ISLAMIC UNIVERSITY OF TECHNOLOGY

UNDER GRADUATE THESIS

Sentiment Analysis of comments having emoticons feedback

Authors:
Md. Mahfuz Ibn Alam(144411)
Mehadi Hasan(144417)

Supervisor:

Dr. Abu Raihan Mostofa Kamal

Professor, Department of CSE

Islamic University of Technology (IUT)

A thesis submitted to the Department of CSE in fulfilment of the requirements for the Degree of B.Sc Engineering in CSE.

Academic Year: 2017-18.

October, 2018

Declaration of Authorship

This is to certify that the work presented in this thesis is the outcome of the analysis and investigation carried out by Md. Mahfuz Ibn Alam and Mehadi Hasan under the supervision of Dr. Abu Raihan Mostofa Kamal in the Department of Computer Science and Engineering (CSE), IUT, Dhaka, Bangladesh. It is also declared that neither of this thesis nor any part of this thesis has been submitted anywhere else for any degree or diploma. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.

Authors:

Md.Mahfuz Ibn Alam
Student ID - 144411

Mehadi Hasan
Student ID – 144417

Supervisor:

Dr. Abu Raihan Mostofa Kamal
Professor
Department of Computer Science and Engineering
Islamic University of Technology (IUT)

Abstract

B.Sc in Computer Science and Engineering

Sentiment Analysis of comments having emoticon feedback

By Md. Mahfuz Ibn Alam (144411) , Mehadi Hasan (144417)

Sentiment analysis refers to the inference of people's views, positions and attitudes in their written or spoken texts. Before the coining of the term, the field was studied under names such as subjectivity, point of view and opinion mining. Nowadays, the field is rapidly evolving due to the rise of new platforms such as blogs, social media and user-generated reviews. Two main research directions can be identified in the literature of sentiment analysis on microblogs. First direction is concerned with finding new methods to run such analysis, such as performing sentiment label propagation on Twitter follower graphs, and employing social relations for user-level sentiment analysis. The second direction is focused on identifying new sets of features to add to the trained model for sentiment identification, such as microblogging features including hashtags, emoticons, the presence of intensifiers such as all-caps and character repetitions etc., and sentiment topic features.

Acknowledgements

In full gratitude, we would like to acknowledge the following individuals who encouraged, inspired, supported, assisted, and sacrificed themselves to help our pursuit of the successful thesis work.

From our academy, we would like to thanks Dr.Abu Raihan Mostofa Kamal, Professor, CSE, IUT for the continuous support of our thesis work and related research, for his patience, motivation, and immense knowledge. His guidance helped us in all the time of research and writing of this thesis.

We would also like to thanks Md. Rafiqul Islam, Assistant Professor, CSE, City University for his insightful comments and encouragement, but also for the hard question which incented us to widen our research from various perspectives.

Last but not the least, we would like to thank our families for supporting us spiritually throughout the writing of this thesis and our life in general.

With Regards,
Md. Mahfuz Ibn Alam (144411)
Mehadi Hasan (144417)

Contents

Table of figures

Chapter 1

Introduction

1.1 Motivation

Sentiment =

- Feelings
- Attitudes
- Emotions
- Opinions

Subjective impressions, not facts.Generally, a binary opposition in opinions is assumed .For/against:

- Like/Dislike
- Good/bad
- Positive/negative

Some sentiment analysis jargon:

– "Semantic orientation"

– "Polarity"

Sometimes referred to as opinion mining, although the emphasis in this case is on extraction

It is very difficult to survey customers who didn't buy the company's laptop Instead, you could use SA to:

A) Search the web for opinions and reviews of this and competing laptops. Blogs, Facebook comments, amazon, tweets, etc.

B) Create condensed versions or a digest of consensus points.

Insights and applications from SA have been useful in other areas

– Politics/political science

– Law/policy making

– Sociology

– Psychology

1.2 Thesis Contribution

The contribution of this thesis is as follows:

1. Data collection: As the comment or user review with reactions feedback is not available so we have to collect the data from Facebook public post using fb graph API. And a python script to fetch the data into csv format.

2. Feature extraction: Adding 6 new feature with lexicon based model using reaction feedback data of the individual comment. So that this feature provide more values on the comment.

1.3 Thesis Outline

The paper is organized as follows: Chapter 2 provides a short overview of Naïve Bayes, Maximum Entropy, SVM, KNN classifier; Chapter 3 discusses the state of the art in the field; Chapter 4 discusses Proposed solution; Chapter 5 shows the results; and Chapter 6 concludes the paper and presents future work.

Chapter 2

Background

2.1 Naïve Bayes:

NB is a probabilistic classifier, where the assignment of a sentiment class c to a given tweet w can be computed as:

$$\hat{c} = \arg\max_{c \in C} P(c|\mathbf{w})$$
$$= \arg\max_{c \in C} P(c) \prod_{1 \leq i \leq N_{\mathbf{w}}} P(w_i|c),$$

where Nw is the total number of words in tweet w, P(c) is the prior probability of a tweet appearing in class c, P(wi |c) is the conditional probability of word wi occurring in a tweet of class c.

In multinomial NB, P(c) can be estimated by P(c) = Nc/N Where Nc is the number of tweets in class c and N is the total number of tweets. P(wi |c) can be estimated using maximum likelihood with Laplace smoothing:

$$P(w|c) = \frac{N(w,c) + 1}{\sum_{w' \in V} N(w'|c) + |V|},$$

where N(w, c) is the occurrence frequency of word w in all training tweets of class c and |V| is the number of words in the vocabulary.

2.2 Maximum Entropy:

The idea behind Maximum Entropy models is that one should prefer the most uniform models that satisfy a given constraint. MaxEnt models are feature-based models. In a two class scenario, it is the same as using logistic regression to find a distribution over the classes. MaxEnt makes no independence assumptions for its features, unlike Naive Bayes.

This means we can add features like bigrams and phrases to MaxEnt without worrying about features overlapping. The model is represented by the following:

$$P_{ME}(c|d,\lambda) = \frac{\exp[\Sigma_i \lambda_i f_i(c,d)]}{\Sigma_{c'} \exp[\Sigma_i \lambda_i f_i(c,d)]}$$

In this formula, c is the class, d is the tweet, and λ is a weight vector. The weight vectors decide the significance of a feature in classification. A higher weight means that the feature is a strong indicator for the class. The weight vector is found by numerical optimization of the lambdas so as to maximize the conditional probability. They use the Stanford Classifier to perform MaxEnt classification. For training the weights they used conjugate gradient ascent and added smoothing (L2 regularization).

Theoretically, MaxEnt performs better than Naive Bayes because it handles feature overlap better. However, in practice, Naive Bayes can still perform well on a variety of problems

2.3 Support Vector Machines (Kernels):

The SVM algorithm is implemented in practice using a kernel. The learning of the hyperplane in linear SVM is done by transforming the problem using some linear algebra, which is out of the scope of this introduction to SVM. A powerful insight is that the linear SVM can be rephrased using the inner product of any two given observations, rather than the observations themselves. The inner product between two vectors is the sum of the multiplication of each pair of input values. For example, the inner product of the vectors [2, 3] and [5, 6] is 2*5 + 3*6 or 28. The equation for making a prediction for a new input using the dot product between the input (x) and each support vector (xi) is calculated as follows:

F(x) = B0 + sum (ai * (x,xi))

This is an equation that involves calculating the inner products of a new input vector (x) with all support vectors in training data. The coefficients B0 and ai (for each input) must be estimated from the training data by the learning algorithm.

2.3.1 Linear Kernel SVM:

The dot-product is called the kernel and can be re-written as:

K(x, xi) = sum(x * xi)

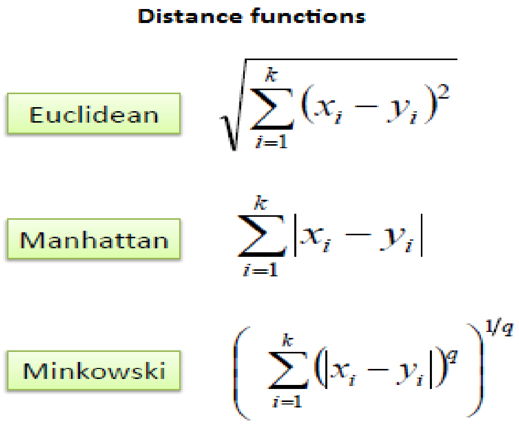
The kernel defines the similarity or a distance measure between new data and the support vectors. The dot product is the similarity measure used for linear SVM or a linear kernel because the distance is a linear combination of the inputs. Other kernels can be used that transform the input space into higher dimensions such as a Polynomial Kernel and a Radial Kernel. This is called the Kernel Trick. It is desirable to use more complex kernels as it allows lines to separate the classes that are curved or even more complex. This in turn can lead to more accurate classifiers.

2.4 K-Nearest Neighbor:

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique.

2.4.1 Algorithm:

A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function. If K = 1, then the case is simply assigned to the class of its nearest neighbor.



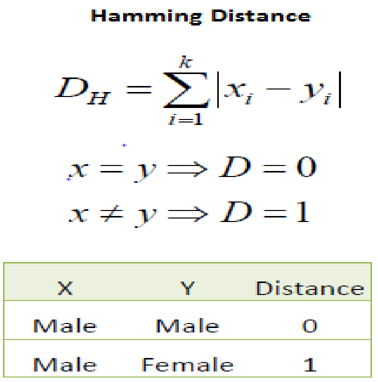
Distance functions

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

Manhattan $\sum_{i=1}^{k}|x_i - y_i|$

Minkowski $\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$

It should also be noted that all three distance measures are only valid for continuous variables. In the instance of categorical variables the Hamming distance must be used. It also brings up the issue of standardization of the numerical variables between 0 and 1 when there is a mixture of numerical and categorical variables in the dataset.



Hamming Distance

$$D_H = \sum_{i=1}^{k}|x_i - y_i|$$

$$x = y \Rightarrow D = 0$$

$$x \neq y \Rightarrow D = 1$$

X	Y	Distance
Male	Male	0
Male	Female	1

Choosing the optimal value for K is best done by first inspecting the data. In general, a large K value is more precise as it reduces the overall noise but there is no guarantee. Cross-validation is another way to retrospectively determine a good K value by using an independent dataset to validate the K value. Historically, the optimal K for most datasets has been between 3-10. That produces much better results than 1NN.

Chapter 3

Literature Review

3.1 Twitter sentiment classification using distant supervision.

Go, A., R. Bhayani, and L. Huang. 2009.

Technical report, Stanford Digital Library Technologies Project.

Approach:

Our approach is to use different machine learning classifiers and feature extractors. The machine learning classifiers are:

1. Naive Bayes
2. Maximum Entropy (MaxEnt)
3. Support Vector Machines (SVM).

The feature extractors are:

1. Unigrams
2. Bigrams
3. Unigrams and bigrams
4. Unigrams with part of speech tags.

3.2 Sentiment analysis of twitter data.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R.

In: Proc. ACL 2011 Workshop on Languages in Social Media, pp. 30–38 (2011)

Proposed Solution:

1. They extend their approach by using real valued prior polarity, and by combining prior polarity with POS. their results show that the features that enhance the performance of our classifiers the most are features that combine prior polarity of words with their parts of speech. The tweet syntax features help but only marginally. (Extended of Barbosa).

In this paper they perform extensive feature analysis and show that the use of only 100 abstract linguistic features performs as well as a hard unigram baseline. (Extended of Go).

Features:

Their features can be divided into three broad categories:

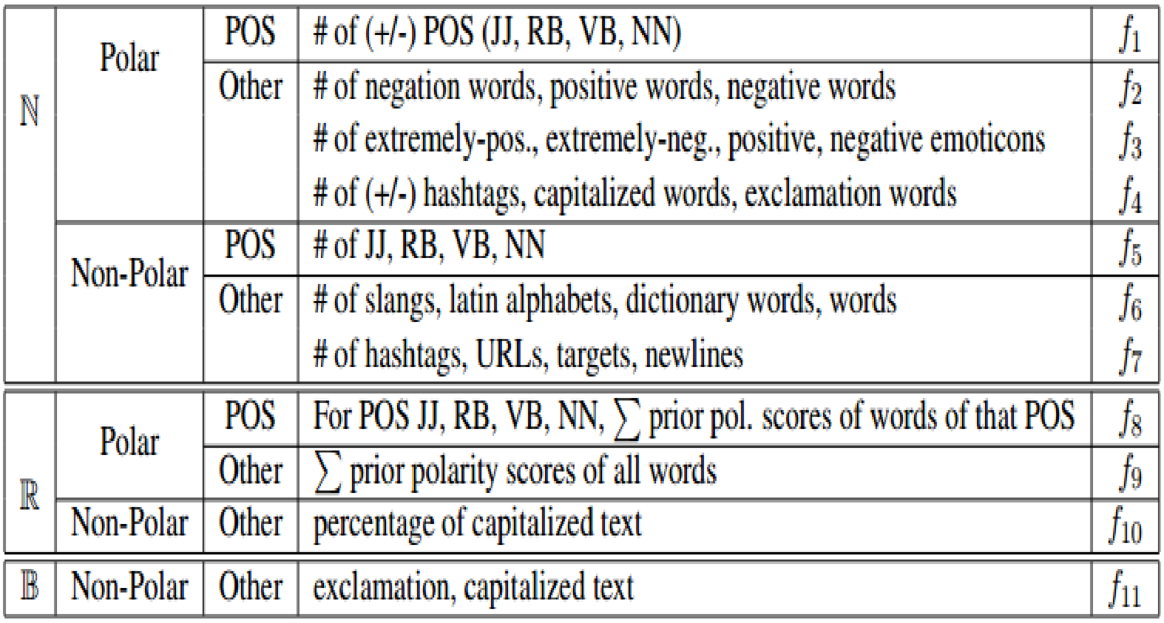
1. Firstly, that are primarily counts of various features and therefore the value of the feature is a natural number ∈ N.
2. Secondly, features whose value is a real number ∈ R. These are primarily features that capture the score retrieved from DAL.
3. Thirdly, features whose values are Boolean ∈ B. These are bag of words, presence of exclamation marks and capitalized text.

Each of these broad categories is divided into two subcategories:

1. Polar features. We refer to a feature as polar if we calculate its prior polarity either by looking it up in DAL (extended through WordNet) or in the emoticon dictionary.
2. Non-polar features. All other features which are not associated with any prior polarity fall in the nonpolar category.

Each of Polar and Non-polar features is further subdivided into two categories:

1. POS. POS refers to features that capture statistics about parts-of-speech of words.
2. Other. Other refers to all other types of features.



$\mathbb{N}$	Polar	POS	# of (+/-) POS (JJ, RB, VB, NN)	f_1
		Other	# of negation words, positive words, negative words	f_2
			# of extremely-pos., extremely-neg., positive, negative emoticons	f_3
			# of (+/-) hashtags, capitalized words, exclamation words	f_4
	Non-Polar	POS	# of JJ, RB, VB, NN	f_5
		Other	# of slangs, latin alphabets, dictionary words, words	f_6
			# of hashtags, URLs, targets, newlines	f_7
$\mathbb{R}$	Polar	POS	For POS JJ, RB, VB, NN, $\sum$ prior pol. scores of words of that POS	f_8
		Other	$\sum$ prior polarity scores of all words	f_9
	Non-Polar	Other	percentage of capitalized text	f_{10}
$\mathbb{B}$	Non-Polar	Other	exclamation, capitalized text	f_{11}

Figure 3.2: Predefined table of polarity

3.3 Semantic sentiment analysis of twitter

H Saif, Y He, H Alani

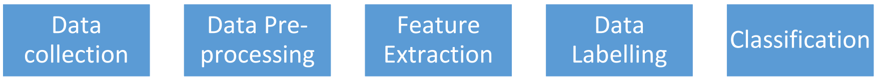
International semantic web conference, 2012 – Springer

Contribution:

1. Introduce and implement a new set of semantic features for training a model for sentiment analysis of tweets.
2. Investigate three approaches for adding such features into the training model by replacement, by argumentation, and by interpolation, and show the superiority of the latter approach.
3. Demonstrate the value of not removing stop words in increasing sentiment identification accuracy.

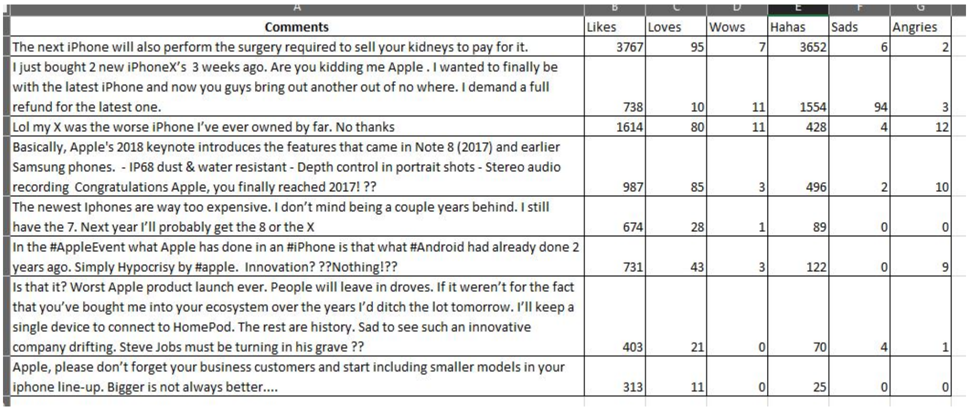
Chapter 4

Proposed Method



4.1 Data Collection:

Using python script and Facebook graph API to get the csv formatted data.



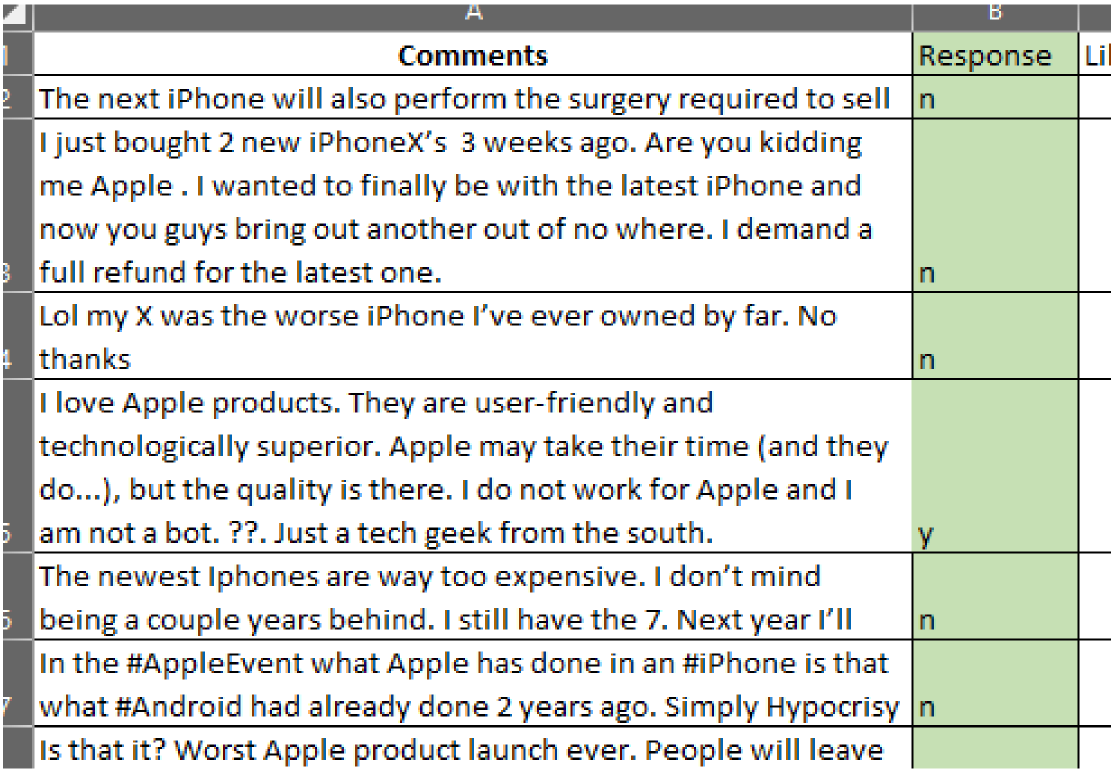
Comments	Likes	Loves	Wows	Hahas	Sads	Angries
The next iPhone will also perform the surgery required to sell your kidneys to pay for it.	3767	95	7	3652	6	2
I just bought 2 new iPhoneX's 3 weeks ago. Are you kidding me Apple . I wanted to finally be with the latest iPhone and now you guys bring out another out of no where. I demand a full refund for the latest one.	738	10	11	1554	94	3
Lol my X was the worse iPhone I've ever owned by far. No thanks	1614	80	11	428	4	12
Basically, Apple's 2018 keynote introduces the features that came in Note 8 (2017) and earlier Samsung phones. - IP68 dust & water resistant - Depth control in portrait shots - Stereo audio recording Congratulations Apple, you finally reached 2017! ??	987	85	3	496	2	10
The newest Iphones are way too expensive. I don't mind being a couple years behind. I still have the 7. Next year I'll probably get the 8 or the X	674	28	1	89	0	0
In the #AppleEvent what Apple has done in an #iPhone is that what #Android had already done 2 years ago. Simply Hypocrisy by #apple. Innovation? ??Nothing!??	731	43	3	122	0	9
Is that it? Worst Apple product launch ever. People will leave in droves. If it weren't for the fact that you've bought me into your ecosystem over the years I'd ditch the lot tomorrow. I'll keep a single device to connect to HomePod. The rest are history. Sad to see such an innovative company drifting. Steve Jobs must be turning in his grave ??	403	21	0	70	4	1
Apple, please don't forget your business customers and start including smaller models in your iphone line-up. Bigger is not always better....	313	11	0	25	0	0

Figure 4.1: Sample training data

4.2Data Pre-Processing:

- Removing comments that only tag Facebook friend
- Manually creating the response for each comment

From around 100 thousands comment → 1,000 usable comments found



	A	B	
1	**Comments**	Response	Lil
2	The next iPhone will also perform the surgery required to sell	n	
3	I just bought 2 new iPhoneX's 3 weeks ago. Are you kidding me Apple . I wanted to finally be with the latest iPhone and now you guys bring out another out of no where. I demand a full refund for the latest one.	n	
4	Lol my X was the worse iPhone I've ever owned by far. No thanks	n	
5	I love Apple products. They are user-friendly and technologically superior. Apple may take their time (and they do...), but the quality is there. I do not work for Apple and I am not a bot. ??. Just a tech geek from the south.	y	
6	The newest Iphones are way too expensive. I don't mind being a couple years behind. I still have the 7. Next year I'll	n	
7	In the #AppleEvent what Apple has done in an #iPhone is that what #Android had already done 2 years ago. Simply Hypocrisy	n	
	Is that it? Worst Apple product launch ever. People will leave		

Figure 4.2: Sample training data with manual response

4.3Feature Extraction:

4.3.1 Extracting Lexicon based feature for the comments:

- **Using LIWC (Linguistic inquiry and word count) tools**

  LIWC generates 92 features for a paragraph

  From them 15 features is used of 3 main category

  - Other grammar (verb, adjective, compare etc.)
  - Affects (positive emotion, negative emotion, anger etc.)
  - Drives (affiliation, achieve, risk, reward etc.

4.3.2 6 new features vector from reactions feedback:

- **Procedure:**
  - Normalization:

    Normalizing data reaction count from 0 to 1.
  - Aging factor:

    X^2 is used to give new reactions a much bigger weight.

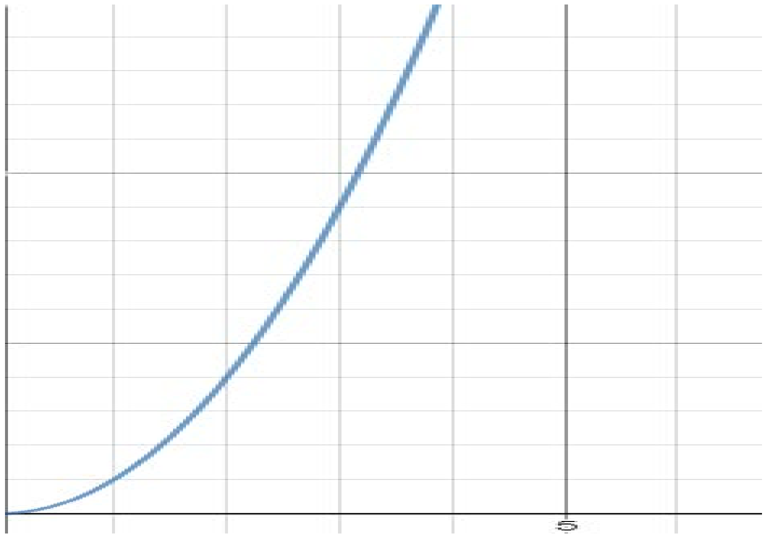


Figure 4.3: aging factor curve

4.4 Classification:

- Support Vector Machine
- K-NN Classifier

We use 5 fold cross validation.

Chapter 5

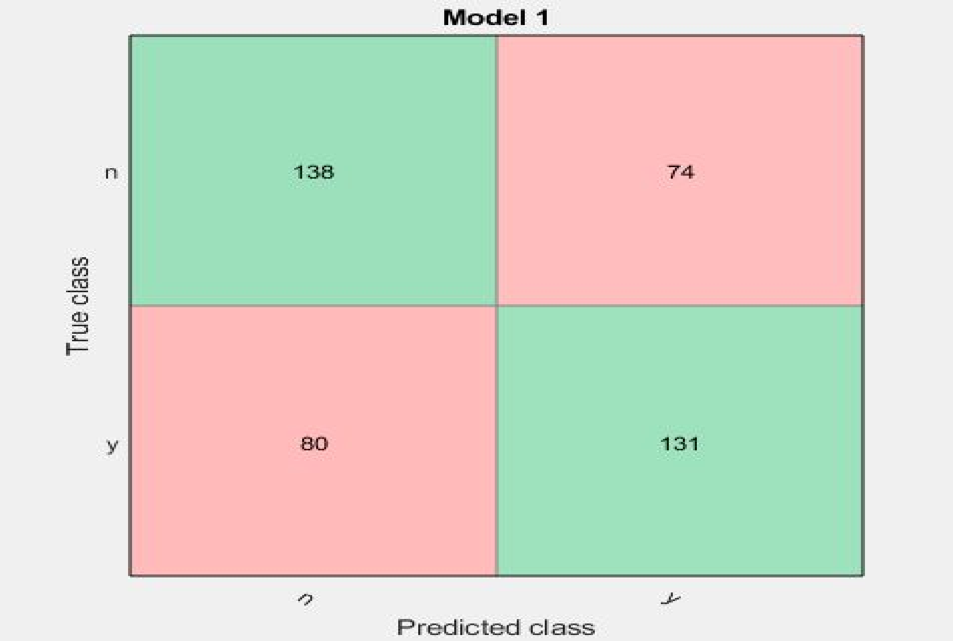
Results:

We find out the result using different classifier. As the extracted feature is the binding the previous method with the new method so we need to evaluate which classification methods suits best.

5.1 Linear SVM:

Linear SVM is the newest extremely fast machine learning (data mining) algorithm for solving multiclass classification problems from ultra large data sets that implements an original proprietary version of a cutting plane algorithm for designing a linear support vector machine. Linear SVM is a linearly scalable routine meaning that it creates an SVM model in a CPU time which scales linearly with the size of the training data set.

The confusion matrix is-



	n	y
n	138	74
y	80	131

Figure 5.1(a): confusion matrix for linear SVM

And the roc curve is-

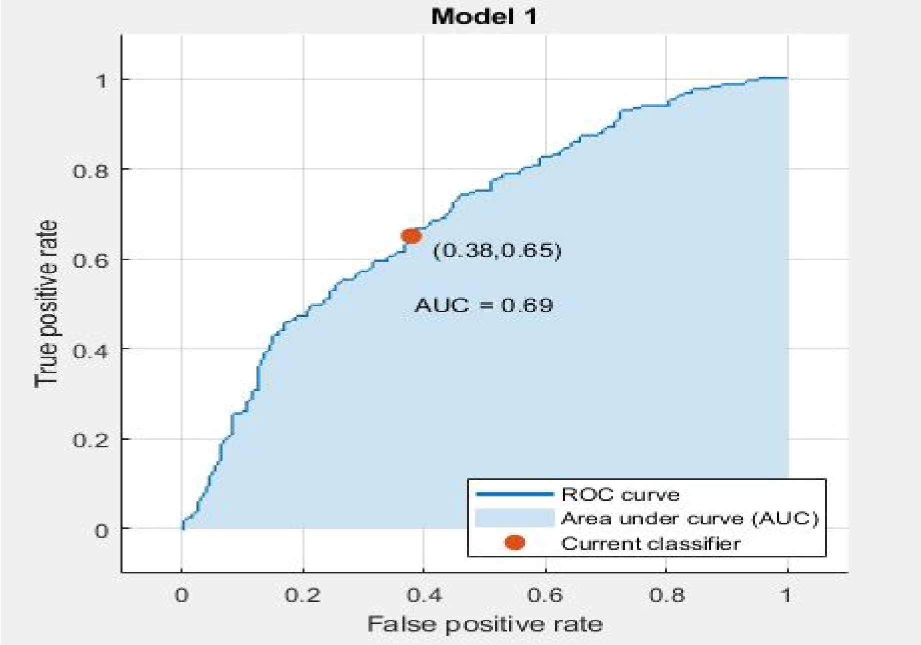


Figure 5.1(b): ROC curve for linear SVM

5.2 Quadratic SVM:

A new quadratic kernel-free non-linear support vector machine (which is called QSVM) is introduced. The SVM optimization problem can be stated as follows: Maximize the geometrical margin subject to all the training data with a functional margin greater than a constant.

The confusion matrix is-

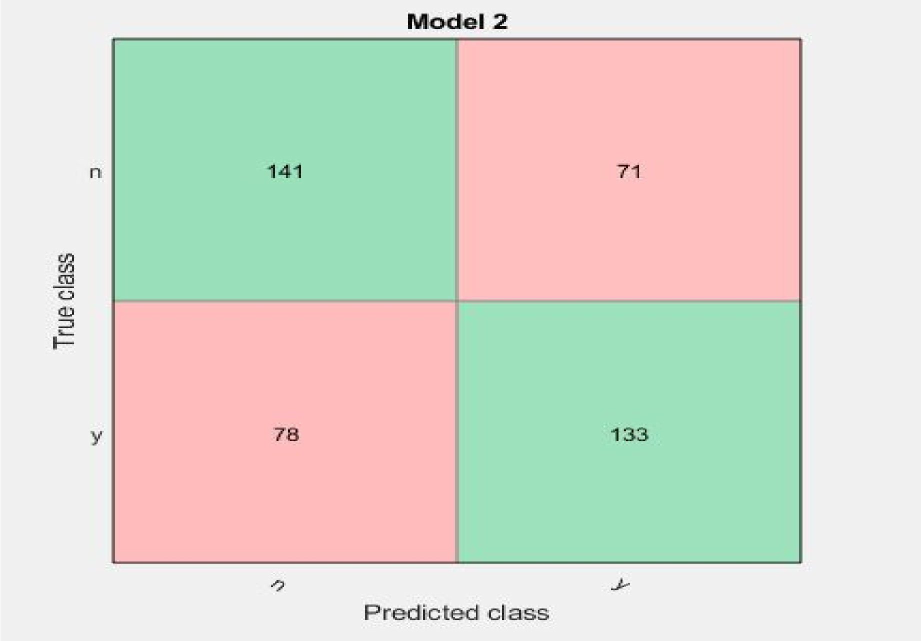


Figure 5.2(a): confusion matrix for quadratic SVM

And the roc curve is-

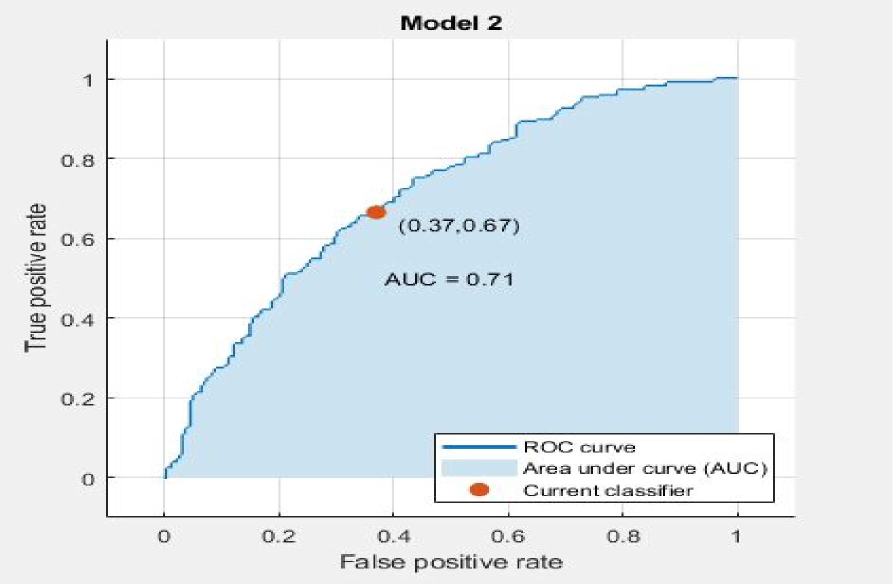


Figure 5.2(b): ROC curve for quadratic SVM

5.3 KNN:

In pattern recognition, the k-nearest neighbors' algorithm is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. We divided KNN into two segments by setting the value of k=5 and k=10 respectively

The confusion matrix is where k=5

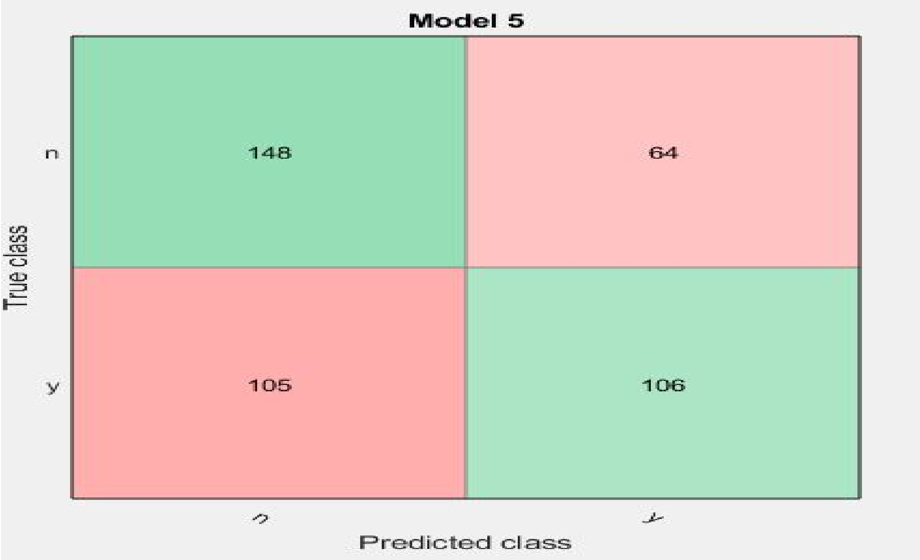


Figure 5.3(a): confusion matrix for KNN where k=5

The confusion matrix is where k=10

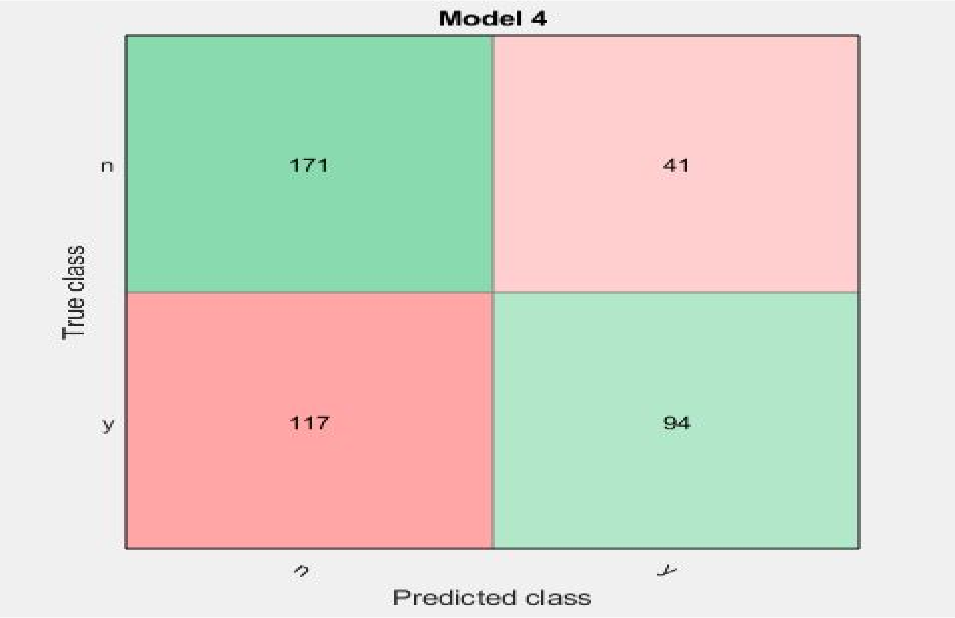


Figure 5.3(b): confusion matrix for KNN where k=10

And the roc curve where k=5 is-

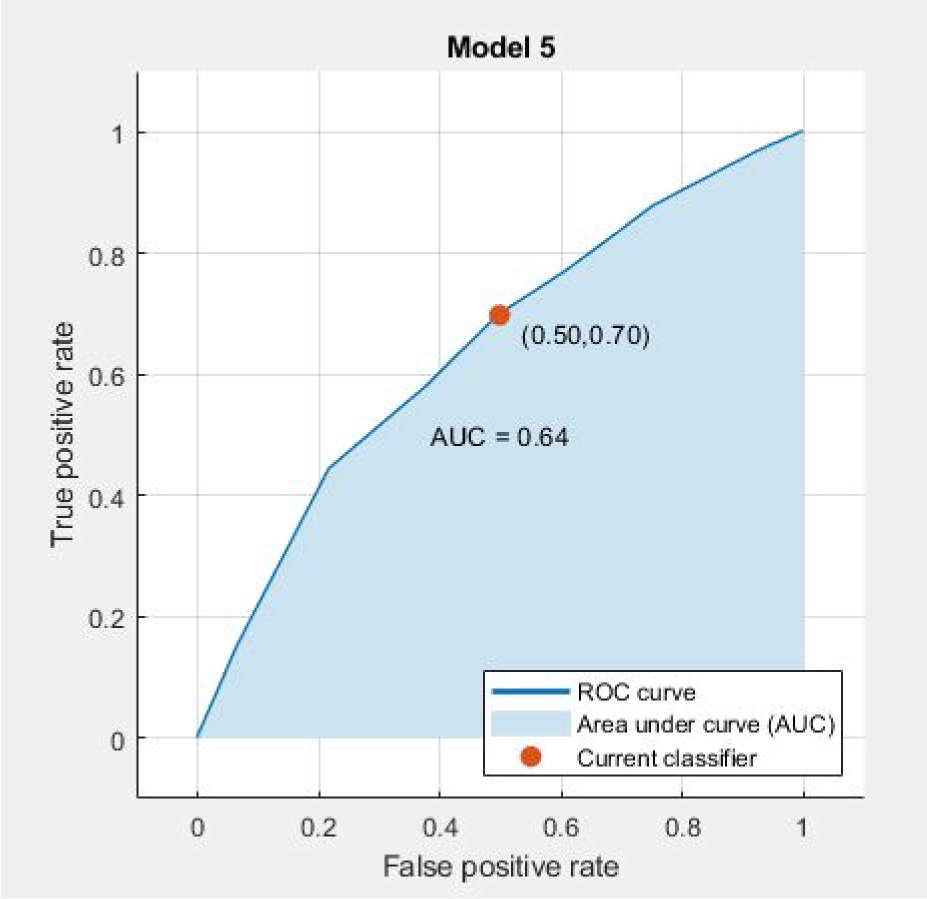


Figure 5.3(c): ROC curve for KNN where k=5

And the roc curve where k=10 is-

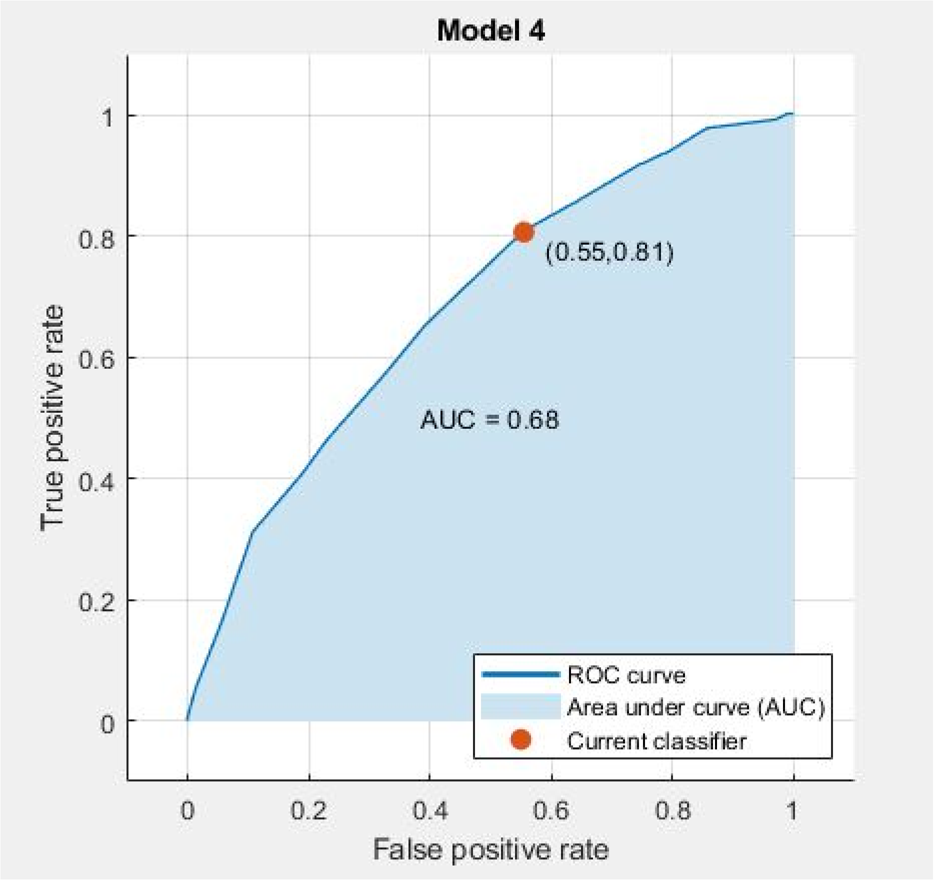
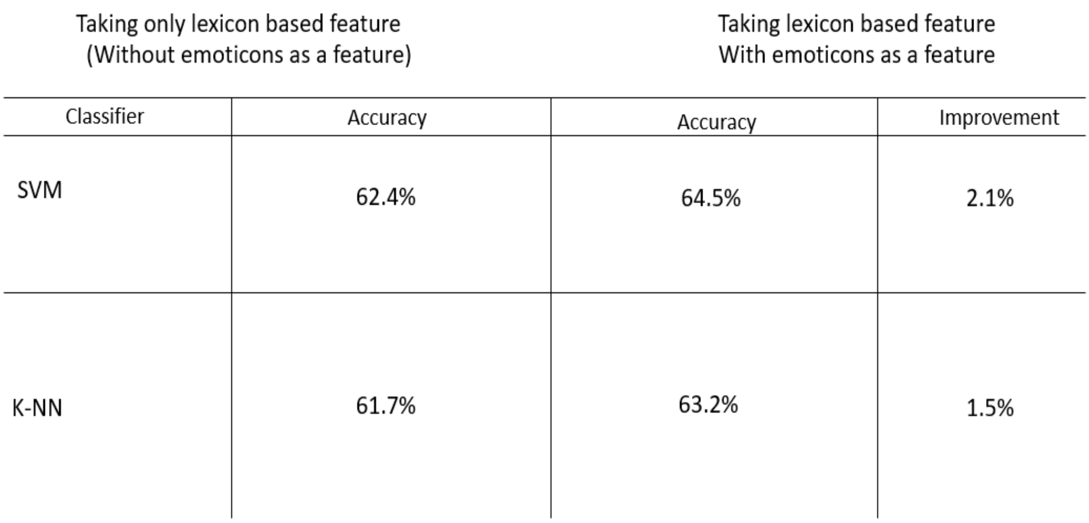


Figure 5.3(d): ROC curve for KNN where k=10

So from all the above result our final result table is-



	Taking only lexicon based feature (Without emoticons as a feature)	Taking lexicon based feature With emoticons as a feature	
Classifier	Accuracy	Accuracy	Improvement
SVM	62.4%	64.5%	2.1%
K-NN	61.7%	63.2%	1.5%

Figure 5.3(e): Result table for all classification

Chapter 6

Conclusion and Future Work

6.1 Conclusion

We modified an existing feature extraction model with some new features and has come up with better accuracy.

The result indicates that our proposed method has a great impact on decision making in term of sentiment analysis.

6.2 Limitations

- Sarcasm detection
- Dataset is not sufficient

6.3 Future Work

As we introduce exactly a new feature with the exciting feature so we got some issues in generating aging factor. And we believe there needs an improvement. Beside only bi-polarity is solved till now. So, the following points need to improve

- We only predict the bi-polarity of the opinion- positive and negative
- Improving aging factor

Bibliography

[1] The psychological meaning of words: LIWC and computerized text analysis methods.

Tausczik, Y.R., & Pennebaker, J.W.
Journal of Language and Social Psychology, 29, 24-54. (2010)

[2] Sentiment analysis of twitter data.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R.
In: Proc. ACL 2011 Workshop on Languages in Social Media, pp. 30–38 (2011)

[3] Sentiment Analysis of Product Reviews: A Review.

Shivaprasad T K, Jyothi Shetty
International Conference on Inventive Communication and Computational Technologies –(ICICCT -2017)