Financial Sentiment Analysis for Predicting Direction of Stocks using Bidirectional Encoder Representations from Transformers (BERT) and Deep Learning Models

Derya Othan, Zeynep Hilal Kilimci and Mitat Uysal

***Abstract*— Stocks are an important type of investment that is affected by the economic crisis. For this reason, forecasting the direction of stocks is a significant for investors, analysts, and researchers. In this study, we propose to predict the direction of stocks in Turkish stock market (BIST100) by employing Turkish texts such as social media platforms. For this purpose, different deep learning methodologies such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short Term Memory Networks (LSTMs) and a new generation natural language processing model, namely Bidirectional Encoder Representations from Transformers (BERT) are employed for classification task. To the best of our knowledge, this is the first study for analyzing the direction of stocks by using both CNN, RNN, LSTM, BERT and texts from social media platforms. Experimental results represent that the utilization of deep learning models and a new generation word embedding model which is called BERT boosts the classification performance. In conclusion, using deep learning algorithms and a new generation of natural language processing model provide a valuable contribution to investors presents 96.26% of accuracy performance by predicting direction of stocks.**

***Keywords*— BERT, deep learning, financial sentiment analysis, stock prediction, text classification.**

I. INTRODUCTION

The economic crises in the world directs the stock markets. With the development of internet and mobile technology, investors are able to predict the direction of investment on many social media platforms shared comments. Recently, user opinions on stocks, news releases on financial sites, and technical analysis have become an important source for obtaining detailed information about the directions in the stock market [1]. Twitter (https://twitter.com) is the most popular social networking service known to share information as it is and connect with others in real time with nearly 350 million active users [2]-[3]. Social media platforms like Twitter allow investors to understand the investor experience, share news, and investor sentiment.

In recent years, deep learning algorithms have been very popular in different research areas such as image/video processing, natural language processing, pattern recognition. The reason behind of to be preferred these models is that generating better predictions compared to the conventional machine learning algorithms. Deep learning models are mainly used to provide automatic feature extraction by training complex features with minimal external support to ensure meaningful representation of data through deep neural networks. In addition, deep learning methods such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), long-term memory networks (LSTMs) [4], the next generation natural language processing (NLP) model Bidirectional Encoder Representations from Transformers (BERT) [5] are used in classification tasks in many areas.

In this study, we propose to estimate the direction of stocks by analyzing individual and organizational user comments including announcements, financial technical analysis as a valuable resource for the investors. For this purpose, Turkish texts are gathered from Twitter by employing web-crawler. After getting textual data, various pre-processing techniques are implemented to remove the influence of dirty data. User comments, news releases, and technical analyzes are classified as negative, or positive by using deep learning models (CNN, RNN, LSTM) and a new generation word embedding model (BERT). To our knowledge, this is the very first attempt to predict the direction of stock market by employing both deep learning models and a new generation word embedding model. In order to demonstrate the contribution of our study, we used Turkish texts aforementioned in our experiments. Experiment results show that the inclusion of deep learning models and BERT improves the classification success of system.

The rest of the article is organized as follows: Section 2 provides a summary of the studies on stock market forecasting. Section 3 presents materials and methods used in the experiments. Experimental results and conclusions are given in Section 4 and Section 5.

II. RELATED WORK

In this section, a brief summary of the literature review of the studies focused on the stock exchange forecast is presented.

In [6], it is shown that artificial neural network models can be used to give direction to investors by giving information about stocks in IMKB. In [6], it is shown that artificial neural network models can be used to give direction to investors by giving information about stocks in Istanbul Stock Exchange (IMKB). In another study [7], stacked auto coders used a deep learning structure to predict the direction of stock prices traded on the USA Nasdaq Stock Exchange. The performance of the model is

This work was supported by Turkey Finance Participation Bank.

D. O. Author is with the Turkey Finance Participation Bank, TR 34764 TURKEY.

Z.H.K. (corresponding author) and M. U. Authors are with Dogus University, Istanbul, TR 34722 TURKEY..

https://doi.org/10.17758/URUAE8.UL12191013

evaluated with accuracy, and F-criterion and the proposed model presents the best performance with the Deep Convolutional Neural Networks (DVM) method. Chen et. al [8] estimate stock returns using historical stock data in China stock exchange with the LSTM model. Compared with the random estimation method, the LSTM model is more successful in estimating stock returns. Gündüz et. al [9] propose to forecast the status of the stocks in BIST100 by employing Evolutionary Neural Network (ESA). The classification performance obtained by ESA was higher than that obtained by chi-square attribute selection and logistic regression classifier.

In another study [10], successful results are obtained by applying deep learning methods to the stock market forecasting problem. In [11], the accuracy rate of the estimation was 94.21% using LSTM deep learning method in Turkish texts. In [12], ARIMA, machine learning algorithms, and LSTM as a deep learning technique are applied to BIST30 stock prices by examining the data of 30 leading companies of Borsa Istanbul. As a result of computational analysis, it is observed that ARIMA performs better than LSTM. Additionally, linear regression outperforms compared to the other machine learning techniques. In [13], the data of 25 leading companies of BIST100 are analyzed and several prediction algorithms are implemented. As a result, it is observed that Random Forest algorithm exhibits the best classification results with 57.37% of accuracy. In [14], the daily movement direction of the three stocks (GARAN, THYAO and ISCTR) which are frequently traded on BIST100 was estimated using deep neural networks. To perform the estimation process, the type of the deep neural network (Convolutional Neural Network) was trained and grading performance was evaluated with accuracy and F-measurement criteria. In the experiments, the movement directions of GARAN, THYAO and ISCTR stocks were estimated with an accuracy ratio of 0.61, 0.578, and 0.574, respectively.

Our study differs in the literature studies mentioned above in that it uses textual data from social media platform. Furthermore, another novelty is that the usage of BERT as a new generation word embedding model in addition to CNN, RNN, LSTM as deep learning models to estimate the direction of stocks. In contrast to literature research, we propose to provide a valuable resource to guide investors' forecasting of stock quotes using new generation models.

III. Proposed Model

A summary of the methods, materials, and proposed framework are presented in this section.

A. Data Collection

In this study, in order to estimate the direction of the Turkish Stock Exchange, the movements of the bank stock whose transaction volume is enormous in BIST100 are examined. For this purpose, we estimate the direction of the stocks of Akbank, Albaraka, Garanti, Halkbank, İş Bankası, Türkiye Sınai Kalkınma Bankası, Vakıfbank and Yapıkredi by analyzing user comments from Twitter. Turkish user comments from Twitter are collected with the label of each share using abbreviations of stocks. These are AKBNK, ALBRK, GARAN, HALKB, ISCTR, SKBNK, TSKB, VAKBN, YKBNK. Selenium browser [15] is used to gather user comments from Twitter environment. This allows us to collect as many tweets as we like without worrying about the limit issue allowed by the Twitter API. A total of 117,136 user reviews are collected between 1 September 2018 and 1 September 2019 with the labels “AKBNK”, “ALBRK”, “GARAN”, “HALKB”, “ISCTR”, “SKBNK”, “TSKB”, “VAKBN”, “YKBNK”.

B. Deep Learning Algorithms

In this study, we first use three commonly used deep learning algorithms such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and Long Short Term Memory Networks (LSTM).

- ***CNN***

Convolutional neural networks (CNNs) are a kind of multilayered sensor (MLP). The use of the CNN structure has advantages such as [16] - [17]:

- Layers deepen more than normal.
- Improves calculation performance of system by using activation functions such as ReLU, dropout, and batch normalization.
- Connections between the network layers increase with the discovery of backpropagation algorithm.

- ***RNN***

Recurrent neural network (RNN) is a type of neural network in which the output from the previous step is fed as input to the current stage. RNN is recommended because of the need to remember words. This problem is solved with the help of hidden layers [4] - [18] - [19]. The most important feature of RNN is the secret state that remembers some information about a sequence. RNN has a “memory” that remembers all the information about the calculated. RNN, unlike other neural networks, reduces the complexity of the parameters. Performs the same task on all inputs or hidden layers to produce the output. Using the same parameters for each input reduces the complexity of the parameters.

- ***LSTM***

Long Short Term Memory networks (LSTM) are a special type of RNN that can learn long term dependencies. The starting point is to provide a solution to the exponential error growth problem using the back propagation algorithm while training deep neural networks. The main reason for this problem is that the values generated by the activation function are constantly in the range of -1, 1, so that these values are given to the back propagation algorithm and multiplied by zero. LSTM, which is developed to avoid the problem of activation function and to better design complex structures with learning algorithms, gives good results in the problems that long-term dependencies and long-term information should be remembered [4] - [20].

C. Deep Contextualized Word Representation Model

Word embedding is the matching of real numbers to vectors using neural network, probabilistic model, or size reduction on the co-existence matrix of words. Word embedding is also known as language modeling, and feature learning technique. Word embedding is also called in different ways such as distributed semantic model, distributed representation, semantic vector space, or vector space model. For example, currencies such as "dollar", "euro", and "pound" are placed in close proximity in terms of their semantics, but the word "monkey" will be located far from these words in word vector. In a broader sense, the currency vector that will be located away from the vector representation of the monkeys is constructed by using word embedding model.

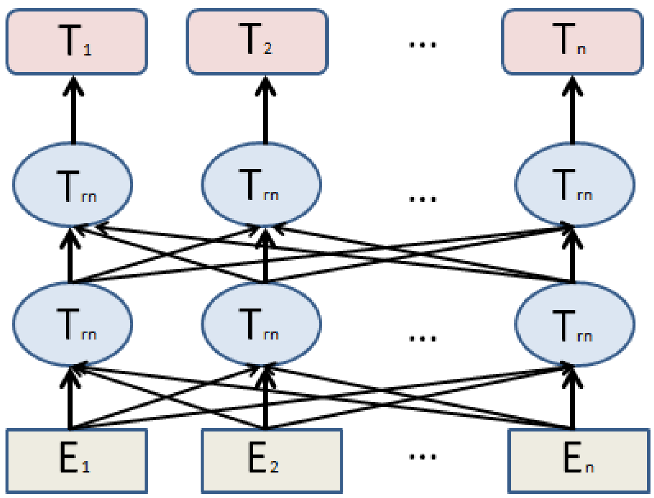

Fig. 1. The architecture of BERT model.

Bidirectional Encoder Representations from Transformers (BERT) is a new generation word embedding model, which means bidirectional encoder representations. Unlike other word embedding models of the BERT model, it is designed to pre-train the dataset in both layers in two directions and to condition the word in both right and left contexts. The BERT model can be used to fine-tune with an additional output layer to create cutting-edge models without the important task of answering questions and language extraction. It is conceptually simple and empirically powerful. In Fig. 1, the architecture of BERT model is presented where the arrows indicate the information flow from one layer to the next. The pink boxes(Article Color printing should be done.) at the top indicate the final contextualized representation of each input word. Input words are demonstrated with *E* that lies from 1 to n.

D. *Proposed Framework*

In this study, user comments from Twitter are collected using with the labels: "AKBNK", "ALBRK", "GARAN", "HALKB", "ISCTR", "SKBNK", "TSKB", "VAKBN", "YKBNK". A total of 117,136 Turkish tweets are gathered with the aforementioned labels using the Selenium crawler and browsing-navigation path to attract user comments from Twitter. Since the collected raw dataset is very dirty, different pre-processing techniques are employed to clean the dataset. In this study, stop-word elimination, removing hashtags, removing URLs, and Turkish Stemmer Library [21] methods are applied. As we focus on supervised machine learning strategy in our study, it is necessary to label the collected Turkish texts. To label the preprocessed dataset, TextBlob is utilized. However, because of lack of Turkish pre-trained dataset, we construct pre-trained dataset by employing user reviews from the website of Hepsiburada (https://www.hepsiburada.com). After labelled dataset is obtained, dataset is splitted into training and test sets, 80% and 20%, respectively. Next, classification task is performed by implementing Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short Term Memory Network (LSTM), and Bidirectional Encoder Representations from Transformers (BERT).

IV. Result

In this study, extensive experiments are conducted to analyze the emotions of user comments related to stocks that have the large volume in the stock market by using new generation word embedding model and deep learning algorithms. Accuracy is used as an evaluation criterion in the experiments to demonstrate the classification performance of each model and the contribution of our study. With the repeated holdout method, we set the training set size to 80% and conduct experiments. This approach is similar to previous literature studies [22], [23] using 80% of training data and 20% for the test. The following abbreviations are utilized for next generation word embedding model, and deep learning algorithms: CNN: Convolutional Neural Network, RNN: Recurrent Neural Network, LSTM: Long Short Term Memory Network, BERT: Bidirectional Encoder Representations from Transformers. The best accuracy results are indicated in bold. First, we analyze the classification performances of each model in terms of stocks that have high volume in BIST100 as shown in Table 1.

TABLE I: The classification performances of each model in terms of stocks that have high volume in BIST100.

Stocks	CNN	RNN	LSTM	BERT
AKBNK	96.22	**96.56**	96.22	94.84
ALBRK	95.48	94.92	94.92	**96.04**
GARAN	92.07	**93.36**	92.44	92.98
HALKB	**94.69**	94.09	94.29	93.70
ISCTR	95.15	95.52	95.15	**96.72**
SKBNK	96.05	96.05	96.05	**97.17**
TSKB	96.79	96.79	96.79	**98.71**
VAKBN	97.26	97.26	97.26	**98.63**
YKBNK	96.17	96.17	96.17	**97.60**

In Table 1, the classification accuracy of deep learning algorithms and a new generation word embedding model are evaluated. BERT is the best performing model when average accuracy results of each model are considered with 96.26% accuracy value. Moreover, the RNN model exhibits more successful classification performance with 95.47% of accuracy than CNN and LSTM. Actually, all deep learning models present very close classification success. As a summary, BERT word embedding model outperforms deep learning models by displaying nearly 1%-2% better accuracy result. Finally, the classification success of deep learning algorithms and the new generation word embedding model are listed as follows: BERT> RNN> CNN> LSTM. Furthermore, BERT generally performs better when stocks are considered except AKBNK, GARAN, and HALKB. 2% improvement for AKBNK 1% enhancement for GARAN are provided by RNN model compared to the BERT. In HALKB, CNN demonstrates the best classification success with 94.69% of accuracy.

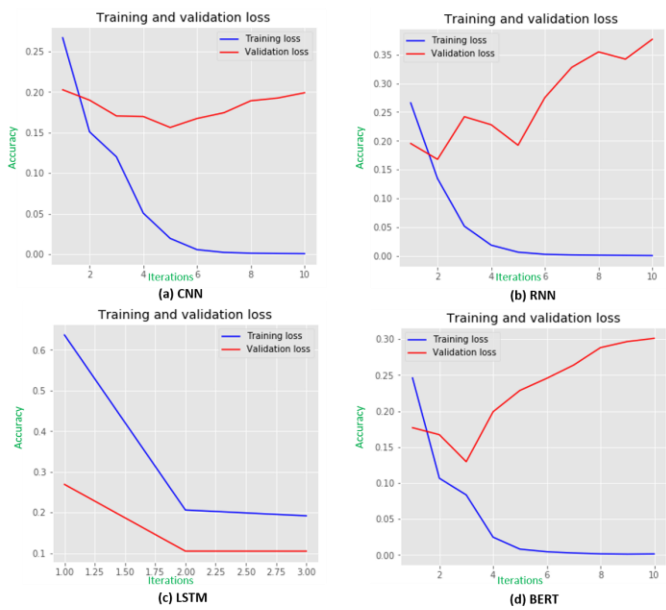

Fig. 2. Training and loss results of each model for ALBRK stock.

In Figure 2, training and loss results of each model for ALBRK stock are also given. It is clearly seen that the loss differences between validation and train sets in terms of number of epochs in Fig. 2. The horizontal axis shows the number of epochs and the vertical axis displays the loss. In general, as the number of epochs increases, the validation loss indicated by the red line decreases, noticeably up to when it reaches an optimum value for CNN, RNN, and BERT models. After reaching an optimum value for each model, the loss value increases as expected except LSTM model. In LSTM, the number of epochs where the lowest loss is observed the interval between 2.00 and 3.00. It can be summarized that there is an optimum value for the number of epochs for each model. The minimum loss value is calculated as between 2.00 and 4.00 in BERT while the optimum value for the number of epochs to get the lowest loss value is obtained the interval between 4.00 and 6.00 in CNN, and RNN. This means the usage of BERT provides the best results that are generated with the less number of epochs compared to the other models.

V. Discussion and Conclusions

In this study, unlike recent researches on forecasting stock market direction, we focus on financial sentiment analysis in social media platforms for determining the direction of stock market by analyzing stocks that have huge volume in BIST100. For this purpose, there is a need to understand and analyze users' comments about financial news on social media platforms to determine the direction of BIST100. In order to implement this, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and Long Short Term Memory Networks (LSTM) are employed as deep learning approaches, and Bidirectional Encoder Representations from Transformers (BERT) is used as a new generation word embedding model. Moreover, to improve the classification performance of the proposed model, we also consider stop-word elimination, removing hashtags, removing URLs, and stemming as preprocessing methods. As a result, the combination of BERT with pre-processing methods has achieved the best classification success with 96.26% of accuracy in order to determine the sensitivity of the users to guide the stocks. This means that the analysis of users' comments will provide investors with a perspective that will guide their investments. As a future study, we plan to develop our model using other new generation word embedding models to further strengthen stock market forecasting.

References

[1] N. Çelik and M. F. Kaya (January 2010). Uç değerler yöntemi ile riske maruz değerin tahmini ve İstanbul Menkul Kıymetler Borsası üzerine bir uygulama. *Bankacılık ve Sigortacılık Araştırmaları Dergisi*, 1(1), pp. 19-32. Available: https://dergipark.org.tr/tr/download/article-file/207744 https://doi.org/10.1501/bsad_0000000002

[2] V.M. Prieto, S. Matos, M. Álvarez, F. Cacheda, J.L. Oliveira. (January 2014) Twitter: a good place to detect health conditions. PloS One 9. e86191. Available: https://doi.org/10.1371/ journal.pone.0086191.

[3] A. Beykikhoshk, O. Arandjelović, D. Phung and S. Venkatesh, “Overcoming Data Scarcity of Twitter: Using Tweets as Bootstrap with Application to Autism-Related Topic Content Analysis”, *ASONAM '15 Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 1354-1361, 2015. https://doi.org/10.1145/2808797.2808908

[4] Z. H. Kilimci, S. Akyokus, (September 2018). Deep Learning-and Word Embedding-Based Heterogeneous Classifier Ensembles for Text Classification. *Complexity*. 2018(9). pp. 10. Available: https://doi.org/10.1155/2018/7130146

[5] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[6] A. Akcan, C. Kartal, (July 2011). İMKB Sigorta Endeksini Olusturan Sirketlerin Hisse Senedi Fiyatlarının Yapay Sinir Ağları İle Tahmini. *Muhasebe ve Finansman Dergisi*. (51). pp. 27-40. Available: https://dergipark.org.tr/tr/pub/mufad/issue/35630/396054

[7] A. C. Türkmen and A. T. Cemgil, “An application of deep learning for trade signal prediction in financial markets,” in Signal Processing and Communications Applications Conference (SIU), IEEE Press, 2015, vol. 23, pp. 2521-2524. https://doi.org/10.1109/SIU.2015.7130397

[8] K. Chen, Y. Zhou, and F. Dai, “A lstm-based method for stock returns prediction: A case study of china stock market,” in Big Data (Big Data), 2015 IEEE International Conference on. IEEE, 2015, pp. 2823–2824. https://doi.org/10.1109/BigData.2015.7364089

[9] Gunduz, H., Yaslan, Y., & Cataltepe, Z. (2017). Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations. *Knowledge-Based Systems*, *137*, 138-148. https://doi.org/10.1016/j.knosys.2017.09.023

[10] A. Hasan, O. Kalıpsız, S. Akyokuş, "Predicting financial market in big data: deep learning," In *2017 International Conference on Computer Science and Engineering (UBMK)*, pp. 510-515. IEEE Press, 2017. https://doi.org/10.1109/UBMK.2017.8093449

[11] N. Pervan, Y. Keleş, "Derin öğrenme yaklaşımları kullanarak Türkçe metinlerden anlamsal çıkarım yapma", Ankara Üniversitesi, Ankara, Türkiye, 2019.

[12] Tekin, S., & Çanakoğlu, E. (2019, April). Analysis of Price Models in Istanbul Stock Exchange. In *2019 27th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE. https://doi.org/10.1109/SIU.2019.8806296

[13] S. Tekin, E. Çanakoğlu, "Prediction of stock returns in Istanbul stock exchange using machine learning methods," In *2018 26th Signal Processing and Communications Applications Conference (SIU)*, pp. 1-4, IEEE Press, 2018. https://doi.org/10.1109/SIU.2018.8404607

[14] H. Gunduz, Z. Cataltepe, Y. Yaslan, "Stock market direction prediction using deep neural networks," In *2017 25th Signal Processing and Communications Applications Conference (SIU)*, pp. 1-4, Press IEEE, 2017. https://doi.org/10.1109/SIU.2017.7960512

[15] A. Bruns, A. Kornstadt, D. Wichmann. (August 2009). Web application tests with selenium. *IEEE software*, 26(5), pp. 88-91. Available: https://ieeexplore.ieee.org/abstract/document/5222802

[16] J. Schmidhuber. (January 2015) "Deep Learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117. Available: https://doi.org/10.1016/j.neunet.2014.09.003

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in ImageNet Classification with Deep Convolutional Neural Networks, 2012, pp. 1097-1105.

[18] Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*.

[19] J. L. Elman. (March 1990). Finding structure in time. *Cogn. Sci*. 14(2). Pp. 179-211. Available: https://doi.org/10.1207/s15516709cog1402_1.

[20] L. Zhang, S. Wang, B. Liu. (July/August 2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 8(4). e1253. Available: https://doi.org/10.1002/widm.1253.

[21] V. Tunali, T. T. Bilgin, "PRETO: A high-performance text mining tool for preprocessing turkish texts," *In Proceedings of the 13th International Conference on Computer Systems and Technologies*, pp. 134-140, ACM Press, 2012. https://doi.org/10.1145/2383276.2383297

[22] K. M. Schneider. (October 2014). On Word Frequency Information and Negative Evidence in Naive Bayes Text Classification. *In: 4th International Conference on Advances in Natural Language Processing*. 3230. pp. 474–485. Available: https://link.springer.com/chapter/10.1007/978-3-540-30228-5_42.

[23] Z. H. Kilimci, Selim Akyokus, Sevinc Ilhan Omurca. (September 2016). "The effectiveness of homogenous ensemble classifiers for Turkish and English texts," *In INnovations in Intelligent SysTems and Applications (INISTA), IEEE Trans, 2016 International Symposium on*, pp. 1-7, 2016. https://doi.org/10.1109/INISTA.2016.7571854

D. OTHAN was born in Kars in Turkey, in 1994. She graduated from Selcuk University Computer Engineering Department in 2017. She began her career as a computer engineer in 2017 Turk Telekom and, since February 2019 to continue her career in Turkey Finans Participation Bank as a Business Analyst. She has been pursuing her career for nearly 3 years. She is a MS.c student at Dogus University. Derya Othan's projects are in Turkey Finans Participation Bank: Requests and research and development of business processes, Preparation of requirement documents and TO-BE processes in accordance with agile methodologies, drawing process drawings and flow charts, Business and Process Analysis tests, implementation Tests, and Testing of User Acceptance Tests, Querying all data in the test environment, Corporate Loan Conversion. Projects in Turk Telekom Company are: Requests and research and development of business processes, Perform of all collection process using the technical analysis(System Requirements), Preparation of related technical and non-techinal document about with software deficiencies and for new demands, Support for analysis teams to make design document, Querying all data in the test environment on the Toad For Oracle and linux platform, Preparation of documents of RDD(Requirement Definition Document) and DDD (Design Definition Document) against the BRD(Business Requirement Document), Supporting the Demo Tests and Integration Tests for finished Projects of Software, Drawing process drawings and flow charts in the Visual Paradigm platform, Support of software development processes for Projects, Demands and Defects. Resolving bugs (Entegration or Software deficiencies etc.) from all customer

Z. H. KİLİMCİ was born in 1985 in Erzurum. In 2008, she graduated from Doguş University Computer Engineering Department. Between 2009 and 2011, she worked as a Software Engineer in Data Mining and CRM Department of the Data Warehouse Department of Denizbank. In 2011, she started to work as a Research Assistant in Computer Engineering Department of Doguş University. After completing her master of science at Computer Engineering Department of Doguş University in 2013, she received her PhD degree in Computer Engineering Department of Kocaeli University in 2018. Since September 2018, she has been working as Assistant Professor at Dogus University. The research fields include text mining, machine learning, community learning, deep learning, artificial intelligence.

M. UYSAL was born in Istanbul,Turkey in 1954. He received the B.S. and M.S. degrees in mechanical engineering from the Technical University of Istanbul, and the Ph.D. degree in System Analysis from Istanbul Technical University, in 1984. From 1977 to 1984, he was a Research Assistant with the System Analysis Department in ITU. Since 1984, he has been an Assistant Professor with the System Analysis Department, in Technical University of Istanbul. He is the author of 30 books on Computer Applications, and more than 70 articles. Dr. Uysal is currently a Professor in Dogus University, Istanbul in Software Engineering Department. His areas of interest include numerical analysis, optimization, data mining and algorithms and software engineering.