out of you #awesomeupdater'''
'@LittleLiverbird oh, the BS ones never last long, but always end up having tons around anyway lol loving Aved...'
'So were now mini celebraties at our local corner shop because of Mr Moyles even the DHL man is wanting a sho...'
'The popular Flip camcorder now ships with new software including Flip Channels & a cool iPhone app. Nice '''
'I am looking forward to ML training next week '''
'is thinking twitter is very weird but thinks he will begin to like it lol '''
'@evzi "you've got me". that still sounds dodgy '''
'@apraalii haha yeah A McFly To Play in Space '''
'There's sun this morning! wow... '''
'I'm actually ready for work!! Money makes me happy! '''
'@TheDarrenxshow No it was so much fun! In thew breaks they were telling so many rude jokes! '''
'is watching The Dark Knight '''
'@wendyg Aye! My bad, I forgot to specify Canada. '''
'Late day to work today, nothing wrong with that '''
'next month, naik level atau game over? '''
'In bed... lazy ass... but hey i'm in london! '''
'The best things in life are free '''
'Work.... Only until 2 '''
'More adults and children now choose Mind-Body Practices. Why don't you try PILATES? www.personalizedpilates.c...'
'is happy today! '''

'in techno class with marion '''
'Anyone know of a great Father's Day Craft for grade 4 students - one that hasn't been done before??? '''
'@IAmBhargava If u r paying the bill, Cable Car rocks!! '''
'soooo sleeepppyyy. cant believe I'm up and it's 8am XD all for No Doubt. '''
'"i vawnt to clean yer vindows surr" "I dont want my FUCKING windows cleaned!" PMSL! i lov...'
'Trying to decide what's for breakfast? Used curriculum sale tonight!!!!! Hopefully I can find a few things ...'
'@aplusk lol. palindromes, right? haha '''
'New Twitter Tuesday blog post is up! http://bit.ly/hywp1'''
'It was just my imagination running away with me... '''
'I am now majoring in family and child sciences '''
'for the very first time I had a 3-days-tour with my whole band ! it was so much fun more more more!!!'''
'@chiefsanjay I already love @ivybean104!!!!'''
'here i am! '''
''Lovebug' by Jonas Brothers...omg I luv this song sooooo much '''
'@myrthu feel free to come visit us at amaliegade (@signaldigital) if you have the time. I can offer a free cu...'
'feels a bit revived ... i wonder if the juice is doing this '''
'@aplusk what's Star backwards... Rats heehehehe morning Ashton!'''
'@KhrystianB you know it! It's my motivation to smile '''
'Glad the others moved into the new offices. Now I have more space here to practice my muay thai kicks #fb'''
'Its so nice to see everyone wearing freedom green. Peace to Iran!'''
'school is closed today, pipes burst yayy'''
'@AwakenToTruth mornin "Mark"! That's right I said it!! LOL '''
'@jonasnessica hahaha fair enough. the "dead" bodies don't look too real though so i'm good.'''
'I'm on mobile trying out! - http://tweet.sg'''
'@shandreen *** hug **'''
'Darn internet. So slow! Good thing there's Sun Alertz '''
'@auerfeld Haha.. Nice LINK there '''
'@christinefarmer fortunately not down... just ... bit like you I guess, only more "meh" then "...'
'@drugaddicteyes '''
'Every 8 seconds a woman gives birth in the US. We have to find this woman and stop her '''
'Don't forget our huge Spring savings currently at http://shooshoosusa.com - buy for your baby or a gift for f...'
'@nilajafever I second that! I'm a put on a show type of dude....get ready '''
'@aussie_ali have the "100 day cough" but all is good in the Webster household. everyone asleep but ...'
'@indiemoviemaker Wow, congrats! Must have been an awesome phone call to get '''
'@slavin Remind me not to give you a big fat smooch when I see you later today at the#140conf '''
'@nozomi83 Hmm ever thought of having some gothic jewellery? black diamonds etc.. '''
'@mikepickard yes it was nice.. and I have been at places that were a lot slower... so they are trying hard a...'
'Getting up eary tomorrow. Should probs sleep i love my beautiful emma '''
'one exam left! yay!!! Hehee can't believe im on holiday next week! Can't wait!'''

BIBLIOGRAPHY

[1] "The world of digital in 2016 | #DMWF series site," in Uncategorized, #DMWF Series Site, 2016. [Online]. Available: http://www.digitalmarketing-conference.com/world-digital-2016/.

[2] J. Kim, J. Yoo, H. Lim, H. Qiu, Z. Kozareva, A. Galstyan, Sentiment Prediction using Collaborative Filtering, Association for the Advancement of Artificial Intelligence, 2013.

[3] F. H. Khan, S. Bashir, and U. Qamar, "TOM: Twitter opinion mining framework using hybrid classification scheme," Decis. Support Syst., vol. 57, pp. 245–257, Jan. 2014.

[4]"Sports betting - statistics & facts," www.statista.com. [Online]. Available: http://www.statista.com/topics/1740/sports-betting/.

[5] "Global gambling market: gross yield 2010-2019 | Statistic", Statista, 2017. [Online]. Available: https://www.statista.com/statistics/253416/global-gambling-market-gross-win/.

[6] K. Toutanova and C. Manning. Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC 2000), pages 63-70, 2000.

[7] G. A. Miller, "Wordnet: A Lexical Database for English", Communication of the ACM, vol. 38, no. 11, pp39-41, 1995.

[8] "The search API," in Twitter, Twitter, 2014. [Online]. Available: https://dev.twitter.com/rest/public/search.

[9] "List of sportspeople by nickname,". [Online]. Available: https://en.wikipedia.org/wiki/List_of_sportspeople_by_nickname.

[10] A. Go, R. Bhayani, L. Huang, Twitter sentiment classification using distant supervision, CS224N Project Report, Stanford, 2009: 1-12.

[11] M. Thelwall. Heart and soul: Sentiment strength detection in the social web with sentistrength. Cyberemotions, pages 1–14, 2013.

[12] H. Saif, Y. He, and H. Alani. "Alleviating data sparsity for twitter sentiment analysis." The 2nd Workshop on Making Sense of Microposts. 2012.

[13] J. Spencer, and G. Uchyigit, " Sentimentor: Sentiment Analysis of Twitter Data, " Proc. of the 1st Int. Workshop on Sentiment Discovery from Affective Data, pp. 56-66, 2012.

[14] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau. "Sentiment analysis of twitter data." Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics, 2011.

[15] S. Zhao, L. Zhong, J. Wickramasuriya, and V. Vasudevan, "Analyzing Twitter for Social TV-Sentiment Extraction for Sports," Future Tv Workshop, 2011.

[16] E. Kouloumpis, T. Wilson and J. Moore, "Twitter Sentiment Analysis: The Good the Bad and the OMG!", In Proc. of the 5th International AAAI Conference on Weblogs and Social Media, 2011. pp. 538-541.

[17] R. Suchdev, P. Kotkar, R. Ravindran, and S. Swamy, "Twitter Sentiment Analysis using Machine Learning and Knowledge-based Approach," IJCA, vol. 103, no. 4, pp. 36-40, Oct. 2014.

[18] X. An, A. R. Ganguly, Y. Fang, S. B. Scyphers, Ann M. Hunter, Jennifer G. Dy, “Tracking Climate Change Opinions from Twitter Data”, Workshop on Data Science for Social Good, 2014.

[19] P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, V. Stoyanov, “SemEval-2016 Task 4: Sentiment Analysis in Twitter”, In: Proceedings of the 10th international workshop on semantic evaluation, pp. 1-18, 2016.

[20] A Pappurajan, SP Victor, “Web sentiment analysis for scoring positive or negative words using Tweeter data”, Int. J. Comput. Appl, vol. 96, pp. 33-37, June. 2014.

[21] E. Cambria, "Affective Computing and Sentiment Analysis", IEEE Intelligent Systems, vol. 31, no. 2, pp. 102-107, 2016.

[22] Y. Rizk, M. Awad, “Syntactic Genetic Algorithm for a Subjectivity Analysis of Sports Articles “, 11th IEEE International Conference on Cybernetic Intelligent Systems, Limerick, Aug 23-25, 2012.

[23] J. Wiebe and E. Riloff, "Creating subjective and objective sentence classifiers from Unannotated texts," in Computational Linguistics and Intelligent Text Processing. Springer Science + Business Media, 2005, pp. 486–497.

[24] V. Vryniotis, "Machine learning platform - text analysis service," [Online]. Available: http://www.datumbox.com.

[25] AlchemyAPI, "Powering the new AI Economy," [Online]. Available: http://www.alchemyapi.com.

[26] S. Motraghi, Twittycity.co.uk, [Online]. Available: http://www.twittycity.co.uk/.

[27] T. Abdulaieve, 'Streamcrab - Realtime, Twitter sentiment analysis engine', Streamcrab.com, [Online]. Available: http://www.streamcrab.com/.

[28] C. Adams, 'Mood Map – How happy is the world?', Moodmap.themaninblue.com, [Online]. Available: http://moodmap.themaninblue.com/.

[29] G. MacKerron and S. Mourato, 'Happiness is greater in natural environments', Global Environmental Change, vol. 23, no. 5, pp. 992-1000, 2013.

[30] M. Eldridge, 'Lymbix · Sentiment Analysis Reinvented', Lymbix.com, [Online]. Available: http://www.lymbix.com/.

[31] Sentimentanalyzer.appspot.com, [Online]. Available: http://sentimentanalyzer.appspot.com/.

[32] J. Perkins, 'Python NLTK Demos and Natural Language Text Processing APIs', Text-processing.com, [Online]. Available: http://text-processing.com/.

[33] Repustate.com, 'Sentiment analysis, social media sentiment, semantic analysis, text analytics | Repustate', [Online]. Available: https://www.repustate.com/.

[34] Opinioncrawl.com, 'Opinion Crawl - sentiment analysis tool for the Web and social media', [Online]. Available: http://opinioncrawl.com/.

[35] A. Mosquera, 'Twit-Fight', Twit-fight.appspot.com, [Online]. Available: http://twit-fight.appspot.com/.

[36] J. Cianciullo, 'Real Time Search - Social Mention', Socialmention.com, [Online]. Available: http://www.socialmention.com/.

[37] V. Jain and R. Rai, Mozvo.com, [Online]. Available: http://www.mozvo.com/.

[38] Semantria.com, 'Semantria | Text Analytics and Sentiment Analysis for Everyone', [Online]. Available: https://semantria.com/.

[39] J. Greenberg, Tweettronics.com, [Online]. Available: http://www.tweettronics.com/.

[40] C. Healey and M. Ramaswamy, 'Tweet Sentiment Visualization App', Csc.ncsu.edu. [Online]. Available: http://www.csc.ncsu.edu/faculty/healey/tweet_viz/tweet_app/.

[41] "Searching applications," 2002. [Online]. Available: http://algs4.cs.princeton.edu/35applications/.

[42] A. Esuli and F. Sebastiani, "SentiWordNet: A publicly available lexical resource for opinion mining," in Proceedings of Language Resources and Evaluation (LREC), 2006.

[43] "UCI machine learning repository,". [Online]. Available: http://archive.ics.uci.edu/ml/.

[44] N. Hajj, M. Awad, "Weighted Entropy Cortical Algorithms for Modern Standard Arabic Speech Recognition", International Joint Conference on Neural Networks (IJCNN), Dallas, TX, August 4–9, 2013.

[45] R. Zafarani, M. Abbasi and H. Liu, "Information Diffusion in Social Media", in Social Media Mining, Cambrdige University Press, 2014, 217-255. [Online] Available: ddml e-book.

[46] "For academics - sentiment140 - A Twitter sentiment analysis tool,". [Online]. Available: http://help.sentiment140.com/for-students/.