Automatic Classification of Film Reviews on Social Media through Opinion Mining

Migena Ceyhan[1]

Zeynep Orhan[2]

Dimitrios Karras[3]

Summary

Digital life has significantly changed human relations. Ideas, emotions and opinions on many topics are shared in virtual environments. If these can be used in different areas, important and useful applications can be developed. This idea is also supported by the increasing number of applications.

The aim of this study is to train a system with the characteristic features of positive and negative movie comments made by users on social media through sentiment and opinion analysis. This system will be able then to automatically classify new comments accordingly, to examine the results in seconds, and to deliver them in an understandable summary format quickly and easily to the computer user, by saving time and manpower.

Keywords: Social media, movie comments, opinion mining, automatic classification

1.Introduction

1.1 Opinion Mining and Sentiment Analysis

Sentiment analysis (SA) is the interpretation of a document as positive, negative or neutral [1]. With the development of internet in recent years, it has become possible to find opinions everywhere. Social networks such as blogs, Facebook, Twitter, news portals and e-commerce sites, have facilitated this situation. Although so many sources of information might be useful, considering their abundance and quantity, these sources become a difficult-to-digest information overload. Instead, it would be useful to provide a more understandable summary, as it can be seen in Figure 1 and Figure 2 [2].

---

[1] Department of Mathematics and Informatics, University of Shkodra "Luigj Gurakuqi", Shkoder, Albania. E-mail: migena.ceyhan@unishk.edu.al

[2] Computer Science Department, Union College, Schenectady, New York, USA

[3] Computer Engineering Department, Epoka University Tirana, Albania

Today, after experiencing a product such as a film, a piece of music or a technological device, people share their opinions about the product and criticise the parts they like or dislike about it on the internet through social media networks, blogs or forums. By examining these reviews, big companies, in particular, identify the positive and negative aspects of the product they have developed and enhance themselves. In addition, consumers get a general idea before using the product by looking at these reviews. However, the review process takes a long time and is very laborious [3].
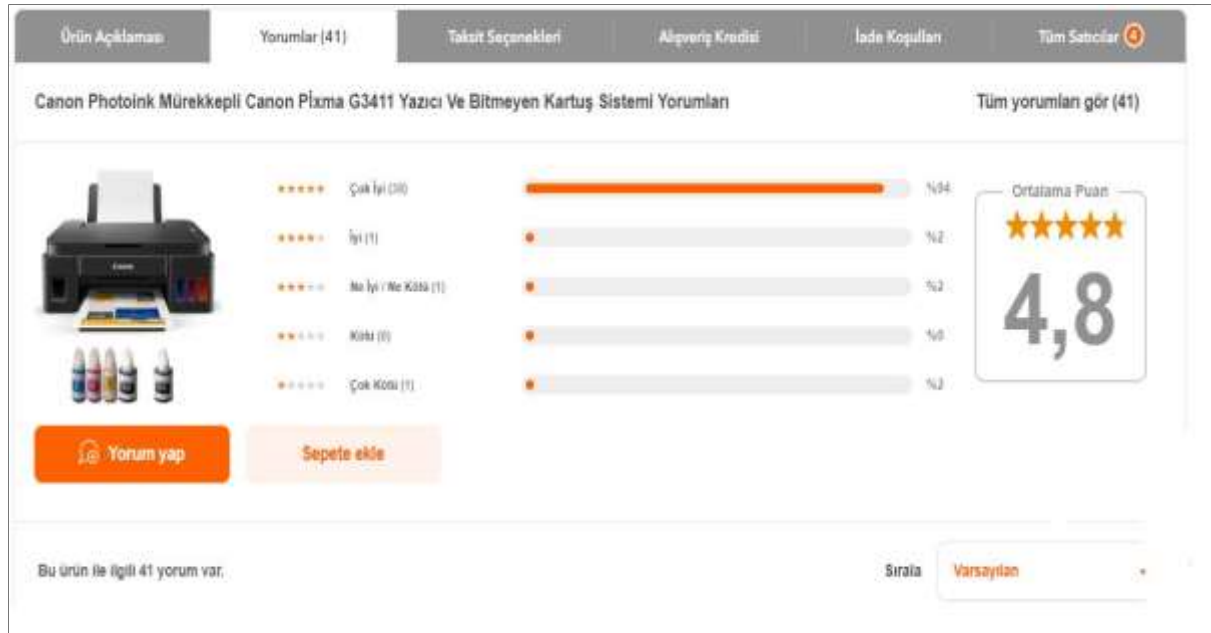

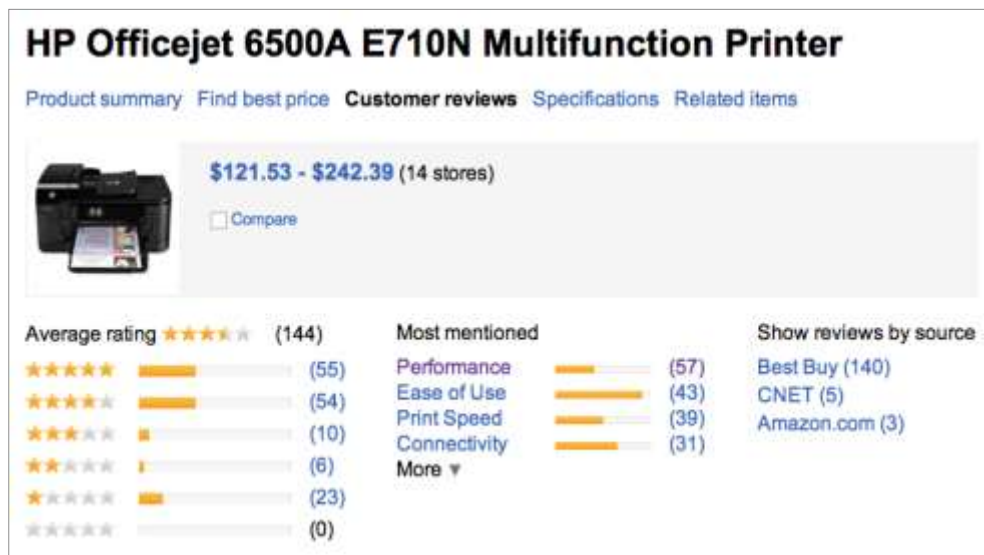Figure 1: Hepsiburada (AllHere) product reviews
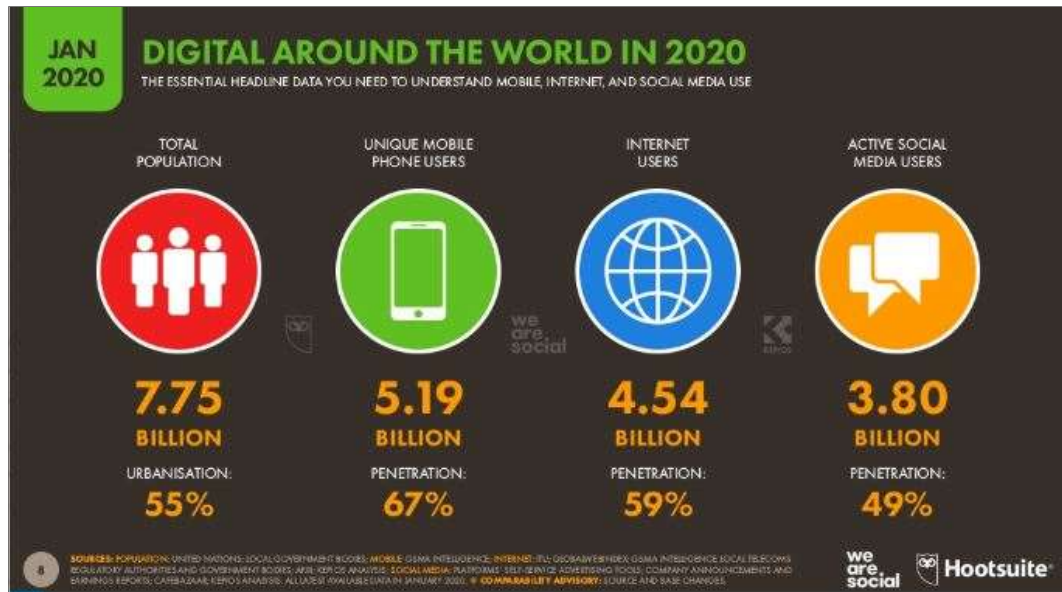

Figure 2: Bing product reviews

Figure 3: Worldwide internet, social media and mobile usage statistics

As it can be seen in Figure 3, the number of people using the digital world has reached billions [3]. The rapid increase in the amount of data on the internet and the impact of internet usage on trade also change the relationship between the producer and the consumer. Now, both producers and consumers share their opinions and experiences on social media. These data are the first source of reference for a subject, product or brand. Considering only Twitter, millions of tweets are posted and new accounts are opened daily. Nonetheless, these incredible numbers exceed the users' tracking capacity, thus, increasing the need for auxiliary tools day by day. At this point, sentiment analysis applications emerge, which interpret and classify information automatically [4].

1.2 The use of Sentiment Analysis

Sentiment analysis is used in many areas [1]. As one of them, the film industry seeks answers to questions such as: Are comments positive or negative; What will the revenue be; and Can the continuation of the film be shot, and, then takes decisions accordingly.

Statistics provide a decent answer to why entrepreneurs and businessmen should include search operations and social networks in their marketing plans and why it is impossible for them to remain in the market without them. Below are some striking data on the topic:

- People who use social networks and social media regularly make up 84% of internet users and this rate is growing rapidly.
- Google (by far) is the most popular search engine. By 2020, Google accounts for over 79% of the total desktop search traffic.
- Google performs more than 5.5 billion searches every day.

- 97% of consumers search for a company online.
- Positive reviews increase the confidence of youth between the ages of 18 and 34 to by 91%.
- 72% of users employing Google search visited a store within eight km from their search location.
- 88% of customers are known to call or visit a store within 24 hours after searching from a mobile device. [5][6]

SA helps to determine people's emotions and opinions about events, services, products, institutions, and other people and so on. While positive feedback encourages the company, negative feedback assumes undertaking deterrent steps. Sometimes firms get help from SA when they want to compare themselves to other firms. Apart from that, it is an important area for marketing managers, politicians, online product managers, entrepreneurs and anyone who needs consumer reviews [6].

1.3 Basic Information on Sentiment Analysis

SA also gets support from many different areas. Natural language processing, text analysis, data mining and artificial intelligence are some of them. The number of studies on the subject has especially increased in recent years. However, studies conducted in Turkish are not so many.

SA can be dealt with at different levels. The s*imple level c*lassifies all available data as a whole only as positive or negative, but also complex and advanced levels are also options [1]. SA may also differ when it comes to handling the data it uses. The simplest level of comparison is document-level. It works on a document A document corresponds to an opinion. Other forms of SA are sentence-level and aspect-level based [4].

Machine learning methods are used for SA. In short, machine learning is the study of providing a computer with a human feature by means of software and it is a sub discipline of artificial intelligence. Its most important categories are supervised, unsupervised and semi-supervised learning. Supervised learning is a difficult and expensive human supported process because it needs labelled data. For an example given in the classification process, it is based on the method of finding the class to which it belongs based on the training criteria set [8].

In this study supervised learning document-level simple classification was used on previously labelled data.

2. Method

2.1 Algorithm

For SA, the algorithm steps are the method applied are as following:

133

- Collect data.
- Extract important and decisive words and phrases used in positive and negative reviews in the data.
- Ensure that data are corrected and audited.
- Convert data into text.
- Perform morphological analysis of data.
- Reduce to one the number of multiple results from morphological analysis.
- Process the texts and extract the features planned to be used.
- Determine assessment methods.
- Type the Naïve Bayes method and try it with the specified features.
- Prepare .csv extension files of above features to use the Naïve Bayes method in Weka program and test Naïve Bayes.
- Use a new method to make a classification with the help of a vocabulary derived from important words or phrases that are important only in positive or negative comments.
- Compare the results in the program and those in the Weka.

Details of steps mentioned in our algorithm are as below:

2.2 Data collection

The data were collected manually. The data were obtained from the www.sinemalar.com and www.beyazperde.com sites. These sites were chosen because they are popular in Turkey and contain enough comments. While for positive data movies with high IMDb (Internet Movie Database) (www.imdb.com) were used, negative data were obtained from low-score movies. 919 positive and 615 negative comments were collected as training data. A total of 50 comments was collected for the test data, with 25 positive and 25 negative comments.

Data information is shown in Table 1 and in Table 2, the first collection of the data is given in an excel line. The comments along with the title of the film, the correction of some spelling mistakes in these comments as well as the words that have a positive or negative effect along with the comment's class information were maintained.

Table 1: Training and Test data information

Comment Type	Total	Training	Test
Positive	919	894	25
Negative	615	590	25

Table 2: Examples from positive and negative training data

Film title	Comment	Corrected comment	Important parts of comment	Positive (1) / Negative (2)
Inception (Başlangıç)	Geleceğin **kült** filmlerinden biridir. **Muhakkak izlemek lazım**, zihinsel arşivde bulundurmak lazım. Oyunculuklar, senaryo **harika**ydı. İlla ki eksik bir şey söylemek gerekirse son sahnelerin fazla uzatıldığını söyleyebilirim, zaten 148 dakika bir film bence (istisnai durumlar dışında) çoktur**. Başarılı** bir film. **Mutlaka** izleyin derim...	One of the **cult** films of the future. It is a **definitely must watch** and should be kept in mental archive. The acting, the script was **wonderful**. If I need to say something negative, I can say the final scenes were too long, I think 148 minutes (apart from exceptional cases) is too much for a film. A **successful** film. You should **absolutely** watch it...	kült (cult), muhakkak izlemek lazım (definitely must watch), harika (wonderful), başarılı (successful), mutlaka (absolutely)	1
Lucy	Cidden çok **değişik** bir film. Ve bence biraz da **saçma** olmuş.	Seriously, it is a very **different** movie. Plus, a little **absurd**, I think.	değişik (different), saçma (absurd)	2

2.3 Data Preprocessing

Important and decisive words and phrases used in the comments were determined by two markers. Later, they were cross checked. Some spelling mistakes were also corrected, but as the number of data was too high and the social media texts contained too many typos, they were not completely fixed. Some of the characters included in the data and some of the unnecessary information were also discarded.

2.4 Morphological Analysis

The data were transcribed. As word roots were to be used, morphological analysis was necessary. Special software was used for this. The data were transferred to the appropriate format for morphological analysis by means of C program.

Table 3: Morphological analysis input and output formats

Morphological analysis input	Morphological analysis output		
\<DOC\>	\<DOC\>	\<DOC\>	
\<TITLE\>	\<TITLE\>	\<TITLE\>	
\<S\>	\<S\>	\<S\>	
xwq**Lucy**xwq	xwq**Lucy**xwq	xwq**Lucy**xwq	*UNKNOWN*
\</S\>	\</S\>	\</S\>	
\</TITLE\>	\</TITLE\>	\</TITLE\>	
\<S\>	\<S\>	\<S\>	
Cidden	Cidden (really)	cidden	+Adverb
çok	çok (many)	çok	+Det
	çok (very)	çok	+Adverb
	çok (a lot of)	çok	+Adj
	çok (ample)	çok	+Postp+PCAbl
değişik	değişik (different)	değişik	+Adj
bir	bir (a)	bir	+Det
	bir (only)	bir	+Adverb
	bir (another)	bir	+Adj
	bir (one)	bir	+Num+Card
film	film (movie)	film	+Noun+A3sg+Pnon+Nom
.	.	+Punc	
\</S\>	\</S\>	\</S\>	
\<S\>	\<S\>	\<S\>	
Ve	Ve (and)	ve	+Conj
bence	bence (mine)	ben	+Noun+A3sg+Pnon+Equ
	bence (as for me)	ben	+Pron+Pers+A1sg+Pnon+Equ
biraz	biraz (a little)	biraz	+Adverb
	biraz (a bit)	biraz	+Adj
da	da (and)	da	+Conj
saçma	saçma (buckshot)	saçma	+Noun+A3sg+Pnon+Nom
	saçma (nonsense)	saçma	+Adj
	saçma (don't sparkle)	saç	+Verb+Neg+Imp+A2sg

olmuş	saçma (irradiate)	saç	+Verb+Pos^DB+Noun+Inf2+A3sg +Pnon+Nom
	olmuş (has became)	ol	+Verb+Pos+Narr+A3sg
	olmuş (done)	ol	+Verb+Pos+Narr^DB+Adj+Zero
</S>			
</DOC>	</S>	</S>	
	</DOC>	</DOC>	

For the morphological analysis, the MORPHOLOGYCAL TAGGER application which includes the Kemal Oflazer's finite state machine prepared for Turkish, which can be accessed form Deniz Yüret's blog, was used. The application can be accessed at http://deniz.yuret.com/turkish/tr-tagger.tgz. Before running the TAGGER, the machine must have Xerox Finite State software. This software can be downloaded from http://www.stanford.edu/~laurik... book software site (with copyright permission). After installing the software, the text must be prepared in a specific format. The sample prepared for this format is available as sample-input.txt in TAGGER input. A sample morphological analysis input and output format for the first line of negative data has been given in Table 3.

2.5 Morphological Disambiguation

Since more than one conclusion could be obtained from morphological analysis, it was necessary to eliminate them. A program was used for this. As with addition of endings many words can be derived from a Turkish word, TAGGER creates all possibilities. Deniz Yüret's MORPHOLOGYCAL DISAMBIGUATOR is an application used to find the result by removing uncertainties within the possibilities. It can be downloaded from http://deniz.yuret.com/turkish/tr-disamb.tgz. The above output of TAGGER is given as input to the DISAMBIGUATOR. For this to work, the perl program needs to be installed. The line produced as output is given in the last column of Table 4.

Table 4: Morphological and disambiguity analysis outputs

Morphological Analysis Output – Disambiguity Analysis Input	Disambiguity Analysis Output
<DOC> <DOC> <TITLE> <TITLE> <S> <S> xwq**Lucy**xwq xwq**Lucy**xwq *UNKNOWN* </S> </S> </TITLE> </TITLE> <S> <S> **Cidden (really) cidden +Adverb** çok (many) çok +Det	<DOC> <DOC> <TITLE> <TITLE> <S> <S> xwq**Lucy**xwqxwq**Lucy**xwq*UNKNOWN* </S> </S> </TITLE> </TITLE> <S> <S> Cidden cidden +Adverb

çok (very)	**çok**	**+Adverb**
çok (a lot of)	çok	+Adj
çok (ample)	çok	+Postp+PCAbl

çok   çok   +Adverb

değişik (different)	**değişik**	**+Adj**

değişik değişik+Adj

bir (a)	**bir**	**+Det**
bir (only)	bir	+Adverb
bir (another)	bir	+Adj
bir (one)	bir	+Num+Card

bir   bir   +Det

film (movie)	**film**	**+Noun+A3sg+Pnon+Nom+Punc**

film   film   +Noun+A3sg+Pnon+Nom +Punc

</S>	</S>
<S>	<S>

</S>   </S>
<S>   <S>

Ve (and)	**ve**	**+Conj**

Ve   ve   +Conj

bence (mine)	ben	+Noun+A3sg+Pnon+Equ
bence (as for me)	**ben**	**+Pron+Pers+A1sg+Pnon+Equ**

bence ben +Pron+Pers+A1sg+Pnon+Equ

biraz (a little)	biraz	+Adverb
biraz (a bit)	**biraz**	**+Adj**

biraz biraz +Adj

da (and)	**da**	**+Conj**

da   da   +Conj

saçma (buckshot)	**saçma +Noun+A3sg+Pnon+Nom**
saçma (nonsense)	saçma +Adj
saçma (don't sparkle)	saç +Verb+Neg+Imp+A2sg
saçma (irradiate)	saç +Verb+Pos^DB+Noun+Inf2+A3sg+Pnon+Nom

saçma saçma +Noun+A3sg+Pnon+Nom

olmuş (has come out)	**ol +Verb+Pos+Narr+A3sg**
olmuş (done)	ol +Verb+Pos+Narr^DB+Adj+Zero

olmuş ol   +Verb+Pos+Narr+A3sg

</S>	</S>
</DOC>	</DOC>

</S>   </S>
</DOC> </DOC>

2.6 Features Extraction

The texts were processed and the features planned to be used were extracted. The used features are described below:

Feature 1. Single word root and Positive/Negative aspect of All words:

First of all, the words were used one by one and independently of each other with their roots and positive status. Positivity here means that if the word is used in its original meaning, the pos attachment is added, if it used with a negative inflection, the neg attachment is added. For example, the root of the word *severim* is sev and because it is in its original meaning, the ***sevPos*** feature will come out here. However, if the word were *sevmem* the feature would be ***sevNeg***.

Feature 2. Single word root, Positive/Negative aspect and Modes of All words:

Here, too, all words were used one by one and independently of each other together with their roots, positive status and some of the mode suffixes. The modes of imperative, necessity and wish were examined. For example, as *sev* is the root of the word, its original meaning and the imperative mode, the feature extracted here is ***sevPosImp***. If the word were *sevmemeliyim* the feature to come out would be ***sevNegNeces***.

Feature 3. Binary word root and Positive/Negative aspect of All words:

It is similar to the first feature. The only difference is that it considers binary phrases instead of words one by one. For example, the sentence *bu filmi sevdim* will result in two features ***buPosFilmPos*** and ***filmPosSevPos***.

Feature 4. Binary word root, Positive/Negative aspect and Modes of All words:

It is similar to the second feature. The only difference is that it looks at binary phrases instead of words one by one. For example, the sentence *bu filmi seyredin* will result in the features ***buPosNoModeFilmPosNoMode*** and ***filmPosNoModeSeyretPosImp***. However, if the word had been *sevmem* the feature would be ***sevNeg***.

Feature 5. Single word root and Positive/Negative formation of Important words:

It is obtained by using the first feature for important words only.

Feature 6. Single word root, Positive/Negative aspect and Mode of Important words:

It is obtained by using the second feature for important words only.

Feature 7. Binary word root and Positive/Negative aspect of Important words:

It is obtained by using the third feature for important words only.

139

Feature 8. Binary word root, Positive/Negative aspect and Mode of Important words:

It is obtained by using the fourth feature for important words only.

2.7 Determining Evaluation Methods

To measure the program's success, Acc-accuracy, P-Precision, and R-Recall as well as the FM-FMeasure, the harmonic mean for P and R, were used. In addition, an error matrix (CM-Confusion matrix) showing the distribution of evaluation by classes was found.

True Positive (TP): Comments you anticipated positive are actually positive.
True Negative (TN): Comments you anticipated negative are actually negative.
False Positive (FP): Comments you anticipated positive are actually negative.
False Negative (FN): Comments you anticipated negative are actually positive.

While calculating the value of Acc, the ratio of all correctly classified samples to all data is checked. P, R, and FM values are calculated separately for each class. While P_c belongs to class c, it is the ratio of the data found as class c to all the data that belong to class c, that is, P_c is the accuracy rate of the values found for c. While R_c belongs to class c, it is the ratio of data found as class c to the all data that belong to class c, that is, R_c shows how much of the class c data can be covered or found as class c. Meanwhile, FM is the harmonic mean of P and R and it is used to prevent systems from giving low R results when P is high, or low P results when R is low; it is calculated as FM=2PR/(P+R). In this article, P, R, FM and Acc values will be shown in percentages.

2.8 Naïve Bayes method

Naïve Bayes method was used in this study, which is one of the supervised learning methods among machine learning methods. Although it is difficult and expensive to prepare the training set for this method, the reason for choosing this method is the ease and high success of testing with Naïve Bayes. It is a basic classification technique successfully used in many areas from e-mail and document classification to sentiment and opinion classification. It also requires less training data and a shorter training time than other alternative methods [8].

NB is suitable in situations with limited processor and memory resources. The NB classifier assumes that the features used in the classification are independent from one another. Although this assumption may be wrong in many cases, there are many theoretical reasons behind this classifier's unreasonable effectiveness [10]. Even if the probability calculations are of low quality, classification decisions are quite good [11].

In the text classification application, words, phrases or their properties are used as features. With the maximum a posteriori (MAP) decision rule the following classifier is obtained:

$$c_{map} = \arg\max_{c \in C}\left(P(c\mid d)\right) = \arg\max_{c \in C}\left(P(c)\prod_{1 \le k \le n_d}P(t_k \mid c)\right)$$

Equation 1

Here t_k is the document's terms/words,

C is the set of probable classes used in classification,

$P(c\mid d)$ is the conditional probability of becoming class c when d is given,

$P(c)$ is the prior probability of class c, and

$P(t_k\mid c)$ is the conditional probability for becoming t_k term/word when class c is given.

Considering the computers' limited capacity to express numbers and some calculation errors, instead of maximizing the multiplication of probabilities, the method of maximising the sum of their logarithm is preferred and the formula is applied as follows:

$$c_{map} = \arg\max_{c \in C}\left(\log P(c) + \sum_{1 \le k \le n_d}\log P(t_k \mid c)\right)$$

Equation 2

In the new formula, instead of the maximum probability class, the class with the maximum log score is selected. Since the logarithm function is a monotone increasing function, the MAP decision rule remains the same. Another issue is that in a certain class some terms/words are not mentioned at all. In this case, the value of conditional probability is 0. If the first rule is used, the product is 0, if the second rule is used, the logarithm 0 is undefined. To prevent this, the Laplace smoothing method is used. Here, B'is the number of elements in the term/word vocabulary V [8]. The formula is changed as below:

$$P(t\mid c) = \frac{T_{ct}+1}{\sum_{t \in V}(T_{ct'}+1)} = \frac{T_{ct}+1}{\sum_{t' \in V}(T_{ct'})+B'}$$

Equation 3

There are some types of NB method. Multinomial NB (MNB), Binarized Multinomial NB (BMNB) and Bernoulli NB (BNB) are some of them. They may give different results because they use different models. MNB is generally used in cases when more than one appearance of terms is important in the classification problem. Topic classification can be given as an example. BMNB is suitable for cases when frequencies do not play an important role in classification. Sentiment analysis is a good field of application for the use of BMNB because in these texts, it is important whether a negative word such as "bad" is used or not, rather than the number of times it is used. Finally, the BNB method is useful for applications such as spam email or unwanted content detection, where it is important that certain words are not used [9].

2.8.1 Multinomial Naïve Bayes Model

In this method, the conditional probability value for t (term/word) is calculated according to the relative frequency value of t' in the documents belonging to class c as below [11]:

$$P\left(t\mid c\right)=\frac{T_{ct}}{\sum_{t\in V}T_{ct'}}$$

Equation 4

For t, multiple uses of documents in class c in the training data are also considered.

The training and testing algorithm of this method is given in Table 5:

Table 5: MNB Training and Testing Algorithm

```
MNBTraining(C,D)
      V← ExtractVocabulary(D)
      N← CountDocuments (D)
      for each cεC
      do Nc← CountDocuments (D,c)
       prior[c]←Nc/N
       textc← JoinClassDocuments(D,c)
       for each tεV
       do Tct← CountTerms(textc,t)
       for each tεV
       do condprob[t][c]←(Tct+1)/∑t'(Tct'+1)
      return V, prior, condprob
MNBTest(C,V,prior,condprob,d)
      W← ExtractTfromDocument (V,d)
      for each cεC
      do score[c]←log prior[c]
       for each tεW
       do score[c]+=log condprob[t][c]
return argmaxcεC score[c]
```

2.8.2 Binarized (Boolean) Multinomial Naïve Bayes Model

Similar to MNB, instead of looking at the frequency/number of t term/word usage, this method only looks at whether it is used or not. Instead of the frequency t, 1 is used, if not, 0 is used [1].

2.8.3 Bernoulli Naïve Bayes Model

For each t in the vocabulary, BNB produces a value of 1, if present in the document, if not, it produces 0 [11]. It is very different from MNB in that it does not look at the number of uses of t but takes into account the t values that are not used. The training and testing algorithm for BNB is given in Table 6:

Table 6: BNB Training and Testing Algorithm

```
BNBTraining(C,D)
     V←ExtractVocabulary(D)
     N←CountDocuments(D)
     for each c∈C
     do Nc←DocumentCount(D,c)
      prior[c]←Nc/N
      for each  t∈V
      do Nct←CountDocumentTerms(D,c,t)
            condprob[t][c]←(Nct+1)/(Nc+2)
     return V, prior, condprob
BNBTest(C,V,prior,condprob,d)
     Vd←ExtractTfromDocument(V,d)
     for each c∈C
     do score[c]←log prior[c]
       for each t∈V
       do if t∈Vd
         then score[c]+=log condprob[t][c]
         else score[c]+=log(1-condprob[t][c])
return argmax_c∈C score[c]
```

2.9 Weka

Weka is the name of software developed for machine learning at the Waikato University. It includes most of the machine learning algorithms and methods that today are widely used [12] [13]. It is possible to make classifications within the existing data sets by using one of the classification algorithms loaded in the WEKA. Moreover, it is possible to use separate sets for testing and validation.

2.9.1 Weka's Naïve Bayes results

NB results of experiments with Weka are given in 7.

7: Weka's CM and Acc results for NB

Measures	References	Appointed							
		Feat0	**Feat1**	**Feat2**	**Feat3**	**Feat4**	**Feat5**	**Feat6**	**Feat7**
TP	25	24	24	24	24	24	23	25	25
FN	0	2	2	2	2	3	3	6	6
FP	0	1	1	1	1	1	2	0	0
TN	25	23	23	23	23	22	22	19	19
Acc	100	94	94	94	94	92	90	88	88

2.9 Program's Results

2.9.1 Naïve Bayes MBNB Results

The NB results of our program with MBNB method are given in Table 8. In this method, it is considered as 1, if the feature is used, and as 0, if it is not used.

Table 8: Program's CM and Acc results (for Binary NB)

Measures	References	Appointed							
		Feat0	**Feat1**	**Feat2**	**Feat3**	**Feat4**	**Feat5**	**Feat6**	**Feat7**
TP	25	24	24	23	23	23	23	24	24
FN	0	2	2	2	2	2	2	7	6
FP	0	1	1	2	2	2	2	1	1
TN	25	23	23	23	23	23	23	18	19
Acc	100	94	94	92	92	92	92	84	86

NB results of our program with MNB method are given in Table 9. In this method, the number of uses for each feature is taken into account.

2.9.2   Naïve Bayes MNB Results

Table 9: Program's CM and Acc results for MNB

Measures	References	Appointed							
		Feat0	Feat1	Feat2	Feat3	Feat4	Feat5	Feat6	Feat7
TP	25	24	24	23	23	23	23	24	24
FN	0	2	2	2	2	3	3	6	8
FP	0	1	1	2	2	2	2	1	1
TN	25	23	23	23	23	22	22	19	19
Acc	100	94	94	92	92	90	90	86	86

2.10 The Dictionary Method

With the method we developed, we made a classification with the help of a dictionary which we extracted from the important words and phrases from positive or negative reviews. Accordingly, we made a simple score calculation for test reviews. If the words in the test review appear in the positive words dictionary, we added it to the positive score as much as this word's rate of appearance in the positive word dictionary. We made the same calculation for the negative words, and assigned the highest scores as the result. These results were also compared to Naïve Bayes. The results of our program with the new method are given in Table 10.

Table 10: Program's CM and Acc results with the New Method

Measures	References	Appointed							
		Feat0	Feat1	Feat2	Feat3	Feat4	Feat5	Feat6	Feat7
TP	25	25	25	24	24	23	23	22	22
FN	0	0	0	1	1	2	2	3	3
FP	0	18	18	8	8	6	5	5	5
TN	25	7	7	17	17	19	20	20	20
Acc	100	64	64	82	82	84	86	84	84

2.11 Comparison of all Results

Table 11 gives the methods' accuracy rates according to features.

Table 11: Methods accuracy rates by features

Accuracy								
Method	**Feat0**	**Feat1**	**Feat2**	**Feat3**	**Feat4**	**Feat5**	**Feat6**	**Feat7**
NBBin	**94**	**94**	92	92	92	92	84	86
NBFR	**94**	**94**	92	92	90	90	86	86
WekaNB	**94**	**94**	**94**	**94**	92	90	88	88
Nmethod	64	64	82	82	84	86	84	84

Table 12, Table 13 and Table 14 give the P, R and FM values of positive and negative classes with the new method, MBNB and MNB methods respectively according to features. The graphs from Figure 4 to Figure 11 show these values in graphs.
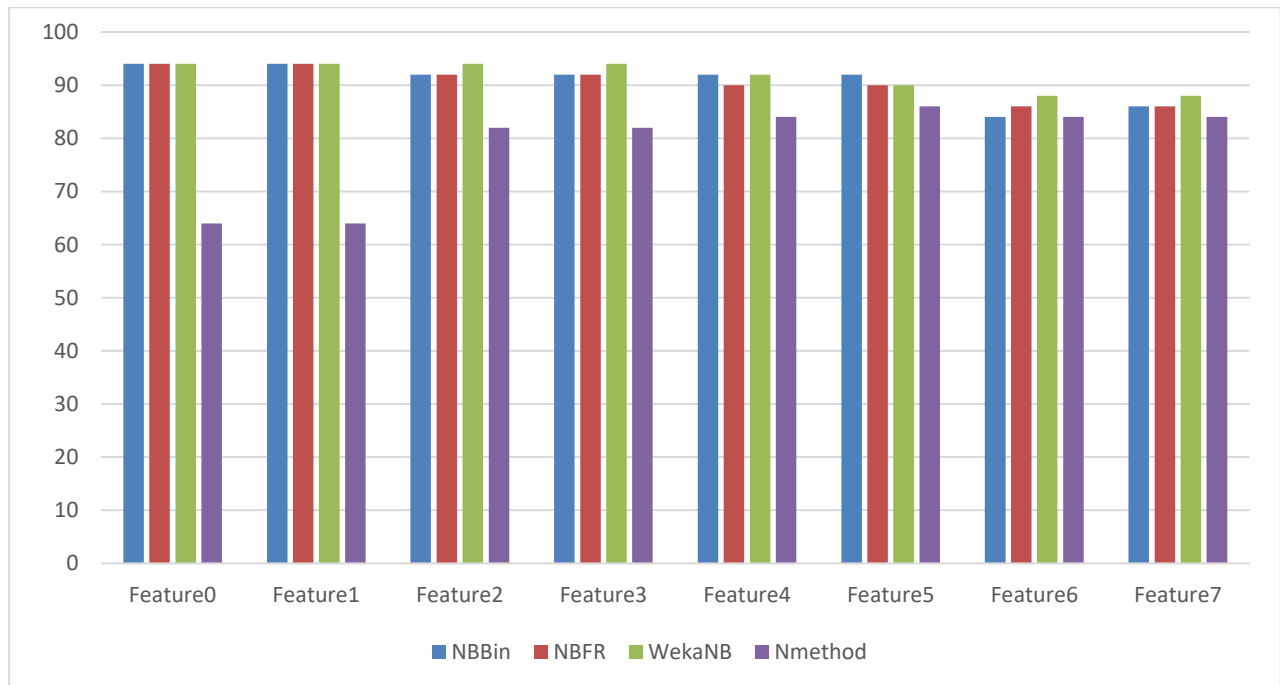
Figure 4: Accuracy rates of methods by features
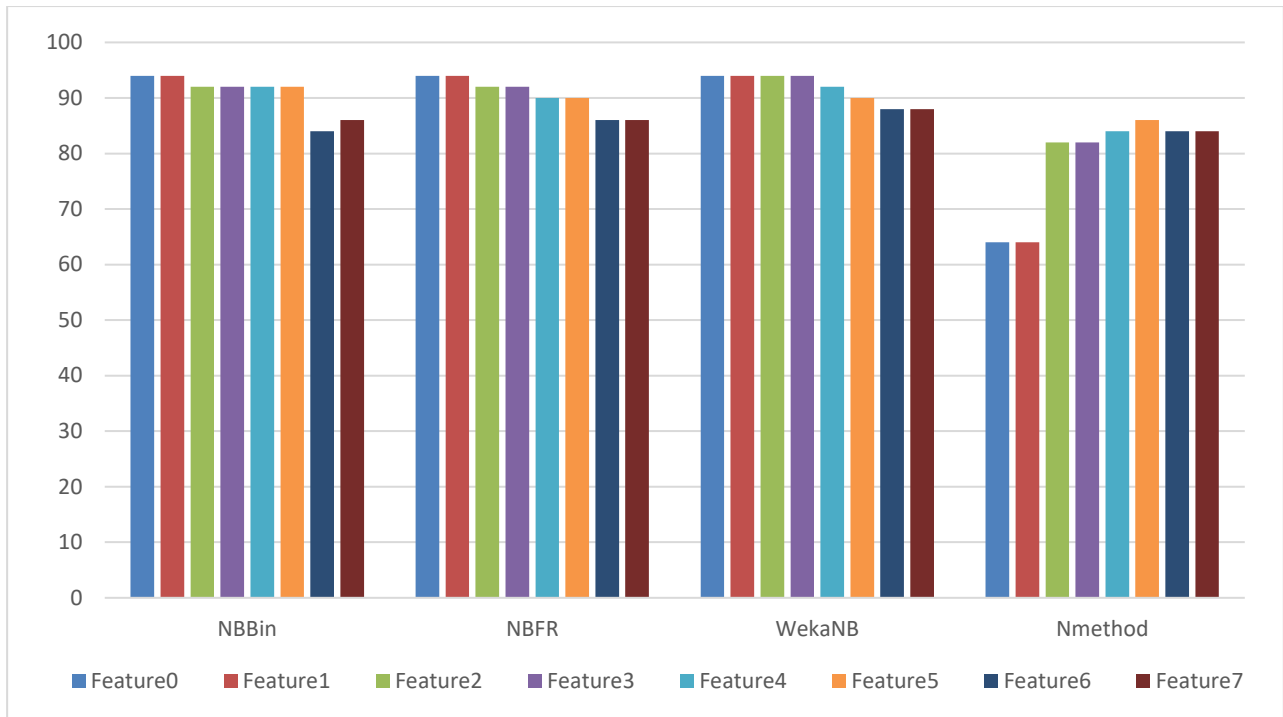
Figure 5: Accuracy rates of features by methods

Table 12: P, R, FM values of positive and negative classes according to new method and features

|	New method							
Measure	**Feat0**	**Feat1**	**Feat2**	**Feat3**	**Feat4**	**Feat5**	**Feat6**	**Feat7**
P_Pos	58	58	75	75	79	82	81	81
P_Neg	**100**	**100**	94	94	90	90	87	87
R_Pos	**100**	**100**	**96**	**96**	92	92	88	88
R_Neg	28	28	68	68	76	80	80	80
FM_Pos	74	74	84	84	85	87	85	85
FM_Neg	44	44	79	79	83	85	83	83
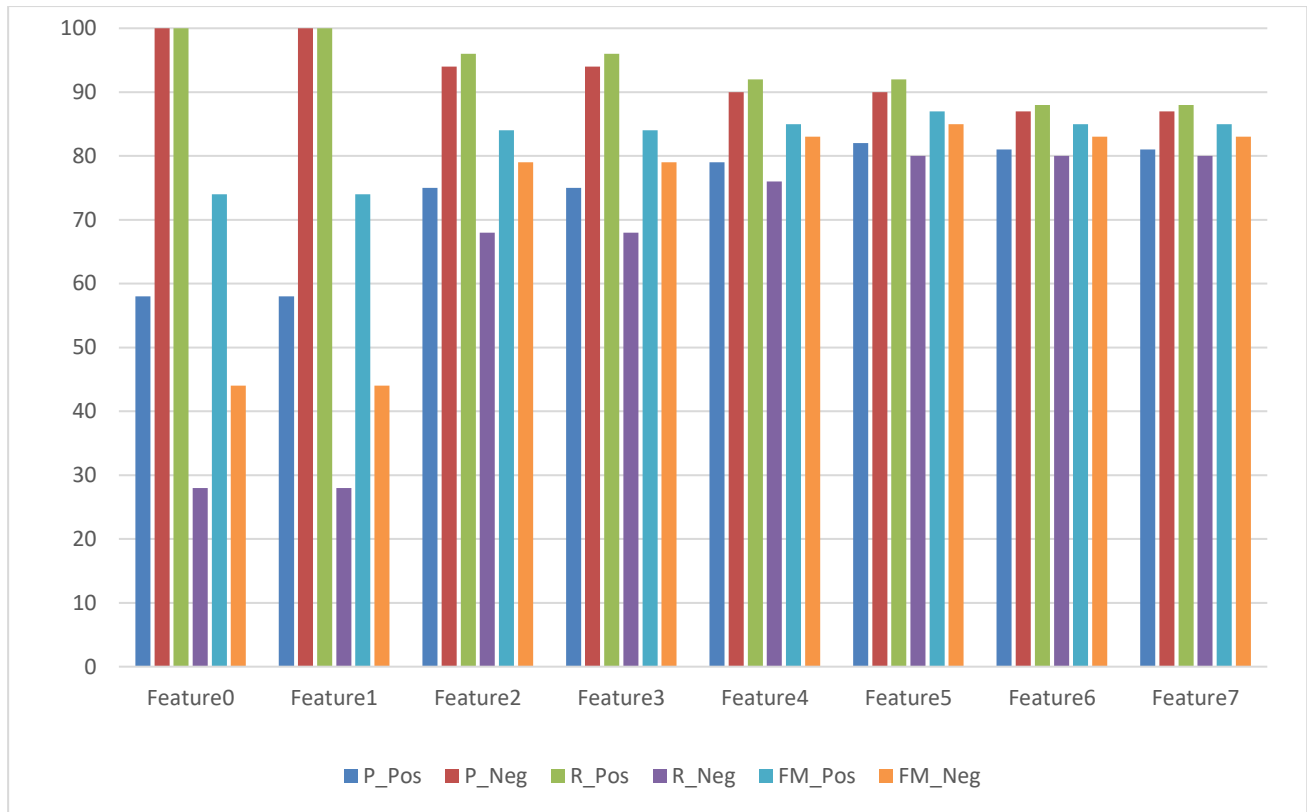
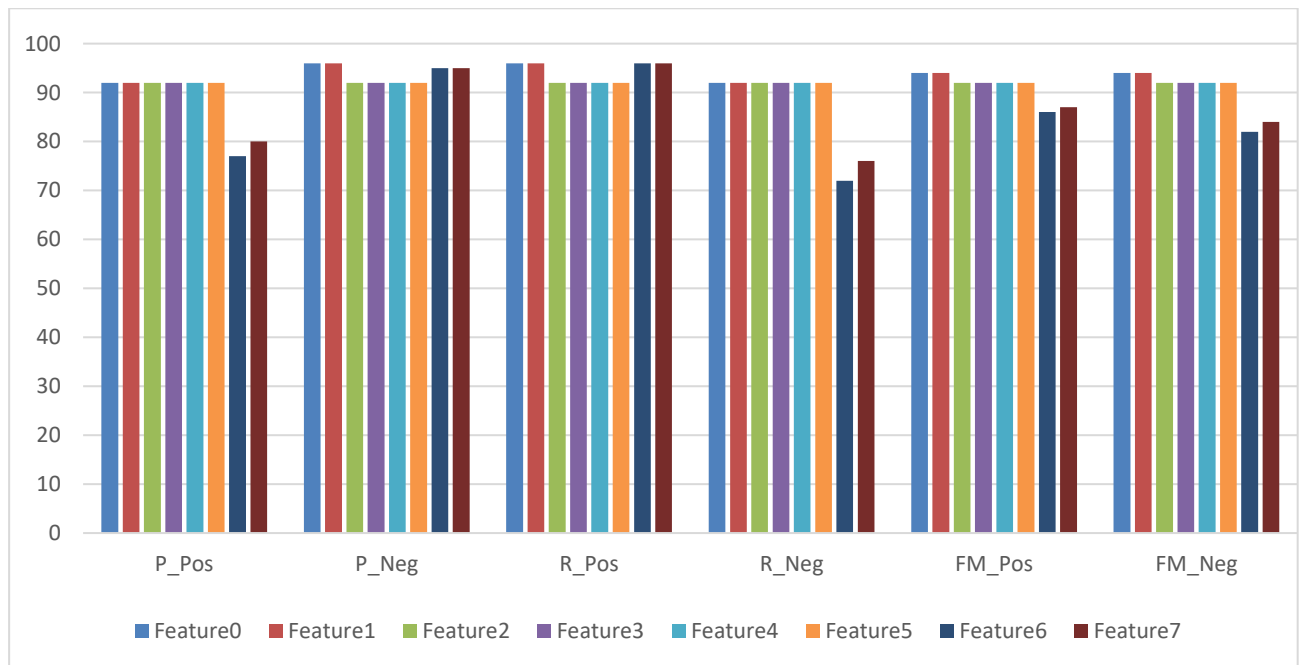Figure 6: P, R, FM values in each feature for positive and negative classes of the new method
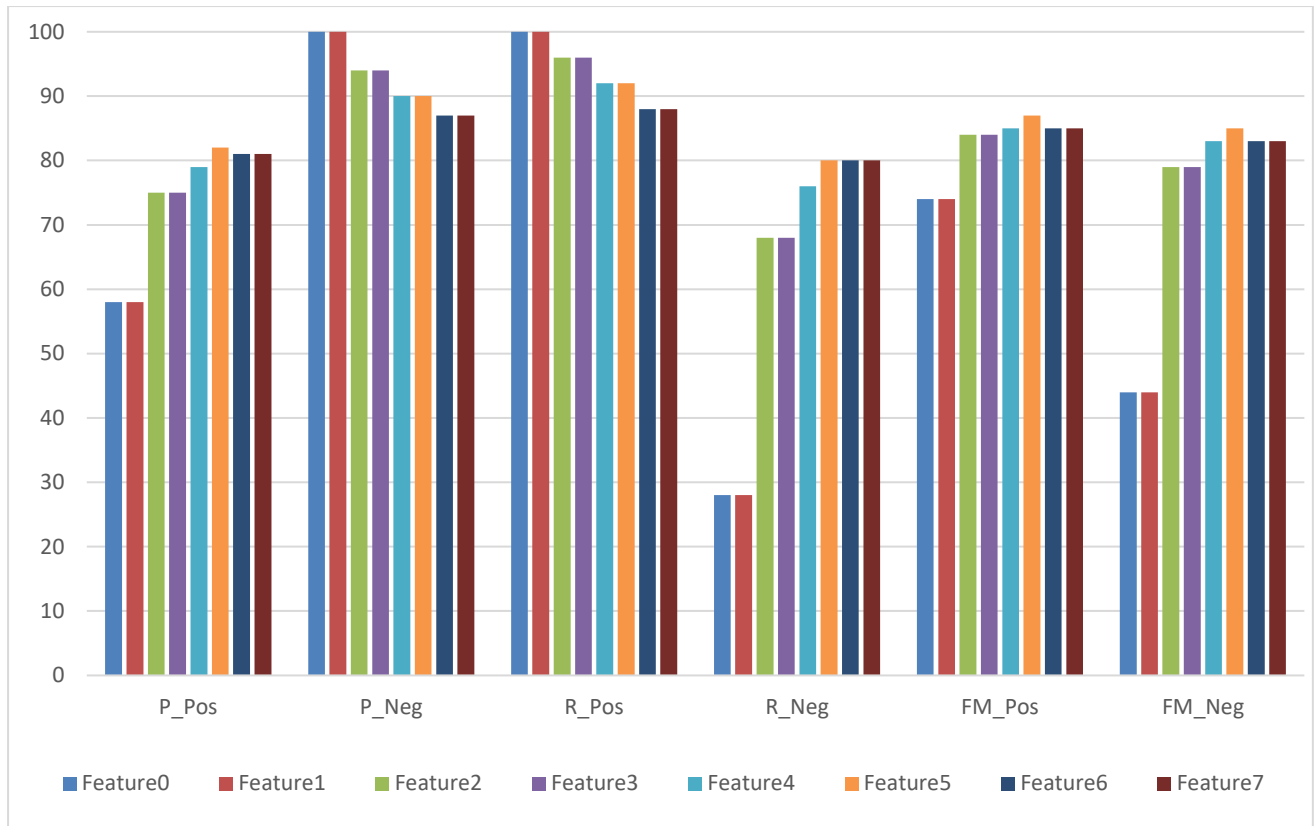
Figure 7: Comparison of the New Method's P, R, FM values for positive and
negative classes by features

Table 13: P, R, FM values of positive and negative classes according to MBNB and
features

| MBNB								
Measure	**Feat0**	**Feat1**	**Feat2**	**Feat3**	**Feat4**	**Feat5**	**Feat6**	**Feat7**
P_Pos	92	92	92	92	92	92	77	80
P_Neg	**96**	**96**	92	92	92	92	95	95
R_Pos	**96**	**96**	92	92	92	92	**96**	**96**
R_Neg	92	92	92	92	92	92	72	76
FM_Pos	94	94	92	92	92	92	86	87
FM_Neg	94	94	92	92	92	92	82	84
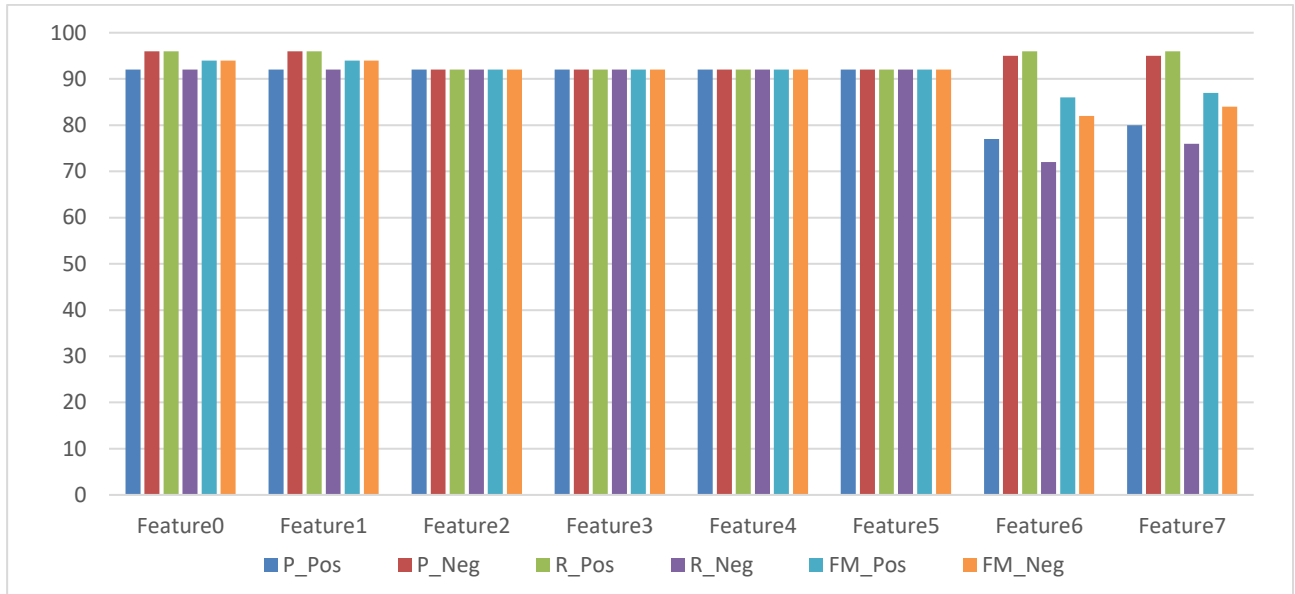
Figure 8: P, R, FM values of each feature in positive and negative classes with the MBNB method

Figure 9: Comparison of the MBNB Method's P, R, FM values for positive and negative classes by features

Table 14: P, R, FM values of positive and negative classes according to MNB and features

MNB								
Measure	Feat0	Feat1	Feat2	Feat3	Feat4	Feat5	Feat6	Feat7
P_Pos	92	92	92	92	88	88	80	80
P_Neg	**96**	**96**	92	92	92	92	95	95
R_Pos	**96**	**96**	92	92	92	92	**96**	**96**
R_Neg	92	92	92	92	88	88	76	76
FM_Pos	94	94	92	92	90	90	87	87
FM_Neg	94	94	92	92	90	90	84	84
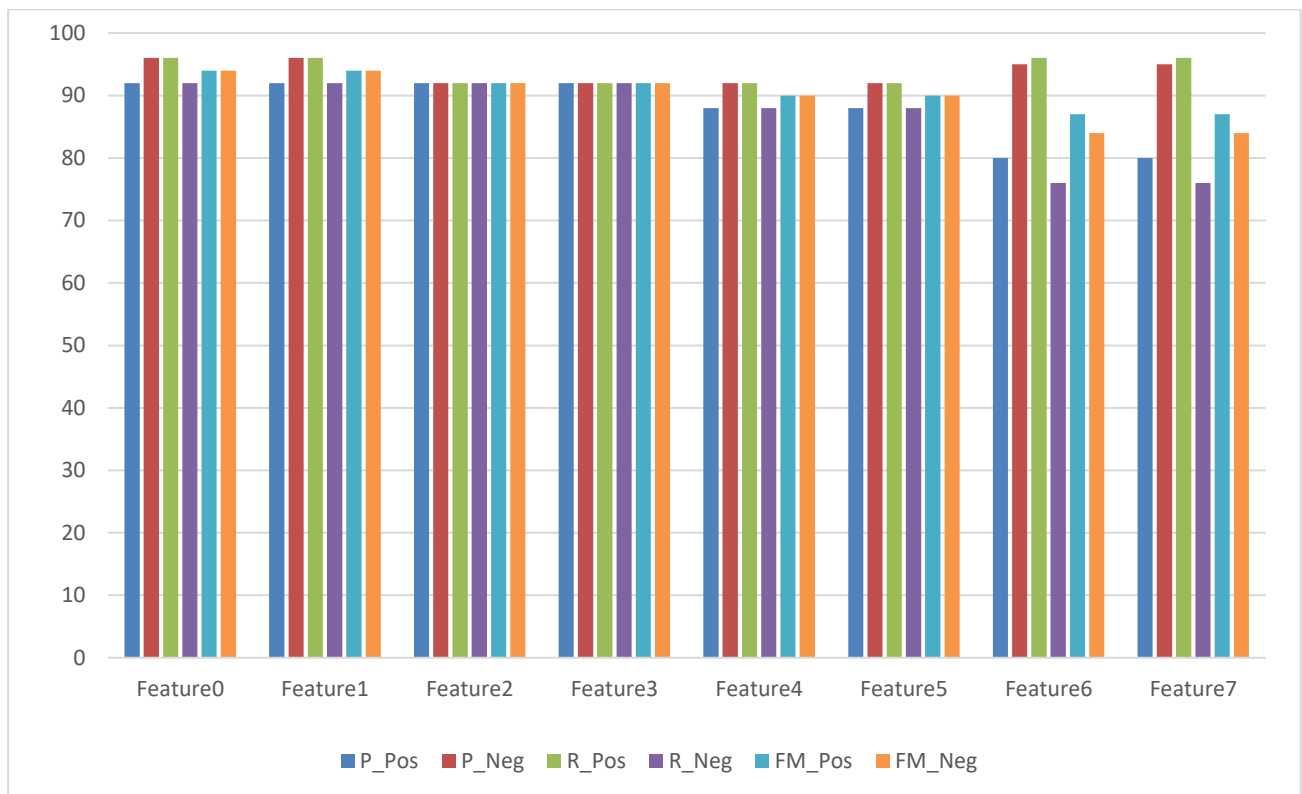
Figure 10: P, R, F values in each feature for positive and negative classes with the MNB method
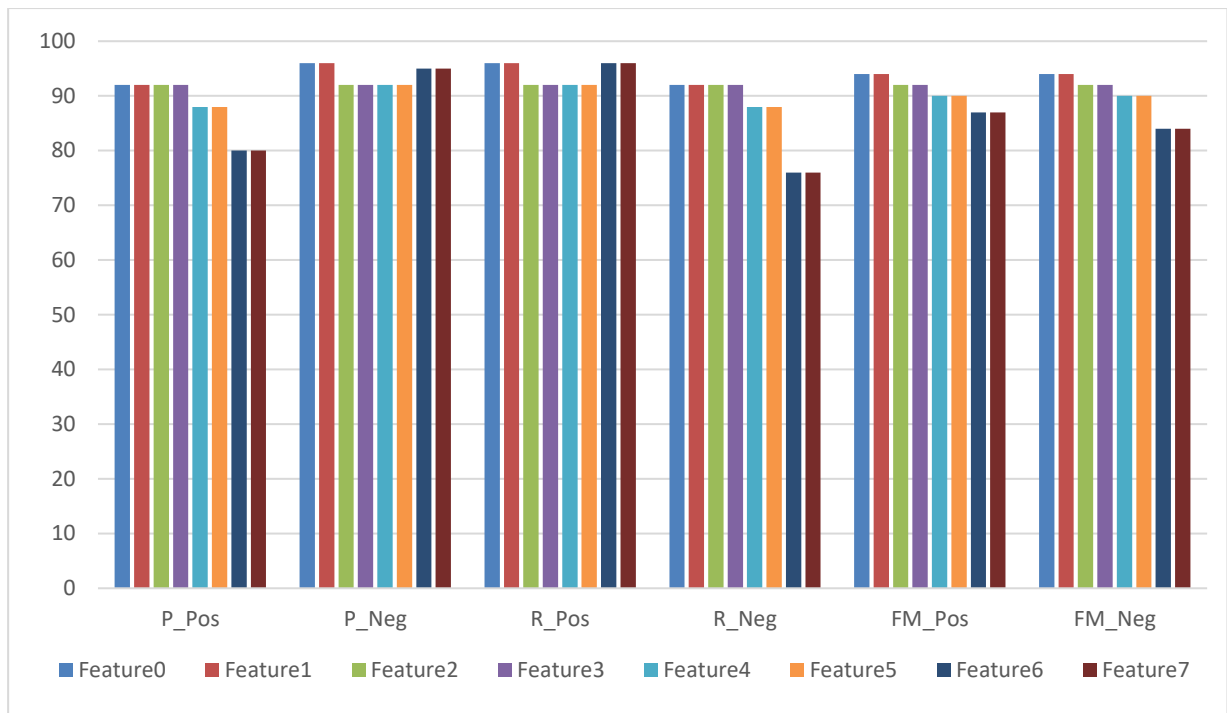
Figure 11: Comparison of the MNB Method's P, R, FM values for positive and
negative classes by features

According to their accuracy, NB methods give approximately the same results. The best results have been obtained for NB methods, with all single root words where 94% accuracy was achieved. When binary words were used, the success rate decreased slightly. When only important words were used, the decrease was higher, and the accuracy rate fell to 84-86%, especially when binary words were used. The fact that single words are more successful than binary words may be attributed to reasons such as the lower number of produced features, not obtaining binary words at the training phase due to their lower probability of occurrence, and the increase in the number of produced features due to production of unnecessary features. In addition, out of the NB methods, MNB and MBNB gave approximate results. The most important factor here is that the comments in the used data are short and the frequency or presence/absence information gives approximately the same result.

The result obtained in the newly proposed method, is different compared to the ones in the full NB methods. When all of the single root words were used, the accuracy rate of 64% was obtained due to the many commonalities in the positive and negative comments. Nonetheless, the aim of this method is not to use all the words. Therefore, when the score was calculated by using the previously produced dictionary of important positive and negative words, as expected, accuracy rates increased from 64% to 84-86%, showing significant improvement. In addition, when looking at the binary phrases in all words, as probability of finding commonalities in the positive and negative is reduced, binary phrases gave better results than single ones in terms of success.

When we look at the P, R and FM values, we encounter the accuracy rates for NB. All singular words gave more balanced P and R values. When important words are used and according to their being single or binary, the P and R values are more unevenly distributed among classes.

In the new method, fluctuations of P, R values among classes are high, especially when all words are used, while when important words are used, sounder results are obtained. Since the difference between scores is very close and no threshold value was determined for this difference, these results can be considered natural. In future studies, this situation can be corrected by first eliminating the common features used in positive and negative reviews with certain methods and then determine the difference between them by selecting the most appropriate level experimentally.

3. Analysis, Discussion and Conclusion

With the advancement of technology and the increase in data in which people express their emotions and opinions in electronic environments, sentiment analysis and opinion mining have become an important area. The exponential increase in the studies in this field in recent years and allocation of astronomical budgets to studies and applications in this field make it even more attractive. Using today's facilities, it is possible to develop many applications from which people can benefit.

While in other countries there is a lot of work in this area, studies on Turkish have just started and are still new. We conducted a study on this developing and exciting field. We obtained important results by making use of film reviews. However, we also noticed that there is a lot to be improved in this regard. Although we applied NB and simple techniques, we obtained very good results with these simple and high performance techniques. In addition, although the manual selection of terms in the dictionary extraction method, which we call the new method, is more difficult, it yielded similar results compared to traditional machine learning methods. However, dictionaries extracted from a large number of reviews will give better results. In order to obtain better results from the dictionary method, important words can be selected more carefully, their numbers can be increased and in particular, binary terms can be selected more meticulously. The important words coming from new reviews can also be added to the dictionary. Similarly, although the method containing modal terms is an original invention, the expected high success was not achieved due to the low number of modal terms mentioned in the reviews; however, the success rate can be increased by selecting features from only small word clusters which contain specific terms instead of using all terms in such methods. When the dictionary extraction method is used together with modal terms and especially with machine learning methods, the hybrid method will yield better results. We believe that these inferences will enlighten us and other researchers of the field.

4. References

1. Jurafsky, D. What is sentiment analysis, https://web.stanford.edu/class/cs124/lec/sentiment.pptx.

2. Hyun Duk Kim, Kavita Ganesan, Parikshit Sondhi, ChengXiang Zhai Comprehensive Review of Opinion Summarization, http://kavita-ganesan.com/sites/default/files/survey_opinionSummarization.pdf

3. Hidayet Takçı, Duygu Analizi (Sentiment Analysis), http://verimadencisi.blogspot.com.tr/2013/08/duygu-analizi-sentiment-analysis.html, 2013.

4. Türkçe otomatik sentiment analizi (otomatik duygu analizi uygulaması, http://www.sentimentanalizi.com/, 2014

5. Allie Gray Freeland, SEO and Social Secrets Every Entrepreneur Must Know and Why, https://www.huffpost.com/entry/6-seo-and-social-secrets-_b_5505296, 2017

6. Roman Daneghyan, 54 SEO Statistics That Will Impact Your Business in 2020, https://www.safaridigital.com.au/blog/seo-statistics-2019, 2019

7. Ender Ahmet Yurt, Duygu Analizi, http://www.slideshare.net/webender/duygu-analizi, 2014

8. Fatih Amasyalı, Yeni Makine Öğrenmesi Metotları ve İlaç Tasarımına Uygulanması, Doktora Tezi, Fen Bilimleri Enstitüsü, Yıldız Teknik Üniversitesi, İstanbul, 2008.

9. Vasilis Vryniotis, Machine Learning Tutorial: The Naïve Bayes Text Classifier, http://blog.datumbox.com/machine-learning-tutorial-the-Naïve-bayes-text-classifier, 2013.

10. Zhang, Harry, The optimality of Naïve Bayes, AA 1.2 (2004): 3.

11. Christopher D. Manning, Prabhakar Raghavan ve Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press, http://nlp.stanford.edu/IR-book/html/htmledition/properties-of-Naïve-bayes-1.html, 2008

12. Weka, http://tr.wikipedia.org/wiki/Weka

13. Ian H. Witten; Eibe Frank, Len Trigg, Mark Hall, Geoffrey Holmes, and Sally Jo Cunningham, Weka: Practical Machine Learning Tools and Techniques with Java Implementations, Proceedings of the ICONIP/ANZIIS/ANNES'99 Workshop on Emerging Knowledge Engineering and Connectionist-Based Information Systems.