

A Novel Hybrid Approach for Sentiment Classification of Turkish Tweets for GSM Operators

Ilkay YELMEN¹, Metin ZONTUL², Oguz KAYNAR³, Ferdi SONMEZ⁴

Abstract— The increase in the amount of content shared on social media makes it difficult to extract meaningful information from scientific studies. Accordingly, in recent years, researchers have been working extensively on sentiment analysis studies for the automatic evaluation of social media data. One of the focuses of these studies is sentiment analysis on tweets. The more tweets are available, the more features in terms of words exist. This leads to the curse of dimensionality and sparsity, resulting in a decrease in the success of the classification. In this study, Gini Index, Information Gain and Genetic Algorithm (GA) are used for feature selection and Support Vector Machines (SVMs), Artificial Neural Networks (ANN) and Centroid Based classification algorithms are used for the classification of Turkish tweets obtained from 3 different GSM operators. The feature selection methods are combined with the classification methods to investigate the effect on the success rate of analysis. Especially, when the SVMs are used with the GA as a hybrid, 96.8% success has been achieved for the classification of the tweets as positive or negative.

Keywords— classification algorithms, feature extraction, genetic algorithms, sentiment analysis, text mining.

I. INTRODUCTION

Due to the developing technology, social media is transforming social values in an innovative way. People use social networks like Facebook, YouTube, Twitter and LinkedIn to share and communicate information. Information on these intelligent social networks is easily accessible anytime, anywhere via a desktop or laptop computer or a mobile device. Moreover, instant updates on profiles, news flow and blog information are available on social media. For this reason, social media attracts many professional groups such as politicians, managers and marketers [1].

Social media offers great opportunities for businesspersons who want to sell their goods and products online. The ease of

sharing and selling products on a global scale by means of social media has created a new economy. Due to the speed of communication people trust social media more and more every day in sharing information regarding their feelings [2]. As there is a huge amount of data about personal feelings on social media, a special data analysis approach called sentiment analysis is required.

Sentiment analysis is the process of identifying the emotions and thoughts of users by analyzing their written expressions. After this identification process, the feelings of the people concerned are separated into categories. Sentiment analysis is the most powerful tool to determine the attitude of a text writer and the polarity of written text [16]. Generally, it identifies written texts as positive, neutral or negative. This identification can be useful in many areas such as customer complaint analysis, product reliability analysis on social media like Facebook or Twitter. In order to produce products that are more reliable and gain more loyal customers, many companies use sentiment analysis on these social media platforms.

The fact that texts in social media are mostly written in colloquial language and both understanding and analyzing these texts in Turkish is somewhat difficult has led researchers to focus their attention on this area. The unsatisfactory studies on sentiment analysis on Turkish social media data have led us to conduct this study. The study aims to find a new method to improve the performance of classification of Turkish texts written in colloquial language, most particularly on social media. Experimental studies have been carried out by support vector machines (SVMs), Artificial Neural Networks (ANNs) and centroid based classification algorithms using Natural Language Processing (NLP) methods. In addition, Gini Index (GI), Information Gain (IG) and Genetic Algorithm (GA) feature selection methods are combined with classification methods in order to construct hybrid models for sentiment classification. Especially, when the SVMs are used with the GA as a hybrid, 96.8% success has been achieved for the classification of the tweets as positive or negative.

The rest of this paper is organized as follows. In Section 2, the related work on sentiment analysis on social media platforms is discussed; then the details of dataset are presented in Section 3. The data preprocessing, feature selection, classification and experimental works are presented in Section 4, 5, 6 and 7, respectively. In the last section results are evaluated.

Ilkay YELMEN is with the Department Computer Engineering, Kadir Has University, Istanbul, 34083, Turkey (phone: +90 506 488 51 10 e-mail: ilkay.yelmen@stu.khas.edu.tr).

Metin ZONTUL is with the Department Software Engineering, Istanbul Aydin University, Istanbul, 34153, Turkey (e-mail: metinzontul@aydin.edu.tr).

Oguz KAYNAR is with the Department of Management Information Systems, Cumhuriyet University, Sivas, 58140, Turkey (e-mail: okaynar@cumhuriyet.edu.tr).

Ferdi SONMEZ is with the Department Computer Engineering, Arel University, Istanbul, 34537, Turkey, (e-mail: ferdisonmez@arel.edu.tr).

II. RELATED WORK

With the widespread usage of social media platforms, forums and blogs as ways of reviewing have emerged as important factors in human life. Researchers have started to focus on these reviews to automatically categorize them into polarity levels such as positive, negative, and neutral since the early 2000s. This research process is known as sentiment analysis. Certain researchers have investigated the utility of linguistic features to detect the sentiment of Twitter messages by evaluating the usefulness of existing lexical resources as well as the emoticon features used in microblogging. They have applied a supervised approach to identify positive, negative, and neutral tweets taken from HASH, EMOT and ISIEVE datasets. They have made many experiments by using n-gram features, lexicon features, part-of-speech features and micro-blogging features. Their experiments have shown that when microblogging features are included; the benefit of emoticon training data is decreased [3].

In a study conducted on Turkish messages on Twitter, the data set was analyzed by text classification methods such as SVM, Naive Bayes, Multinomial Naive Bayes and KNN algorithms to determine whether the messages were positive or negative. Prior to the classification the features represented by the Vector Space model were obtained in two different ways, as word bag (BoW) and N-Gram model, and the effect of this condition on the classification results was investigated. The root finding, stop words and repetition of the letters applied to the BoW model during the feature extraction phase were not applied to the N-Gram model. In this study, the attributes for the N-Gram model were extracted at the character level rather than at the word level, unlike the studies done in the literature. According to the experimental results, the character level N-Gram model on the generated data gave better results for all classifiers in the BoW model [4].

For the sentiment classification of Turkish political columns, four supervised machine learning algorithms of Maximum Entropy, SVM, Naïve Bayes, and the character based N-Gram Language Model were compared by Kaya, Fidan and Toroslu (2012). They discussed the sentiment classification problem in the political news domain in detail. After several experiments using unigram, stemmed unigrams, unigrams + adjectives and unigrams + effective words, they observed that the N-Gram Language Model and Maximum Entropy outperformed the Naïve Bayes and SVM. In conclusion, accuracies of 65% to 77% were obtained in all models with different features [5].

One of the most recent studies on this topic aimed to investigate the potential benefit of the concept of multiple classifier systems (MCSs) on the Turkish sentiment classification problem and propose a novel classification technique. In the experiment three classifiers, namely Naive Bayes, SVM and Bagging were with vote algorithm. Experimental results have shown that MCSs increase the performance of individual classifiers on Turkish sentiment classification data sets and meta classifiers contribute to the

power of these MCSs. The proposed approach which is MCS has achieved better performance than Naive Bayes, which was reported to be the best individual classifier for these datasets. As a result, SVMs and parameter optimization of individual classifiers were recommended when developing MCS-based prediction systems [6].

One of the most important problems of sentiment analysis on social media is labelling huge amounts of instances. In a recent study, in order to cope with this problem, researchers applied active learning to a framework containing two ensemble approaches: a probabilistic algorithm and a derived version of the Behavior Knowledge Space (BKS) algorithm. Moreover, they used the Shannon Entropy approach for the selection of training data during the active learning process. Ultimately, they compared this approach with the maximum disagreement method and random selection of instances. As a result, it was indicated that the former method gave better results with less iteration on Cornell movie review dataset [7].

A. Go, R. Bhayani and L. Huang used a distant supervision method to automatically classify the sentiment of tweets as positive and negative. Different machine learning classifiers such as Naive Bayes, Maximum Entropy, and SVM's were used along with feature extractors such as unigrams, bigrams, unigrams and bigrams, and unigrams with part of speech tags. Moreover, the emoticons at the end of each tweet were used to determine the tweet's sentiment. Tweets ending with “:.)” or “:D” were labelled as positive tweets, and tweets ending with “:(” or “:-)” as negative. Their algorithm was implemented and included in the Twitter API which enables users to classify tweets and integrate sentiment analysis classifier into web applications [13].

Some scientific studies have placed research on the selection of features in the foreground and have tried to increase the classification success rate. One of these studies aimed to achieve a high level of performance for classifying English tweets according to sentiment information. Authors have proposed a feasible solution that improves the level of accuracy. They developed a novel feature combination scheme that specifically utilizes the sentiment lexicons and the extracted tweet unigrams of high IG. Performance was evaluated using six popular machine learning classifiers. Eventually, the Naive Bayes Multinomial (NBM) classifier achieved the accuracy rate of 84.60% [12].

In the literature, many studies have shown high quality results for feature selection methods [23, 24, 25, 26, 27]. However, until today, the focus has been on either reducing faulty data or selecting more representative features for effective classification. This leads to the important research question of which step should be taken first when both steps are critical to improving the mining performance. For many large scale and related datasets, both preprocessing steps should be applied. The reason for this is that, there is usually no exact number of variables agreed upon in most domain problems, and all the variables collected for a specific domain may not be informative. Furthermore, some data samples in a

given large dataset may be regarded as noisy. Therefore, in order to develop a more effective model, feature selection and instance selection should both be considered [28-29].

Gupta, Reddy and Ekbal suggested a method for selecting features for sentiment classification and text using PSO for aspect-based sentiment analysis [30]. The success of the proposed method depends upon a reduced set of features and sometimes suffers in the event of unlabeled product reviews. Additionally, Zhu, Wang, and Mao suggested a GA and conditional random forest based hybrid method to classify sentiments [31].

A Naïve Bayes based framework, which classifies tweets as positive or negative and links them to the related news items, was developed by Kulcu and Dogdu to classify Turkish tweets and news items. They have used NLP techniques of stemming and morphological analysis, and bag-of words method in order to map the classification process and for linking tweets to news items with Zemberek NLP library. The results of experiments on Turkish tweets indicated that Naive Bayes performs well in classifying tweets in Turkish [8].

In a study in which classification was done using the word embedding method, four different Turkish sector tweet datasets were used. SVM and Random Forests classifiers were used in the classification process. At the end of the research, results were compared and better classification results were obtained in sector based tweet classification compared to general tweets. The accuracy rates achieved are: 89.97% for the banking sector, 84.02% for the football sector, 73.86% for the telecom sector, 63.68% for the retail sector and overall accuracy as 74.60% [21].

In another study, the word embedding technique was used as the feature representation. On the other hand, SVM was used as the classifier for the Turkish tweet dataset. It was shown that the proposed approach enhanced sentiment classification accuracy and significantly reduced the dimension of tweet representation. The best results were obtained using the Dvot fusion technique with an accuracy rate of 80.05% [22].

One important point in sentiment analysis is the representation of texts). Instead of traditional methods, supervised term weighting methods (TF, TFID, D1, D2, F1, F2, RF, and KL) that include terms' distribution of classes have been used by Cetin and Amasyali that they compared term weighting methods in different dimensions on two Turkish datasets. As a conclusion, they determined that supervised term weighting methods are more successful and applicable [9].

Due to a great number of review documents, various feature selection methods have been used by researchers to eliminate non-valuable features. On the other hand, there are not many studies on feature selection methods for sentiment analysis of Turkish texts. F. Akba, A. Uçan, E. Sezer and H. Sever [14] investigated the performances of feature selection methods for Turkish sentiment analysis. They applied the IG and Chi square feature selection methods to select the most valuable features in their experiments. Boynukalın [15] used the

Weighted Log Likelihood Ratio Ranking method for sentiment analysis.

In one of the most recent studies, a new feature selection method called Query Expansion Ranking based on query expansion term weighting methods was proposed. The Query Expansion Ranking method was compared with the Chi Square method and Document Frequency Difference on four Turkish product review datasets by using the NBM classifier. In conclusion, it was shown that the proposed method increased the performance of the classification in terms of accuracy and time [10].

In another recent study on Turkish sentiment analysis, various machine learning approaches were compared using the famous hotel reservation web site, booking.com. Buket ve Ercan applied the dictionary-based method, SentiTFIDF, which differs from conventional methods in terms of logarithmic differential term frequency and term presence distribution usage [11]. The results were assessed using the area under a ROC curve (AUC). It was indicated that better classification results were obtained when a document term matrix was used as an input rather than a TFIDF matrix. The Random Forest classifier gave the best results with an AUC value of 89% on both positive and negative comments. It was demonstrated by researchers that feature selection methods help to improve the accuracy of classification with fewer features [11].

One research used the Maximum Entropy Modeling classification algorithm over Turkish data set to compare the performance of four feature selection methods. Thus, the effects of the Ant Colony Optimization, Chi-square, IG, and Query Expansion Ranking methods over the success of sentiment analysis of Turkish Twitter data were evaluated. According to the experimental results, the Ant Colony Optimization and Query Expansion Ranking methods outperformed the other feature selection methods for sentiment analysis [20].

In this study, we have proposed a hybrid approach to classify the Turkish tweets as negative and positive with more accuracy. After NLP preprocessing, the GI, IG and GA have been utilized orderly for dimension reduction. SVMs, ANNs and centroid based classification algorithms have been used on the dataset after each feature selection method. Especially, when the SVMs are used as with the GA a hybrid 96.8% success has been achieved for 3 different data sets originated from 3 different GSM operators' followers tweets.

III. DATASET

During the data collection phase, tweets from followers of A, B and C GSM operators were fetched using twitter4j API. The GSM company name was used as the keyword and a total of 8379 tweet data were collected. Detailed information of the collected data is shown in Table 1.

TABLE 1. DATASET

Data Set	Number of Negative Tweets	Number of Positive Tweets	Total Number of Tweets
Operator A	3718	559	4278
Operator B	2157	798	2956
Operator C	702	442	1145

The tweets were distributed as three different users label them as positive or negative as shown in Fig. 1. If two or three users say positive or negative for a tweet, then this value is accepted as a class label in order to improve the quality of the dataset.

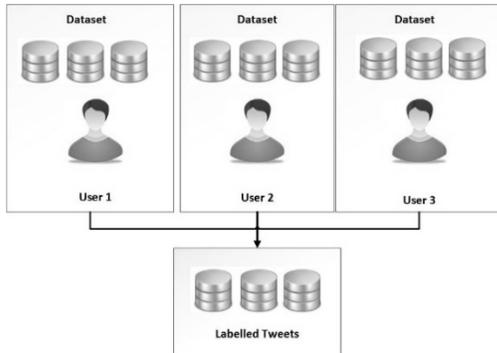


Figure 1. Data Labelling

IV. DATA PROCESSING

During the data processing phase, links, usernames, punctuation marks, stop words and retweets in related tweets were removed. In addition, the same sentences were deleted and all words were converted to lower case as shown in Fig. 2.

Next, using the ITU Natural Language Processing tool the word correction process was performed on the data that had completed the normalization process. Prior to this step, a filter program was used to ensure that the data was at a certain standard. Later, the words that were misspelled in the data set were replaced with the correct spellings using a programme.

Lastly, the words were broken down to their roots in order to increase the success rate in singularizing and classifying the expressions in the texts. Since rooting the misspelled words would be wrong, stemming was applied with the program previously prepared using the Zembek library after the word correcting step.

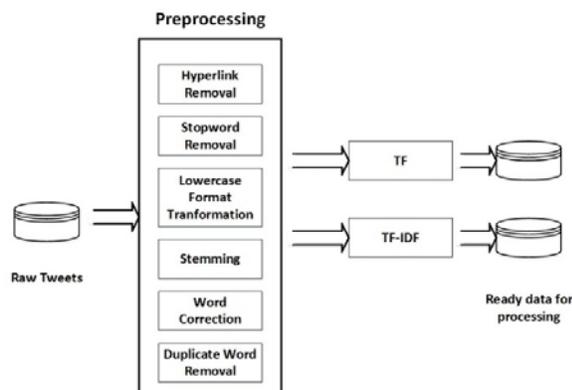


Figure 2. Data preprocessing

After unigram transformation, TF (Term Frequency) and TF-IDF (Term Frequency-Reverse Document Frequency) weighting methods were applied with the following formulas for the weighting of the features.

$$W_{TF}(c, d) = TF(c, d) \quad (1)$$

$$DF(c, d) = |\{d \in D : c \in d\}| \quad (2)$$

$$W_{IDF}(c, d) = \log \frac{N}{DF(c, d)} \quad (3)$$

$$W_{TF-IDF}(c, d) = TF(c, d) * IDF(c, d) \quad (4)$$

V. FEATURE SELECTION

In both text classification and sentiment analysis, as the number of documents or texts with opinions increase so do the different words used as features. This increase in word count leads to the curse of dimensionality and sparsity, resulting in a decrease in the success of the classification. For this reason, feature selection for the sentiment analysis methods is inevitable. Feature selection can be defined as the process of selecting features from the candidate feature set in a way that the selected features provide the greatest contribution to the classification performance. There are many methods used in feature selection, and three of the most popular ones have been used in the experiments conducted within this paper. These methods are GI, IG and GA.

GI and IG have been used in this study because their computational costs are low and their implementations are easy. In addition to these, GA has been used to get better results than dimension reduction.

The GI is a feature selection method developed as an alternative to the IG method. This method puts features in order by calculating the gain for each feature just like the IG method. However, it does not use the entropy value. In the first step of the GI method, the class label value of the data set and the GI for each feature are calculated. The gain value for each feature is then calculated by subtracting the GI calculated for that feature from the GI calculated for the class labels. Finally, the features whose gain values are below the defined threshold value are excluded from the data set and a new data set is created. The calculation of GI for the class labels is shown in Equation 5 and Equation 6. The formulations show the calculation of the GI for each property.

$$Gini = \prod_{i=1}^n p(class = i) \quad (5)$$

$$Gini(k) = \sum_{i=1}^n p(value = i) \times \prod_{j=1}^m \frac{N(value=i \& class=j)}{N(value=i)} \quad (6)$$

n refers to the number of classes in Equation 5. However in Equation 6, n corresponds to the number of different values for k variable and m represents the number of classes [18].

In this study by using the GI calculation, 200 features have been selected based on 2 different class labels: positive and negative.

IG is an entropy-based method used to calculate the estimated loss when the data set is divided into features. Entropy is a value between 0 and 1 that determines the irregularity or uncertainty of the system. The entropy value approaching 1 indicates that the system contains more information. At the first stage of the IG method, the entropy value for the class labels of a given data set is calculated as shown in Equation 7.

$$E = -\sum_{i=1}^n \left(\log_2 \frac{ns(i)}{N} \right) * \frac{ns(i)}{N} \quad (7)$$

In Equation 7, n , $ns(i)$ and N refer to the number of classes, the sample size for class i and the total sample size, respectively.

In the second step of the IG method, after the entropy value is calculated for each feature in the data set, the IG is calculated by subtracting each value from the value obtained in the first step. The IG indicates the post-split representation value of the data set. Therefore, this value is expected to be great. When properties are selected using the IG method, variables that are insufficient to identify the system are removed from the data set. The remaining variables are used to train the system. Equation 8 shows the calculation of the entropy value for each feature and the IG value calculation is shown in Equation 9.

$$E(i) = \sum_{k=1}^n \frac{ns(k)}{N} * \sum_{m=1}^{nc} \frac{nsc(k,m)}{ns(i)} \left(\log_2 \frac{nsc(i,m)}{ns(i)} \right) \quad (8)$$

$$B(i) = E(i) - E \quad (9)$$

$E(i)$, n , $ns(k)$, N , nc , $nsc(k,m)$, $B(i)$ and E correspond to the entropy value for feature i , the number of different values for feature i , the sample size for feature i having value k , the total sample size, the number of classes in dataset, the sample size for feature i having value k representing class m , IG and entropy value calculated in equation 7, respectively [19].

In our study, the entropy calculation has been performed on IG and 200 features were determined with respect to 3 different data sets.

As the final feature selection method, the GA has an important place in this study. This algorithm, receiving the initial population generated in the process of feature selection shown in Fig. 3 below, evaluates each individual (chromosome) of the population through the fitness function. Here the stop criterion, i.e. the number of iterations, is checked. The crossover and mutation procedures are performed on selected individuals until the GA ends. These operators create a new population and return to the evaluation phase and then continue until they reach the stopping criterion. When the stopping criterion is met, the GA obtains the best classification accuracy and a subset of the closest or most appropriate features.

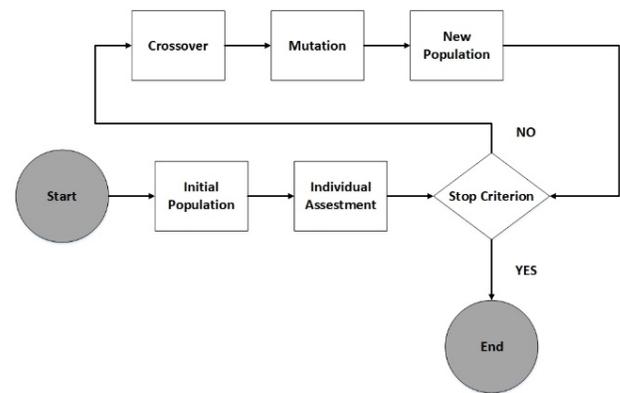


Figure 3. Feature selection with GA

In the implementation of GA for a specific problem solution, three significant design decisions should be considered: how to encode the candidate solutions on the GA chromosome, how to define the objective function for the evaluation of each solution quality and how to specify the GA run parameters. In the first step, a binary mask vector is combined with the weights originated from the training results. This combination is encoded on the GA chromosome by indicating 0 for unselected and 1 for selected features for the classification, respectively (see Fig. 5). Then, the predicted output is calculated for the selected features by using the input dataset and activation functions. In the next step, to find the least costly subset of features, a fitness function suitable with genetic search is used as in Equation (10):

$$f = (1 - e) * \left(m - \frac{F_s}{F_t} \right) \quad (10)$$

In Equation 10, F_t , F_s and e refer to the total number of features, the number of subset features, and the classification error rate with the feature subset F_s , respectively. m is a tuning parameter with a value greater than 1, compromising between minimizing the number of features in the subset and maximizing the classification rate.

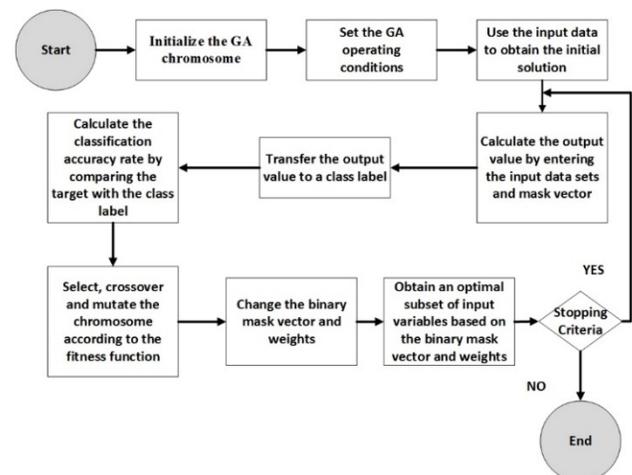


Figure 4. GA optimization process

After determining the GA run parameters empirically, the GA will change the binary mask vector and weights in any supervised model such as neural network or SVMs to find the optimal solution based on the stopping criteria continuously by maximizing the classification accuracy and minimizing the number of binary bits in the mask vector [17]. The GA optimization process is visualized in detail in Fig. 4.

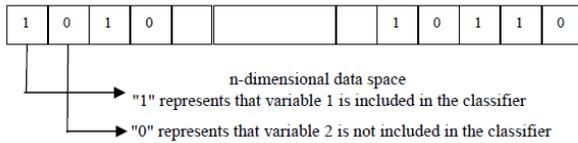


Figure 5. n-dimensional binary mask vector, comprising a set of the GA chromosome for a GA-based feature selection method

VI. CLASSIFICATION

Our aim in this work is to examine the sentiment classification of Turkish tweets by using machine learning techniques and feature selection methods together. We have experimented with three different algorithms: ANN, Centroid Based Algorithm (CBA) and SVM. Furthermore, we have used three different feature selection algorithms in our experiments. These are IG, GI and GA. The sentiment analysis process is visualized in Fig. 6.

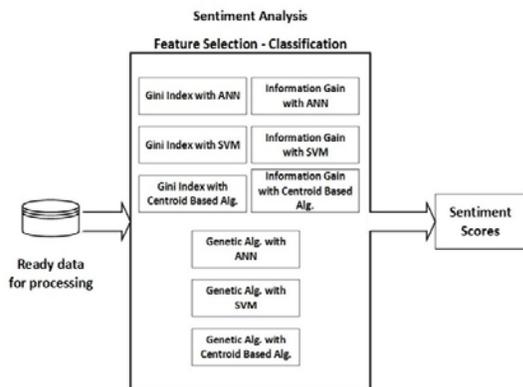


Figure 6. Sentiment analysis process

VII. EXPERIMENTAL WORK

The raw data collected from Twitter was first of all refined from unnecessary expressions and stop words. Then, after the spell correction and stemming processes, the duplicated words and sentences were eliminated to obtain good quality in the classification of tweets in terms of positive and negative polarity.

After the preprocessing above, in all experiments in this work conducted within the scope of data mining classification studies, 75% of the data was devoted to training and 25% was devoted to the test set. In the TF and TF-IDF matrices, composed of the smoothed features, experimental works were performed by using 20 fold cross validation and the 200 highest features among all the features.

In this study classification experiments have been carried out on Turkish tweet data using three different classification

algorithms with three different feature selection methods as shown in the tables below. Using the TF, the SVM classification algorithm has surpassed ANN with 40 iterations and 20 hidden layers (optimized values) and CBA by giving the accuracy values of 90.0%, 91.2% and 90.8% in the Operator A, B and C data sets, respectively as shown in Table 2.

Table 3 and Table 5 below show that the best results have been found when SVM and GA have been used as a hybrid method with both TF and TF-IDF. In addition Fig. 8 and Fig. 10 below show the ROC curve of the best accuracy values of TF and TF-IDF respectively. When Table 2 is compared with Table 7, it is seen that little improvements for ANN and CBA have been achieved using the feature selection methods, but the results are still below SVM's. Table 4 shows that using TF-IDF with only 3 classification algorithms slightly reduces the success of SVM. On the other hand, Fig. 7 and Fig. 9 below show the ROC curve of the lowest accuracy values of TF and TF-IDF, respectively.

Table 2. Accuracy Values of 3 Classification Algorithms with TF

Data Set	Categorization Technique	SVM	ANN	CBA
Operator A	TF	90.0%	87.0%	61.9%
Operator B	TF	91.2%	74.2%	63.1%
Operator C	TF	90.8%	74.8%	74.5%

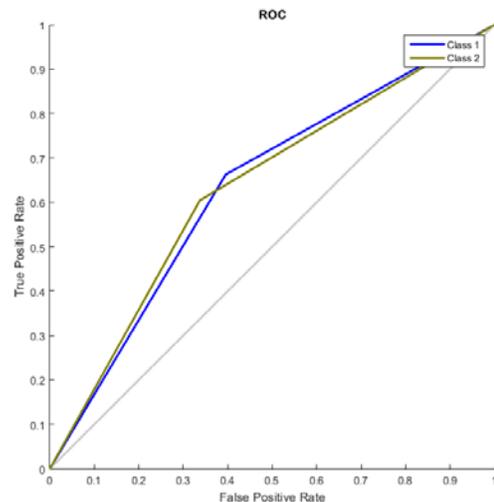


Figure 7. ROC curve of Operator A (TF CBA)

Table 3. Accuracy Values of 3 Classification Algorithms and GA with TF

Data Set	Categorization Technique	SVM + GA	ANN + GA	CBA + GA
Operator A	TF	93.8%	86.9%	86.9%
Operator B	TF	95.9%	73.4%	74.2%
Operator C	TF	94.2%	76.4%	74.5%

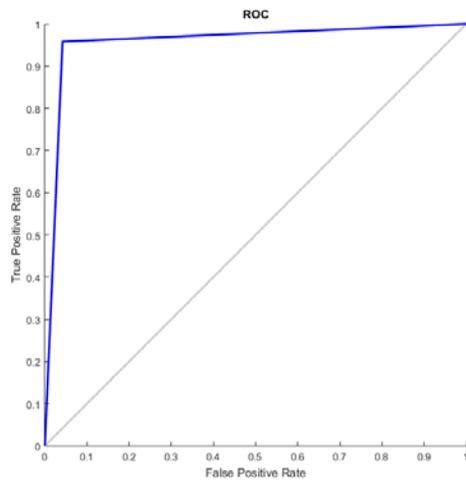


Figure 8. ROC curve of Operator B (TF SVM + GA)

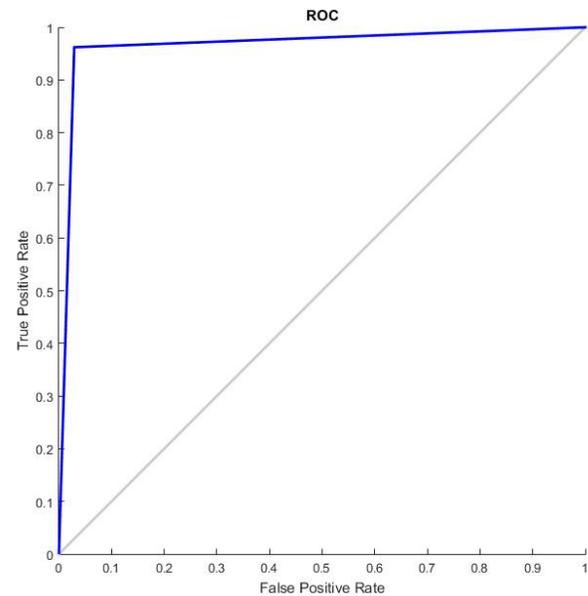


Figure 10. ROC curve of Operator B (TF-IDF SVM + GA)

Table 4. Accuracy Values of 3 Classification Algorithms with TF-IDF

Data Set	Categorization Technique	SVM	ANN	CBA
Operator A	TF-IDF	89.8 %	86.5 %	69.7 %
Operator B	TF-IDF	91.0 %	75.6 %	69.6 %
Operator C	TF-IDF	90.2 %	75.5 %	76.9 %

When the experimental results in Tables 6, 7, 8, and 9 are examined, the application of the GI and IG algorithms on SVM does not increase the success rate significantly. On the contrary, with GA as a non-deterministic feature selection method, the success rate has been increased in all three data sets.

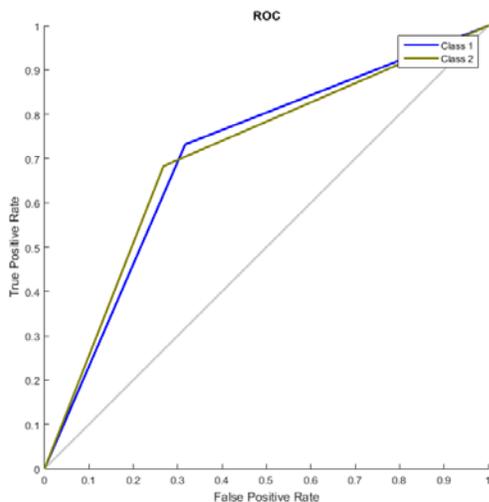


Figure 9. ROC curve of Operator B (TF-IDF SVM + GA)

In general, when we compare the GI and IG feature selection methods, we see that CBA gives better results in TF-IDF instead of the TF technique as shown in Table 8 and Table 9.

TABLE 6. Accuracy Values of 3 Classification Algorithms and Gini Index Algorithm with TF

Data Set	Categorization Technique	SVM + GI	ANN + GI Alg.	CBA + GI Alg.
Operator A	TF	89.2%	87.9%	62.6%
Operator B	TF	89.8%	76.0%	66.7%
Operator C	TF	88.6%	80.8%	79.4%

TABLE 7. ACCURACY VALUES OF 3 CLASSIFICATION ALGORITHMS AND INFORMATION GAIN ALGORITHM WITH TF

Data Set	Categorization Technique	SVM + Info. Gain Alg.	ANN + Info. Gain Alg.	CBA + Info. Gain Alg.
Operator A	TF	88.8%	87.7%	62.3%
Operator B	TF	89.0%	77.3%	66.4%
Operator C	TF	89.2%	77.3%	80.4%

Table 5. Accuracy Values of 3 Classification Algorithms and GA with TF-IDF

Data Set	Categorization Technique	SVM + GA	ANN + GA	CBA + GA
Operator A	TF-IDF	95.4%	86.9%	87.4%
Operator B	TF-IDF	96.8%	73.3%	73.5%
Operator C	TF-IDF	95.0%	76.4%	75.5%

TABLE 8. ACCURACY VALUES OF 3 CLASSIFICATION ALGORITHMS AND GINI INDEX ALGORITHM WITH TF-IDF

Data Set	Categorization Technique	SVM + GI Alg.	ANN + GI Alg.	CBA + GI Alg.
Operator A	TF-IDF	88.8%	87.2%	77.3%
Operator B	TF-IDF	89.2%	75.8%	77.3%
Operator C	TF-IDF	89.5%	78.7%	81.1%

TABLE 9. ACCURACY VALUES OF 3 CLASSIFICATION ALGORITHMS AND INFORMATION GAIN ALGORITHM WITH TF-IDF

Data Set	Categorization Technique	SVM + Info. Gain Alg.	ANN + Info. Gain Alg.	CBA + Info. Gain Alg.
Operator A	TF-IDF	89.2%	87.6%	75.8%
Operator B	TF-IDF	89.8%	75.2%	75.5%
Operator C	TF-IDF	90.0%	76.6%	80.1%

Finally, by considering all experiments carried out on the 3 GSM operator data sets consisting of Turkish texts written in colloquial language and obtained from Twitter, the highest success rate of classification has been achieved using SVM with GA with both TF and TF-IDF.

VIII. CONCLUSION

It is very important to have features that describe the data set properly or to discard irrelevant features to make an effective classification. Within the scope of this study, 3 different classification algorithms (SVM, ANN and CBA) have been applied together with feature selection methods (GI, IG and GA) on the preprocessed tweet data of followers of GSM operators for sentiment analysis. The best classification results have been achieved by using SVM classification with GA feature selection on both TF and TF-IDF term weighting methods.

The best hybrid model (SVM with GA) where(in which) the GA tries to find the most appropriate subset of attributes with high accuracy and small dimension can be used to obtain successful results in sentiment analysis on any texts written in daily speech Turkish language(Turkish texts written in colloquial language).

ACKNOWLEDGEMENT

We would like to express our special appreciation and thanks to Turkish Airlines for the financial support.

REFERENCES

- [1] Thadani Dimple R. and Christy MK Cheung, Online social network dependency: Theoretical development and testing of competing models, In System Sciences (HICSS), 44th Hawaii International Conference on IEEE, 2011, pp.1-9.
- [2] Neti Sisira, Social media and its role in marketing, International Journal of Enterprise Computing and Business Systems vol.1, no.2, pp. 1-15, 2011.
- [3] Kouloumpis Efthymios, Theresa Wilson and Johanna D. Moore. , Twitter sentiment analysis: The good the bad and the omg!, Icwsm, vol.164, no.11, pp.538-541, 2011.
- [4] Çoban Önder, Barış Özyer and Gülşah Tümüklü Özyer., Sentiment analysis for Turkish Twitter feeds, 23th In Signal Processing and Communications Applications Conference (SIU), 2015, pp. 2388-2391.
- [5] Kaya Mesut, Guven Fidan and Ismail H. Toroslu, Sentiment analysis of turkish political news, Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent

- Technology-Volume 01. IEEE Computer Society, p. 174-180, 2012.
- [6] Catal Cagatay and Mehmet Nangir, A sentiment classification model based on multiple classifiers, Applied Soft Computing vol.50, pp.135-141, 2017.
- [7] Aldoğan Deniz and Yusuf Yaslan, A comparison study on active learning integrated ensemble approaches in sentiment analysis, Computers & Electrical Engineering, vol.57, pp.311-323, 2017.
- [8] Kulcu Sercan and Erdogan Dogdu, A Scalable Approach for Sentiment Analysis of Turkish Tweets and Linking Tweets to News, In: Semantic Computing (ICSC), 2016 IEEE Tenth International Conference on. IEEE, 2016, pp.471-476.
- [9] Çetin Mahmut and M. Fatih Amasyali, Supervised and traditional term weighting methods for sentiment analysis, 21st Signal Processing and Communications Applications Conference (SIU), 2013, pp.1-4.
- [10] Parlar Tuba and Selma Ayşe Özel., A new feature selection method for sentiment analysis of Turkish reviews, Innovations in Intelligent Systems and Applications (INISTA), 2016 International Symposium on. IEEE, 2016, pp.1-6. G. R. Faulhaber, "Design of service systems with priority reservation," in *Conf. Rec. 1995 IEEE Int. Conf. Communications*, pp. 3-8.
- [11] Oğul Burçin Buket and Gönenç Ercan, Sentiment classification on Turkish hotel reviews, Signal Processing and Communication Application Conference (SIU), 2016, pp.497-500.
- [12] Yang Ang, et al, Enhanced Twitter Sentiment Analysis by Using Feature Selection and Combination, In Security and Privacy in Social Networks and Big Data (SocialSec), 2015 International Symposium on IEEE, 2015, pp. 52-57.
- [13] Go Alec, Richa Bhayani and Lei Huang, Twitter sentiment classification using distant supervision, CS224N Project Report, Stanford, vol.1, no.12, 2009.
- [14] Akba Fırat, et al, Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews, In 8th European Conference on Data Mining 2014, Vol. 191, pp. 180-184.
- [15] Boynukalin Zeynep, Emotion analysis of Turkish texts by using machine learning methods, M.Sc. dissertation, Middle East Technical University, 2012.
- [16] Liu Bing and Lei Zhang, A survey of opinion mining and sentiment analysis, Mining text data, pp.415-463, 2012.
- [17] Li Te-Sheng, Feature selection for classification by using a GA-based neural network approach, Journal of the Chinese Institute of Industrial Engineers, vol.23, no.1, pp.55-64, 2006.
- [18] Shang Wenqian, et al, A novel feature selection algorithm for text categorization, Expert Systems with Applications, vol. 33, no.1, pp.1-5, 2007.
- [19] Uğuz Harun, A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm, Knowledge-Based Systems vol.24, no.7, pp.1024-1032, 2011.
- [20] Parlar Tuba, Esra Saraç and Selma Ayşe Özel, Comparison of feature selection methods for sentiment

- analysis on Turkish Twitter data, 25th In Signal Processing and Communications Applications Conference (SIU), 2017, pp.1-4.
- [21] Ayata Değer, Murat Saraçlar and Arzucan Özgür, Turkish tweet sentiment analysis with word embedding and machine learning, 25th In Signal Processing and Communications Applications Conference (SIU), 2017, pp.1-4.
- [22] Hayran Ahmet and Mustafa Sert, Sentiment analysis on microblog data based on word embedding and fusion techniques, In Signal Processing and Communications Applications Conference (SIU), 2017, pp. 1-4.
- [23] Gunal Serkan and Rifat Edizkan, Subspace based feature selection for pattern recognition, Information Sciences, vol.178, no.19, pp.3716-3726, 2008.
- [24] Kuri-Morales Angel and Fátima Rodríguez-Erazo, A search space reduction methodology for data mining in large databases, Engineering Applications of Artificial Intelligence, vol.22, no.1, pp.57–65, 2009.
- [25] Piramuthu Selwyn, Evaluating feature selection methods for learning in data mining applications, European journal of operational research, vol.156, no.2, pp.483-494, 2004.
- [26] Tsai Chih-Fong, Feature selection in bankruptcy prediction, Knowledge-Based Systems, vol.22. no.2, pp.120-127, 2009.
- [27] Wang Jeen-Shing and Jen-Chieh Chiang, A cluster validity measure with outlier detection for support vector clustering, IEEE Transactions on Systems, Man, and Cybernetics – Part B Cybernetics, vol.38, no.1, pp.78–89, 2008.
- [28] Fragoudis Dimitris, Dimitris Meretakakis and Spiros Likothanassis, Integrating feature and instance selection for text classification, In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2002, pp. 501 506.
- [29] Derrac Joaquín, Salvador García and Francisco Herrera, A survey on evolutionary instance selection and generation, International Journal of Applied Metaheuristic Computing, vol.1 no.1, pp.60–92, 2010.
- [30] Gupta Deepak Kumar, Kandula Srikanth Reddy and Asif Ekbal, Pso-ament: Feature selection using particle swarm optimization for aspect based sentiment analysis, In International Conference on Applications of Natural Language to Information Systems, 2015, pp. 220-233.
- [31] Zhu Jian, Hanshi Wang and JinTao Mao, Sentiment classification using genetic algorithm and conditional random fields, In 2nd IEEE international conference on information management and engineering (ICIME), 2010, pp.193–196.