

# The Analysis of Text Categorization Represented With Word Embeddings Using Homogeneous Classifiers

Zeynep Hilal Kilimci

Department of Computer Engineering  
Dogus University  
Acibadem, İstanbul, 34722, Turkey  
hkilimci@dogus.edu.tr

Selim Akyokus

Department of Computer Engineering  
Medipol University  
Beykoz, İstanbul, 34810, Turkey  
sakyokus@medipol.edu.tr

**Abstract**— Text data mining is the process of extracting and analyzing valuable information from text. A text data mining process generally consists of lexical and syntax analysis of input text data, the removal of non-informative linguistic features and the representation of text data in appropriate formats, and eventually analysis and interpretation of the output. Text categorization, text clustering, sentiment analysis, and document summarization are some of the important applications of text mining. In this study, we analyze and compare the performance of text categorization by using different single classifiers, an ensemble of classifiers, a neural probabilistic representation model called word2vec on English texts. The neural probabilistic based model namely, word2vec, enables the representation of terms of a text in a new and smaller space with word embedding vectors instead of using original terms. After the representation of text data in new feature space, the training procedure is carried out with the well-known classification algorithms, namely multivariate Bernoulli naïve Bayes, support vector machines and decision trees and an ensemble algorithm such as bagging, random subspace and random forest. A wide range of comparative experiments are conducted on English texts to analyze the effectiveness of word embeddings on text classification. The evaluation of experimental results demonstrates that an ensemble of algorithms models with word embeddings performs better than other classification algorithms that uses traditional methods on English texts.

**Keywords**— Classifier ensembles; deep learning; text data mining; word embeddings.

## I. INTRODUCTION

Text mining is a major research field because of very large amounts of unstructured information in the form of text available in different fields. New and better methods are needed to discover information and knowledge from unstructured texts. The aim of text categorization methods is to classify a given text document into one of predefined categories by using machine learning algorithms. In text categorization, the first step is to use supervised learning techniques to build classifiers from a set of training documents. The classifiers learn and constitute a model of relationship between features and class labels. Once trained, the classifiers can be used to predict the category of a new text document from a test dataset. A text categorization task generally consists of the following steps: tokenization, document parsing, stemming, stop words removal, and representation of documents in convenient

formats, reduction of features, selection of classifier algorithms, training and testing. The bag of words model is a frequent method used to represent the text documents. In bag of words model, a set of documents is represented as a document-term matrix where each row corresponds to a document and each column corresponds to a term (word). Each cell of the matrix contains a weight that expresses the significance of a term according to the related document and/or document collection. Different term weighting methods such as Boolean, term frequency (TF) and term frequency-inverse document frequency (TF-IDF) can be employed to represent each weight in document-term matrix. Naïve Bayes (NB), decision trees (DT), artificial neural networks (ANN) k-nearest neighbor (k-NN), and support vector machines (SVM) are some of the most widespread and frequently used classification algorithms on text categorization area.

Ensemble learning approach uses a set of multiple classifiers to achieve a better performance than a single classifier. Ensemble learning based classifiers are also named multiple classifier systems, ensemble-based classifiers, classifier ensembles, or just ensemble systems. An ensemble-based classifier is a group of combined single classifiers. Each single classifier is named an individual, base or weak learner. On the training phase, each individual classifier is trained separately with a given training data set. An ensemble classification task consists of two steps: ensemble generation and integration (aggregation, combination or fusion). In ensemble generation step, a diverse set of individual classifiers is produced by using the training data set. In the integration step, the predictions of the trained individual classifiers are combined to obtain a joint decision of the whole system. Consequently, the main objective in ensemble approach is to generate many classifiers and integrate their outputs such that the consolidation of classifiers enhances the performance of classification system. The performance of an ensemble system is related with the diversity of the base classifiers that constitute the ensemble and the integration strategy. Each base classifier must display of some level of diversity among themselves. Data, parameter, and structural diversity are three different approaches to ensure the diversity of ensemble system [1]. Different data subsets are produced for each individual learner by employing data resampling techniques in data diversity. Parameter diversity is ensured with the usage of different

parameter settings for each different base learner. The structural diversity is also provided with the different learning algorithms for each base learner. The usage of different learning algorithms in the structural diversity is called heterogeneous ensemble system. The first two types of diversity mentioned above is also called homogeneous ensemble system because of the usage of the same learning algorithm for each of base learners.

The success of machine learning algorithms largely depends on the representation of features for a given problem. This is especially very important in text mining and natural language processing because of large number of features. In recent years, representation learning is a very popular approach that enables to learn representations of terms/words, called word embeddings, from a collection of text [2, 3]. A word embedding is a learned representation of a term/word obtained from a collection of text by using learning algorithms. Word embeddings are usually real-valued dense vectors that represent the significance of a word and its relationships with other words used in similar contexts. Word embeddings became very popular with the introduction of Word2vec by Mikolov et al. in 2013 [4]. Word2vec algorithm takes a collection of text documents as its input and produces a set of vectors for each term in the collection by using a two-layer neural network. Each vector consists of real-valued numbers in several hundred dimensions. Typically, the dimensions range from 100 to 300. Word2vec enables grouping or placement of vectors in a new vector space for words that appears in nearby locations or in a similar context.

In this paper, we try to harness the power of Word2vec representation learning method with ensemble systems to increase the text classification accuracies. Word embeddings are obtained from the training data through word2vec tool by using shallow neural networks. Two variants of naive Bayes (NB) namely, multivariate Bernoulli NB (MVNB), and multinomial NB (MNB), support vector machine (SVM), and decision tree (DT) are used as single classifiers. We tested performance of the single classifiers by using original features and word embeddings at the first step. After that, ensemble methods bagging, random subspace, and random forest are applied to see the effect of a diverse set of an ensemble of homogeneous ensemble classifiers. Extensive experiments on widely used English datasets show that the proposed model enhances the classification success of the homogeneous ensemble system.

The rest of the paper is organized as follows: Related researches on the use of word embeddings, and ensemble systems on text classification field are presented. In the next section, the proposed framework is expressed. Then, experiment setup and results are given in the subsequent sections. Finally, the paper is concluded with a discussion and conclusion.

## II. RELATED WORK

There has been little research to solve text categorization problem using ensemble systems. We review some of research that utilizes ensemble methodologies for text categorization. One of early studies on ensemble systems is accomplished by Larkey and Croft in [5]. Authors make use of heterogeneous ensemble system by consolidating k-nearest-neighbor,

relevance feedback and Bayes classifiers on medical text documents. In a recent study [17], Onan et al., analyze the predictive performance of ensemble learning methods on text documents by evaluating keywords. They carry out different feature extraction algorithms to acquire the keywords. After that, authors concentrate on five different ensemble techniques and four different base classifiers to obtain the diversity of ensemble system. In [18], authors concentrate on ensemble of multiple classifiers on Turkish and English datasets. They implement a comparative analysis of the impact of the ensemble techniques for text categorization field in both Turkish and English languages. The same type of base classifiers but diversified training sets are employed through various ensemble algorithms such as bagging, boosting, random subspace and random forest. Multivariate Bernoulli naïve Bayes is selected as a base classifier due to its superior classification performance mentioned in this work.

The other recent work [19] focuses on the heterogeneous ensembles that employs different types of individual classifiers. Authors propose to demonstrate the effectiveness of heterogeneous classifier ensembles using naïve Bayes, support vector machine and random forest algorithms on Turkish texts. They conclude that heterogeneous ensemble systems enhance the classification success on Turkish text and propose to be tried with similar studies on the other agglutinative languages. In [20], authors analyze the classifier ensembles with extended space forests by extending feature space with various feature selection techniques on Turkish texts. They utilize random combinations of features and significant features with gain ratio to extend the feature space. Naïve Bayes as a classification algorithm and Bagging, Random Subspace and Random Forest as ensemble algorithms are evaluated in the experiments. They report that the enhancement of the feature space and the usage of ensemble strategies significantly contributes the classification performance.

Recent research on the use of word embedding representations and deep learning especially focuses on sentiment analysis of short texts. In [21], authors use a simple convolutional neural network model to analyze the sentiment of Twitter data. The other work [22] analyze the importance of keywords to evaluate the semantics between them. Long-short term memory and gated recurrent are performed as deep learning models on IMDB and SemEval-2016 datasets. Enhancement in accuracy results prove that the usage of keywords and deep learning models contribute to the classification success. The study [23] proposes a novel recurrent random walk network by employing posted tweets and social relationships which is called heterogeneous microblog sentiment classification. This approach implements the training procedure by using deep neural networks with random-walk layer and back-propagation technique. Experiment results boost the classification success compared to the state-of-the-art studies. The study [24] is constructed on word and character level embedding of Twitter dataset by utilizing deep learning models such as long-short term memory and convolutional networks. The comparison in character-based architecture with long short-term memory embedding, convolutional embedding, convolutional embedding freeze, convolutional character level embedding and conventional

support vector machine algorithm is presented. Especially, it is observed that the classification success of character-based convolutional architecture is noteworthy with respect to others.

### III. PROPOSED APPROACH

We try increase the text classification performance by using both word embeddings and ensemble strategies. Word embeddings is a representation learning method that maps words in a text collection into vectors of real numbers. Word embedding representation method enables to capture semantic and syntactic similarity, and relationships among words in a given context. Because of its representation characteristics of word embeddings, they are frequently employed in NLP tasks, and sentiment analysis. There are many methods for getting word embeddings like neural networks, dimension reduction methods, co-occurrence matrices, and probabilistic models. Word embeddings became very popular after Word2vec method developed by Tomas Mikolov at Google in 2013 [4]. Tomas Mikolov produced a word embedding toolkit which includes a group of related methods that are utilized to make word embeddings faster than the previous approaches. The Word2vec model uses shallow two-layer neural networks to process text to acquire the linguistic contexts of words. The output of the system is a set of feature vectors which compose a vector space with several hundred dimensions. The related words appear in neighboring positions in the vector space. As a first step, we used Word2vec toolkit to generate word-embeddings from our datasets.

After the word embedding is obtained through word2vec tool, the training phase of the ensemble system is implemented. For this purpose, multivariate Bernoulli naïve Bayes (MVNB), multinomial naïve Bayes (MNB), support vector machines (SVM), decision trees (DTs) are employed as learning algorithm to choose a base learner for the homogeneous ensemble system. Then, bagging (BG), random subspace (RS), and random forest (RF) are utilized as ensemble strategies to provide the diversity. BG and RS ensemble methods use MVNB as a base classifier which performs as the best algorithm among other single classifiers. The RF method uses decision trees as base classifiers.

Based upon Bayes' theorem, the naïve Bayes algorithm is a simple probabilistic model which assumes the independence of features (words) from each other. There are two different variants of naïve Bayes classifier commonly applied for text categorization. These two variants are named as multivariate Bernoulli naïve Bayes (MVNB) and multinomial naïve Bayes (MNB). In MVNB, each document is represented by a vector with binary values that can become 1 or 0 depending upon the occurrence or absence of a word in the document. In MNB, the weights of each document vector are denoted with the frequency of words that is counted in the document [6]. Support vector machines (SVM) are commonly used supervised learning algorithms that use a decision plane to separate set of instances of two different categories. Though SVM is a linear

classifier, it can be also used for a non-linear classification problems using the kernel trick that implicitly maps instances of a dataset into high-dimensional feature spaces [7]. Different kernel functions can be applied to produce a set of diverse classifiers with different decision boundaries. Decision trees (DTs) are simple and supervised learning algorithms that enable easy interpretation and visualization of relations among features. Besides, it can be used to construct regression trees and other classification models [11].

Bagging is one of the most common ensemble methods to ensure the data diversity. A subset from the entire training dataset is randomly selected with replacement by resampling to achieve diversity. Each data subset is used to train a different base learner. A final decision of ensemble system is maintained by applying a majority vote [11]. Random subspace (RS) ensemble is like bagging but a random subset of features is chosen from the dataset instead of instances [8, 12]. Given a dataset with  $m$  features, RS randomly chooses  $m^*$  features where  $m^* < m$ . The procedure of selection is repeated  $K$  times to get  $K$  different feature subsets. Then each of  $K$  feature subset is trained with  $K$  base classifiers. The final decision of the ensemble system is obtained by consolidating decisions of  $K$  different base classifiers by a voting scheme. Random forests are a group of decision tree classifiers. A mixture of bagging and random subspace methods corresponds to the random forest approach. Selection of training subsets for each individual decision tree constitutes the bagging part. Splitting of each node in decision tree is implemented on the random subset of features which corresponds to the random subspace approach. This method enables additional randomness to the algorithm in addition to bagging. Random forest endures against overfitting owing to randomness applied in both sample and feature spaces [9].

### IV. EXPERIMENT SETUP

There are four widely-used benchmark datasets with different sizes and properties to observe the classification performances of ensemble learning on datasets represented with Word2vec representation learning method. These include three versions of Newsgroups<sup>1</sup> dataset. The first one is the original 20 Newsgroups dataset and it is called as 20News-19997. The second one is named 20News-18828 and it covers a smaller number of documents than the original dataset where duplicate postings are taken out. Besides, posting headers are removed except from and subject headers. It encloses roughly 20,000 documents and twenty different categories. The third one is a small subset of the original dataset composed of 100 postings per class and it is called as mini newsgroups<sup>2</sup>. All of them are divided into twenty different categories. The last dataset is the WebKB<sup>3</sup> which includes web pages collected from computer science departments of different universities. These web pages are constituted of seven categories (faculty, student, staff, course, project, department and other) and approximately 8,300 pages. Another version of WebKB is called WebKB4 where the

<sup>1</sup> <http://people.csail.mit.edu/people/jrennie/20Newsgroups>

<sup>2</sup> <http://kdd.ics.uci.edu/databases/20newsgroups/20newsgroups.html>

<sup>3</sup> <http://www.cs.cmu.edu/~textlearning>

number of categories is reduced to four categories that are used in our experiments.

TABLE I. CHARACTERISTICS OF THE DATASETS

Dataset	C	D	V
20News-19997	20	19,997	43,553
20News-18828	20	18,828	50,570
Mininews	20	2,000	13,943
WebKB4	4	4,199	16,116

Characteristics of the datasets without any preprocessing are given in Table 1 which lists the number of classes ( $|C|$ ), the number of documents ( $|D|$ ) and the vocabulary size ( $|V|$ ). We focus on the frequent terms whose document frequency is greater than two and do not carry out any stop-word filtering or stemming methods to avoid from any bias that can be introduced by stop list or stemming algorithms.

Experiments are performed with the training set size 80% of the data for training and the rest for testing. The widely-applied 5x2 cross validation approach is implemented on each dataset. This approach is like the previous works [6, 19, 13, 14, 15, 16] where they use 80% of data for training data and 20% for test. In ensemble learning, the number of base classifiers is generally adjusted between 50 and 100 [13, 14]. We take the number of base learners as 100 owing to its success. Finally, the classification accuracies are utilized as the evaluation criteria in our experiments. Besides, word2vec model used in experiments employs a 200-dimensional vector to represent words and the training window is set to 5.

## V. EXPERIMENT RESULTS

At first, word embedding is extracted from the original datasets through word2vec. Then, we observe the classification success of individual classifiers on both original datasets and word embedding of them to determine the base classifier of ensemble system.

TABLE II. THE CLASSIFICATION ACCURACIES OF SINGLE CLASSIFIERS ON ORIGINAL DATASETS

Dataset	MVNB	MNB	SVM	DT
20News-19997	75.68±1.02	76.06±0.84	61.84±1.13	54.15±0.77
20News-18828	88.52±0.42	86.26±0.36	84.18±0.96	79.82±0.72
Mininews	87.25±1.10	77.90±0.61	83.00±0.42	77.12±0.56
WebKB4	84.10±1.12	85.64±1.17	89.85±0.96	77.45±1.07
avg	<b>83.89±0.92</b>	81.47±0.75	79.72±0.87	72.14±0.78

The learning methods are multivariate Bernoulli naïve Bayes (MVNB), multinomial naïve Bayes (MNB), support vector machines (SVM), and decision trees (DTs) are compared to determine the base learner. MVNB exhibits the superior classification performance among other machine learning algorithms on both versions of datasets. For this reason, MVNB is specified as a base learner of the ensemble system with 83.89% average accuracy on original datasets and 85.41%

mean accuracy on dataset represented with word embeddings as seen in Table 2 and Table 3. We can summarize the classification success order of the individual classifiers like this: MVNB > MNB > SVM > DT. Furthermore, the usage of word embedding enhances the classification success about 2% in terms of mean accuracy results of MVNB compared to the original datasets.

TABLE III. THE CLASSIFICATION ACCURACIES OF SINGLE CLASSIFIERS ON WORD EMBEDDINGS

Dataset	MVNB	MNB	SVM	RF
20News-19997	77.13±0.78	76.80±0.55	66.13±0.45	60.32±0.63
20News-18828	89.45±0.34	86.71±0.60	85.44±0.87	80.20±0.80
Mininews	88.70±0.94	80.44±0.92	86.23±1.00	78.14±1.06
WebKB4	86.35±0.47	87.36±1.02	90.55±0.74	79.51±0.52
avg	<b>85.41±0.63</b>	82.83±0.77	82.09±0.77	74.54±0.75

After determination of the base learner, ensemble algorithms bagging, random subspace, and random forest are employed to get decisions of an integrated set of diverse classifiers for both original datasets and dataset with word embeddings as shown in Table 4 and Table 5. As an ensemble algorithm, bagging is not efficient to boost the classification performance for all datasets on both original datasets and datasets with word embeddings. Bagging method decreases the accuracy results approximately 1% for 20News\_18828 and Mini news while it slightly enhances the success of MVNB classifier for 20News-19997 and WebKB4 in Table 4. Bagging accuracies are close to classification results of base learner MVNB for all datasets with word embeddings in Table 5. In other words, the utilization of bagging as an ensemble algorithm does not provide a notable improvement on the classification performance.

TABLE IV. THE CLASSIFICATION ACCURACIES OF ENSEMBLE ALGORITHMS ON ORIGINAL DATASETS

Dataset	MVNB	BG	RS	RF
20News-19997	75.68±1.02	76.13±0.74	77.91±0.81	80.44±0.36
20News-18828	88.52±0.42	87.63±0.97	88.63±0.77	92.02±0.73
Mininews	87.25±1.10	86.78±0.65	88.24±1.04	90.10±0.65
WebKB4	84.10±1.12	84.75±0.86	85.32±0.83	88.75±1.19
avg	83.89±0.92	83.82±0.81	85.03±0.86	<b>87.83±0.73</b>

Random subspace improves roughly 2% of the accuracy results in comparison with both bagging algorithm and base learner of the system (MVNB) in terms of averaged accuracy results. While random subspace provides nearly 1-2% improvement compared to the performance of the single classifier, the classification success of it also ranges from 1% to 2% in proportion to bagging method on original datasets as observed in Table 4. Except WebKB4 dataset, random subspace algorithm provides approximately 2% enhancement on datasets represented with word embeddings compared to the base learner and bagging method as seen in Table 5. On the other hand, random subspace and random forest demonstrate obvious

classification performance compared to the bagging algorithm and MVNB.

TABLE V. THE CLASSIFICATION ACCURACIES OF ENSEMBLE ALGORITHMS ON WORD EMBEDDINGS

Dataset	MVNB	BG	RS	RF
20News-19997	77.13±0.78	78.24±0.35	79.90±0.67	83.27±0.47
20News-18828	89.45±0.34	89.11±0.71	91.42±0.90	95.10±0.88
Mininews	88.70±0.94	88.06±0.50	90.56±1.15	92.73±0.40
WebKB4	86.35±0.47	86.00±0.29	86.88±0.53	89.27±1.04
avg	85.41±0.63	85.35±0.46	87.19±0.81	<b>90.09±0.70</b>

A close inspection of Table 4 and Table 5 makes clear that the random forest algorithm is better than other algorithms in terms of the classification success and it overwhelmingly outperforms others for all datasets. Random forest provides the considerable contribution to the classification success and it maintains minimum 2% and maximum 5% enhancement as compared to the success of the base classifier, bagging, and random subspace algorithms. The lowest and highest accuracy improvements differ according to the datasets 1% - 6% for 20News-19997, 2% - 6% for 20News-18828, 2% - 4% for Mini news, and maximum 3% increment for WebKB4 dataset with word embeddings. Moreover, enhancements on accuracy results are varied according to the datasets 1% - 5% for 20News-19997, maximum 4% improvement for 20News-18828, maximum 3% rise for Mini news, and 1% - 4% increment for WebKB4 on original datasets as shown in Table 4. Especially on 20News-19997 and 20News-18828 datasets, 6% rise shows that a significant contribution to the classification success is obtained with the utilization of ensemble strategies and word embeddings. Considering the mean values of the accuracies, the classification success of ensemble systems can be ordered as: RF > RS > BG. As a result, we can say that random forest should be chosen as the best ensemble algorithm among others to improve the classification success on text data categorization field.

## VI. DISCUSSION AND CONCLUSIONS

In this work, we analyzed the effectiveness of Word2vec representation learning integrated with homogeneous classifier ensembles on supervised text categorization. For this purpose, word2vec model is initially constructed from the original datasets to obtain word embeddings that provide a representation that enables a syntactic and semantic relationships among words (features) in nearby locations. Second, learning algorithms multivariate Bernoulli naive Bayes, multinomial naive Bayes, support vector machines, and decision trees are employed to decide the base classifier of the ensemble system. After that, ensemble models bagging and random subspace that use the multivariate Bernoulli naïve Bayes as a base classifier and random forest algorithms are applied on four widely-used datasets represented with word-embeddings to evaluate their classification successes.

We acquired noteworthy experimental results which point out that the choice of bagging as an ensemble technique does

not provide major contribution compared to the random subspace and random forest algorithms on both original datasets and datasets with word embeddings. The random forest algorithm is the best ensemble algorithm that represents the highest classification success on both versions of the datasets. The random subspace is the subsequent best ensemble technique that is better than BG. Thus, it can be concluded that the random forest algorithm is the best ensemble method that advances the overall classification performance of the system. Eventually, the overall performance order of the ensemble systems is as follows: RF > RS > BG >= MVNB. It is significant to emphasize that the use of word embeddings representation instead of original datasets together with ensemble models enhances the classification performance of the system even further.

As well as the classification success of proposed system, execution time analysis is also evaluated in terms of training time. More training time is necessitated for both the word embedding models and classifier of mixtures compared to the original dataset and individual classifiers. When the Word2vec word embedding model and homogenous classifier ensembles are employed, the training time for all datasets and experiments is around 32h24min using 12 threads in a Intel® Xeon® E5-2643 3.30 GHz machine. If original datasets and homogenous classifier ensembles are utilized instead of Word2vec word embedding model, the training time for all datasets and experiments is about 15h45min. We consider that performing comprehensive experiments for a GPU environment can have a great impact on the training time performance.

In conclusion, extensive experiments by using different ensemble algorithms, datasets, and word embeddings demonstrate that the consolidation of representation learning-based features and ensemble models enhances the overall classification performance of the system in text categorization. In future, we plan to build heterogeneous ensemble system that utilize different types of individual classifiers and apply deep learning algorithms on different datasets encoded with different representation learning methods.

## REFERENCES

- [1] Ren Y., Zhang L. and Suganthan P.N., "Ensemble Classification and Regression-Recent Developments, Applications and Future Directions: Review Article", IEEE Computational Intelligence Magazine, Vol. 11, No. 1, pp. 41-53, 2016.
- [2] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. 2003. A neural probabilistic language model. *J. Mach. Learn. Res.* 3 (March 2003), 1137-1155.
- [3] Y. Bengio, A. Courville, P. Vincent, "Representation learning: a review and new perspectives", IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798-1828, 2013.
- [4] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'13)*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.), Vol. 2. Curran Associates Inc., USA, 3111-3119.
- [5] Larkey L.S. and Croft W. B., "Combining Classifiers in Text Categorization," In *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1-17, 1996.

- [6] McCallum A. and Nigam K. A., " Comparison of Event Models for Naive Bayes Text Classification," In Workshop on Learning for Text Categorization, pp. 41-48, 1998.
- [7] Burges C.J.C., "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining on Knowledge and Discovery, vol. 2, no. 2, pp. 121-167, 1998.
- [8] Ho T.K., "Random Subspace Method for Constructing Decision Forests", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 8, pp. 832-844, 1998
- [9] Breiman L., "Random Forests", Machine Learning, Vol. 45, No.1, Pp. 5-32, 2001.
- [10] Rennie J.D.M., Shih L., Teevan J. and Karger D.R., "Tackling the Poor Assumptions of Naive Bayes Text Classifiers", International Conference on Machine Learning, pp. 616-623, 2003.
- [11] Rokach L. and Maimon O., "Top-Down Induction of Decision Trees Classifiers - A Survey", IEEE Transactions on Systems, Man and Cybernetics Part C, Vol. 35, No. 4, pp. 476-487, 2005.
- [12] Panov P. and Džeroski S., "Combining Bagging and Random Subspaces to Create Better Ensembles", International Conference on Intelligent Data Analysis, pp. 118-129, 2007.
- [13] Amasyalı M.F. and Ersoy O.K., "Classifier Ensembles with the Extended Space Forest", IEEE Transactions on Knowledge and Data Engineering, Vol. 26, pp. 549-562, 2014.
- [14] Adnan M.N., Islam M.Z. and Kwan P.W.H. , "Extended Space Decision Tree", International Conference on Machine Learning and Cybernetics, pp. 219-230, 2014.
- [15] Kilimci Z.H., Ganiz M.C., "Evaluation of Classification Models for Language Processing", International Symposium on INnovations in Intelligent SysTems and Applications, pp. 1-8, 2015.
- [16] Kilimci Z.H., Akyokus S., "N-Gram Pattern Recognition using Multivariate Bernoulli Model with Smoothing Methods for Text Classification", IEEE Signal Processing and Communications Applications Conference; pp. 79-82, 2016
- [17] Onan A., Korukoglu S. and Bulut H., "Ensemble of Keyword Extraction Methods and Classifiers in Text Classification", Expert Systems with Applications, Vol. 57, pp. 232-247, 2016.
- [18] Kilimci Z.H., Akyokus S. and Omurca S.I., "The Effectiveness of Homogenous Ensemble Classifiers for Turkish and English Texts", International Symposium on INnovations in Intelligent SysTems and Applications, pp. 1-7, 2016.
- [19] Kilimci Z.H., Akyokus S. and Omurca S.I., "The Evaluation of Heterogeneous Classifier Ensembles for Turkish Texts", International Symposium on INnovations in Intelligent SysTems and Applications, pp. 1-5, 2017.
- [20] Kilimci Z.H. and Omurca S.I., "A Comparison of Extended Space Forests for Classifier Ensembles on Short Turkish Texts", International Academic Conference on Engineering, IT and Artificial Intelligence, Pp. 96-104, 2017.
- [21] Liao S., Wang J., Yu R., Sato K. and Cheng Z., "CNN for Situations Understanding Based on Sentiment Analysis of Twitter Data", Procedia Comput Science, Vol. 11, pp. 376-381, 2017.
- [22] Hu F., Li L., Zhang Z., Wang J. and Xu X., "Emphasizing Essential Words for Sentiment Classification Based on Recurrent Neural Networks", Journal of Computer Science and Technology, Vol. 32, No. 4, pp. 785-795, 2017.
- [23] Zhao Z., Lu H., Cai D., He X. and Zhuang Y., "Microblog Sentiment Classification via Recurrent Random Walk Network Learning", International Conference on Artificial Intelligence, pp. 3532-3538, 2017.
- [24] Becker W., Wehrmann J., Cagnini H.E.L. and Barros R.C., "An Efficient Deep Neural Architecture for Multilingual Sentiment Analysis in Twitter", International Conference Florida on Artificial Intelligence Research Society, pp. 246-251, 2017.