**Syracuse University**

**IST-736 Assignment 6**

ThulasiRam RuppaKrishnan

IST 736

Professor Ami Gates

Contents

[Introduction 3](#_Toc17211630)

[Analysis and Models 4](#_Toc17211631)

[**About the data** 4](#_Toc17211632)

[**Models** 8](#_Toc17211633)

[Results 22](#_Toc17211634)

[Conclusion 24](#_Toc17211635)

## 

## **Introduction**

Deception is an act or statement which misleads, hides the truth, or promotes a belief, concept, or idea that is not true. It is often done for personal gain or advantage. Deception can involve dissimulation, propaganda, and sleight of hand, as well as distraction, camouflage, or concealment. There is also self-deception, as in bad faith. It can also be called, with varying subjective implications, beguilement, deceit, bluff, mystification, ruse, or subterfuge.

Deceit and dishonesty can also form grounds for civil litigation in tort, or contract law (where it is known as misrepresentation or fraudulent misrepresentation if deliberate) or give rise to criminal prosecution for fraud. It also forms a vital part of psychological warfare in denial and deception.

**Deception Recognition**

Deception is a socially pervasive psycholinguistic phenomenon from lies during legal trials to fabricated online product reviews. Its detection in human communication has long been of great interest in real-life situations involving law enforcement, national security, and business, etc. The techniques employed for the detection of deception are varied, ingenious, and often dramatic from the ancient Chinese method of spitting dry rice to the modern polygraph. Deception detection has also been the subject of investigation within psychology, social science, and linguistics, where it has mainly been based on qualitative and quantitative observations of gesture, facial expression and voice analysis. Nonetheless, very little scientific work has been done to date on the fundamental theoretical underpinnings of systems for automatically detecting deception in text. The proliferation in recent years of fake online reviews meant to deceive consumers has heightened the interest in automatic deception filtering systems.

Source: <https://core.ac.uk/download/pdf/54849184.pdf>



Deception is potentially disruptive in everyday communication, information seeking, and decision making. It is a message knowingly and intentionally transmitted by a sender to foster a false belief or conclusion by the perceiver. it is widespread phenomenon and often undetected especially in electronic environments where credibility assessments are complicated by the absence of many traditional cues such as verifiable credentials or face-to-face contact. The need arises for decision support tools capable of alerting users to potentially deceptive content.

Human Judgments and Automated Deception Detection In interpersonal psychology and communication studies, human respondents are often asked to distinguish deceptive statements from truthful ones. People are notoriously unreliable in this task. In a meta-analytical review of over 100 experiments with over 1,000 participants, DePaulo, and colleagues (1997) determined an unimpressive mean accuracy of 54%, slightly above chance. Recently attainable with natural language processing (NLP) and machine learning, automated approaches show promise in distinguishing a general sense of deception with success rates slightly higher than those of humans. The underlying mechanism is to identify reliable verbal cues which show linguistic differences when people deceive as compared to when they tell the truth. For instance, 3 out of 18 verbal cues derived from Statement Validity Analysis techniques (amount of detail reported, coherence, and admissions of lack of memory) used in law enforcement for credibility assessments tested as statistically reliable (Porter & Yuille,1996). Other studies report that, compared to truth-tellers, deceivers produce more total words and sense-based words show lower cognitive complexity, use more negative emotion words, use more extreme positive emotions and more general knowledge references, but use fewer certainty and hesitation words. When implemented in decision support tools, three standard classification algorithms (neural nets, decision trees, and logistic regression) achieved 74% accuracy (Fuller, Biros, & Wilson, 2009). Another approach adapted a pre-existing psycholinguistic lexicon (Linguistic Inquiry Word Count (LIWC)) and achieved an average 70% classifier accuracy in binary lie-truth text categorization (Mihalcea & Strapparava, 2009). With more evidence for reliable verbal cue combinations, human abilities to spot deception can be complemented, if not enhanced.

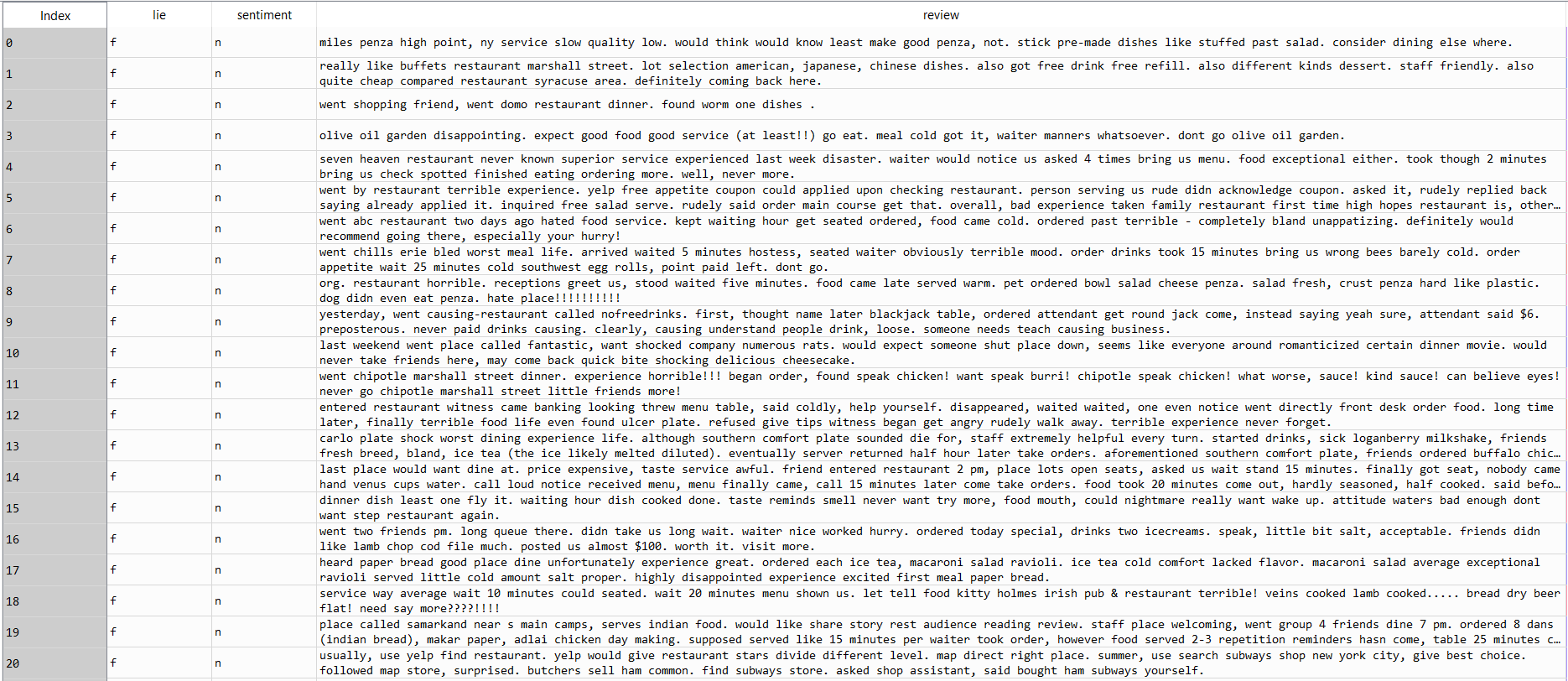
Source: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/meet.2011.14504801098>

## **Analysis and Models**

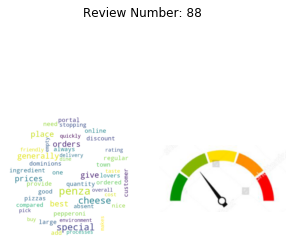
### **About the data**

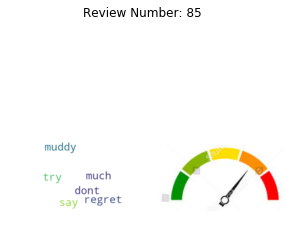
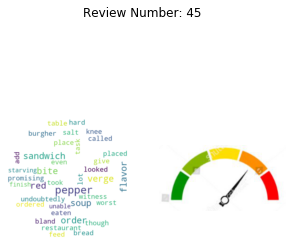
The dataset for deception analysis consists of 90 observations and 3 columns out of which one is labeled for fake/truth and the 2nd one is labeled for sentiment (positive/negative). The third variable(column) contains actual review submitted by public for a restaurant`. **Table 1.1** represents the sample dataset of reviews and its labels.

In the exploratory data analysis, word clouds are generated for every review and the positivity and negativity is tagged by using Sentiment Intensity analyzer. **Figure 1.1** shows word cloud and tagged sentiment intensity by each review. Also, **Figure 1.2 and 1.3** shows word count distribution for each review, and the labels are differentiated by color. Fake and negative sentiment tags are marked using red whereas true and positive sentiment tags are marked in blue.

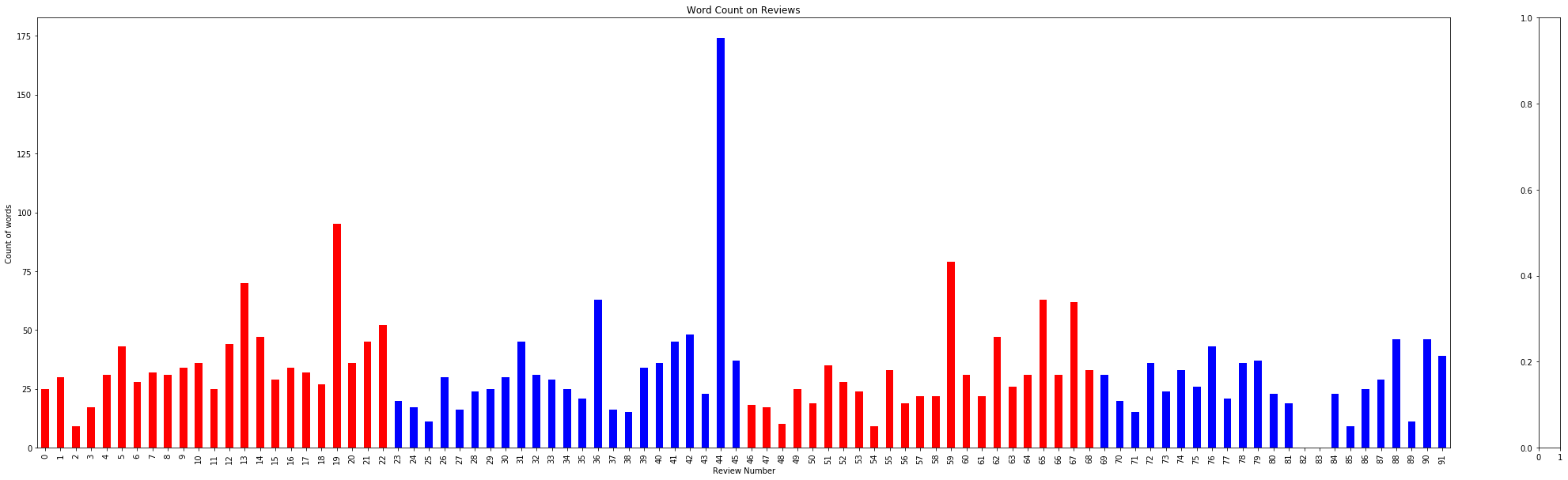


**Table 1.1 A section of Input Data**

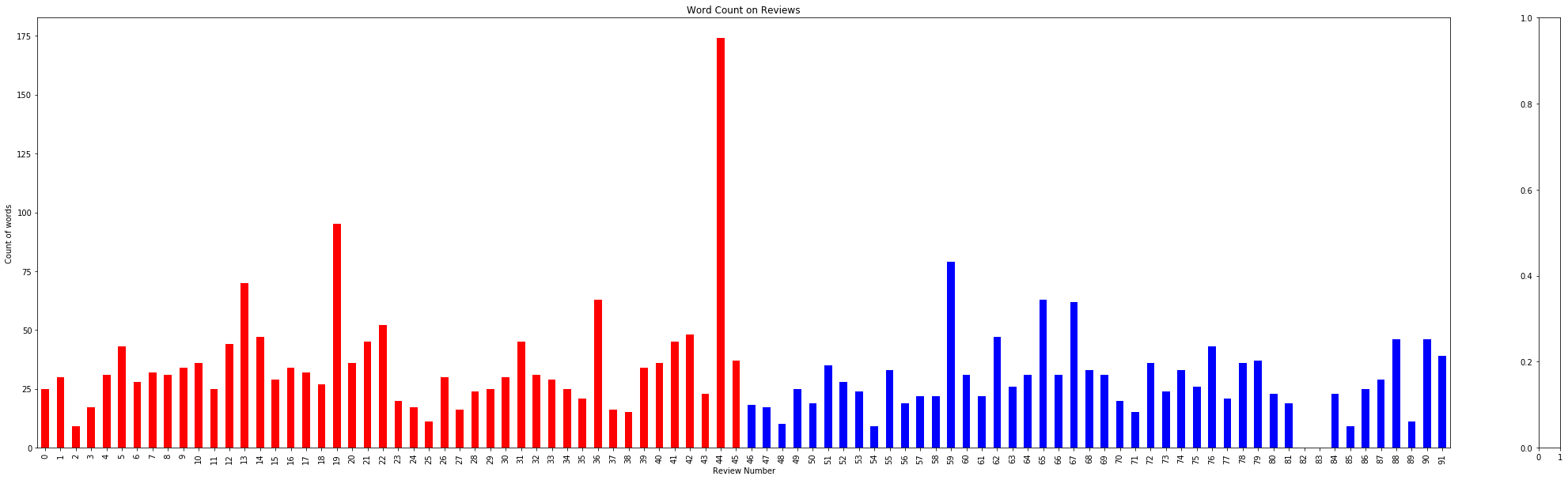
 

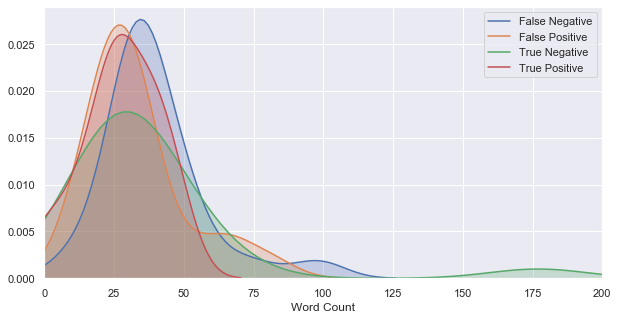
**Figure 1.1 Sample word Clouds from Individual reviews**



**Figure 1.2 Word count distribution for fake and true reviews**



**Figure 1.3 Word count distribution for positive and negative reviews**



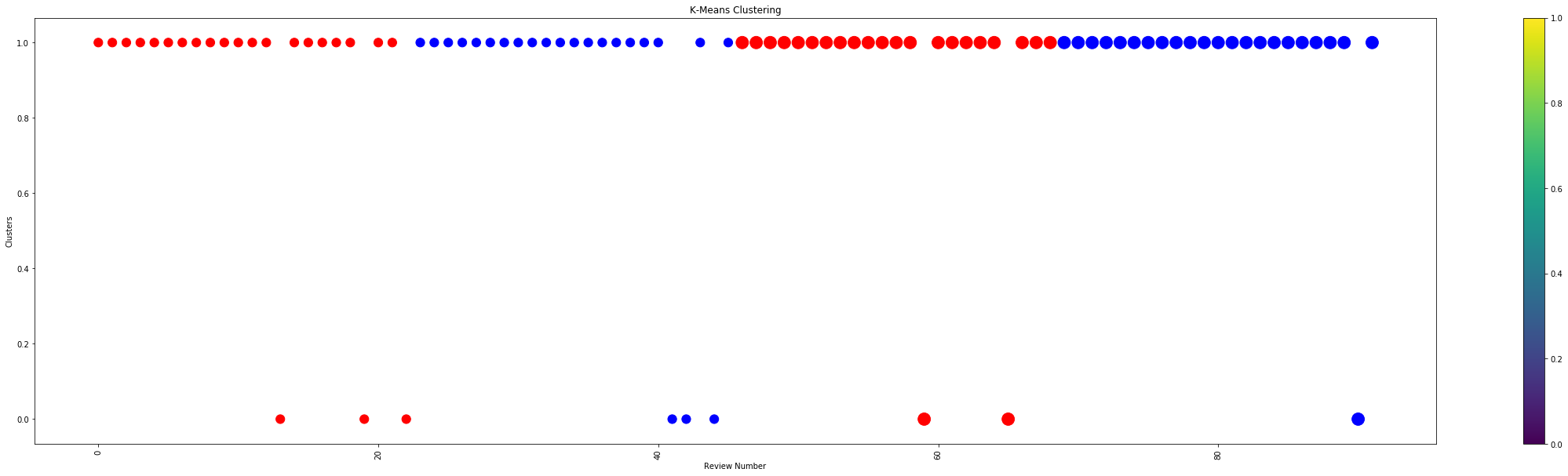
**Figure 1.4 Density curve for each category**

**K-means clusters**

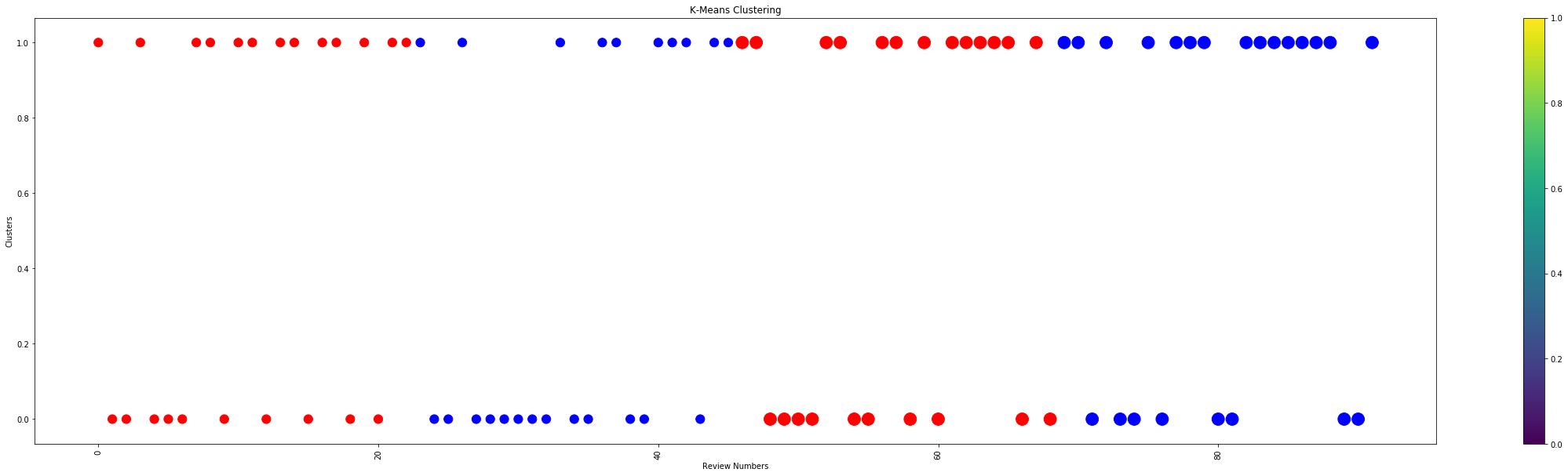
Clustering is performed after vectorization and the observation are shown in the below figures.

1. **Figure 1.4** represents the clusters formed by using word frequency count without normalization
2. **Figure 1.5** represents the clusters formed by using term frequencies
3. **Figure 1.6** represents the clusters formed by using TFIDF

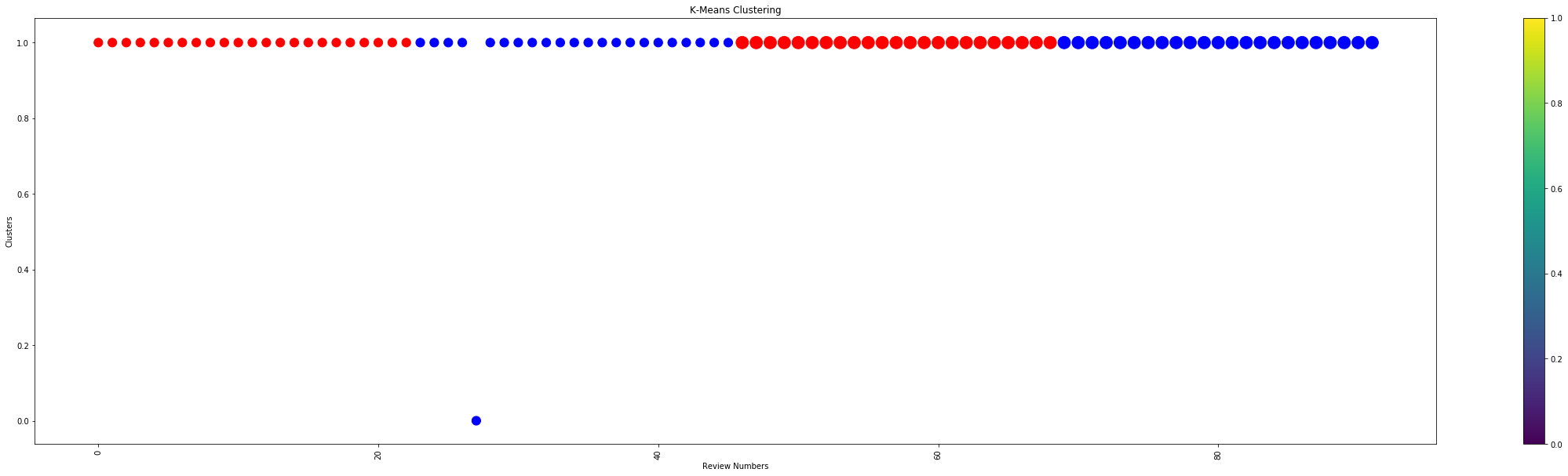
Clusters formed by using frequency count without normalization and TFIDF weightage doesn’t show healthy clusters and the term frequency clusters formed clusters with almost equal number of observations



**Figure 1.4 represents the clusters formed by using word frequency count**



**Figure 1.5 represents the clusters formed by using term frequencies**



**Figure 1.6 represents the clusters formed by using TFIDF**

### **Models**

In this exercise, models are developed using Naïve Bayes, SVM and Kmeans Clustering to compare their efficiency and accuracy in classifying deception and sentiments on a text document.

#### **Naïve Bayes Classification**

It is a classification technique based on Bayes’ Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a specific feature in a class is unrelated to the presence of any other feature. For example, a fruit may be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as ‘Naive’.

Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

**Bayes theorem**

Bayes theorem provides a way of calculating posterior probability P(c|x) from P(c), P(x) and P(x|c). Look at the equation below:



* P(c|x) is the posterior probability of class (c, target) given predictor (x, attributes).
* P(c) is the prior probability of class.
* P(x|c) is the likelihood which is the probability of predictor given class.
* P(x) is the prior probability of predictor.

**Classification based on conditional probability**

To classify whether players will play or not based on weather condition using Naïve Bayes classification approach

Likelihood table Frequency Table are derived by finding the probabilities like Overcast probability = 0.29 and probability of playing is 0.64.

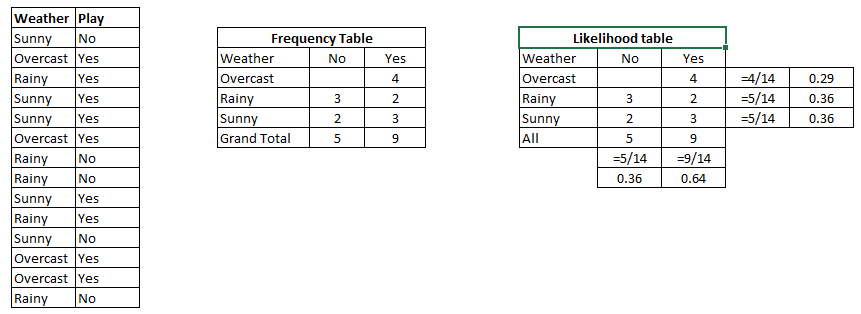
[](https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Bayes_41.png)

Table 2.1

Using Naive Bayesian equation, the posterior probability for each class is calculated. The class with the highest posterior probability is the outcome of prediction.

Say if we want to find out if the Players will play when the weather is sunny?

To solve the above discussed method of posterior probability.

P (Yes | Sunny) = P (Sunny | Yes) \* P(Yes) / P (Sunny)

Here we have P (Sunny |Yes) = 3/9 = 0.33, P(Sunny) = 5/14 = 0.36, P(Yes)= 9/14 = 0.64

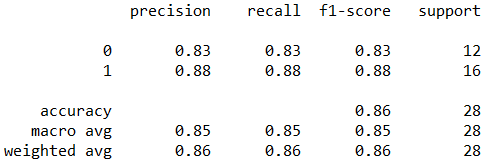
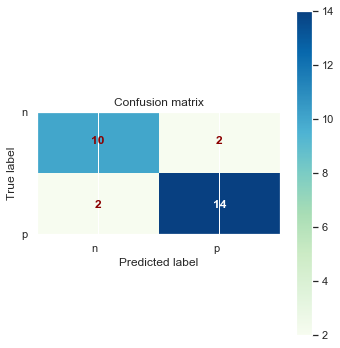
Now, P (Yes | Sunny) = 0.33 \* 0.64 / 0.36 = 0.60, which has higher probability.

Naive Bayes uses a similar method to predict the probability of different class based on various attributes.

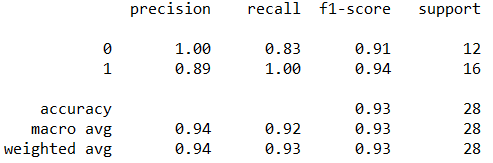
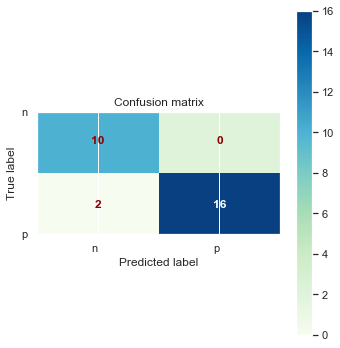
Models are generated for both sentiment prediction and deception detection using all the below vectorization technique and their performance are measured.

**Model 1.1: MultinomialNB and Bernoulli using unigram and Boolean vectorization**

**Figure 2.1 and 2.2** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using unigram and Boolean vectorization in sentiment polarity prediction

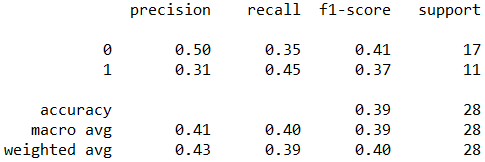
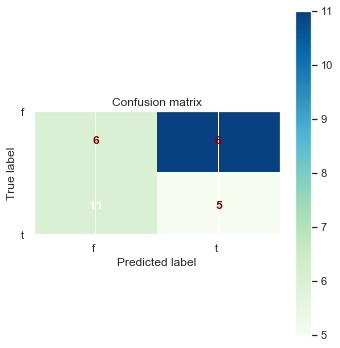
 

**Figure 2.1 MultinomialNB Classification report and Confusion Matrix**

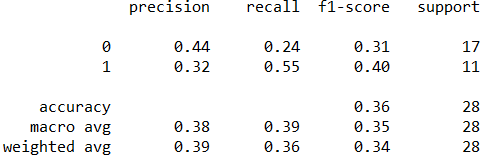
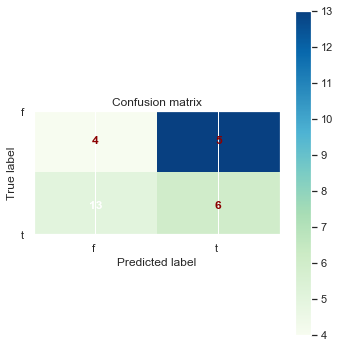
 

**Figure 2.2** **Bernoulli Classification report and Confusion Matrix**

**Figure 2.3 and 2.4** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using unigram and Boolean vectorization in deception prediction

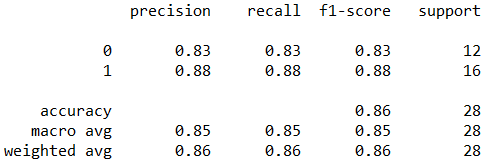
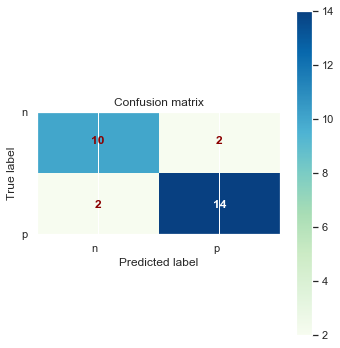
**Figure 2.3 MultinomialNB Classification report and Confusion Matrix**

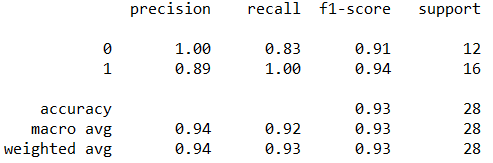
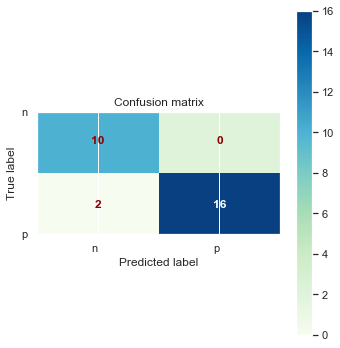
**Figure 2.4** **Bernoulli Classification report and Confusion Matrix**

**Model 1.2: MultinomialNB and Bernoulli using unigram and Frequency vectorization**

**Figure 2.5 and 2.6** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using unigram and frequency vectorization in sentiment polarity prediction

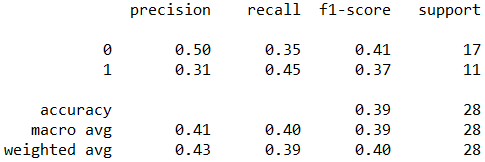
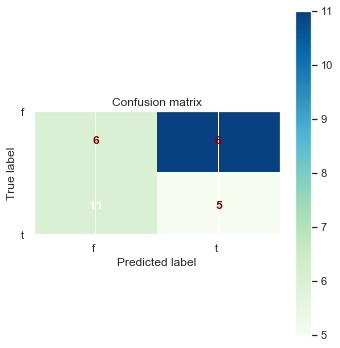
 

**Figure 2.5 MultinomialNB Classification report and Confusion Matrix**

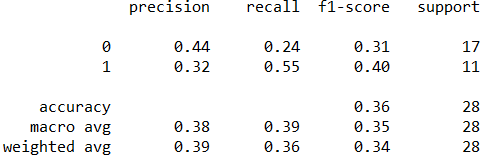
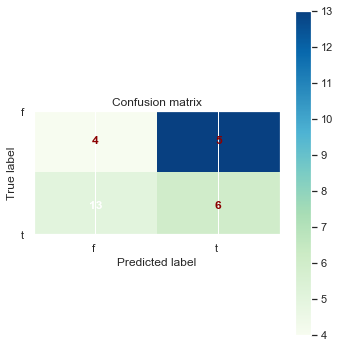
 

**Figure 2.6** **Bernoulli Classification report and Confusion Matrix**

**Figure 2.7 and 2.8** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using unigram and frequency vectorization in deception prediction

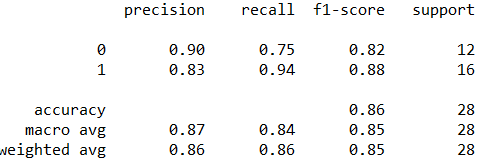
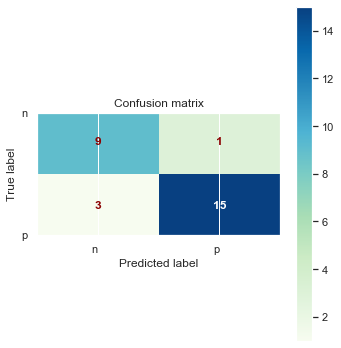
**Figure 2.7 MultinomialNB Classification report and Confusion Matrix**

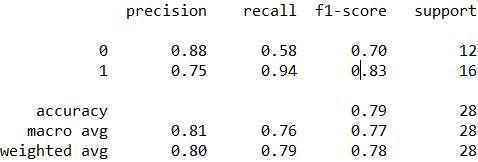
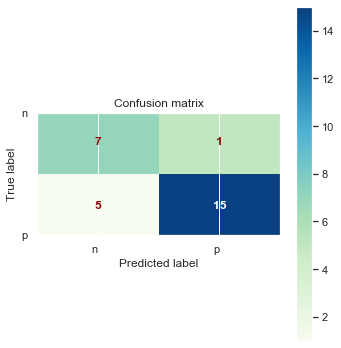
**Figure 2.8** **Bernoulli Classification report and Confusion Matrix**

**Model 1.3: MultinomialNB and Bernoulli using unigram and TF vectorization**

**Figure 2.9 and 2.10** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using unigram and TF vectorization in sentiment polarity prediction

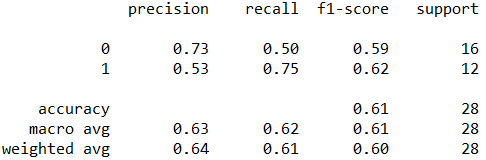
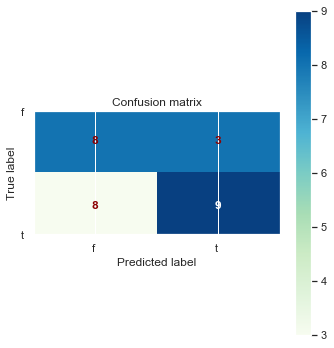
 

**Figure 2.9 MultinomialNB Classification report and Confusion Matrix**

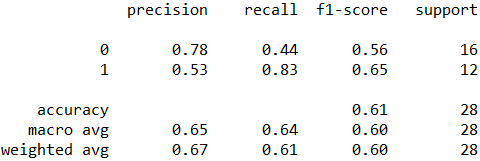
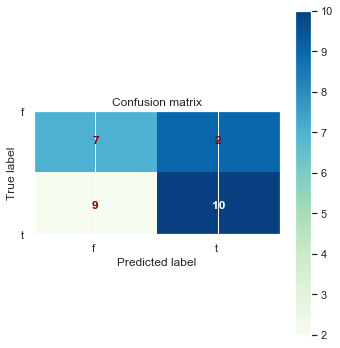
 

**Figure 2.10** **Bernoulli Classification report and Confusion Matrix**

**Figure 2.11 and 2.12** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using unigram and TF vectorization in deception prediction

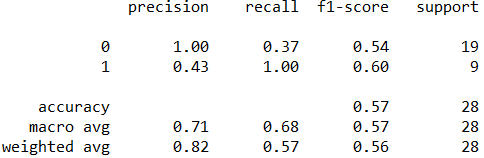
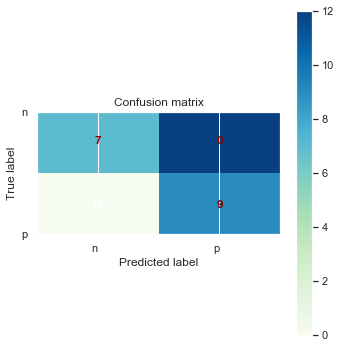
**Figure 2.11 MultinomialNB Classification report and Confusion Matrix**

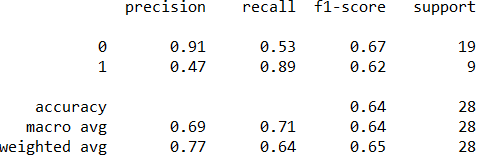
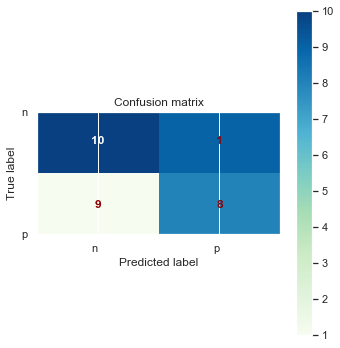
**Figure 2.12** **Bernoulli Classification report and Confusion Matrix**

**Model 1.4: MultinomialNB and Bernoulli using unigram and TFIDF vectorization**

**Figure 2.13 and 2.14** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using unigram and TFIDF vectorization in sentiment polarity prediction

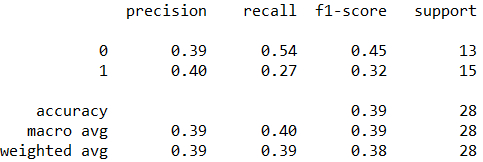
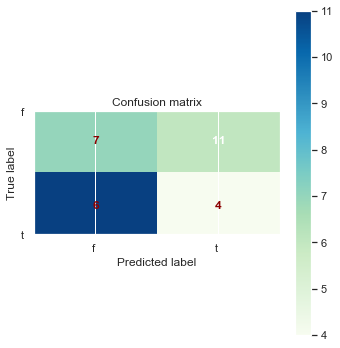
 

**Figure 2.13 MultinomialNB Classification report and Confusion Matrix**

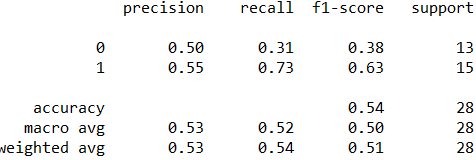
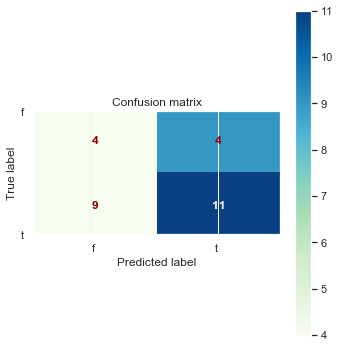
 

**Figure 2.14** **Bernoulli Classification report and Confusion Matrix**

**Figure 2.15 and 2.16** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using unigram and TFIDF vectorization in deception prediction

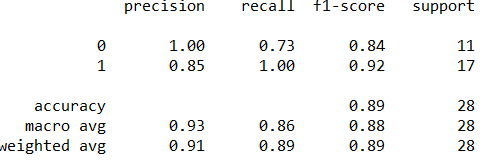
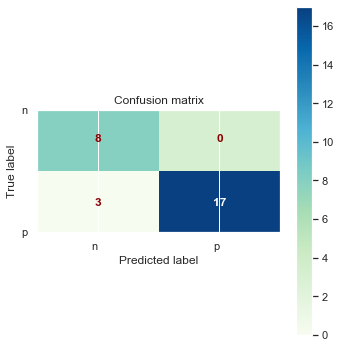
**Figure 2.15 MultinomialNB Classification report and Confusion Matrix**

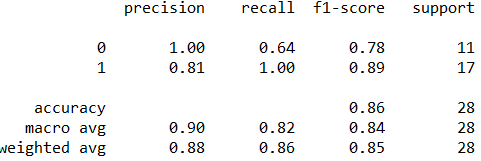
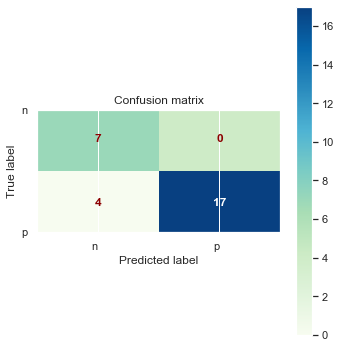
**Figure 2.16** **Bernoulli Classification report and Confusion Matrix**

**Model 1.5: MultinomialNB and Bernoulli using gram(1,2) and Boolean vectorization**

**Figure 2.17 and 2.18** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using gram(1,2) and Boolean vectorization in sentiment polarity prediction

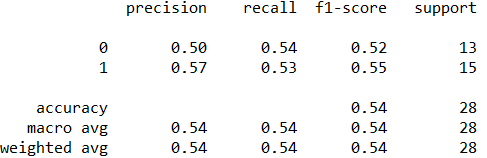
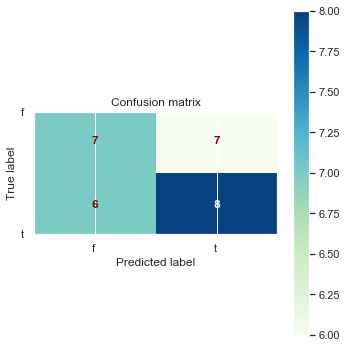
 

**Figure 2.17 MultinomialNB Classification report and Confusion Matrix**

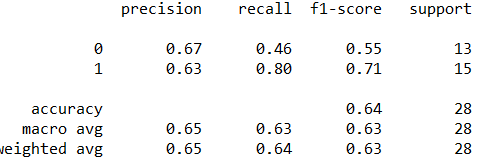
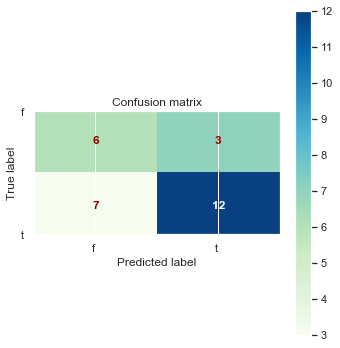
 

**Figure 2.18** **Bernoulli Classification report and Confusion Matrix**

**Figure 2.19 and 2.20** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using gram(1,2) and Boolean vectorization in deception prediction

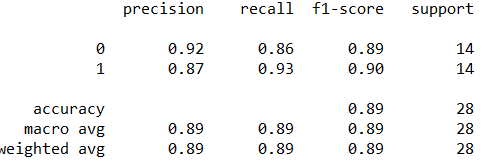
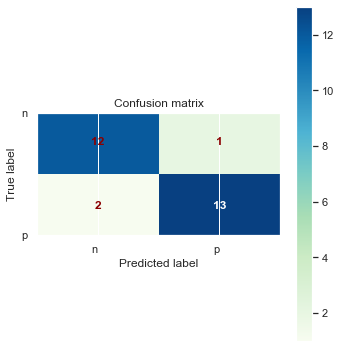
**Figure 2.19 MultinomialNB Classification report and Confusion Matrix**

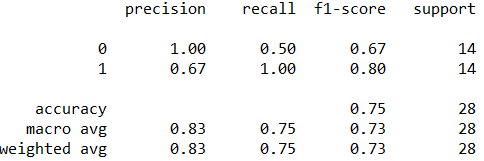
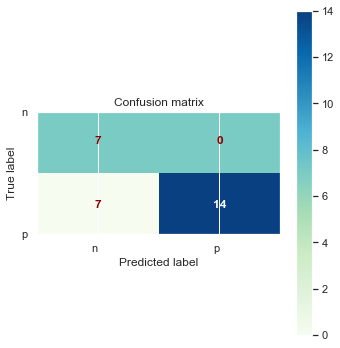
**Figure 2.20** **Bernoulli Classification report and Confusion Matrix**

**Model 1.6: MultinomialNB and Bernoulli using gram(1,2) and Frequency vectorization**

**Figure 2.21 and 2.22** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using gram(1,2) and frequency vectorization in sentiment polarity prediction

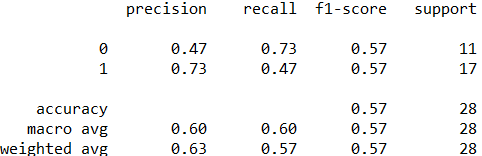
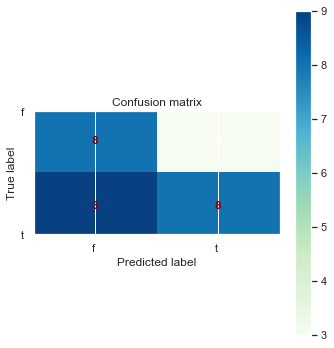
 

**Figure 2.21 MultinomialNB Classification report and Confusion Matrix**

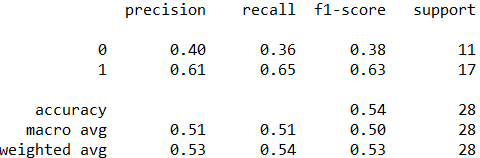
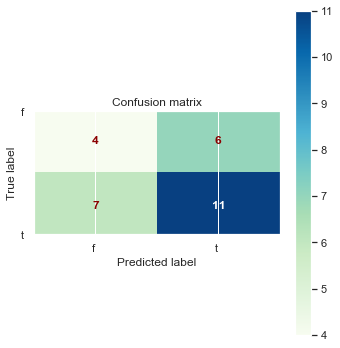
 

**Figure 2.22** **Bernoulli Classification report and Confusion Matrix**

**Figure 2.23 and 2.24** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using gram(1,2) and frequency vectorization in deception prediction

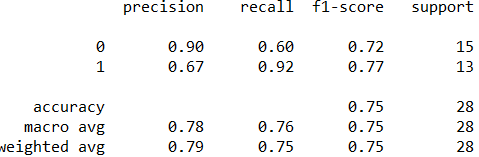
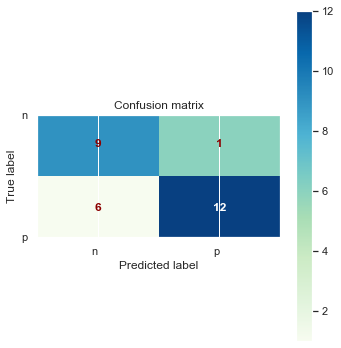
**Figure 2.23 MultinomialNB Classification report and Confusion Matrix**

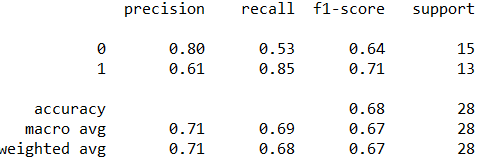
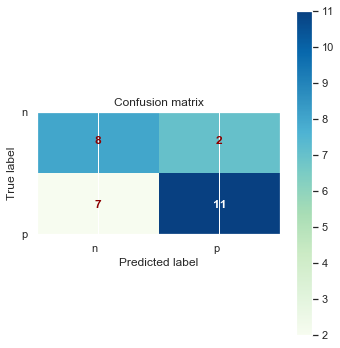
**Figure 2.24** **Bernoulli Classification report and Confusion Matrix**

**Model 1.7: MultinomialNB and Bernoulli using gram(1,2) and TF vectorization**

**Figure 2.25 and 2.26** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using unigram and TF vectorization in sentiment polarity prediction

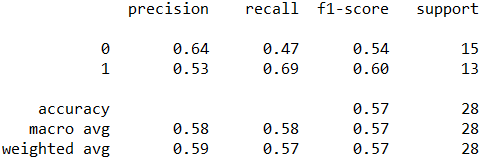
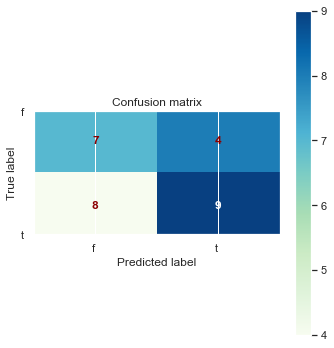
 

**Figure 2.25 MultinomialNB Classification report and Confusion Matrix**

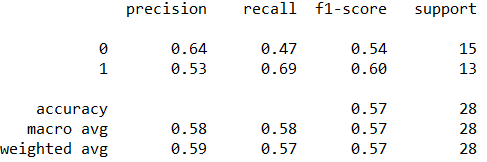
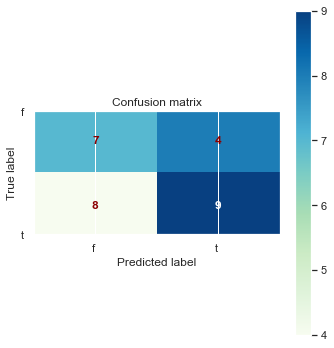
 

**Figure 2.26** **Bernoulli Classification report and Confusion Matrix**

**Figure 2.27 and 2.28** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using gram(1,2) and TF vectorization in deception prediction

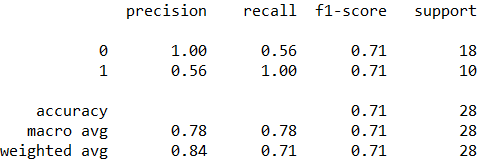
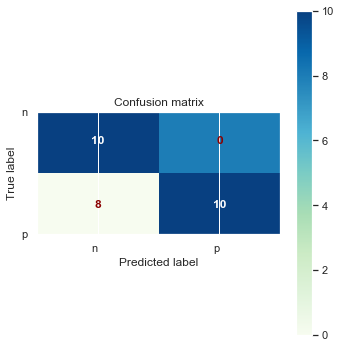
**Figure 2.27 MultinomialNB Classification report and Confusion Matrix**

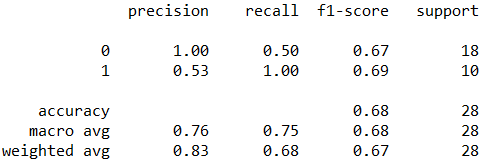
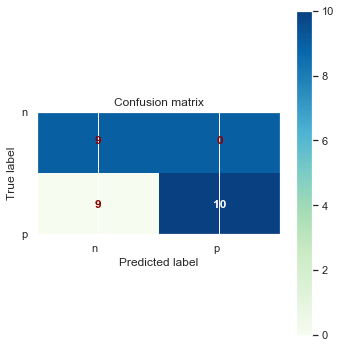
**Figure 2.28** **Bernoulli Classification report and Confusion Matrix**

**Model 1.8: MultinomialNB and Bernoulli using gram(1,2) and TFIDF vectorization**

**Figure 2.29 and 2.30** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using gram(1,2) and TFIDF vectorization in sentiment polarity prediction

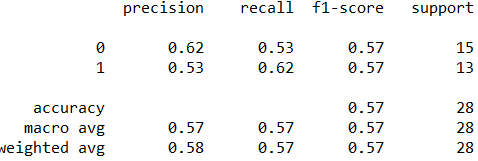
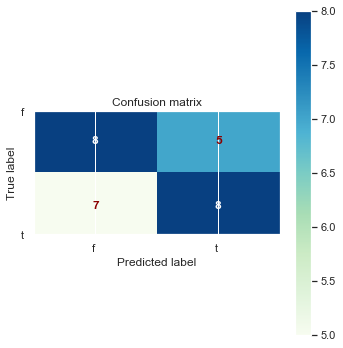
 

**Figure 2.29 MultinomialNB Classification report and Confusion Matrix**

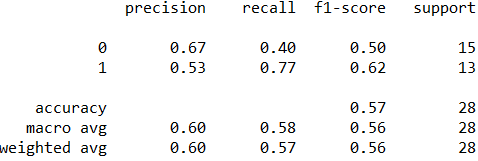
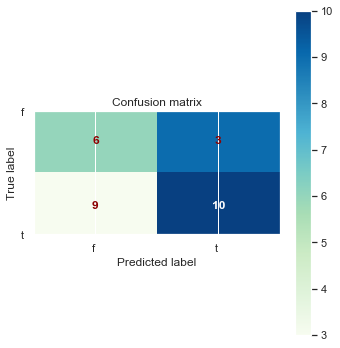
 

**Figure 2.30** **Bernoulli Classification report and Confusion Matrix**

**Figure 2.31 and 2.32** Shows the confusion matrix and the classification report of the Naïve Bayes MultinomialNB and Bernoulli algorithm using gram(1,2) and TFIDF vectorization in deception prediction

**Figure 2.31 MultinomialNB Classification report and Confusion Matrix**

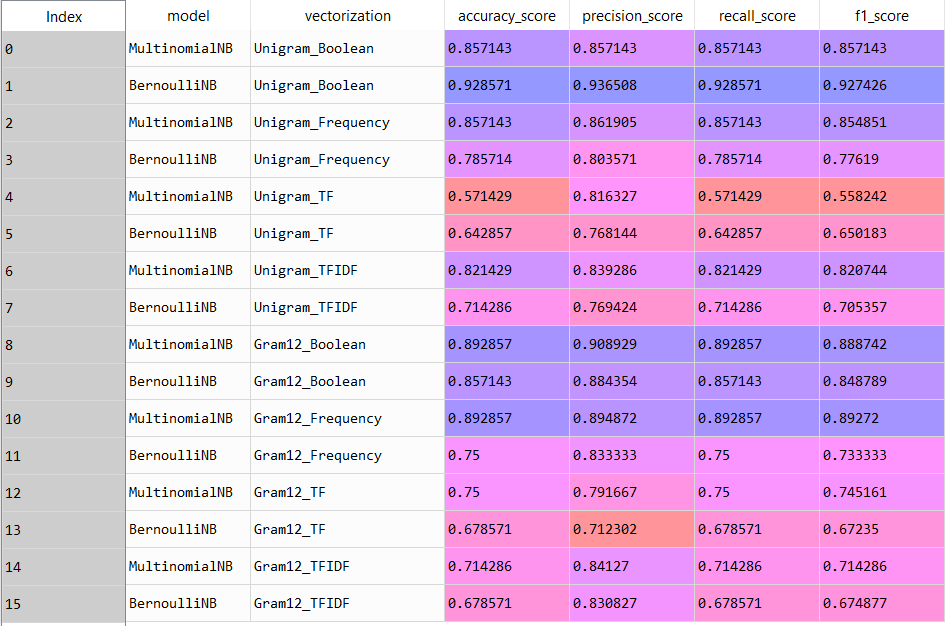
**Figure 2.32** **Bernoulli Classification report and Confusion Matrix**

## **Results**

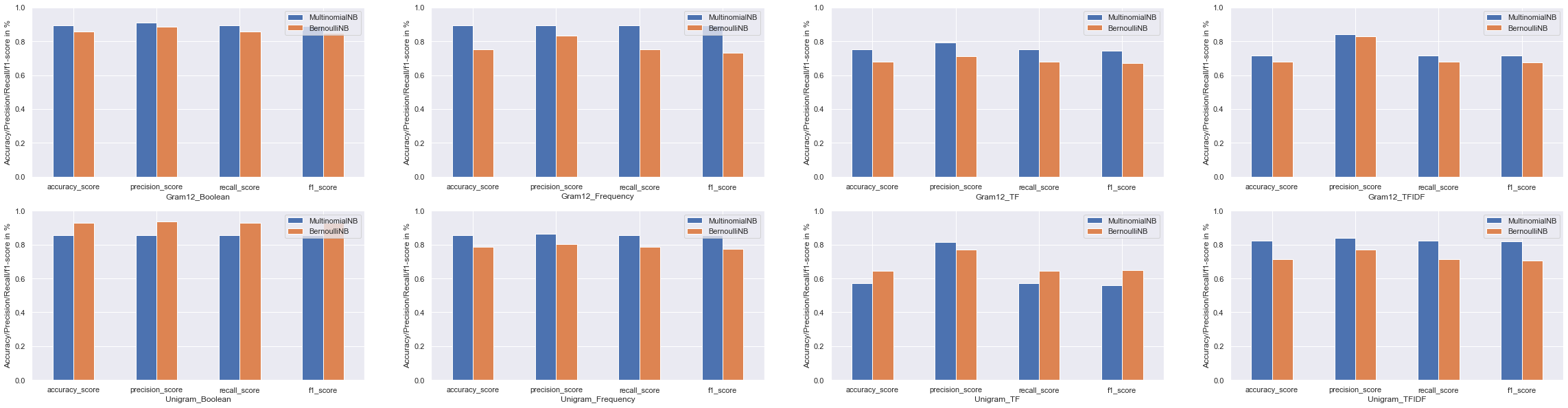
Accuracy, precision and recall using MultinomialNB and Bernoulli are tabulated in **Table 3.1 and Table 3.2** for sentiment and Lie detection respectively

**Sentiment polarity prediction**

Based on the experimental results, it appears the model generated using Boolean vectorization is more accurate when compared to another vectorization technique. Also, it is observed that Multinomial algorithm performing better than Bernoulli for all vectorization method except for Boolean on unigram vectors. The highest accuracy of 92% in predicting sentiment polarity is obtained using Bernoulli and unigram Boolean vectorization (**Figure 3.1**).



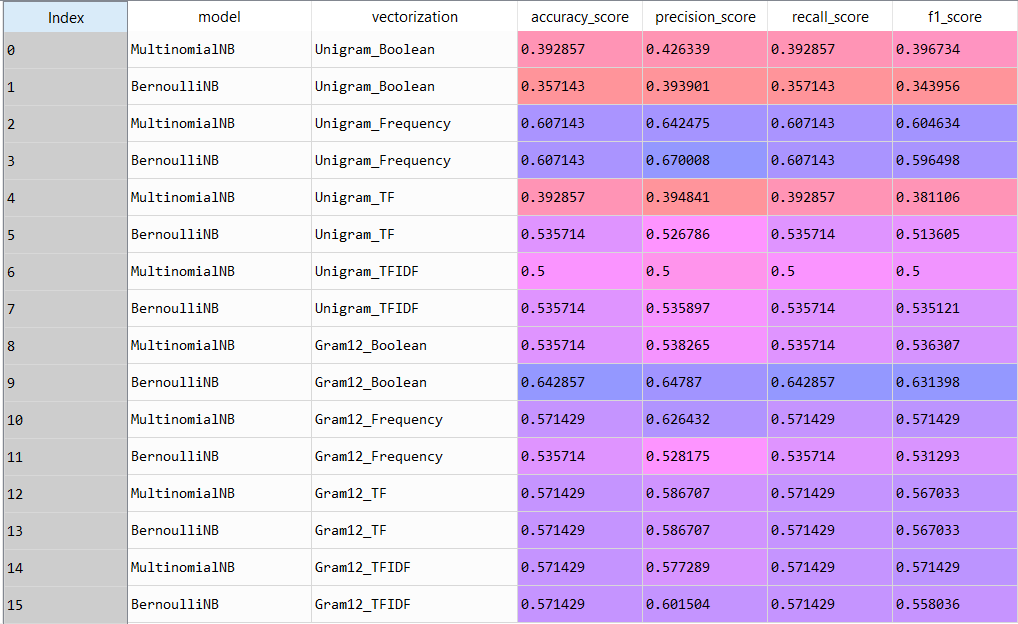
**Table 3.1 Model performance comparison for sentiment analysis**



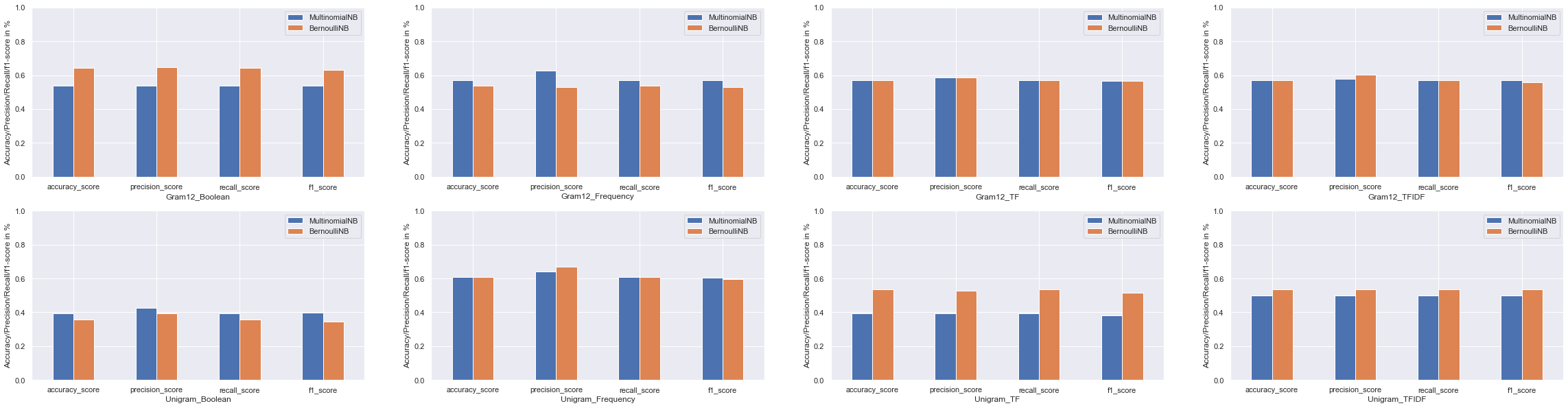
**Figure 3.1 Model performance comparison for sentiment analysis**

**Deception detection**

Based on the experimental results, it appears the model generated using Boolean and gram (1,2) vectorization is more accurate when compared to another vectorization technique. Also, it is observed that Multinomial algorithm and Bernoulli showing mixed results where both doing good in different vectorization techniques. The highest accuracy of 64% in detecting deception is obtained using Bernoulli and gram (1,2) Boolean vectorization (**Figure 3.2**).



**Table 3.2 Model performance comparison for deception detection**



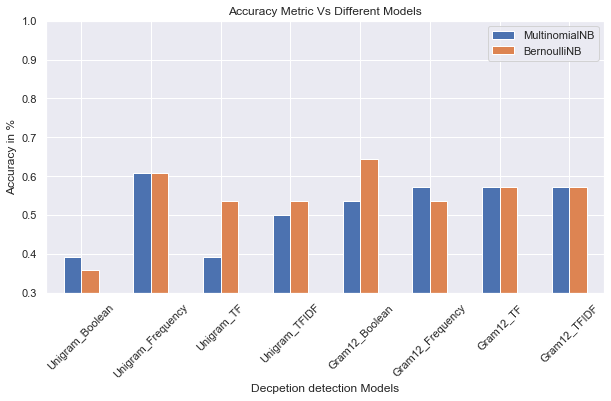
**Figure 3.2 Model performance comparison for deception detection**

## **Conclusion**

Dealing with humans is one of the most difficult and important challenges for computing, and automatic deception detection using computers is one of the more challenging tasks among human computer interactions. Deception is a common human behavior, and deception detection has been a huge interest during human history. If we can solve the deception detection task or even improve tasks performance, then it will help in many fields, such as business, jurisprudence, law enforcement, and national security.

<https://pdfs.semanticscholar.org/f42b/6376836a0669d5b28a8e3ef8bc906304b900.pdf>

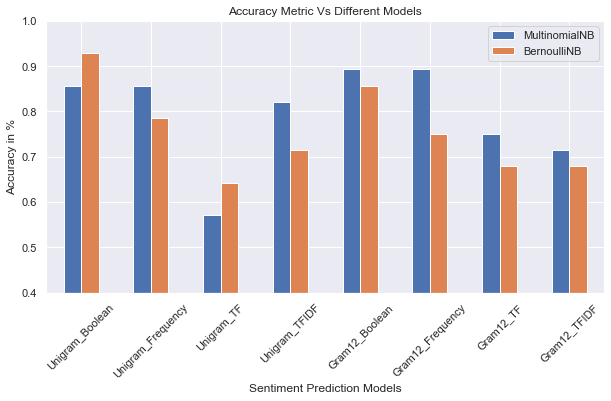
Current approaches for detecting deception build upon the experiment of computer scientist based on the psychological theory of deception, and there are some findings and successful detections of deception. For the feature extraction, a variety of techniques were introduced for extracting lexical features, acoustic and prosodic features. For classification, various kinds of classifiers were used for detecting deception. There have been some successful approaches in recent years for deception detection, such as emotional words type, filled pause features, and acoustic and prosodic features. However, by comparing the performance of different approaches, we can find that even the best performing approach is still not quite accurate on detecting deception of speech.(**Figure 4.1**) It’s not only because detecting detection is difficult, but also we think there are quite number of cues that nobody has tried or are less adopted, such as emotion, personality and cultural difference.



**Figure 4.1 Experimental Results on Deception Detection**

Feedbacks are generated by writer on websites, social networks, blogs, online portals, reviews, opinions, recommendations, ratings. This writer generated sentiment content can be about books, people, hotels, products, research, events, etc. These sentiments become very beneficial for businesses, governments, and individuals. A bulk of this writer generated content require using the text mining techniques and sentiment analysis.

Although, the sentiment polarity prediction results are more accurate (**Figure 4.2**) than deception detection, there are several challenges faced the sentiment analysis and evaluation process. These challenges become obstacles in analyzing the accurate meaning of sentiments and detecting the suitable sentiment polarity.

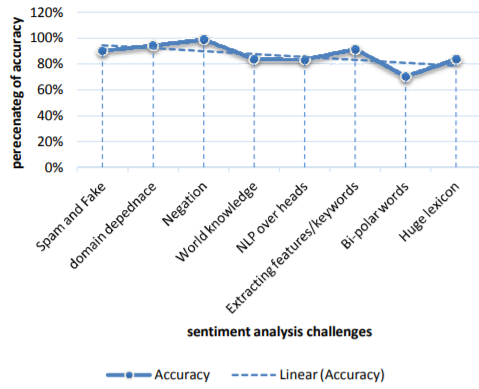


**Figure 4.2 Experimental Results on Sentiment Polarity Detection**

An article published in Journal of King Saud University · April 2016 by Doaa Mohey El-Din shows the improvement in accuracy for each sentiment analysis challenges related to the second comparison (**Figure 4.3**).

Negation has the highest accuracy percentage that can support the result of the first comparison because there is no research in sentiment don't need to understand the negative reviews whether explicit or implicit. And the least score in accuracy is bi-polar words research.

<https://www.researchgate.net/publication/301649355_A_Survey_on_Sentiment_Analysis_Challenges>



**Figure 4.3 Highest accuracy to each sentiment analysis challenge**