

T.R.
GEBZE TECHNICAL UNIVERSITY
FACULTY OF ENGINEERING
DEPARTMENT OF COMPUTER ENGINEERING

**US DOLLAR/TURKISH LIRA EXCHANGE RATE
FORECASTING MODEL BASED ON DEEP
LEARNING METHODOLOGIES AND TIME
SERIES ANALYSIS**

ALP EMİR BİLEK

**SUPERVISOR
DR. BURCU YILMAZ**

**GEBZE
2022**

T.R.
GEBZE TECHNICAL UNIVERSITY
FACULTY OF ENGINEERING
COMPUTER ENGINEERING DEPARTMENT

**US DOLLAR/TURKISH LIRA EXCHANGE
RATE FORECASTING MODEL BASED ON
DEEP LEARNING METHODOLOGIES
AND TIME SERIES ANALYSIS**

ALP EMİR BİLEK

SUPERVISOR
DR. BURCU YILMAZ

2022
GEBZE

 <p>GEBZE TECHNICAL UNIVERSITY</p>	<p>GRADUATION PROJECT JURY APPROVAL FORM</p>
--	--

This study has been accepted as an Undergraduate Graduation Project in the Department of Computer Engineering on 31/08/2021 by the following jury.

JURY

Member

(Supervisor) : Dr. Burcu YILMAZ

Member : Prof. Dr. Hasari ÇELEBİ

ACKNOWLEDGEMENT

To those who contributed to the preparation of this report, I would like to express my sincere thanks to my teacher Dr. Burcu YILMAZ and Gebze Technical University for supporting this project. In addition, I would like to express my respect and love to my family, who supported me in every way during my education, and to all my teachers who set an example for me with their lives.

Alp Emir BİLEK

CONTENTS

Acknowledgement	iv
Contents	v
List of Figures	vi
List of Symbols and Abbreviations	vii
Abstract	viii
Introduction	ix
Literatur Review	x
Project Scheme and Purpose	xii
Requirements	xiii
1 Dataset	1
1.1 Data parser	1
1.2 Test Data	1
2 Deep Learning Model	2
2.1 What is Deep Learning?	2
2.2 LSTM	2
2.2.1 LSTM - Text Classification	3
2.3 Model Training	4
3 Word Embeddings Model	5
3.1 SentenceBert	5
4 Experiments and Results	6
5 Conclusion	7
Resources	8
Appendices	9

LIST OF FIGURES

1	LSTM sample diagram.	xi
2	SentenceBert sample diagram.	xi
3	Scheme diagram.	xii
4	Design diagram.	xii
1.1	Sample labeled data.	1
2.1	LSTM	2
2.2	LSTM - Text Classification	3
2.3	LSTM - Text Classification Sample Accuracy Graph	3
2.4	LSTM - Text Classification Accuracy Graph	4
3.1	Sample SentenceBert Model Output	5
4.1	Test datas accurany rate with LSTM - Text Classification.	6
4.2	Test datas accurany rate with LSTM - Word Embeddings.	6

LIST OF SYMBOLS AND ABBREVIATIONS

Symbol or

Abbreviation : Explanation

FSA : Financial sentiment analysis

LSTM : Long short-term memory

SBert : SentenceBert

ABSTRACT

Exchange rate forecasting has been an important topic for investors, researchers, and analysts. In this project, FSA are proposed to form a predicting model for US Dollar/Turkish Lira exchange rate. For this purpose, it was aimed in this project to compare traditional methods and an experimental method.

As the traditional method, a deep learning model that makes predictions on the data will be used. The most suitable one will be selected from many traditional deep learning models. As an experimental new method, the word embeddings of the input data will be obtained and the resulting vectors will be trained on the deep learning model used in the traditional method.

To briefly explain word embeddings, they can be expressed as the numerical meaning of a text. That is, the values of word embedding vectors of the text with two close meanings should be close to each other. This similarity can also be evaluated with cosine similarity.

For this project, which was made in order to understand the difference between these two models, the dataset was obtained by collecting the Turkish daily news on the dollar rate. The training of the two models was done with the collected data. Finally, with the usage of the proposed model, any user who wants to make a US Dollar/Turkish Lira exchange rate increment or decrement forecast will be able to make a more consistent and strong exchange rate forecast.

The result decided whether our experimental new method is useful or not by comparing the predicted results between the two trained models.

Keywords: deep learning; exchange rate prediction; financial sentiment analysis; word embedding;

INTRODUCTION

Today, social networks have an important place in our daily life. Users evaluate free time using social platforms like Twitter, Facebook or Instagram. Although these habits when viewed negatively, social media platforms continue to be used by millions of people, research area and data source for researchers. The main reason for this is the generation of data. on these platforms millions of different types, in large volumes, such as images, texts and videos. very quickly at the same time. While this data flow power on Twitter is so high, it is inevitable that tweets about the economy will have an effect on the exchange rates.

Exchange rates are considered one of the most important investment tools for any country. Countries and companies use exchange rate, which is one of the most important economic variables. This makes the exchange rate and exchange rate one of the largest and most important financial markets in the world. For this reason, exchange rates can be affected by many developments in the markets and economy in a positive way or in a negative way.

A tweet about exchange rates or economy by an economist, a minister, or a president of government can affect exchange rates. Also a daily economy new can affect the same. After discussing why exchange rates are important and economic news and tweets have an impact on the exchange rates, we can talk about how to proceed with this project.

In this project, we will first start by finding the news and tweets that have such an impact on exchange rates, and then labeling them as an increase or decrease in the exchange rate and creating the dataset. Then, we will research traditional prediction deep learning models that will work on our dataset and choose the most suitable model for our project. The next step is research about what word embedding is, what it means and why it is important for our project. After we have an idea about word embeddings, we will research the methods of obtaining these vectors and choose the most suitable one for our project.

LITERATUR REVIEW

In the days when this project was developed, the dollar rate was experiencing serious fluctuations against the Turkish lira. The most visible reasons for this were seen as interest rate cuts and wrong economy management.

Twitter has a serious power in Turkey. With this power, it is possible to change the agenda or stay on the agenda for a long time. The best way to observe the effects of the interest rate cut and the bad economy will be to look at the tweets of the accounts with a high number of followers on the economy on twitter. For these reasons, the dataset was constructed from tweets and news on the economy.

In this part, it was decided which model to use as a result of the researches on the structure and working mechanics of deep learning models. It has been observed that the LSTM - Text Classification model, which is one of the deep learning models, is generally used.

In addition to Twitter, news sites are a good way to follow the statements of the president and ministers. Because many statesmen prefer to make statements to news sources instead of tweeting.

For the experimental new method, the first target was set to obtain word embeddings. It has been observed that generally 2 different methods are used to obtain word embeddings and these are Word2Vec and SentenceBert. Among these two methods, SentenceBert method was preferred because it is a newer method and has more resources.

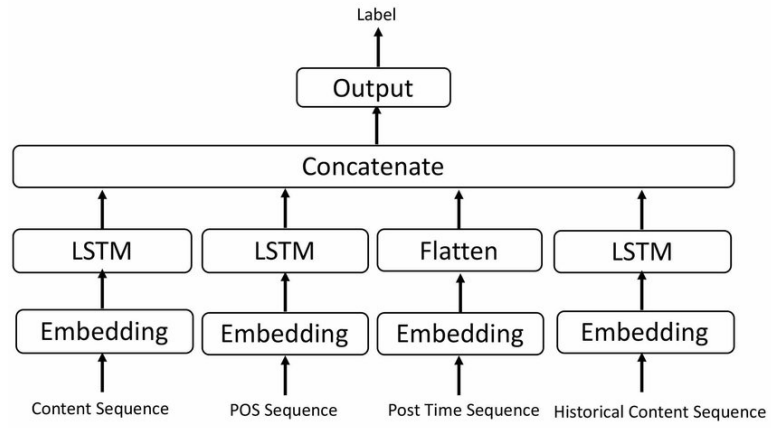


Figure 1: LSTM sample diagram.

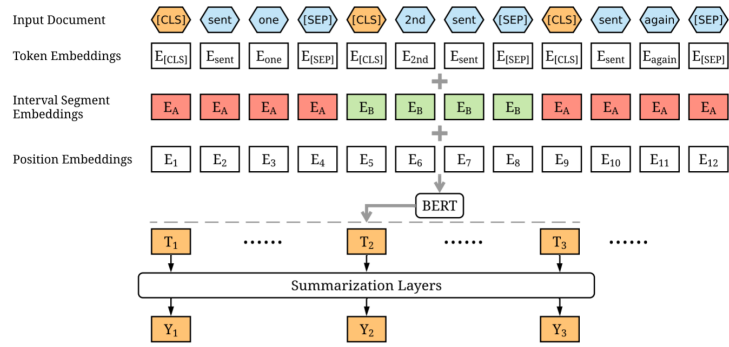


Figure 2: SentenceBert sample diagram.

PROJECT SCHEME AND PURPOSE

As mentioned before, the dollar exchange rate will be forecasted with 2 different methods. The reason why it is done with 2 different methods is to find out, if there is a better experimental method than the traditional method.

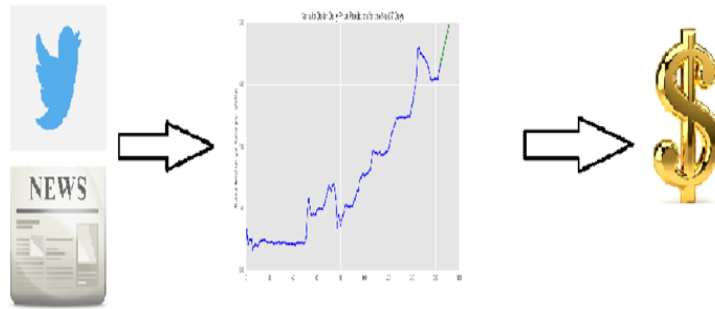


Figure 3: Scheme diagram.

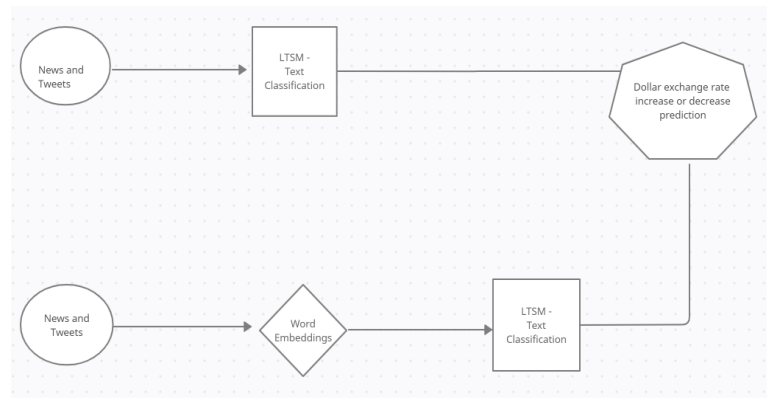


Figure 4: Design diagram.

REQUIREMENTS

Briefly, if we list the things that need to be done to achieve the result in the project;

- Collecting daily news and tweets and turning them into a dataset.
- Labeling whether the datas in the dataset has a increase or decrease effect on the dollar exchange rate.
- Implementation of the decided LSTM - Text Classification model.
- Implementation of the decided SentenceBert model.
- Preparation of the same test data for the two methods.
- Making the required library installations for both models;
 - pip 20.1.1
 - python 3.7.8
 - scikit-learn 1.0.1
 - tensorflow 1.14.0
 - transformers v4.6.0
 - pyTorch 1.6.0
 - numpy
 - scipy
 - tokenizers

1. DATASET

The dataset to be used for the two models will be the same because in this project we are examining the difference between the two methods.

As mentioned earlier, there are many factors on the dollar exchange rate, some of which are economy news and tweets about the economy. Economy news and tweets about the dollar that may be important were tried to be collected manually from 1 September to 10 January.

This collected data had to be labeled before it was added to the dataset. Labeling process was done as 1 if it causes an increase on the dollar rate of the instant data, and 0 if it causes a decrease. It is very difficult to predict exactly how much the exchange rate will be using deep models because if it were easy, these models would be used by everyone and huge profits could be made.

[illegible]

Figure 1.1: Sample labeled data.

1.1. Data parser

The data consists of 2 columns, label and content. Therefore, the data must be parsed in order to be given as input to the deep learning model.

1.2. Test Data

After both models are trained, they will be tested with the same data because this is how we can tell the difference. For a healthy test, the test data should not be in the main dataset A different test case has also been made for the main dataset. The prediction results are saved in a csv file as the actual result and the model's prediction for each data.

2. DEEP LEARNING MODEL

2.1. What is Deep Learning?

Deep learning is a type of machine learning and artificial intelligence that acts like the way humans gain certain types of information. It includes deep learning, statistics and predictive modelling. Responsible for collecting, analyzing and interpreting large amounts of data; deep learning makes this process faster and easier.

In this project, researches were carried out on the deep learning model that will make predictions on the news, and as a result, it was decided to work with LSTM.

2.2. LSTM

Long Short Term Memory networks – usually just called “LSTMs” – are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Hochreiter Schmidhuber (1997), and were refined and popularized by many people in following work.¹ They work tremendously well on a large variety of problems, and are now widely used.

LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn!

All recurrent neural networks have the form of a chain of repeating modules of neural network. In standard RNNs, this repeating module will have a very simple structure, such as a single tanh layer

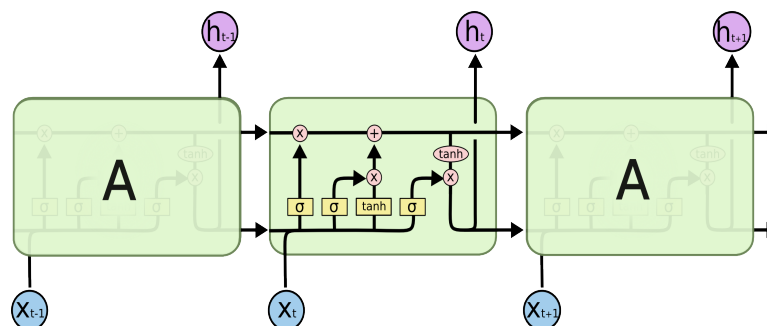


Figure 2.1: LSTM

2.2.1. LSTM - Text Classification

Text classification is the process of breaking the text into words. This is the analysis of the separated word groups. Using NLP, text classification helps analyze text and then assigns a set of predefined tags or categories based on the topics of the texts. As a result of these sequences, it makes predictions according to the subject of the texts. In this project, the labels will change according to the increase or decrease of the dollar rate. Our expectation from this model is that training is done with these labels and then accuracy is high in the predictions it makes.

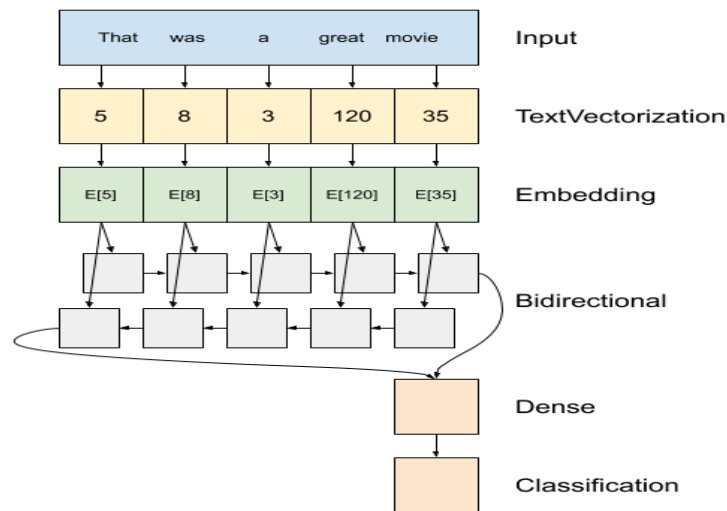


Figure 2.2: LSTM - Text Classification

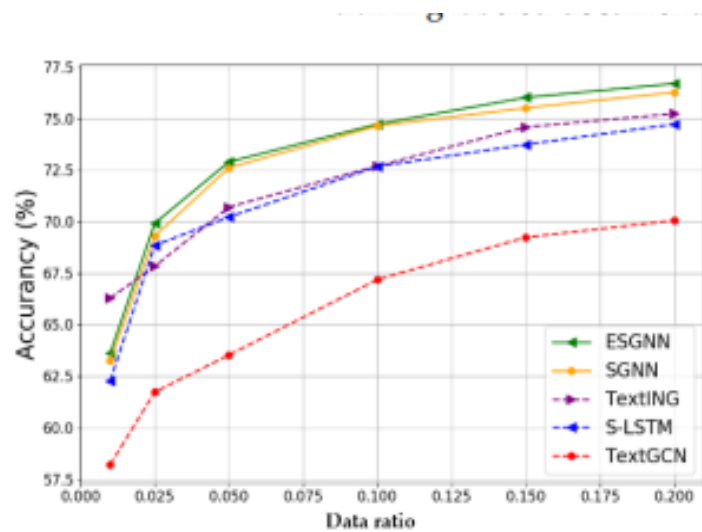


Figure 2.3: LSTM - Text Classification Sample Accuracy Graph

2.3. Model Training

The model designed in this project is an LSTM - Text Classification model. This model has been trained with daily news and tweets and data on what impact it will have on the dollar rate. The model was trained with 50 epochs and 32 batch sizes. LSTM - Text Classification, an estimation success rate of %80-%90 was obtained.

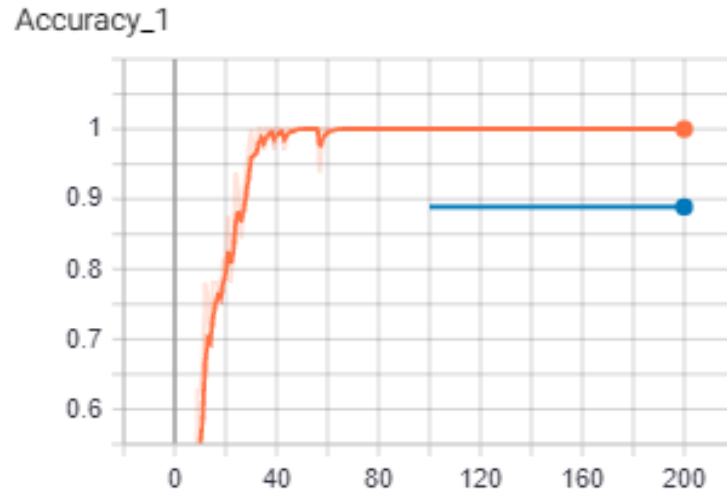


Figure 2.4: LSTM - Text Classification Accuracy Graph

3. WORD EMBEDDINGS MODEL

A word embedding is a learned representation for text where words that have the same meaning have a similar representation. Word embeddings are in fact a class of techniques where individual words are represented as real-valued vectors in a predefined vector space. Each word is mapped to one vector and the vector values are learned in a way that resembles a neural network, and hence the technique is often lumped into the field of deep learning.

3.1. SentenceBert

SentenceTransformers is a Python framework for state-of-the-art sentence, text and image embeddings. The initial work is described in our paper Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. You can use this framework to compute sentence / text embeddings for more than 100 languages. These embeddings can then be compared e.g. with cosine-similarity to find sentences with a similar meaning. This can be useful for semantic textual similar, semantic search, or paraphrase mining. The framework is based on PyTorch and Transformers and offers a large collection of pre-trained models tuned for various tasks. Further, it is easy to fine-tune your own models.

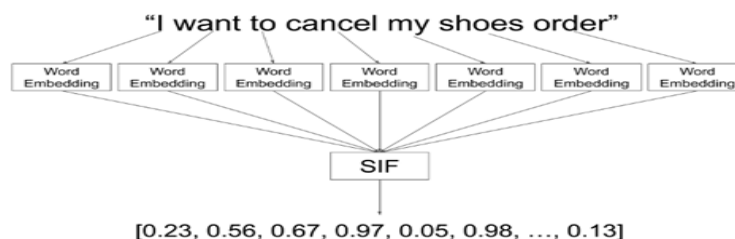
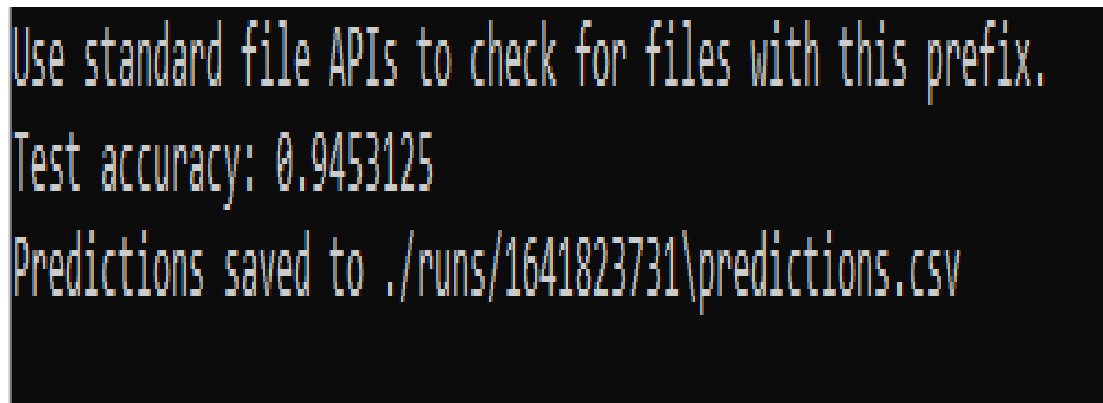


Figure 3.1: Sample SentenceBert Model Output

4. EXPERIMENTS AND RESULTS

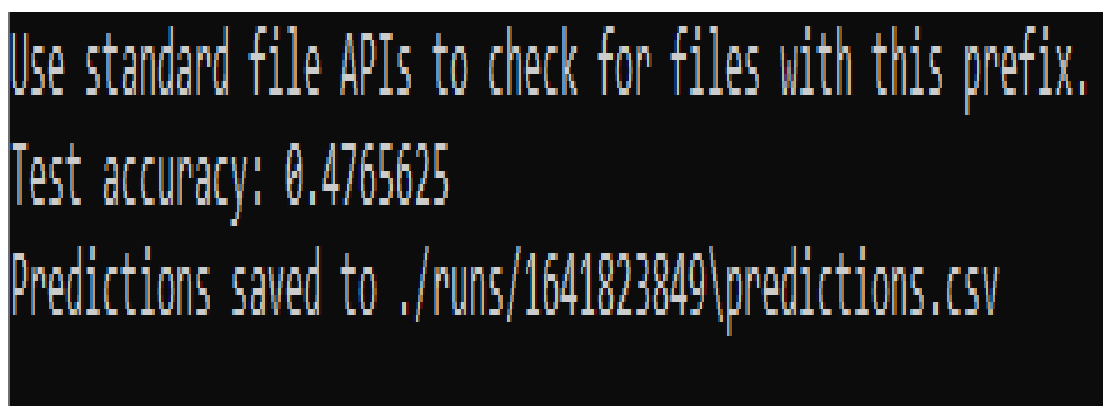
After completing the same model but different data type training for both methods, experiments were performed with the same test data.

As a result of the tests performed with the traditional method, LSTM - Text Classification, an estimation success rate of %80-%90 was obtained. As a result of the tests made with the experimental new method LSTM - Word Embeddings, an estimation success rate of %40-%50 was obtained.

A terminal window with a black background and yellow text. The text displays the instruction 'Use standard file APIs to check for files with this prefix.', the test accuracy '0.9453125', and the file path for predictions './runs/1641823731\predictions.csv'.

```
Use standard file APIs to check for files with this prefix.  
Test accuracy: 0.9453125  
Predictions saved to ./runs/1641823731\predictions.csv
```

Figure 4.1: Test datas accurany rate with LSTM - Text Classification.

A terminal window with a black background and yellow text. The text displays the instruction 'Use standard file APIs to check for files with this prefix.', the test accuracy '0.4765625', and the file path for predictions './runs/1641823849\predictions.csv'.

```
Use standard file APIs to check for files with this prefix.  
Test accuracy: 0.4765625  
Predictions saved to ./runs/1641823849\predictions.csv
```

Figure 4.2: Test datas accurany rate with LSTM - Word Embeddings.

5. CONCLUSION

After collecting the news and tweets that have an impact on the economy, we first trained these news in our deep learning model, LSTM -Text classification, and then we trained in the same model by removing the word embeddings of these news.

To perform the same experiments for the 2 trained models, we first tested on a small dataset. We then used the same test data for medium and large datasets. The accuracy rate was close to each other for these 3 different experiments. As a result of the tests performed with the traditional method, LSTM - Text Classification, an estimation success rate of %80-%90 was obtained. As a result of the tests made with the experimental new method LSTM - Word Embeddings, an estimation success rate of %40-%50 was obtained.

Based on these ratios, it can be said that whether the experimental new methods, which is the aim of this project, are better or not, it has been clearly seen that a better result can be obtained with the traditional method.

RESOURCES

- <https://pythonclass.in/gensim-doc2vec-tutorial.php>
- <https://medium.com/@mishra.thedeepak/doc2vec-simple-implementation-example-df2afbbfbad5>
- Kelime Vektörü Yöntemlerinin Model Oluşturma Sürelerinin Karşılaştırılması
Araştırma Makalesi/Research Article
- <https://www.sbert.net>
- <https://engineering.talkdesk.com/what-are-sentence-embeddings-and-why-are-they-useful-53ed370b3f35>
- <https://medium.com/dair-ai/tl-dr-sentencebert-8dec326daf4e>
- https://www.researchgate.net/publication/336996965_Sentence-BERT_Sentence_Embeddings_using_Siamese-BERT-Networks
- <https://engineering.talkdesk.com/what-are-sentence-embeddings-and-why-are-they-useful-53ed370b3f35>
- https://www.researchgate.net/publication/318018787_LSTM_Recurrent_Neural_Networks_Sh
- <https://medium.com/dair-ai/tl-dr-sentencebert-8dec326daf4e>
- <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- A Sequential Graph Neural Network for Short Text Classification /Research Article

APPENDICES

Appendix 1: Dataset Construction

If it is desired to run it with data other than the uploaded dataset, the .csv file must be read and written with a code, not manually (csv.reader() and csv.writer()). Same thing should be done for test data.

Appendix 2: Train and Test

When training the model, dataset must be given as an argument.

- `python train.py -data_file=dataset.csv -clf=lstm`

After the model is trained, the model version and test data file must be given as arguments for testing.

- `python test.py -test_data_file=dataset.csv -run_dir=/1641823731 -checkpoint=clf-10000`