BAŞKENT UNIVERSITY
INSTITUDE OF SOCIAL SCIENCES
DEPARTMENT OF TECHNOLOGY AND KNOWLEDGE MANAGEMENT
MASTER'S OF TECHNOLOGY AND KNOWLEDGE MANAGEMENT WITH THESIS

TWITTER SENTIMENT ANALYSIS WITH TURKISH TWEETS FOR EXCHANGE RATE PREDICTION

SUBMITTED BY

YUSUF FURKAN KOÇ

MASTER THESIS

SUPERVISED BY

ASST. PROF. NURCAN ALKIŞ BAYHAN

ANKARA - 2022

BAŞKENT ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

YÜKSEK LİSANS ÇALIŞMASI ORİJİNALLİK RAPORU

Tarih: 28/07/2022

Öğrencinin Adı, Soyadı: Yusuf Furkan KOÇ

Öğrencinin Numarası: 22010325

Anabilim Dalı: Teknoloji ve Bilgi Yönetimi

Programı: Teknoloji ve Bilgi Yönetimi

Danışmanın Unvanı/Adı, Soyadı: Dr. Öğr. Üyesi Nurcan ALKIŞ BAYHAN

Tez Başlığı: Twitter Sentiment Analysis with Turkish Tweets for Exchange Rate Prediction

Yukarıda başlığı belirtilen Yüksek Lisans çalışmamın; Giriş, Ana Bölümler ve Sonuç Bölümünden oluşan, toplam 44 sayfalık kısmına ilişkin, 28/07/2022 tarihinde şahsım/tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 15'tirr. Uygulanan filtrelemeler:

1. Kaynakça hariç

2. Alıntılar hariç

3. Beş (5) kelimeden daha az örtüşme içeren metin kısımları hariç

"Başkent Üniversitesi Enstitüleri Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Usul ve Esaslarını" inceledim ve bu uygulama esaslarında belirtilen azami benzerlik oranlarına tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Öğrenci İmzası:…….

ONAY

Tarih: 28/07/2022

Öğrenci Danışmanı Unvan, Ad, Soyad, İmza:

Dr. Öğr. Üyesi Nurcan Alkış Bayhan

……

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Asst. Prof. Nurcan Alkış Bayhan even though it is not enough to show my gratitude to her. I am very thankful to her patience, innovativeness, open-mindnesses and mostly to her great knowledge. I am very grateful that she gave me the chance to work with her.

I would like to thank also to Prof. Dr. Hakkı Okan Yeloğlu for his contribution to me and my master's degree education. I would also like to thank jury members, Assoc. Prof. Mehmet Güray Ünsal and Prof. Dr. Ersin Karaman for their valuable comments and evaluations in the jury.

I would like to thank my family, especially to my brother Mustafa Koç because of his great contributions and support to me and this work. Also, I would like to thank my brother Mehmet Koç who believes in me and always supports me.

Finally, I would like to thank my friends; Kübra Asena Gelişli and M. Tunç Sağlam for their kind friendship and support.

ÖZET

YUSUF FURKAN KOÇ, DÖVİZ KURU TAHMİNİ İÇİN TÜRKÇE TWEETLER İLE TWITTER DUYGU ANALİZİ, BAŞKENT ÜNİVERSİTESİ, SOSYAL BİLİMLER ENSTİTÜSÜ, TEKNOLOJİ VE BİLGİ YÖNETİMİ TEZLİ YÜKSEK LİSANS PROGRAMI, 2022

Türkiye gibi gelişen ülkelerde yabancı döviz kurları, ticari ve günlük hayatta alınacak kararlarda önemli bir etmendir. Karar vericilerin döviz kurlarını doğru öngörmesi ve bu öngörüler kapsamında hareket etmesi önem arz etmektedir. Türkiye gibi döviz kurlarının karmaşık ve dalgalı olduğu ülkelerde döviz kurlarını doğru tahmin edebilmek oldukça zordur ve geçmişte bu kapsamda birçok çalışma yürütülmüştür. Literatürde yer alan çalışmaların en yenileri ise makine öğrenmesi yöntemleriyle yapılan çalışmalardır. Bu çalışmada ise duygu analizi yöntemiyle daha doğru tahminler yapılıp yapılamayacağı ölçülmek istenmiştir. Çalışma kapsamında, Dolar, Sterlin ve Euro'nun Türk Lirası karşısında değerini tahmin etmek için Twitter'dan bu döviz kurlarına ilişkin 168.000'in üzerinde Tweet'in duygu analizi yapılmıştır. Duygu analizi sonuçları ve döviz kur değerleri destek vektör regresyonu yöntemi ile analiz edilmiştir. Çalışma başlangıcı COVID-19 pandemi salgınının başlangıcı olan 1 Kasım 2019'dan başlayarak 1 Mart 2022 olarak belirlenmiştir. Duygu analizi yönteminin sonuçlarını ölçebilmek için aynı veriler öncesinde destek vektör regresyonu ve, naive tahmini ve rastgele yürüyüş modelleri ile tahmin yapılmıştır ve bu veriler duygu analizi ile yapılan değerler ile kıyaslanmıştır. Sonuçlar, Twitter duygu analizinin döviz kuru tahminlerinin doğruluk oranını artırdığını ve döviz fiyatları ile negatif Tweetler arasında pozitif ve güçlü bir ilişki olduğunu ve döviz fiyatları ile pozitif Tweetler arasında negatif bir ilişki olduğunu ortaya koydu. Ayrıca atılan Tweet sayısı ile para birimi fiyatı arasında pozitif bir ilişki olduğu çıkan sonuçlar arasındadır.

Anahtar Kelimeler: Duygu analizi, Döviz tahmini, Makine öğrenmesi

ABSTRACT

YUSUF FURKAN KOÇ, TWITTER SENTIMENT ANALYSIS WITH TURKISH, TWEETS FOR EXCHANGE RATE PREDICTION, BAŞKENT UNIVERSITY, INSTITUTE OF SOCIAL SCIENCES, DEPARTMENT OF TECHNOLOGY AND KNOWLEDGE MANAGEMENT MASTER'S PROGRAM WITH THESIS, 2022

In developing countries such as Turkey, foreign exchange rates are important in business and daily life decisions. It is important for decision-makers to predict exchange rates accurately and act within the scope of these predictions. Since exchange rates are complex and volatile, it is very difficult to accurately predict exchange rates, and many studies have been carried out in this context in the past. The newest trends in the literature are the research which are performed with machine learning methods. In this thesis study, it was aimed to determine whether more accurate predictions can be performed with the sentiment analysis method or not. Within the scope of the study, the value of Dollar, Sterling and Euro against Turkish Lira was predicted with sentiment analysis of more than 168,000 Tweets related to these exchange rates from Twitter. Sentiment analysis results and exchange rate values were analyzed by the support vector regression method. The time limits of the study were determined as 1 March 2022, starting from 1 November 2019, the beginning of the COVID-19 pandemic outbreak. In order to compare the results of the sentiment analysis method, the same data were predicted with support vector regression, naive forecasting and random walk models, and these data were compared with the values made by sentiment analysis. The results revealed that Twitter sentiment analysis increase the accuracy rate of exchange rate predictions and there is a positive and strong relationship between currency prices and negative Tweets, and there is a negative relationship between currency prices and positive tweets. Also, there is positive relation between number of Tweets posted and currency price.

Keywords: Sentiment analysis, Currency prediction, Machine learning

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
API	Application programing interfaces
BSE	Bombay Stock Exchange
CHF	Confederatio Helvetica Franc
EOS	Electro Optical System Cryptocurrency
EUR	Euro
GBP	Great Britain pound
HKD	Hong Kong Dollar
INR	Indian Rupee
JPY	Japanese Yen
MAE	Mean absolute error
MAPE	Mean absolute percentage error
NLP	Natural Language Processing
NSE	National Stock Exchange
PKR	Pakistani Rupee
RMSE	Root mean squared error
RWF	Random walk forecast
SVM	Support vector machine
SVR	Support vector regression
TRY	Turkish Lira
USD	United States Dollar
XRP	Ripple

1. INTRODUCTION

The foreign exchange rate is one of the most critical economic indicator for countries, especially developing countries like Turkey. It is critical not only for governments but also for companies, organizations, investors and individuals, since the exchange rate affects everyone's daily financial decisions. Making decisions based on exchange rate movement can cause serious amount of money saving, especially for international trades. Therefore, exchange rate prediction is critical for countries, companies, investors, and individuals.

In the past, many researchers studied exchange rate prediction with traditional statistical methods such as linear regression, naïve forecast and time series modelling (regARIMA model). However, it is hard to achieve accurate predictions because of the complex and volatile structure of exchange rate data. The exchange rate is affected by many indicators. Some of the most important indicators include the country's political state, and economic performance, especially in countries like Turkey. Because the exchange rate is highly dependent on political relations, it makes more challenging to make accurate predictions. Therefore, traditional prediction methods could be insufficient and incapable of accurate prediction. New generation methods like machine learning, language processing and sentiment analysis are more effective for more accurate prediction. There are several prediction methods; sentiment analysis is one of the newest methods. Any new information, or news available in the market affect exchange rate (Chethan & Sangeetha, 2020). Moreover, Twitter is used not only for currency predictions but also for cryptocurrency predictions. A study found that the number of tweets is a significant driver of next day trading volume (Shen et al., 2019). This research focuses on currency prediction with sentiment analysis of Twitter data, which is an important indicator of the public's reaction toward policies about exchange rates.

In this thesis, currencies, U.S. Dollar, British Sterling, and Euro against Turkish Lira were collected between 1st of November 2019 and 1st of March 2022, which was the beginning of the COVID-19 Pandemic, and more than 168.000 Tweets in Turkish about these currencies were collected for prediction. In Turkey, the most common currencies are U.S. Dollar, Euro, and British Sterling. Turkish people are more concerned and react faster to these so these currencies were selected for this study for prediction. This study aims to

determine whether Twitter sentiment analysis improves the accuracy of currency predictions. The advantage of studying the exchange rate is that, due to the higher volume, the possibility of manipulation due to Twitter misinformation is expected to be lower than in stock market analysis (Ozturk & Ciftci, 2014). Several analyses were made to answer the research question and whole study process is described in the following chapters of the thesis.

The Second chapter of this thesis presents the background of the study, which states definitions of the exchange rate, prediction methods that used in previous researches, and the scopes of previous researches. Third chapter of the thesis is Methodology, which includes, scope of the study, the exchange rate and Twitter data collection methods, analyze methods which are, random walk forecast (RWF), support vector regression (SVR) and naïve forecast. The chapter of the thesis presents the results. This chapter, states all the analysis results. The detailed examination and statistical results of currency and Twitter data, the results of currency predictions without sentiment analysis with SVR, naïve forecast and RWF, contents of sentiment analysis and the results of currency predictions with sentiment analysis with SVR methods stated in this chapter of the study. The results of these discussed in the fifth chapter of the thesis. The last chapter shows the conclusion and the possible future research in that could come from this study.

2. BACKGROUND OF THE STUDY

2.1. Exchange Rate

The exchange rate is one of the important economic indicators of the countries. All of the investment decisions of corporations or governments are made based on the exchange rate. The exchange rate prediction has critical importance for governments, companies, investors, and almost anyone who does business. However, foreign exchange prediction is a very difficult task due to the nonlinear and volatile nature of this problem. The currency exchange rate data are time series data, which are affected by various economic and political factors. Making accurate prediction of exchange rate provides important economic benefit for decision makers. So that exchange rate prediction takes the attention of economic research (Yasir et al., 2019).

The exchange rate is the value of one nation's currency versus the currency of another nations. In other words, it can be defined as the value of a country's money against other countries (Chen, 2022). There are some tangible and intangible factors that affect currencies values, which are inflation rate, interest rate, current account deficit/balance of payments, government debt, terms of trade, recession, political stability, and speculation (Twin, 2022). The inflation rate is the change of purchasing power over time. The interest rate is the amount of interest due per period, as a proportion of the amount lent, deposited, or borrowed. The current account deficit/balance of payment reflects the balance of trade and earnings on foreign investment. The difference between a country’s export and import values can cause a deficit or excess of foreign money and it fluctuates the exchange rate of its domestic currency. Government depts, terms of trade, and recession are other tangible factors that affect the currency rate of a country. The intangible factors that affect currency rate are political stability and speculation. A country's political state and economic performance can affect its currency strength. A country with less risk for political confusion is more attractive to foreign investors, as a result, drawing investment away from other countries with more political and economic stability. A country with reliable financial and trade policy prevents from uncertainty in the value of its currency. But a country tend to political confusions may see a depreciation in exchange rates. The other important intangible factor that affects currency rate is speculations. Political events or

changes in commodity prices may cause a decrease in currency value. If speculators take actions like selling off or collecting that currency based on their expectations. For this reason, sentiments in the financial markets can heavily influence foreign exchange rates. Also, foreign investors make their investments based on these expectations. This study focusses on estimating intangible factors that affect currency rate based on tweets that could cause speculations and political confusion or tension.

The majority of the academic work on exchange rates did not consider discussions of the general public. There are a limited number of studies that consider public discussions and exchange rates examine the role of media or institutions of the finance sector on exchange rates or they examine tactical public communications of politicians (Janetzko, 2014). The studies that focus on the mining of public opinions and emotions generally focused to Twitter which is one of the widely used social platform. Because social media especially Twitter plays a vital role in the financial prediction of Chief Executive Officers (Yasir et al., 2020). Especially in short term investments and decision, decision makers consider Twitter as a financial predictor because not only public opinion exist on Twitter but also financial analysists, financial authorities of the governments and even presidents exist on Twitter. Furthermore, public authorities use Twitter as a new way of sharing public announcements.

2.2. Twitter

Twitter is a popular social networking platform created by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams in 2006, in United States. On Twitter, users can post messages restricted to 280 characters, photos, and videos which are also restricted to 140 seconds, and these posts are named as “tweets” (Iqbal, 2022). In the fourth quarter of 2020 twitter has 186 million daily users in worldwide and approximately 20% percent of these users from United States, more than 500 million tweets shared per day and company has 3.7 billion USD annual revenue (Iqbal, 2022). In 2021, Turkey has 16.25 million twitter active users which is approximately 19% of the total population.

Twitter is a quick and effective way for people to announce their messages to millions of other people. So that, not only people use twitter but also companies and even politicians use Twitter actively to announce their opinions or messages. Moreover, these also political parties have their official Twitter accounts. When the accounts with the most

followers on Twitter are examined, the 44th President of the United States, Barack Obama has 130.5 million followers, and former President Obama has the highest number of followers on Twitter in the world. One of the other politicians has high numbers of followers is the 45th President of the United States, Donald Trump who has 88.8 million followers. In the top 10 most-followed Twitter accounts there are 3 politicians, and it shows that Twitter is one of the most effective ways for politicians to spread their messages to masses (Iqbal, 2022). Also in Turkey, there is a similar situation, the President of Turkey, Recep Tayyip Erdoğan has 18.3 million followers, and it is the most followed account in Turkey. That situation reveals that a tweet of a president has a similar impact with public announcements of government even more mostly tweets are more effective than public announcements.

Because of the politicians consider Twitter as a way of public speech, Twitter became a common and important data source for researches. There is considerable amount of researches that take Twitter as a main data source. In the field of economics, researchers use Twitter data to measure effect of public speech for predicting crypto-currencies, interest rate and currency rates. Some of the important researches that take Twitter as a main data source are mentioned Table 1. The research methods mentioned in detail in following titles of the research.

2.3 Sentiment Analysis of Tweets

The sentiment analysis is mining the emotions from texts, widely used by companies to determine customer feedback from surveys and comments. The sentiment analysis used for determining the indicators of the comments or tweets via categorizing the tweets into two groups, which is negative and positive words within a scale. In other words, sentiment analysis is a language processing technique to evaluate an opinion or sentiment in a tweet (Sarlan, Nadam, & Basri, 2014). There are two main methods for extracting sentiment in selected tweets that are machine learning-based method and the lexicon-based method.

2.3.1 Machine learning based method

Machine learning methods rely on pre-directed classification approaches where sentiment detection is classified as positive and negative. In this method, front adjectives are taken into consideration such as negative adjectives or intensifications. The common application process of the machine learning approach; apply a part of speech tagger to each

tweet post Collect all the adjectives for entire tweet posts, make a popular word set composed of the top N adjectives, navigate all the tweets in the experimental set to a number of positive words, number of negative words and presence, absence or frequency of each word (Sarlan et al., 2014). After all processes of the machine learning sentiment analysis are prepared the method should be detected. There are different types of method for machine learning sentiment analysis and every method have a different accuracy rate based on dataset. Most common methods are, Support Vector Machine (SVM), Naïve forecast, Natural Language Processing (NLP) and Artificial Neural Network (ANN). Because every method has different accuracy rates, these methods should be tested, and detecting the method that has the highest accuracy rate, and sentiment analysis should be made with that method.

2.3.1.1. Support vector machine (SVM)

SVM is a machine learning classification algorithm for detecting to sentiments of the texts. Researches show that SVM has greater than 80% accuracy in detecting the sentiment in the tweet data set (Sarlan et al., 2014). Common opinion and majority of the research show that however the accuracy of the analysis can change based on the dataset, still, SVM is one of the most accurate methods for sentiment analysis. In SVM the words are classified based on sentiments. For time series regression there is a sub-method of SVM, which is support vector regression (SVR).

2.3.1.2. Naïve bayes

The Naïve bayes is an efficient method for text type and large data. The algorithm counts the number of each word that appears in a text in the category and divides that by the number of words appearing in that category. This is referred to as conditional probability. In this case, the probability that a word will appear in a particular category (McAteer, 2014).

Naïve bayes assumes that the occurrence of all words in a text is independent. Though the approach is simple, it is well-proven technique that has shown to be effective when compared to more sophisticated algorithms (McAteer, 2014).

2.3.1.3. Natural language processing (NLP)

Natural language processing (NLP) is a machine learning based artificial intelligence that helps computers understand and manipulate human language. NLP draws from many disciplines, including computer science and computational linguistics, in its pursuit to fill the gap between human communication and computer understanding. Sarlan (2014) mentioned that NLP is the computer science that involves making computers understand human language and communicate with the real world.

2.3.1.4. Artificial neural network (ANN)

Sarlan (2014) explain Artificial Neural Network (ANN) as a machine learning based technique that compounds group of artificial neurons. It will process information using the connections approach to computation. ANN is used in finding the relationship between input and output or to find patterns in data.

2.3.2. Lexicon-based method

Lexicon-based model classifies the words positive and negative. The predefined words list associated with specific sentiments. Lexicon methods vary depending on the context in which they are created and include calculating the orientation for a document from the semantic orientation of texts or phrases in documents. Besides, also states that a lexicon sentiment is to detect word-carrying opinion in the corpus and then to predict opinion expressed in the text. has shown the lexicon methods which have a basic paradigm (Sarlan et al., 2014). Determining a tweet is positive or negative, calculating the total score of the words in that tweet, if the total score is greater than zero, it shows that tweet has positive sentiment, If the total score is negative that tweet has negative sentiment. If total score equals to zero that means that tweet has neutral sentiment.

2.4. Time Series Forecasting without Sentiment Analysis

Time series forecasting methods without sentiment analysis are less complicated forecasting approaches that can be contain dependent or independent variable. There are different types of forecasting methods most common methods are RWF and linear regression method but more complex forecasting methods such as naïve bayes and SVR which is more effective version of SVM on time series data methods can also be applied on

time series data. The accuracy of the forecast can differ based on types of data, so detecting the most effective method for forecast at least two different types of methods should be tested and compared.

2.5. Twitter and Exchange Rate Prediction Studies in the Literature

The Twitter sentiment analysis is an important prediction method for researchers. Twitter sentiment analysis method has widely usage area, as mentioned in Table 2.1, exchange rate prediction, interest rate prediction and cryptocurrency prediction. The researches mainly study on prediction with Twitter sentiment analysis have higher accuracy than traditional statistical prediction methods. The earliest research made 2014 by Janetzko and Öztürk & Çiftçi. Since 2014, Twitter sentiment analysis still commonly using by researches for analyzing different topics as in the mentioned Table 2.1.

Table 2.1. Studies in the literature

Scope of the Study	**Currency**	**Analysis Method**	**References**	**Data Year**	**Country**
An Intelligent Event-Sentiment-Based Daily Foreign Exchange Rate Forecasting System	PKR/USD, GBP/USD, HKD/USD	Support Vector Regression,	Yasir, et al., 2019	2008-2018	Hong Kong, Pakistan, UK
Using Twitter to Model the EUR/USD Exchange Rate	EUR/USD	RWF, regARIMA	Janetzko, 2014	2012-2013	
An Efficient Deep Learning Based Model to Predict Interest Rate Using Twitter Sentiment	Interest Rate	SVM, Linear Regression	Yasir, et al., 2020	2010-2019	UK, Turkey, China, Mexico
Twitter Sentiment Analysis to Predict Bitcoin Exchange Rate	Bitcoin	Naïve Bayes	McAteer, 2014	2014	
Exchange rate disconnect and private information: What can we learn from Euro-Dollar tweets?	EUR/USD	Lexicon	Gholampour & Wincoop 2019	2013-2017	
A Sentiment Analysis of Twitter Content as a Predictor of Exchange Rate Movements	USD/TRY	Sentiment Analysis, Regression Analysis,	Ozturk & Çiftçi, 2014	2013	Turkey
Sentiment Analysis of Twitter Data to Examine the Movement of Exchange Rate and Sensex	USD/INR, BSE, NSE	Naïve Bayes, SVM	Chethan & Sangeetha, 2020	2019	India
A Twitter-Based Prediction Tool for Digital Currency	Cryptocurrencies	SVM	McCoy, 2015.	2015	USA
Big Data in Financial Markets: Twitter Analysis of Forex Data	Forex, USD/TRY, EUR/TRY	RWF, Pearson Coorelation	Alkoç, 2019	2014-2019	Turkey
The predictive power of public Twitter sentiment for forecasting cryptocurrency prices	Bitcoin, Ethereum, XRP, Bitcoin Cash, EOS, Litecoin, Cardano, Stellar and TRON	Lexicon	Kraaijeveld, & Smedt, 2020	2017-2018	
Bitcoin Price Prediction Using Sentiment Analysis on Twitter	Bitcoin	Naïve Bayes, Logistic Regression	Köksal, et al., 2021	2014 – 2016	Turkey
Exchange Rate Prediction from Twitter's Trending Topics	EUR/USD,GBP/USD, CHF/USD,JPY/USD	NLP, Lexicon	Özcan, 2016	2013-2015	

As mentioned in the Table 2.1, the researchers that featured on twitter sentiment analysis show that most common analysis methods are SVM and naïve bayes methods. For example, Özcan (2016), Köksal (2021), McCoy (2015), Chethan & Sangeetha (2020) and Yasir (2020) preferred SVM and naïve bayes as their research methods. The other analysis methods that widely used, random walk model, lexicon model, natural language processing model and other traditional statistical methods such as Pearson Correlation and regression analysis. Majority of the researches mentioned that machine learning based sentiment analysis have higher accuracy rate on prediction against to traditional statistical prediction methods. The SVM approach has higher accuracy among machine learning approaches.

Also as indicated in Table 2.1, Özcan (2016), Gholampour & Wincoop (2019) and Janetzko (2014) took Euro and U.S. Dollar as research data. The majority of the researches mainly focus on Euro and U.S. Dollar currencies also in literature there are researches focus on USD/TRY, EUR/TRY, GBP/USD, HKD/USD, PKR/USD, CHF/USD, JPY/USD and interest rate. In addition to these official currencies, there are a considerable amount of research that focuses on digital or virtual currencies as known as cryptocurrencies. Especially about Bitcoin which is the most common and most expensive cryptocurrency among the cryptocurrencies.

3. METHODOLOGY

3.1. Scope of the Study

The main of this study is to predict the British Sterling, U.S. Dollars and Euro currencies against to Turkish Lira based on Tweets that were related to these currencies. When the related literate is examined, as can be seen in Table 2.1, most of the studies in this field examine the EUR/USD exchange rate. However, there are a few research examine USD/TRY which did not use SVM method to examine these currencies as mentioned Table 1. Also, there are not any studies about Turkish Lira, British Sterling currencies. For these reasons this study aims to predict USD/TRY, EUR/TRY and GBP/TRY currencies by analyzing with SVR in order to contribute to the related literature

All these Tweets were collected in Turkish language and analyzed from Turkish people's perspective. Also, when the related literature is examined, Tweets analysis in Turkish language is relatively low. So, this research will contribute also the research dealing with Turkish tweet analysis. The time limits of the study were determined according to COVID-19 pandemic which was 1st of November 2019 and upper limit was 1st of March 2022.

3.2. Data Collection

In the scope of this thesis, two types of data were collected and used for analysis, which are currency data and Twitter data. Twitter data was collected via Twitter API (application programing interfaces) and currency data was collected via www.investing.com website (Investing, 2022). Both data were collected for GBP/TRY, EUR/TRY and USD/TRY currencies.

3.2.1. Currency Data

The currency rates that take into consideration are, GBP/TRY, USD/TRY, and EUR/TRY. The time series that targeted the beginning of the Covid-19 pandemic, that were November 1, 2019 to March 1, 2022 for each currency and data were collected by a daily base. Weekends were excluded from the data set. The currency data was collected and stored in Microsoft Office Excel's csv format from www.investing.com (Investing, 2022) which is a financial markets platform providing real-time data, quotes, charts,

financial tools, breaking news, and analysis around the world. There are 608 data in a daily base for each currency. Also, 12 days' currency data were collected for each currency between 2nd of March and 17th March 2022 for accuracy test.

3.2.2. Twitter

A total of number of 168.808 tweets were collected in this study which were obtained through the academic study application to the Twitter API and a screen of Twitter API is given in Figure 1. The time series of the tweets were selected same as the currency data which was between beginning of the Covid-19 pandemic, that are November 1, 2019, to March 1, 2022. The tweets were selected based on keywords and hashtags. The retweets, mentions and replies extracted from data to have more effective data set.

For U.S. Dollar and Turkish Lira currency rate, a total number of 97663 tweets were collected which contain “tl dolar” or “Türk Lirası Dolar” keywords. Tweet collection screen of Twitter API for Dollar is represented in Figure 3.1. For Euro and Turkish Lira currency rate, a total number of 66838 tweets were collected which contain “tl euro” or “Türk Lirası Euro” keywords. For British Sterling and Turkish Lira currency rate a total number of 4307 tweets were collected which contain “tl sterlin” or “Türk Lirası Sterlin” keywords. All Tweets were collected in Turkish language and sentiment word set was created by common positive and negative words mentioned in these Tweets. The positive and negative word set was created according to the meanings used in the Tweets, although some words seem positive, they are used in a negative sense in the Tweets and therefore are included in the negative word set. Retweets, mentions and replies were excluded from the data set.

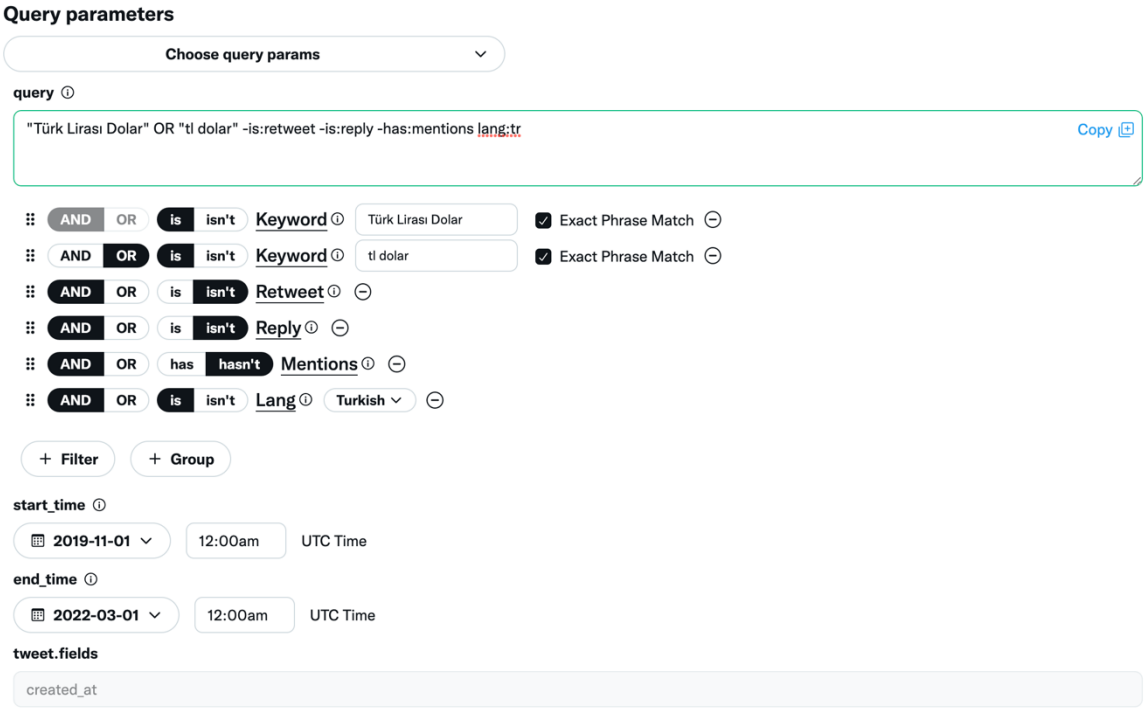

Figure 3.1. Twitter API screen view

3.3. Data Analysis Methods

In this research currency data prediction analysis was performed by RWF, naïve forecast and SVR method. These methods were selected based on previous researches which were mentioned on Table 2.1 in this domain. The highest accuracy for currency prediction was obtained with RWF, naïve forecast and SVR. Naïve forecast methods are used for basic prediction without Twitter data and SVR method are used for both with and without Twitter data.

3.3.1. Random walk forecast (RWF)

RWF is one of the simplest and the most effective forecasting method on univariable time series data. This model assumes that in each period the variable moves away from its previous value by one random step, and the steps are independently and equally distributed. A random walk model has 2 methods with drift (mean of difference) or without drift according to whether the distribution of step sizes has a non-zero mean or a zero mean. At period n, the k-step-ahead forecast that the random walk model without drift gives for the variable Y, and d is the estimated drift (Nau, 2014).

$$\hat{Y}_{n+k} = Y_n$$

Equation 3.1. RWF without drift

$$\hat{Y}_{n+k} = Y_n + k\hat{d}$$

Equation 3.2. RWF with drift

In this research, RWF used with drift for predicting currency data for without sentiment analysis.

3.3.2. Support vector machine (SVM)

SVM is one of the most common supervised machine learning algorithms that analyze data used for both regression and classification. The goal of the SVM algorithm is to create the hyperplane which is the decision boundary that can discriminate n-dimensional space into classes so that the new data point can be categorized truly (Reddy, 2018). In sentiment analysis, SVM works for the classification of the words based on their sentiments and group by positive, negative, and neutral.

The mathematical intuition behind SVM:

W: a vector which is perpendicular to the hyperplane.

X: distance from the selected point to zero is considered a vector

C: distance of vector w from origin to decision boundary

- w . x = c (the point line on the decision boundary)
- w . x > c (for positive samples)
- w . x < c (for negative samples)

SVM has different type of approaches such as linear SVM, non-linear SVM, soft margin and Kernels, these are main approaches can differ based on type of the data. Reddy (2018) mentioned that SVM draws the hyperplanes by transforming the data with the mathematical functions such as Kernels. The kernel is also divided into different titles within itself, these are ANOVA, Bessel, sigmoid, radial basis function (RBF), polynomial. The adjustment parameter Kernel's RBF is for non-linear problems and it is also a general-purpose kernel used when there is no prior knowledge about the data. Kernel’s linear

approach is for linear separable problems such as just positive and negative (Reddy, 2018). The Figure 3.2 shows, simple clustering example of SVM.

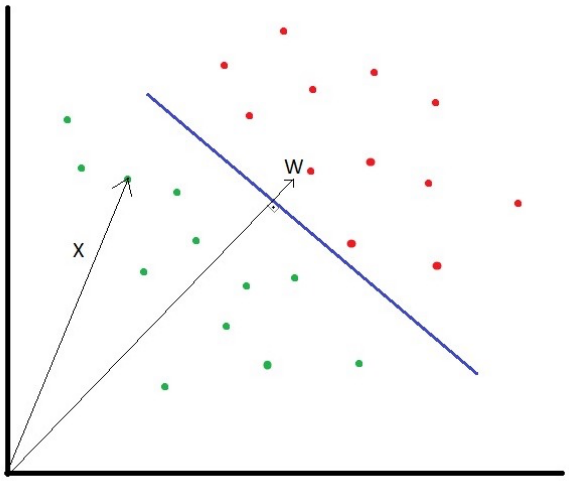

Figure 3.2. Clustering graph of SVM

The use of SVM for regression models are known as support vector regression (SVR). The SVR uses the same principles as the SVM for classification, with only a few differences. Because output is a real number, it's difficult to predict the information which has infinite possibilities. In the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM which would have already requested from the problem. Besides to this, there is also more complicated reason, the algorithm is more complicated therefore to be taken in consideration. However, the main idea is always the same: to minimize error, individualizing the hyperplane which maximizes the margin, keeping in mind that part of the error is tolerated (Sayad, n.d.).

The Figure 3.3 shows the negative and positive hyperplanes. ($w.x - b \leq \varepsilon$) vector is positive hyperplane, ($w.x - b \leq -\varepsilon$) is negative hyperplane and ($w.x - b = 0$) vector is optimal hyperplane (Sayad, n.d.).

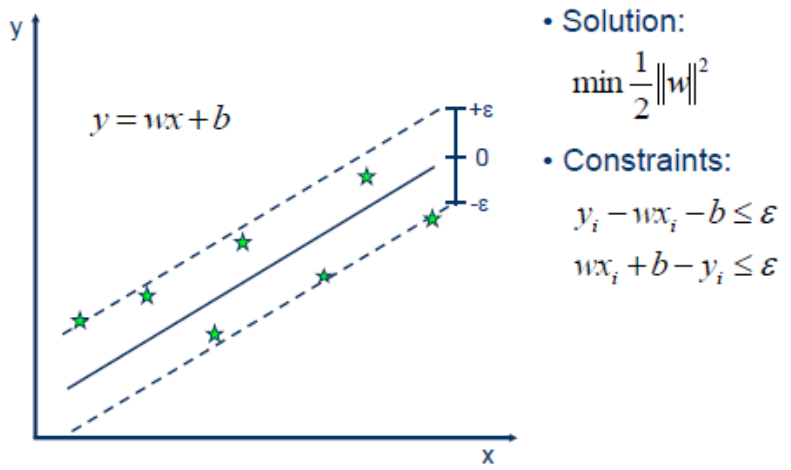

Figure 3.3 SVR clustering graph (Sayad, n.d.)

For nonlinear data set as shown in Figure 3.4, it is not possible to draw single line or hyperplane can classify the point correctly. For these types of data, Kernel's approach increases the dimension and find a decision boundary that clearly divides the data points.

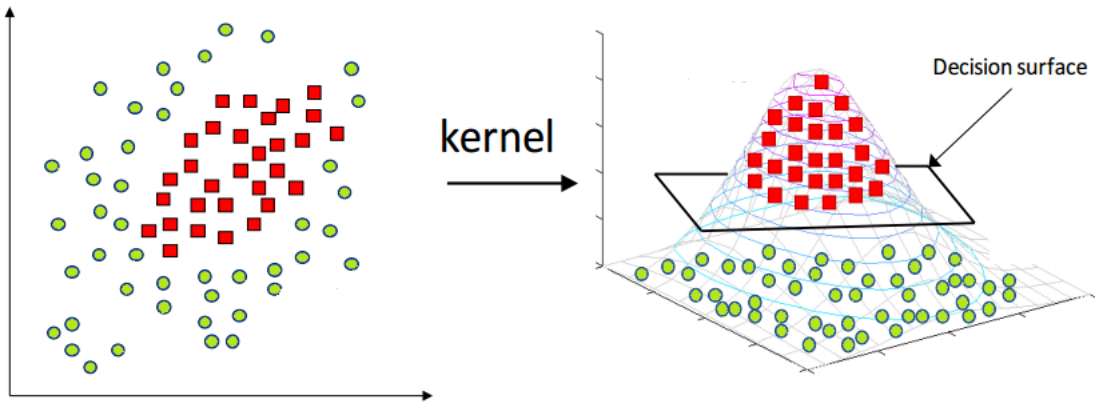

Figure 3.4. Kernels in SVM (Saini, 2021)

3.3.3. Naïve forecasting

Naive forecasting is one of the most basic forecasting methods for predicting historical datas. Naïve forecasting is the technique in which the last period's data used for the next period's forecast without predictions or adjusting the factors. Naïve forecast does not include the moving averages method, linear trend forecasting, and exponential smoothing and only focuses on previous data. Method also estimates confidence interval

based 5% or 20% depends of users need. Naïve forecasting commonly used for sales predictions or historical economic data.

$$Y_t = Y_{t-1}$$

Equation 3.3. Naïve Forecasting

3.4. Data Analysis Tools: R & R Studio

All collected twitter and currency data was analyzed on R-Studio which is an interface of R program. The R is a language, similar to widely used Python. R is an open-source language and environment for statistical computing and graphics and runs on almost every software system such as UNIX, Windows and MacOS. R provides a wide variety of statistical and graphical techniques such as linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering and etc., and is highly extensible. One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formula where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control (R Website).

4. RESULTS

4.1 Data Description

4.1.1 Currency data

The currency data was collected between November 1, 2019, and March 1, 2022, which was the beginning of the Covid-19 pandemic. Total 608 daily exchange rates were collected for each USD/TRY, EUR/TRY, and GBP/TRY exchange rate in CSV format. To measure the test's accuracy, 12 days' exchange rates were collected between March 2 and March 17, 2022, and weekends' data were not included in the main data.

The Figure 4.1 shows USD/TRY, EUR/TRY and GBP/TRY exchange rates between 1st of November 2019 and 1st of March 2022. The graph codes are given in Appendix 2. This graph shows that the Turkish Lira has decreased among U.S Dollar, Euro, and British Pound during this time period. The exchange rates are highly volatile and show upward trend over time. Especially on December 17, 2021, all exchange rate showed an unexpected increase and immediate decrease which decrease result of Turkish Governments' new policy against currency increase. These non-trend fluctuations responded by people with Tweets. The descriptive statistical analysis for each currency is given in Table 4.1.

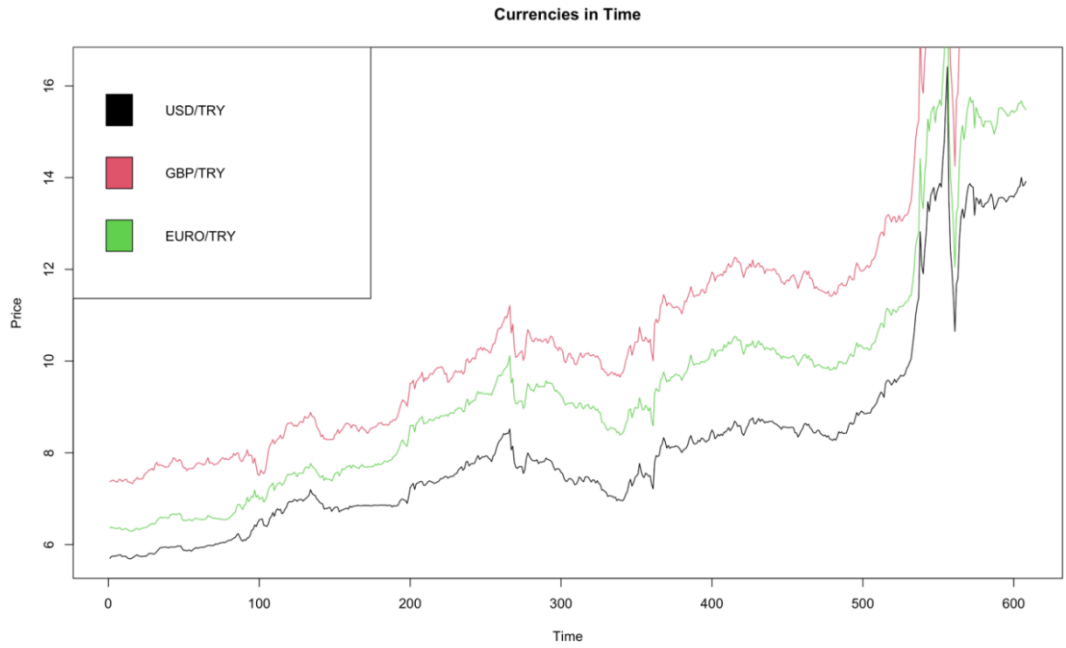

Figure 4.1. Currencies changes in time

Table 4.1. Statistical analysis of currencies

	Min Value	1st Quarter	Median	Mean	3rd Quarter	Max Value
USD/TRY	5.691	6.853	7.691	8.188	8.559	16.413
EUR/TRY	6.293	7.662	9.126	9.475	10.232	18.445
GBP/TRY	7.329	8.587	10.246	10.926	11.902	21.728

4.1.2 Twitter data

Within the scope of the Twitter data, a total number of 168.808 tweets were collected, which included 97663 tweets for USD/TRY, 66838 for EUR/TRY and 4307 tweets for GBP/TRY. Each data set was analyzed according to word distribution and "TL" and "Dollar" were the most common words for each data set even in EUR/TRY and GBP/TRY tweets. The detailed content analysis of each currencies' Tweets were located in 4.3.2 Content analysis section.

4.2. Currency Prediction without Twitter Sentiment Analysis

The mean absolute percentage error (MAPE) is used to calculate the accuracy rate. Mean absolute percentage error is a method for measuring of prediction accuracy of a forecasting method in statistics. The Equation of mean absolute percentage error is given in Equation 4.1. The currency prediction without sentiment analysis codes were given in Appendix 3.

$$MAPE = \sum_{t=0}^{n} \left| \frac{observed_t - predicted_t}{observed_t} \right| \text{ x } \frac{100}{N}$$

Equation 4.1. Equation of MAPE

According to Lewis (1982, pg.40), less than 10% MAPE rate was accepted as highly accurate forecasting. Between 10-20% was accepted as good forecast but higher error rates were defined as reasonable and unacceptable forecast. But MAPE score should be less than 5% in general for acceptance of accurate forecast.

The daily exchange rate data was collected between November 1, 2019, and March 1, 2022, for the British pound, Euro, and U.S Dollar and there are 608 daily exchange rate price for each currency. The test aims to forecast 12 future data which are between March

2, 2022, and March 17, 2022. The prediction without sentiment analysis does not include any other dependent or independent variable the test covers only past data. For the prediction without sentiment analysis, naïve forecast, RWF, and SVR were applied. Although SVR method is not considered as an effective forecasting method like naïve and random walk for time series, it has considerable amount of accuracy rate for time series data especially for short term prediction. However, SVR has more than 90% accuracy, publicly accepted methods like naïve and RWF have higher accuracy like more than 97%.

The Table 4.2 shows the real values of each currency and the predicted values for 12 days by both RWF and SVR. When real values compared with predicted ones based on mean absolute percentage error method RWF method makes closer predictions for EUR/TRY and GBP/TRY rates. On the other hand, for the USD/TRY rate SVR has lower error margin than RWF. The error margin of SVR on USD/TRY 2.89%, 8.98% for EUR/TRY and 6.57% for GBP/TRY. The RWF's average accuracy rate is 97.37%. For GBP/TRY rate, RWF has lowest error margin 2.2%, for EUR/TRY rate error margin 2.23% and USD/TRY 3.45%. Those results show that time series SVR give higher accuracy than RWF for USD/TRY, but for other exchange rates RWF give higher accuracy rate.

Table 4.2. Predicted values of currencies without sentiment

Date	USD/TRY				EUR/TRY				GBP/TRY			
	Real Value	RWF	Naive	SVR	Real Value	RWF	Naive	SVR	Real Value	RWF	Naive	SVR
2.03.22	14,01	13,93	13,92	13,78	15,58	15,50	15,48	15,14	18,78	18,51	18,54	18,16
3.03.22	14,10	13,94	13,92	13,79	15,60	15,51	15,48	15,03	18,81	18,58	18,54	18,09
4.03.22	14,19	13,96	13,92	13,81	15,50	15,53	15,48	14,92	18,76	18,60	18,54	18,01
7.03.22	14,37	13,97	13,92	13,84	15,59	15,54	15,48	14,81	18,82	18,62	18,54	17,94
8.03.22	14,48	13,98	13,92	13,88	15,78	15,56	15,48	14,69	18,97	18,64	18,54	17,87
9.03.22	14,64	14,00	13,92	13,93	16,21	15,57	15,48	14,56	19,29	18,65	18,54	17,80
10.03.22	14,81	14,01	13,92	14,01	16,27	15,59	15,48	14,44	19,38	18,67	18,54	17,75
11.03.22	14,76	14,02	13,92	14,10	16,10	15,60	15,48	14,31	19,23	18,69	18,54	17,69
14.03.22	14,78	14,04	13,92	14,21	16,16	15,62	15,48	14,19	19,21	18,71	18,54	17,66
15.03.22	14,70	14,05	13,92	14,35	16,10	15,63	15,48	14,07	19,17	18,73	18,54	17,63
16.03.22	14,61	14,07	13,92	14,51	16,12	15,65	15,48	13,96	19,20	18,75	18,54	17,62
17.03.22	14,70	14,08	13,92	14,71	16,31	15,66	15,48	13,86	19,34	18,76	18,54	17,63

The Figure 4.2-4.3-4.4 show that all prediction methods have greater than 90% accuracy in time series data without sentiment analysis, but these methods lose their

efficiencies when an unexpected increase or decrease occurred. That obviously appeared on SVR forecast graphs when flotation occurred on graph between 500-600 time period.

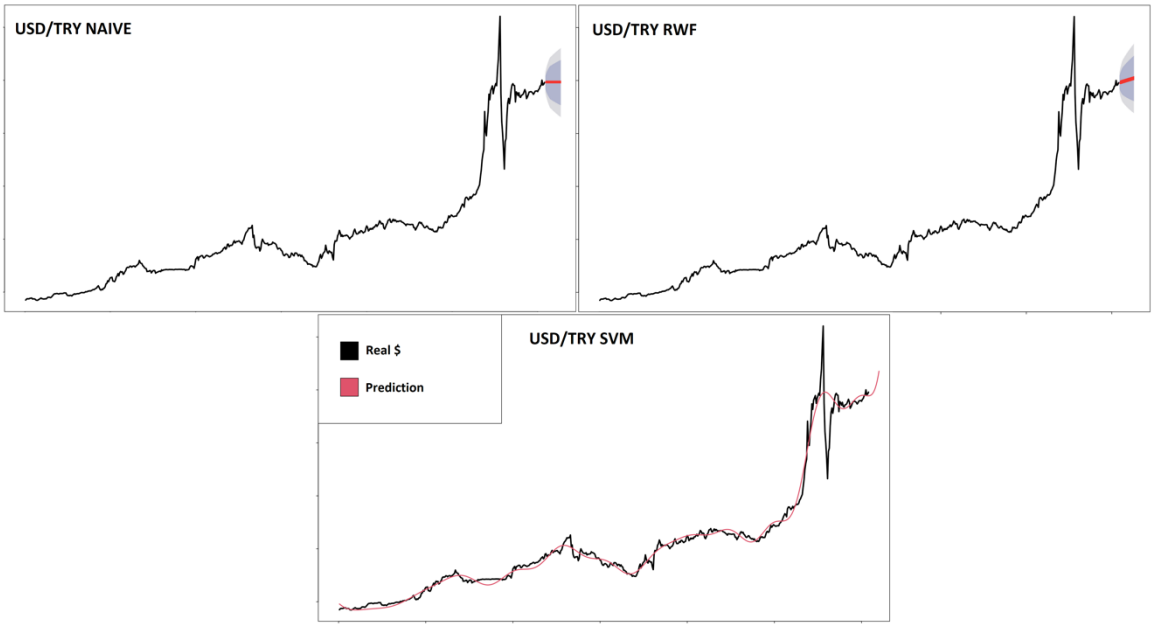

Figure 4.2. USD/TRY predictions without sentiment

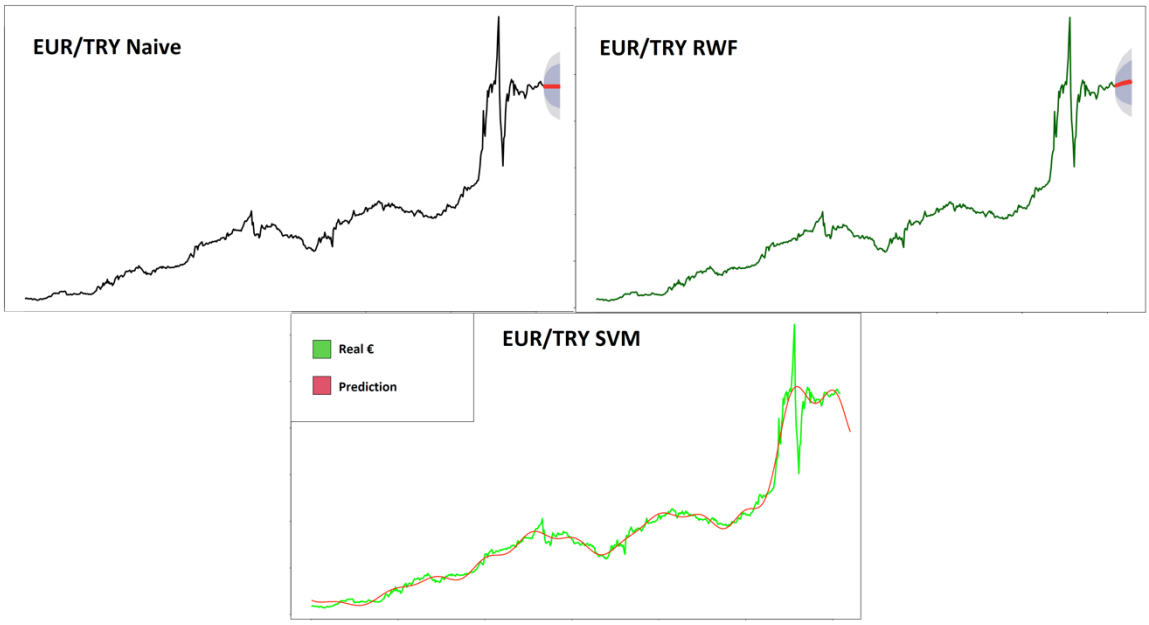

Figure 4.3. EUR/TRY predictions without sentiment

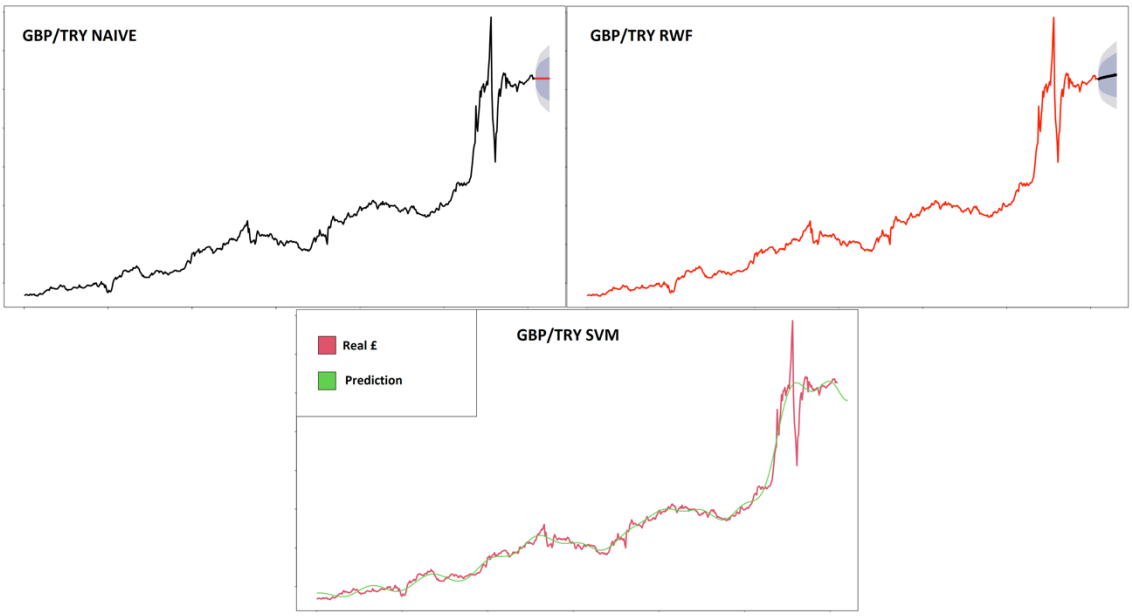

Figure 4.4. GBP/TRY predictions without sentiment

When the results of naïve forecast and RWF are compared (Table 4.3) the results show that RWF method has higher accuracy with approximately 97.7% on Euro and Pound exchange rate prediction. The highest accuracy rate with approximately 97% reached with SVR for dollar.

Table 4.3. Accuracy rates of predictions without sentiment analysis

	USD			EUR			GBP		
	SVR	**Naive**	**RWF**	**SVR**	**Naive**	**RWF**	**SVR**	**Naive**	**RWF**
MAPE	**2,9856**	4.0604	3.4589	8,9829	2.8467	**2.2709**	6.5775	2.8009	**2.1771**

4.3. Currency Prediction with Twitter Sentiment Analysis

4.3.1 Data preparing process

To analyze Tweets, data should be cleaned and prepared because Tweets should be on the monotype to make successful analysis. All data preparing process made with R program and the related codes are given in Appendix 4. All processes of data preparation for sentiment analysis is give in Figure 4.5. During the data preparation, first, data were converted into data frame and the words in the Tweets are separated into lines. Then website links were removed. After that, special characters were removed such as #, @, <, >, [,], /, ^, “the” and other common special characters used in Tweets. All numbers were removed and all letters were turned to lower case. Lastly, the stop words were cleaned

from Tweets (Stop words do not change the meaning of the sentence in a positive or negative way, such as conjunctions, question suffixes. Every language has different stop words set and Turkish stop words set were created and removed from Tweets). (Turkish Stop words used in this study is given in Appendix 1)

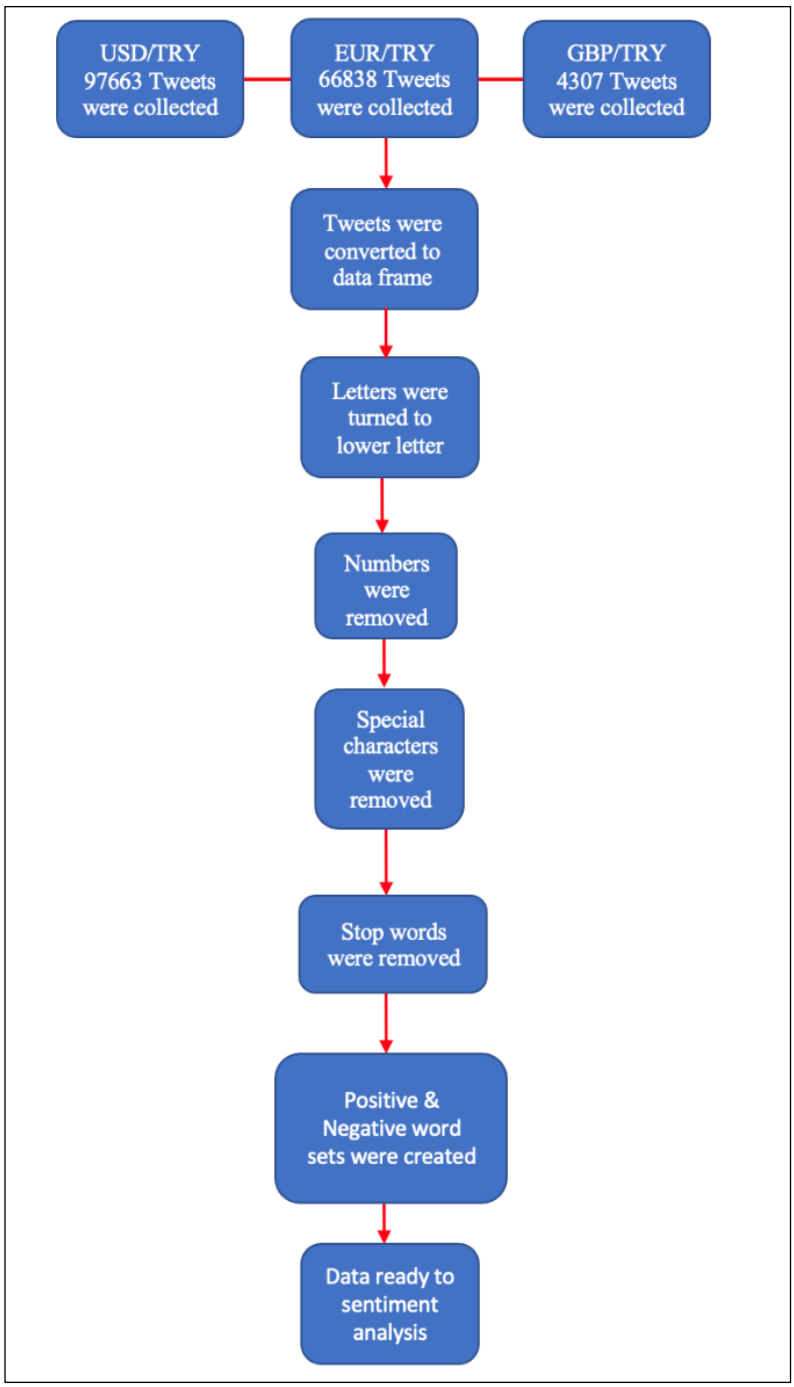

Figure 4.5 Data preparation process for sentiment analysis

4.3.2 Content analysis

After all data sets were prepared into proper type, Tweet contents were analyzed based on and the most common words and amounts of words used. The mostly used word lists were given in in Figure 4.6-4.7-4.9-4.10-4.12-4.13.

4.3.2.1 Dollar – Turkish Lira

In Figure 4.6, the words that were used more than 3150 times in USD/TRY data set, were listed. The prominent words in this list, “lost” (kaybetti), “raise” (zam) and “collapsed” (çöktü). These words are some of the most commonly used as negative sentiment words.

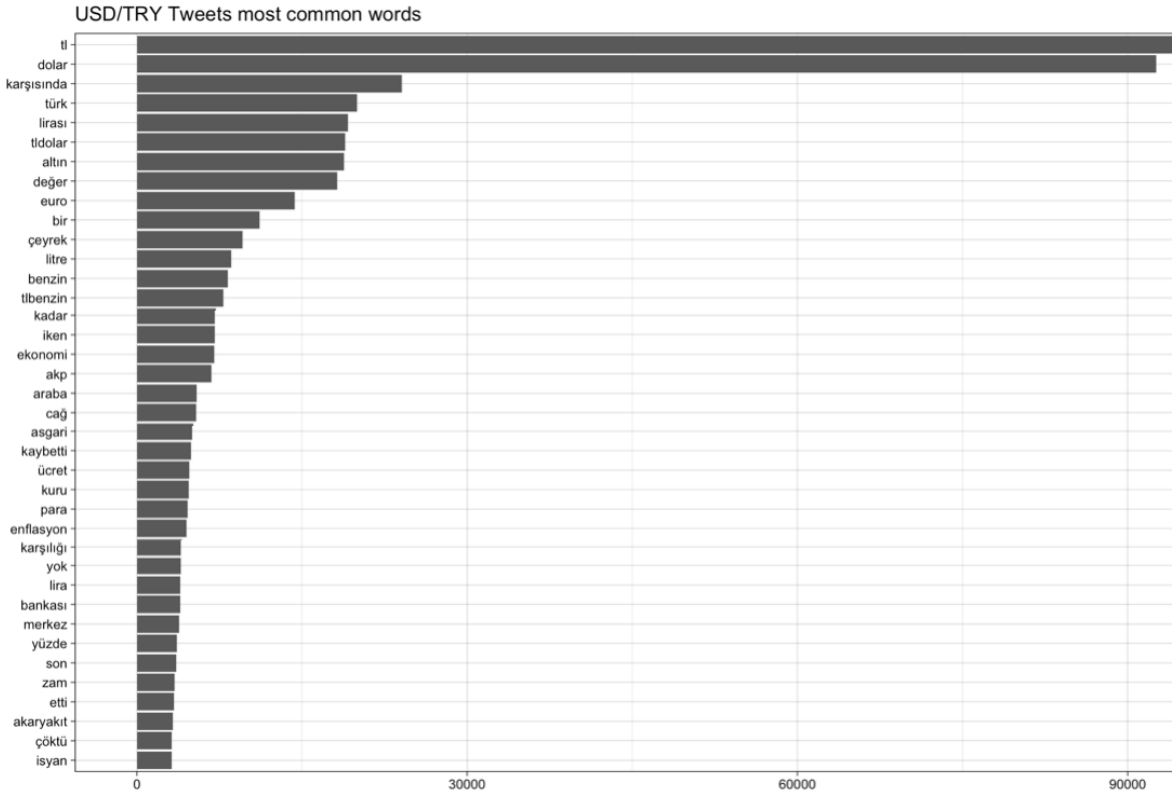

Figure 4.6. USD/TRY most common words

In Figure 4.7, word cloud of USD/TRY tweet data set shows some featured words such as crisis (kriz), poverty (yoksulluk), meltdown (erime), rebellion (isyan), unemployment (işsizlik), AKP (government party of Turkey), record (rekor) and other similar words emphasize negative sentiment words. These word sets reveal parallel attitude with exchange rate over time, it is expected because USD/TRY was 5.69 in 1st of November 2019, and 13.91 in 1st of March 2022. This is a very negative situation for the

Turkish people and it was expected that Turkish people's tweet contain mostly negative sentiment words.

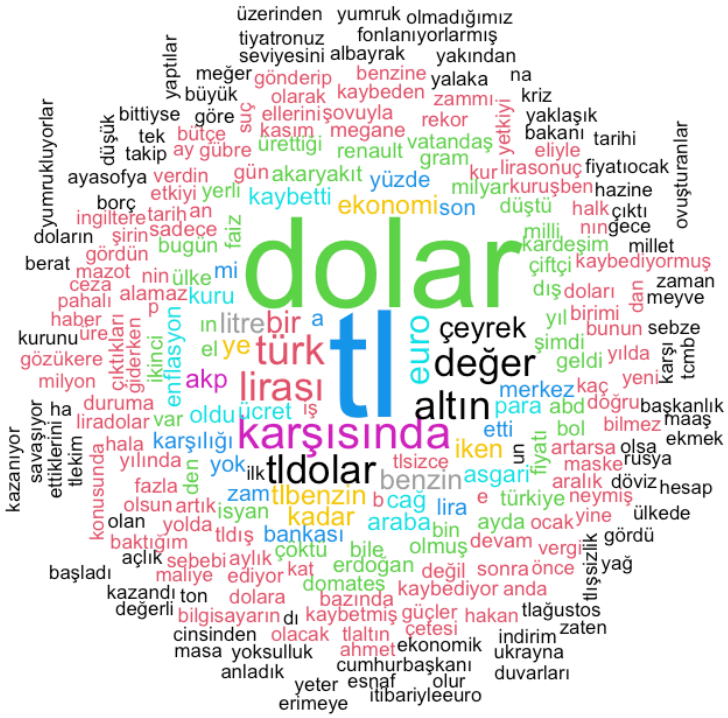

Figure 4.7. Word cloud of USD/TRY tweets

The Figure 4.8 shows USD/TRY Tweets frequencies in time and red line shows USD/TRY rate in time. There are some significant peaks on the graph. Especially the most frequent time in the graph which is 4th quarter of the 2021 on 21.12.2021, 3968 Tweets were tweeted, it coincides with the significant fluctuation in the exchange rate graph. Every peak of Tweet frequency coincides with currency fluctuations. The highest number of Tweets in a single day occurred on 24th of February 2020, totally 4423 Tweets were posted and in February 2020, USD/TRY exceeds 6TL for the first time in its history.

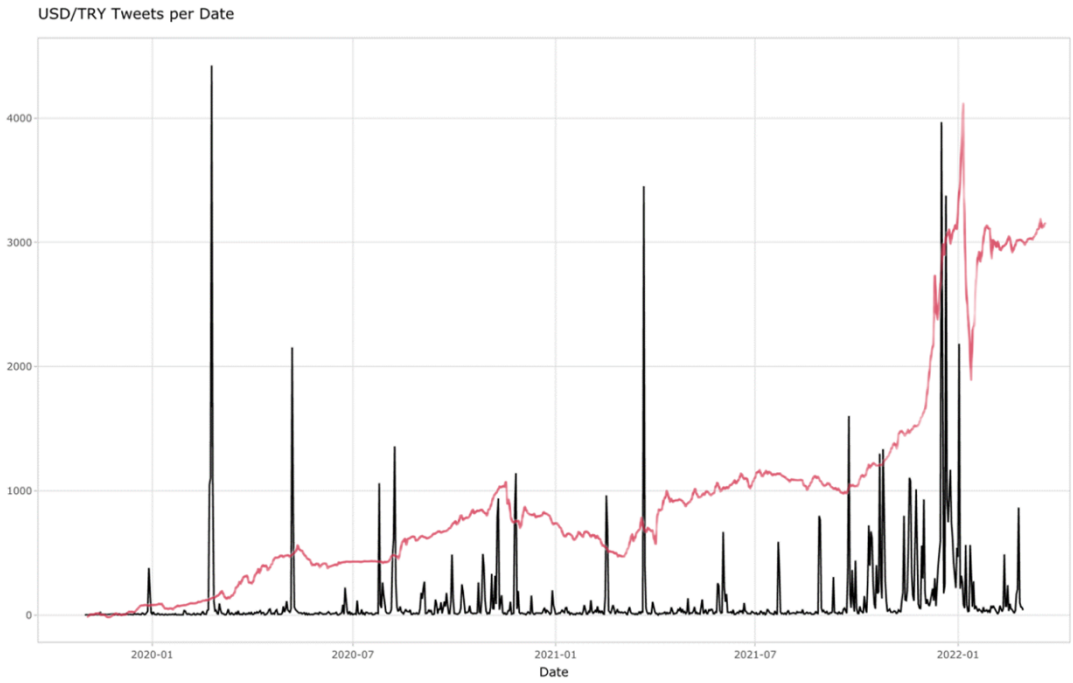

Figure 4.8. USD/TRY tweet frequencies per date

4.3.2.2 Euro – Turkish Lira

The Figure 4.9 show words that use more than 2500 times in EUR/TRY tweets. When the figure is examined, the second most used word is “dollar” is which used more than 50000 times, however tweets were collected for Euro vs Turkish Lira currency, so this situation leads to the conclusion that the dollar is a currency that Turkish people more concern about and affects their lives more than other currencies. Besides to these in the same time period in the tweets collected for these exchange rates, Turkish people tweeted 30825 more tweets about the Dollar than about Euro.

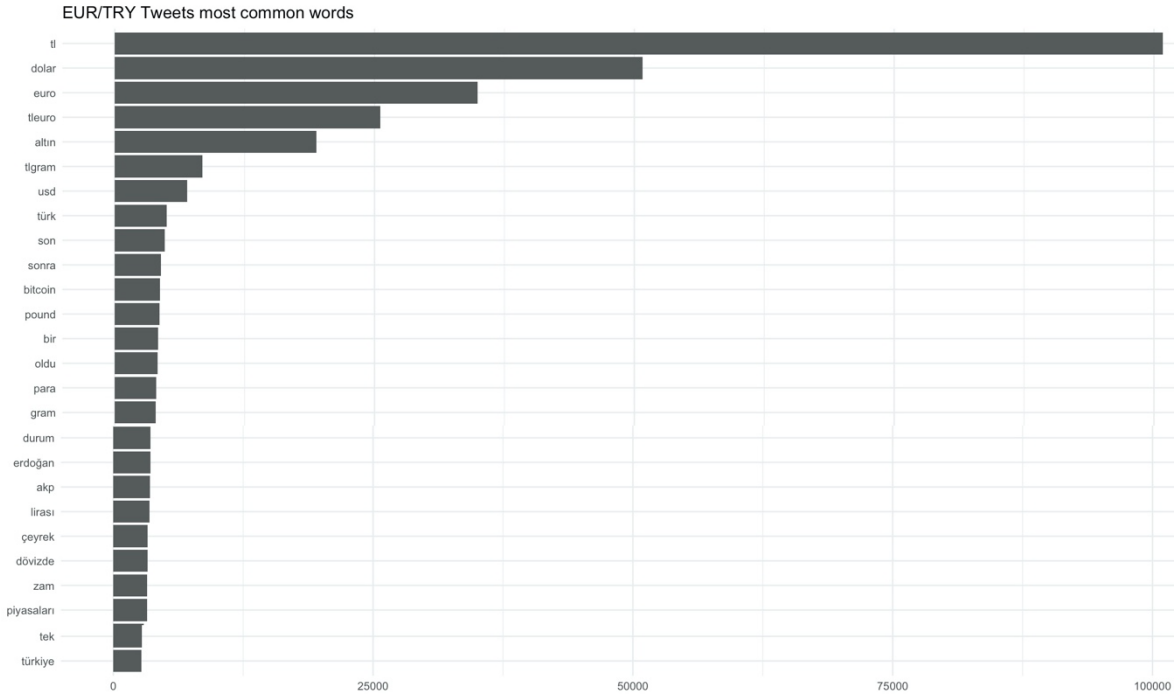

Figure 4.9. EUR/TRY most common words

The Figure 4.10 shows the word cloud of EUR/TRY tweets. Some featured words according to the word cloud are interest (faiz), raise (zam), poverty (yoksulluk), inflation (enflasyon) and other similar words emphasize negative sentiment words. There are only a few positive words in word cloud, such as discount ("indirim") and great ("harika").

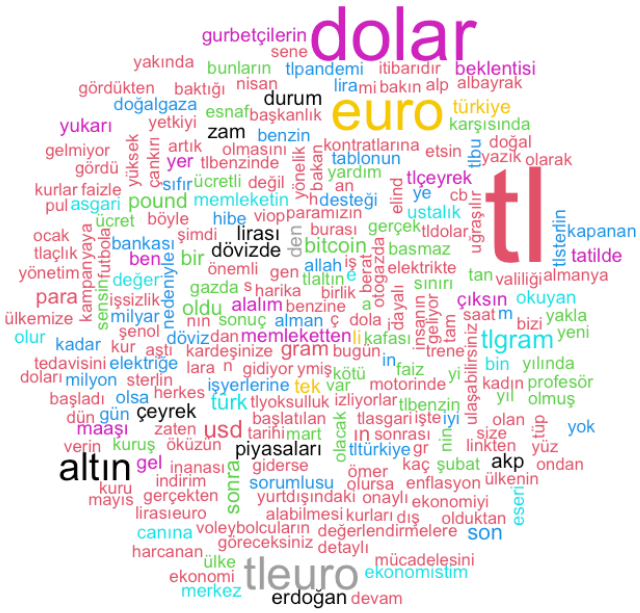

Figure 4.10. Word cloud of USD/TRY tweets

The Figure 4.11 shows Euro Turkish Lira Tweets frequency and green line shows currency graph. There is an obvious increase trend which coincides with currency graph. In 3 periods, significant peaks were observed. These periods coincide with 3 major increases in currency period, as seen on the graph. Since the EUR/TRY currency is highly complex and volatile, Euro Turkish Lira Tweets have very similar trend that are responsive to changes in currency. As foreseen, Tweets are seeming suitable for the make an accurate prediction.

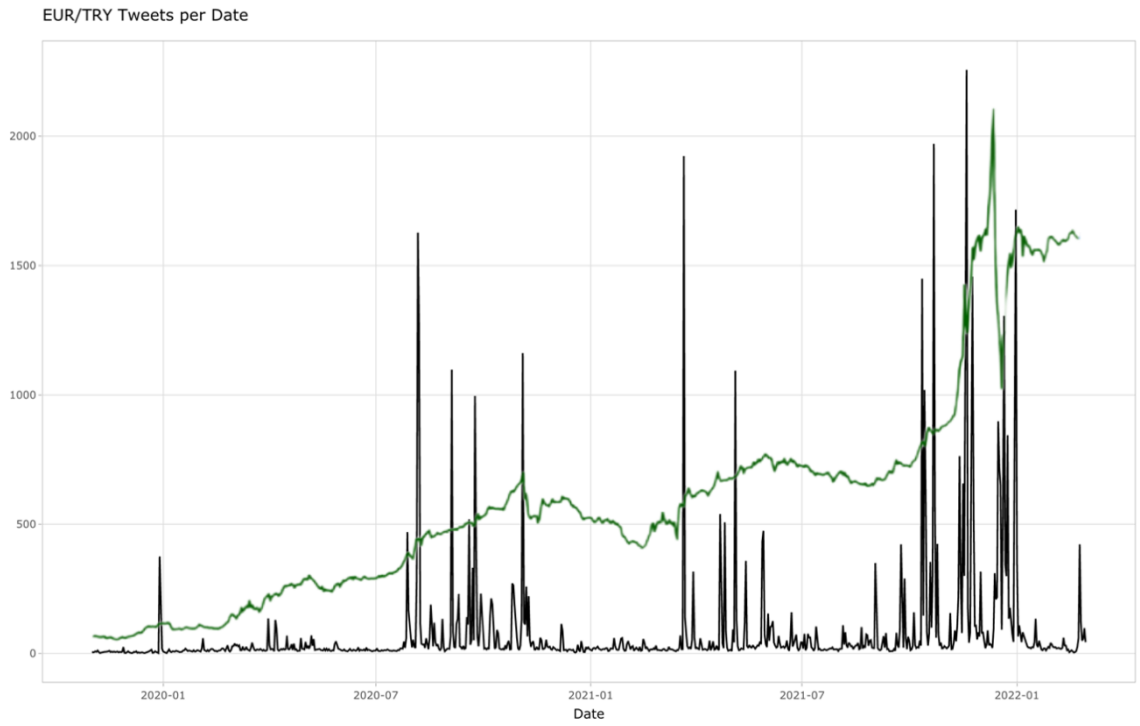

Figure 4.11. EUR/TRY tweet frequencies per date

4.3.2.3 Sterling – Turkish Lira

The Figure 4.12 shows the words which were used more than 200 times in tweets of GBP/TRY currency. When the figure is examined, the most common two words are same with other currencies' data sets, which are "dollar" and "tl". So, it is proven that the dollar is the most concerned currency for Turkish people. Also, the number of tweets show that sterling has least concern among Turkey because there are only 4307 tweets between beginning of COVID-19 pandemic and March 2022. One of the featured commonly used word is "lost" (kaybetti) with 1445 times.

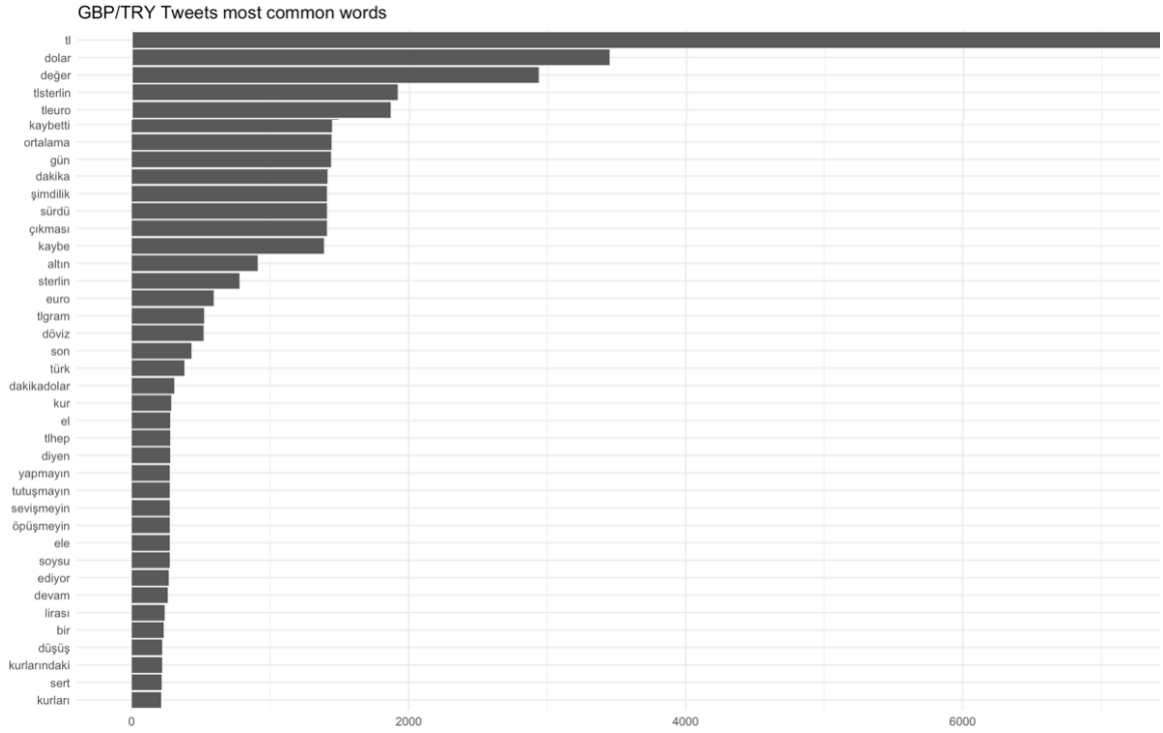

Figure 4.12. GBP/TRY most common words

The Figure 4.13 shows word cloud of most commonly used words of GBP/TRY exchange rate. Again, in these tweet data set, poverty, unemployment, interest, crisis, and inflation are featured negative sentiment words.

According to all these tweets, gold price (altın) is one of the most commonly used word. Totally 39152 times people mentioned about gold in their tweets about all the three currency rates. Also, bitcoin, benzene, petrol, Ukraine and Erdoğan who is President of Turkey are other widely used words in all Tweet data.

Figure 4.13. Word cloud of GBP/TRY tweets

The Figure 4.14 shows Sterling Turkish Lira Tweet frequency and red line shows currency change in time, there is only two significant peaks first one, is 23rd of November 2021, total 779 Tweets were tweeted, second peak is 16th December 2021 and total 626 Tweets were tweeted. There are 2 other periods that Tweet amount increases, these periods same periods with currency increase. According to Sterling Tweets, it is difficult to make an obvious and meaningful interpretation. The reason for this should be lack of sufficient number of Tweets, because compared to other currencies there are only 4307 Tweets were collected. Although there is an increasing trend, it may not be enough to make an accurate prediction because the currency graph highly volatile.

Figure 4.14. GBP/TRY tweets frequency

4.3.3 Sentiment analysis

The negative and positive words set were created from contents of collected Tweets. Sentiment words should be related with topic to assure highest relevance and totally 44 words were selected as positive and negative sentiment. Table 4.4 shows the Turkish and English translation of sentiment words used in sentiment analysis.

Table 4.4. Sentiment words

Negative		Positive	
Words (TR)	**Words**	**Words (TR)**	**Words**
kaybediyor	loosing	düşüyor	falling
kayıp	loss	düşüş	fall
eriyor	meltdown	iyi	good
olumsuz	negative	güçlü	strong
değersiz	worthless	kazandı	gain
artış	increase	korudu	kept
erimeye	meltdown	koruyor	remains
zam	raise	olumlu	positive
kötü	bad	değerli	valuable
çöktü	collapsed	düştü	fell
isyan	rebellion	indirim	discount
kriz	crisis	harika	great
kaybetti	lost	müjde	good news
kaybı	loss	güzel	nice
eridi	melted	mutlu	happy
değersizleşme	devaluation		
yoksulluk	poverty		
devalüasyon	devaluation		
perişan	miserable		
artarsa	increase		
kaybetmiş	lost		
kaybeden	loser		
açlık	poverty		
bittik	end		
iflas	bankruptcy		
işsizlik	unemployment		
enflasyon	inflation		
zarar	deprediation		
mutsuz	unhappy		

The Figure 4.15 show most common positive and negative sentiment words in USD/TRY Tweets. While the most common negative sentiment words are lost and inflation, the most common positive sentiment words are fell and discount. It's observed that there is enormous amount of difference between usage of positive and negative words.

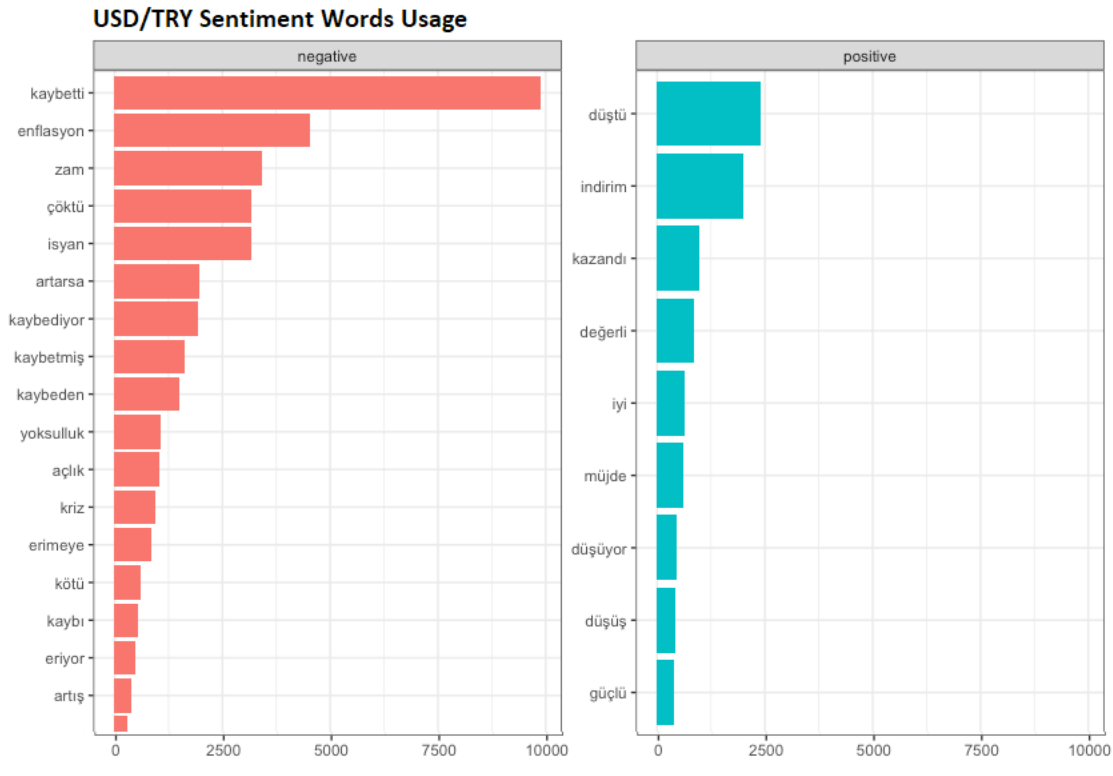

Figure 4.15. USD/TRY sentiment words

The Figure 4.16 shows positive and negative Tweets frequency. The graph in the upper side shows positive Tweets and below one shows negative Tweets. Negative Tweets reach 6000 in peak but approximately 1000 positive Tweets only in peak time. The difference between positive and negative Tweets reveals the increase of U.S. dollar against to Turkish Lira. Also increase of negative Tweets and decrease of positive Tweets in time, have a similar trend with USD/TRY rate. It is expected that increase of USD/TRY will affect Turkish people's attitude in negative way.

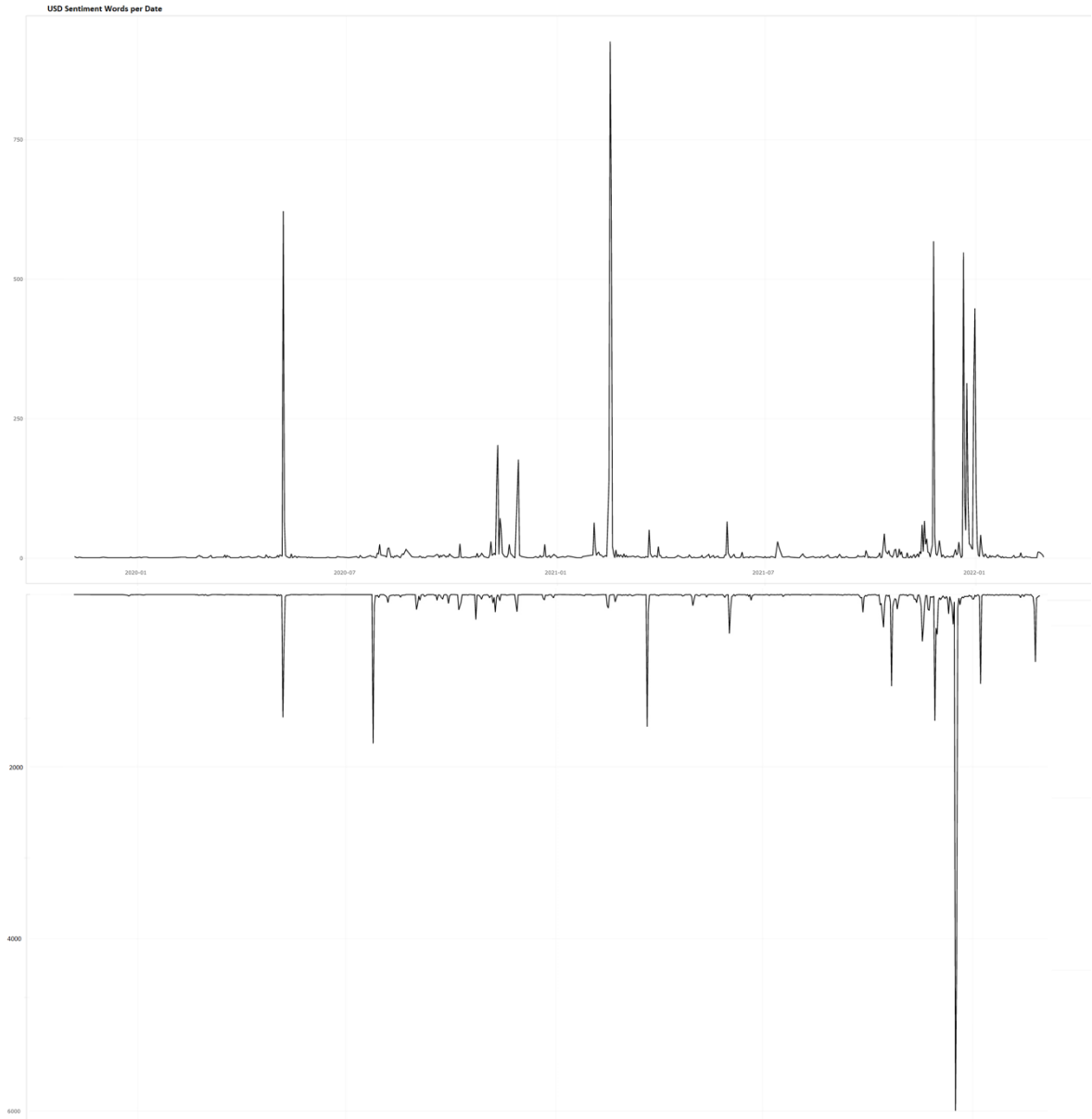

Figure 4.16. USD/TRY sentiments per date

The Figure 4.17 shows most common used sentiment words in EUR/TRY Tweets. The highest amount of negative word used is raise with more than 3000 times, for positive sentiment word is discount with used more than 1000 times. There is similar situation with USD/TRY, negative words used more than positive words. That situation coincides with EUR/TRY' exponential increase graph.

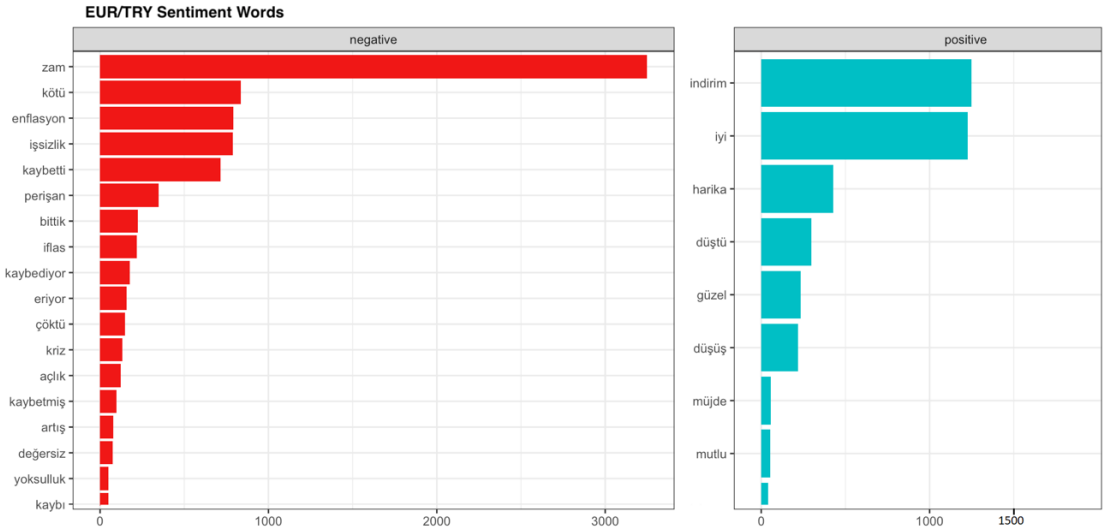

Figure 4.17. EUR/TRY sentiment words

The Figure 4.18 shows frequency of negative and positive Tweets, unlike other currencies EUR/TRY negative and positive Tweets frequencies have a disposition that does not fit perfectly into the trend of the currency. The frequency of Tweets without separate negative and positive, has more accurate trend with currency change.

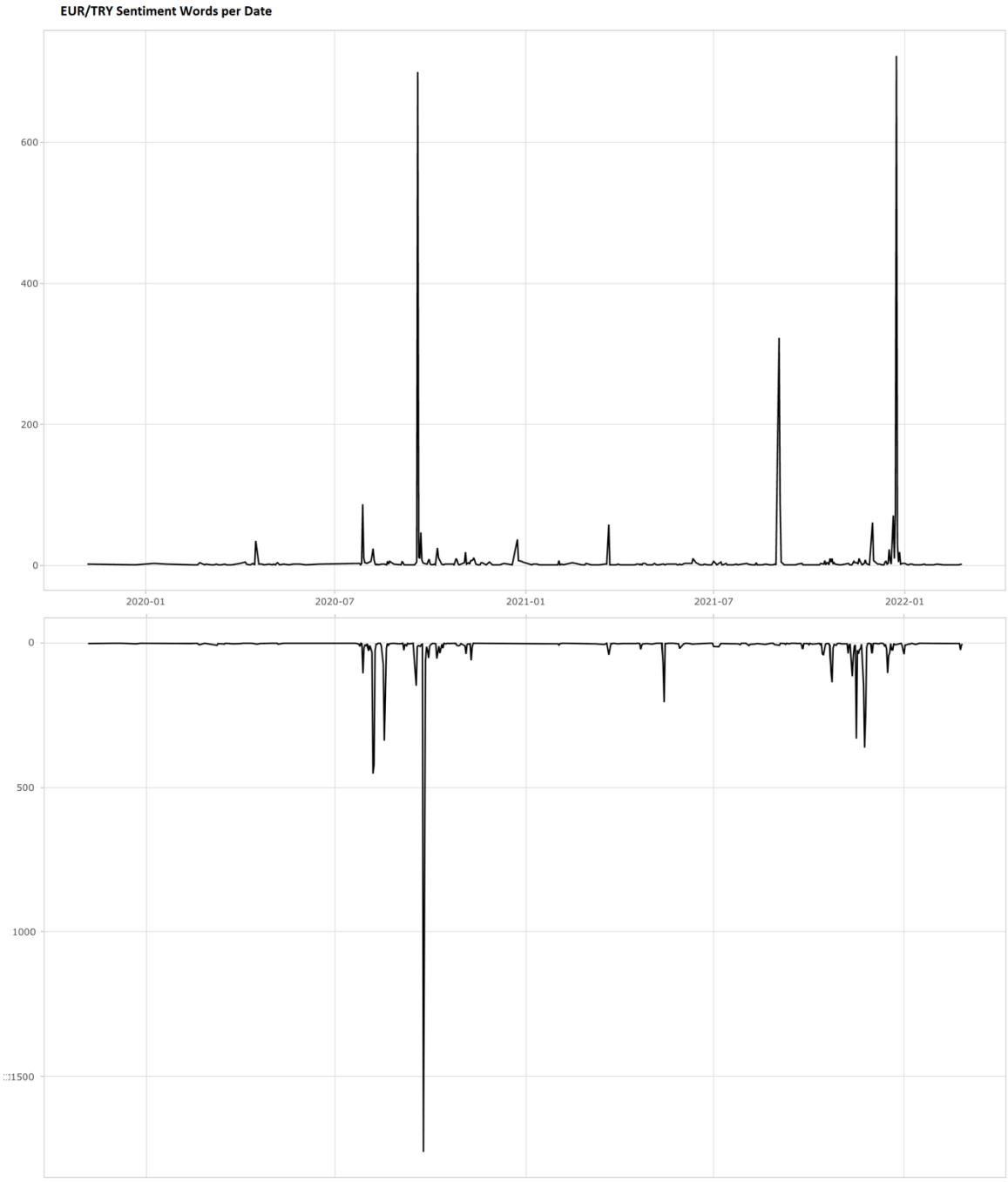

Figure 4.18. EUR/TRY sentiments per date

The Figure 4.19 shows most commonly used sentiment words in GBP/TRY Tweets. The most used negative word is “lost” with nearly 3000 times and for the positive word “fall” is the most used word which is used less than 300 times. As mentioned before, there aren’t enough data to make accurate prediction in GBP/TRY Tweets. But it is obvious that negative attitude towards GBP/TRY rate is seen in the figure above.

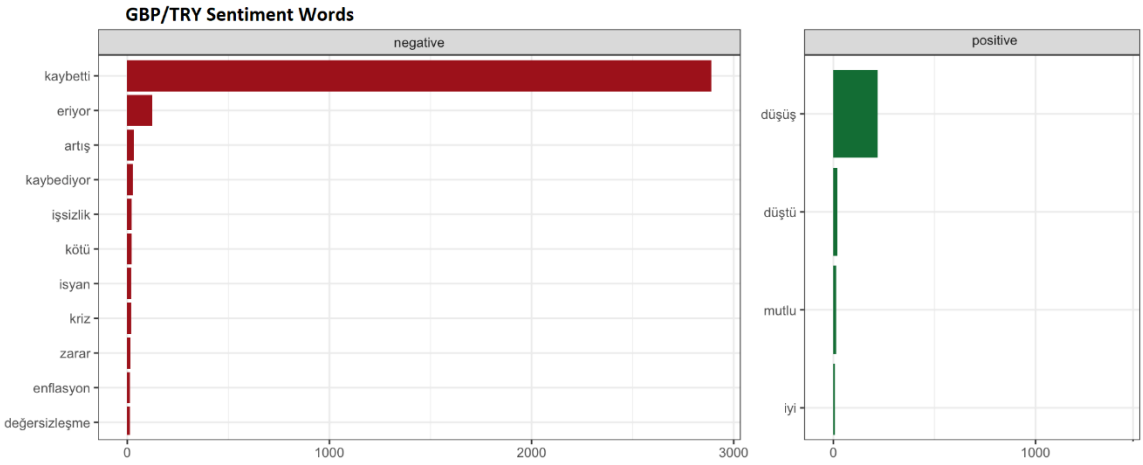

Figure 4.19. GBP/TRY sentiment words

The Figure 4.20 shows negative and positive Tweets frequencies about British Sterling and Turkish Lira in time. The highest number of negative Tweets were tweeted on 23rd of November 2021, that the day of the excessive increase which is 12.47% occurred in GBP/TRY rate. The highest number of positive Tweets were posted on 24th of December 2021, which is the day GBP/TRY decrease more than 11% in two days.

Although there are not enough Tweets, the effect of sentiment analysis can be clearly seen because the times when positive and negative Tweets are posted and the changes in the exchange rate coincide.

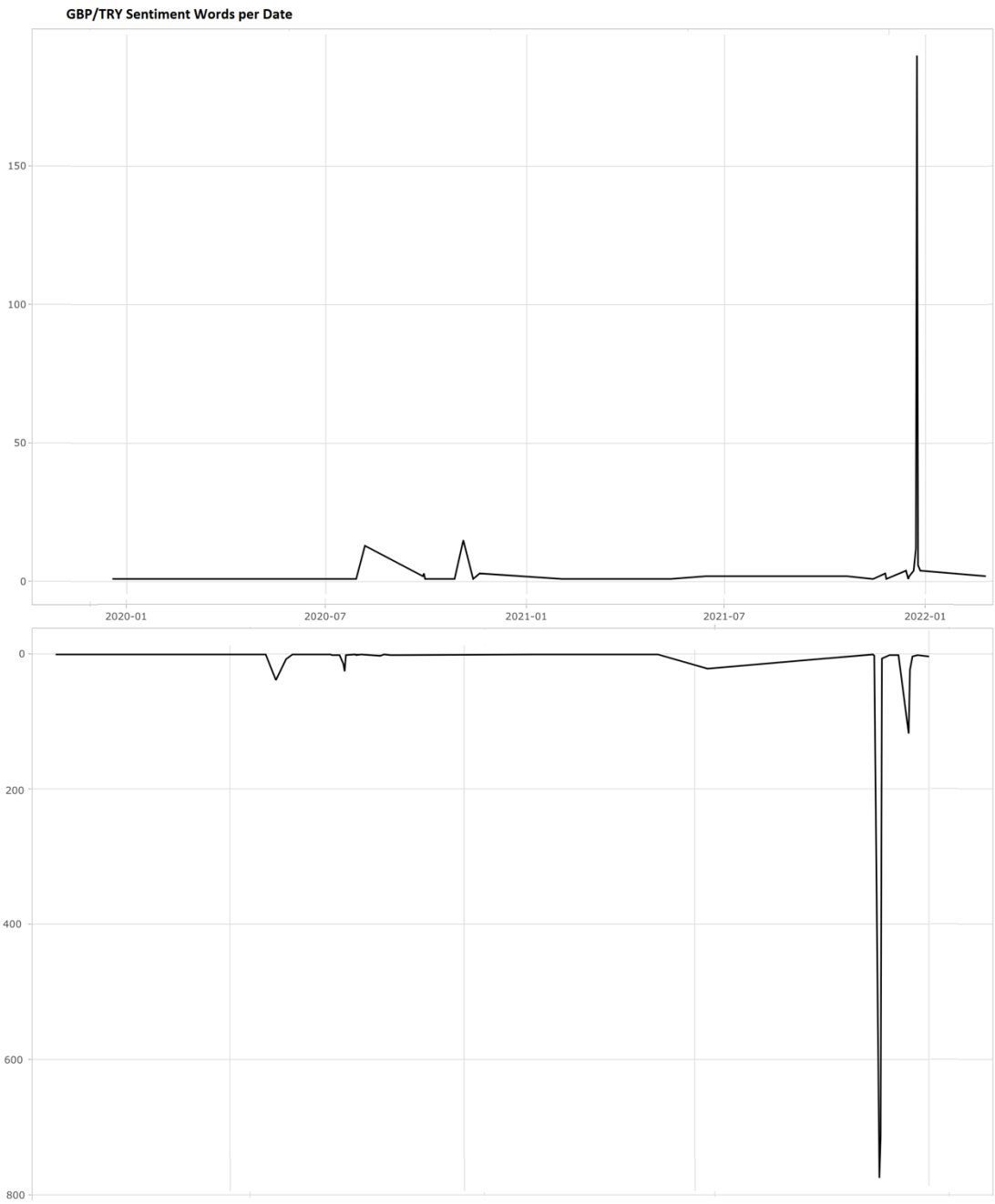

Figure 4.20. GBP/TRY sentiments per date

4.3.4. SVR analysis to predict exchange rate with sentiment analysis

After sentiment analysis of each currency type, all Tweets were scored positive and negative (each positive and negative words scored as +1 or -1). Then the scores were sorted based on days posted for all three exchange rates. Each day from the beginning of the data set which is November 1st had a sentiment score and these scores used as a variable input for SVR prediction. After scores were created, there are two main variables

for prediction first one is currency data and the second one is sentiment score. SVR analysis were performed via R program, and 12 days predicted values created and stated in Table 4.5. After predicted values created, these values compared with real values and accuracy test made with MAPE. R codes stated in Appendix 4.

When it is compared, currency prediction with sentiment analysis has more accurate results than prediction without sentiment analysis. Especially SVR with sentiment analysis for USD/TRY rate has the highest accuracy among all other methods. According to the results given in Table 4.3, SVR method give highest accuracy in USD/TRY rate. Accuracy rate increased and reached approximately 97.46% which is a highly accurate score for a volatile and unstable exchange rate. when sentiment scores were included. Not only in USD/TRY, but also EUR/TRY and GBP/TRY exchange rates had higher accuracy rate with SVR with sentiment analysis. SVR with sentiment analysis give 93.61% for EUR/TRY. All the results comparing accuracies are presented in Table 4.5.

The sentiment analysis of GBP/TRY exchange rate has increased accuracy rate by less than 0.1% and reached to 93.495% which is the lowest accuracy rate among other sentiment analysis scores. This result was not surprising since, there are not enough Tweets posted for GBP/TRY rate as compared to other currency types.

Table 4.5. Predicted values of currencies with sentiment analysis

	USD/TRY		EUR/TRY		GBP/TRY	
Date	**Real**	**Prediction (SVR)**	**Real**	**Prediction (SVR)**	**Real**	**Prediction (SVR)**
2.03.2022	14,0084	13,8593	15,5787	15,1868	18,7755	18,1415
3.03.2022	14,0950	13,8625	15,5947	15,1385	18,8112	18,0826
4.03.2022	14,1845	13,8698	15,4980	15,0825	18,7590	17,9614
7.03.2022	14,3666	13,8752	15,5915	14,9835	18,8242	17,9840
8.03.2022	14,4762	13,8914	15,7776	14,9079	18,9682	17,9996
9.03.2022	14,6378	13,9612	16,2114	14,9235	19,2912	17,8753
10.03.2022	14,8131	14,1493	16,2692	14,9460	19,3800	17,7844
11.03.2022	14,7595	14,2063	16,0952	14,9202	19,2334	17,7548
14.03.2022	14,7754	14,3106	16,1628	14,8501	19,2080	17,7013
15.03.2022	14,7002	14,4363	16,0982	14,7155	19,1661	17,6055
16.03.2022	14,6077	14,5807	16,1152	14,5916	19,2018	17,5801
17.03.2022	14,6960	14,6812	16,3082	14,7502	19,3405	17,5424

The Table 4.6 shows all accuracy results of this research which are SVR without sentiment analysis, naïve forecast, RWF and SVR with Twitter sentiment analysis. For

USD/TRY highest accuracy achieved by SVR with Twitter sentiment analysis, for EUR/TRY highest accuracy rate achieved by RWF without and for GBP/TRY highest accuracy rate achieved by RWF method.

Table 4.6. Accuracy tests of predictions (MAPE results)

	USD/TRY			**EUR/TRY**			**GBP/TRY**		
	SVR	Naive	RWF	SVR	Naive	RWF	SVR	Naive	RWF
MAPE	**2.986**	4.060	3.459	8.983	2.847	**2.271**	6.577	2.801	**2.177**
MAPE for SVR with Twitter sentiment analysis	**2.539**			**6.386**			**6.505**		

5. DISCUSSION

This chapter of thesis aims to discuss the findings and compare them with the previous studies. This research aims to measure, whether it is possible to make more accurate exchange rate prediction with sentiment analysis of Twitter data or not. In this scope, several tests and analysis have been performed with the collected currency and Twitter data.

Two major tests were conducted. The first one is the currency prediction without sentiment analysis, in which only the currency data was taken as a dependent variable and analyzed with 3 different methods, which are RWF, naïve forecast and SVR. The second test is the currency prediction with Twitter sentiment analysis, in which sentiment scores and currency data taken as variables and analysis was conducted with SVR method. These methods were chosen according to previous researches mentioned in Table 2.1. Yasir (2019 and 2020), Chethan & Sangeetha (2020) and McCoy (2015) used SVM in their researches and gained higher accuracy. Also, only a few studies worked on Turkish Lira which are Ozturk & Ciftci (2014) and Alkoç (2019). There aren't any study works on both SVM and Turkish Lira with Turkish Tweets.

The results differed for each currency type. For USD/TRY highest accuracy was achieved by SVR methods in both with and without sentiment analysis, and highest accuracy was achieved which was 97.46% with sentiment analysis. These findings imply that understanding and analyzing the public's reaction correctly could improve future predictions.

For EUR/TRY rate, highest accuracy was achieved by RWF method, which was 97.73%. When Twitter sentiment analysis with SVR method was applied, accuracy was increased by 2.6% and reached 93.61% compared to the SVR without sentiment analysis. However still RWF methodology without sentiment analysis has higher accuracy. But the results show that Twitter sentiment analysis improve accuracy, especially on upsurges and slumps, public reactions happen immediately.

In Sterling Turkish Lira currency rate, sentiment analysis increased accuracy less than 0.1% and still RWF have higher accuracy with 97.82% which is a quite high accuracy rate for a single variable test. The increase on accuracy with the sentiment analysis proves

that sentiment analysis of Twitter data, a good indicator of currency movement. The accuracy increase can be considered as a successful result considering the insufficient Twitter data, because for GBP/TRY currency only 4307 Tweets were collected which was quite low number according to other currency Tweets.

When consider all accuracy tests of prediction average accuracy is 94.86%, which is quite accurate and acceptable result for currency prediction. Especially considering USD/TRY accuracy which is 97.46%, the results are highly accurate. The reason for having higher accuracy than other currencies is collecting more Tweets about USD/TRY. The findings, show that if the number of Tweets increases, accuracy rate will increase. Although time interval for this research is approximately 2.5 years and total 3 currencies were targeted, number of Tweets only 168.808, because only Tweets in Turkish were collected.

In previous research such as Köksal and friends (2021), about bitcoin prediction with sentiment analysis, accuracy rate is 94.16% with random forest regression and this rate support that sentiment analysis is an effective method for prediction although the analysis was performed with total 3737 Tweets (Köksal et al., 2021). Also, in other research about bitcoin prediction, the accuracy rate is 81.39% accuracy which is a low rate to be considered a successful prediction (Pant et al. 2018). However, Köksal's (2021) research achieved highest accuracy with random forest regression. Kumaş (2021) achieve highest accuracy with SVM method and accuracy rate was 73%. In different research about Twitter sentiment analysis on exchange rate movements shows that Twitter sentiment analysis can correctly predict 58% of increases in the exchange rate and 82% of decreases (Öztürk & Çiftçi, 2014). In research about Recurrent Neural Network (RNN) based bitcoin price prediction with Twitter sentiment analysis, overall price prediction accuracy using RNN is found to be 77.62% (Pant et al., 2018).

In Alkoç and Sütcü's research (2019), it is mentioned that; there is a positive and strong relationship between the number of Tweets and currency prices. When currency price increases, the number of Tweet posted increases. Based on this inference, the positive relation between number of tweets and currency prices obvious on Figure 4.8-4.11-4.14, which show that currency price increases, the number of Tweets increases immediately. Also, there are other studies focuses on number of Tweets to predict Bitcoin prices. For example, Shen et al., (2019) found that the number of tweets is a significant driver of next

day trading volume and realized volatility. Based on this, it is possible to conclude that, there is a positive and strong relationship between currency prices and negative Tweets, and there is a negative relationship between currency prices and positive tweets as seen in Figure 4.8-4.11-4.14. The results show that people are more sensitive against to increase about exchange rate, and people tend to post Tweets about negative situations.

All the results and researches about sentiment analysis deduce and prove that Twitter is a good predictor and a rich seam of information for prediction if it is properly examined. The reason for this, people have tendency to response daily life's issue on social media, and especially in Turkey exchange rate is an important part of both professional and daily life. Also as already mentioned in previous sections, nowadays government officials actively use social media especially Twitter as an easy and quick way of public speech, which makes Twitter an official and an accepted communication channel.

6. CONCLUSION AND FUTURE WORKS

6.1. Conclusion

This research aims to investigate whether it is possible to make more accurate exchange rate prediction by Twitter sentiment analysis or not. Daily data of USD/TRY, EUR/TRY and GBP/TRY currency rate, ranging from 1st of November 2019 to 1st of March 2022 were collected. Tweets about these currencies were used as input parameters of the model analysis.

There were two main predictions made with three different methods. To compare effects of sentiment analysis on accuracy, firstly predictions made without sentiment analysis with SVR, naïve forecast and RWF. These methods selected from previous researches which mentioned in Table 2.1. The second prediction made based on sentiment analysis scores of collected Tweets and prediction made with SVR method.

The sentiment analysis was performed by using 168.808 Tweets. The obtain sentiment scores, Tweets were analyzed with predetermined sentiment words which were collected from words used in Tweets. All negative words scored as -1 and all positive words scored as +1, then every Tweet had a score that determine Tweet was positive or negative, and Tweets that don't have any positive or negative score, considered as neutral Tweet and scored as 0. After scoring process is completed, Tweet score data set used as a variable for SVR prediction.

Results assert that SVR with sentiment analysis have higher accuracy than SVR without sentiment analysis, in another saying sentiment analyze has improved the predictive power of SVR method. On the other hand, RWF outperforms the SVR and naïve forecast on GBP/TRY and EUR/TRY currencies, but for USD/TRY currency SVM had highest accuracy for prediction. Therefore, Twitter sentiment analysis can be used with SVR to enhance the accuracy of economic predictions. These prediction models can guide investors, government authorities and decision makers to manage their future plans.

6.2. Future Works

For future research, a model-based analysis from the Tweets of government presidents, ministers, government authorities, policy makers, major investors, big company

managers and others, that can influence exchange rate movements, and may prove more accurate results. Unlike current model, in such model, effects of Tweets on exchange rate will be observed. A model like this, could be more effective and more easily observable on cryptocurrencies, because cryptocurrencies are more volatile and complex than exchange rates. Also, cryptocurrencies are tied to investors movements.

Furthermore, a similar study can be done to observe the exchange rate change in different time intervals such as effects of length of time interval on accuracy rate can be observed.

REFERENCES

Alkoç, M. (2019). *Big Data in Financial Markets: Twitter Analysis of Forex Data.* (Unpublished doctoral thesis). Marmara University, İstanbul, Turkey

Alkoç, M., & Sütcü, C. S. (2019). Büyük Veri ile Foreks Algısının Twitter Üzerinden Gündem Belirleme Kuramı Bağlamında Araştırılması. *İstanbul Sosyal Bilimler Dergisi*, 82-103.

Chauhan, N. S., (2022, April 8). *Naïve Bayes Algorithm: Everything You Need to Know.* KDNuggets. Retrieved from (https://www.kdnuggets.com/2020/06/naive-bayes-algorithm-everything.html)

Chen, J. (2022, July 1). *Exchange Rate.* Investopedia. Retrieved from https://www.investopedia.com/terms/e/exchangerate.asp

Chethan, N., & Sangeetha, R. (2020). Sentiment Analysis of Twitter Data to Examine the Movement of Exchange Rate and Sensex. *Journal of Computational and Theoretical Nanoscience*, *17*(8), 3323-3327.

Gholampour, V., & van Wincoop, E. (2019). Exchange rate disconnect and private information: What can we learn from euro-dollar tweets? *Journal of international Economics*, *119*, 111-132.

Investing. (n.d.). *EUR/TRY – Euro Turkish Lira*. Currencies. Retrieved March 10, 2022, from https://www.investing.com/currencies/eur-try

Investing. (n.d.). *GBP/TRY – British Pound Turkish Lira.* Currencies. Retrieved March 10, 2022, from https://www.investing.com/currencies/gbp-try

Investing. (n.d.). *USD/TRY - US Dollar Turkish Lira.* Currencies. Retrieved March 10, 2022, from https://www.investing.com/currencies/usd-try

Iqbal, M. (2022, June 30). *Twitter Revenue and Usage Statistics.* Business od Apps. Retrieved from https://www.businessofapps.com/data/twitter-statistics/

Janetzko, D. (2014). Using Twitter to Model the EUR/USD Exchange Rate. *ArXiv, abs/1402.1624*.

Köksal, B., Erdem, G., Türkeli, C., & Öztürk, Z. K. (2021). Twitter'da Duygu Analizi Yöntemi Kullanılarak Bitcoin Değer Tahminlemesi. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi,* 9(3), 280-297.

Kraaijeveld, O., & De Smedt, J. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, *65*, 101188.

Kumaş, E. (2021). Türkçe Twitter Verilerinden Duygu Analizi Yapılırken Sınıflandırıcıların Karşılaştırılması. *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*, *2*(2), 1-5.

McAteer, C. (2014). *Twitter Sentiment Analysis to Predict Bitcoin Exchange Rate*. Unpublished master's thesis, University of Dublin.

McCoy, M., (2015). *A Twitter-Based Prediction Tool for Digital Currency.* Unpublished master's thesis, Southern Illinois University, Carbondale

Ozturk, S.S., & Çiftçi, K. (2014). A Sentiment Analysis of Twitter Content as a Predictor of Exchange Rate Movements. *Review of Economic Analysis, 6*, 132-140

Özcan, F. (2016). Exchange Rate Prediction from Twitter's Trending Topics.

Pant, D. R., Neupane, P., Poudel, A., Pokhrel, A. K., & Lama, B. K. (2018, October). Recurrent neural network based bitcoin price prediction by twitter sentiment analysis. In *2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)* (pp. 128-132). IEEE.

Redd, V. (2018). *Sentiment Analysis using SVM*. Retrieved, March 1, 2022, from https://medium.com/@vasista/sentiment-analysis-using-svm-338d418e3ff1

Saini, A. (2021). *Support Vector Machine (SVM): A Complete guide for beginners.* Retrieved, March 1, 2022, from https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/

Sarlan, A., Nadam, C., & Basri, S. (2014, November). Twitter sentiment analysis. In *Proceedings of the 6th International conference on Information Technology and Multimedia* (pp. 212-216). IEEE.

Sayad, S. (n.d.). *Support Vector Machine - Classification (SVM).* Retrieved August 9, 2022, from https://www.saedsayad.com/support_vector_machine.htm

Shen, D., Urquhart, A. & Wang, P. (2019). Does twitter predict Bitcoin? *Economic Letters*. 118-122.

Smith, M. (2018). *Support Vector Machine*. Retrieved, March 1, 2022, from https://rpubs.com/masmit11/533879

Twin, A. (2022, July 5). *6 Factors That Influence Exchange Rates*. Investopedia. https://www.investopedia.com/trading/factors-influence-exchange-rates/

What is R? (n.d.) Retrieved March 10, 2022 from https://www.r-project.org/about.html

Yasir, M., Afzal, S., Latif, K., Chaudhary, G. M., Malik, N. Y., Shahzad, F., & Song, O. Y. (2020). An efficient deep learning based model to predict interest rate using twitter sentiment. *Sustainability*, *12*(4), 1660.

Yasir, M., Durrani, M. Y., Afzal, S., Maqsood, M., Aadil, F., Mehmood, I., & Rho, S. (2019). An intelligent event-sentiment-based daily foreign exchange rate forecasting system. *Applied Sciences*, *9*(15), 2980.

APPENDIX

Appendix 1: Turkish Stop words

- acaba
- ama
- aslında
- az
- bazı
- belki
- biri
- birkaç
- birşey
- biz
- bu
- çok
- çünkü
- da
- daha
- de
- defa
- diye
- eğer
- en
- gibi
- hem
- hep
- hepsi
- her
- hiç
- için
- ile
- ise
- kez
- ki
- kim
- mı
- mu
- mü
- nasıl
- ne
- neden

- nerde
- nerede
- nereye
- niçin
- niye
- o
- sanki
- şey
- siz
- şu
- tüm
- ve
- veya
- ya
- yani

Appendix 2: R Codes for Currency Data Description

```
> library(caret)
> library(e1071)
> library(ggplot2)
> library(klaR)
> library(ISLR)
> library(InformationValue)
> library(forecast)
> library(fpp2)
> library(TTR)
> library(dplyr)
> library(naivebayes)
> library(knitr)
> install.packages("fpp2")
> install.packages("TTR")
> install.packages("dplyr")
> install.packages("knitr")
> usd <- read_excel("Desktop/USD_TRY.xls")
> euro <- read_excel("Desktop/EUR_TRY.xls")
> gbp <- read_excel("Desktop/GBP_TRY.xls")
> cur <- read_excel("Desktop/currency.xls")
>
> plot(usd$Price, col=2, type="l")
> plot(euro$Price, col=rainbow(300))
> plot(gbp$Price, col=27, type="l")
```

```
>
> plot(cur$USD, type="l", col=1, lwd=1, xlab="Time", ylab="Price",
  main="Currencies in Time")
> lines(cur$GBP, col=2, lwd=1)
> lines(cur$EURO, col=3, lwd=1)
> legend("topleft", c("USD/TRY","GBP/TRY", "EURO/TRY"), fill = 1:3)
>
> summary(cur$EURO)
> summary(cur$GBP)
> summary(cur$USD)
>
> mean(usd$Price)
> mean(gbp$Price)
> mean(euro$Price)
```

Appendix 3: R Codes for Currency Prediction without Sentiment Analysis

```
> monthly_data <- unclass(test)
> months <- 1:608
> months
> monthly_data
> FK <- data.frame(months,test$Price)
> colnames(FK)<-c("x","y")
> svm_usd <- svm(y ~ x,data=FK, type="eps-
  regression",kernel="radial",cost=10000, gamma=10)
> svmodel
> nd <- 1:620
> nd
> usd_pred <- predict(svm_usd, newdata=data.frame(x=nd))
> View(usd_pred)
> usd_pred
> ylim <- c(min(FK$y), max(FK$y))
> xlim <- c(min(nd),max(nd))
> plot(FK$y, col=1, lwd=3, ylim=ylim, xlim=xlim, type="l", xlab="Time",
  ylab="Price")
> par(new=TRUE)
> lines(usd_pred, col=2, ylim=ylim, xlim=xlim, lwd=2)
> legend("topleft", c("Real $", "Prediction"), fill = 1:2, main = "USD/TRY SVM")
> plot(FK$y, col=1, type="l", lwd=3, xlab="Time", ylab="Price", main = "USD/TRY
  SVM")
> lines(usd_pred, col=2, lwd=2)
> legend("topleft", c("Real $", "Prediction"), fill = 1:2)
```

```
> plot(xxx$Price, col=1, type="l", xlab="Time", ylab="Price", main = "USD/TRY")
> lines(usd_pred, col=2)
> legend("topleft", c("Real $", "Prediction"), fill = 1:2)
> mean(usd$Price)
> mean(usd_pred)
>
> gb <- cur$GBP
> daily_gbp <- unclass(gb)
> View(daily_gbp)
> days <- 1:608
> GB <- data.frame(days,cur$GBP)
> colnames(GB)<-c("x","y")
> svm_gbp <- svm(y ~ x,data=GB, type="eps-
  regression",kernel="radial",cost=10000, gamma=10)
> svm_gbp
> e_pre <- 1:620
> gbp_pred <- predict(svm_gbp, newdata=data.frame(x=e_pre))
> summary(euro_pred)
> ylim <- c(min(GB$y), max(GB$y))
> xlim <- c(min(e_pre),max(e_pre))
> plot(GB$y, col=18, lwd=3, type="l", xlab="Time", ylab="Price", main =
  "GBP/TRY SVM")
> lines(gbp_pred, col=19, lwd=2)
> legend("topleft", c("Real £", "Prediction"), fill = 18:19)
> fur=c(mean(gbp_pred), mean(cur$GBP))
> gbp_pred
> summary(gbp_pred)
>
> eur <- cur$EURO
> eur
> daily_data <- unclass(eur)
> View(daily_data)
> days <- 1:608
> EU <- data.frame(days,cur$EURO)
> colnames(EU)<-c("x","y")
> svm_euro <- svm(y ~ x,data=EU, type="eps-
  regression",kernel="radial",cost=10000, gamma=10)
> e_pre <- 1:620
> euro_pred <- predict(svm_euro, newdata=data.frame(x=e_pre))
> View(euro_pred)
> euro_pred
> ylim <- c(min(EU$y), max(EU$y))
> xlim <- c(min(e_pre),max(e_pre))
```

```
> plot(EU$y, col="green", type="l", lwd=3, xlab="Time", ylab="Price", main =
  "EUR/TRY SVM")
> lines(euro_pred, col="red", lwd=2)
> legend("topleft", c("Real €", "Prediction"), fill = 3:2)
>
> library(forecast)
> naive_usd <- naive(cur$USD, h=12)
> rwf_usd <- rwf(cur$USD, h=12, drift = TRUE)
> rwf_usd
>
> naive_eur <- naive(cur$EURO, h=12)
> naive_eur
> rwf_eur <- rwf(cur$EURO, h=12, drift = TRUE)
> rwf_eur
>
> naive_gbp <- naive(cur$GBP, h=12)
> naive_gbp
> rwf_gbp <- rwf(cur$GBP, h=12, drift = TRUE)
> rwf_gbp
>
> realUSD <- c(14.0084, 14.095, 14.1845, 14.3666, 14.4762, 14.6378, 14.8131,
  14.7595, 14.7754, 14.7002, 14.6077, 14.696)
> realUSD
> realEURO <- c(15.5787, 15.5947, 15.498, 15.5915, 15.7776, 16.2114, 16.2692,
  16.0952, 16.1628, 16.0982, 16.1152, 16.3082)
> realEURO
> realGBP <- c(18.7755, 18.8112, 18.759, 18.8242, 18.9682, 19.2912, 19.38,
  19.2334, 19.208, 19.1661, 19.2018, 19.3405)
> realGBP


> plot(rwf_eur, col="dark green", main = "EUR/TRY RWF", lwd=3)
> plot(rwf_gbp, col="red",main = "GBP/TRY RWF", lwd=3)
> plot(rwf_usd, main = "USD/TRY RWF", lwd=3)
> plot(naive_eur,main="EUR/TRY NAIVE", lwd=3)
> plot(naive_gbp, main="GBP/TRY NAIVE", lwd=3)
> plot(naive_usd, main="USD/TRY NAIVE", lwd=3)
>
> accuracy(naive_usd, realUSD)
> accuracy(rwf_usd, realUSD)
> accuracy(rwf_eur, realEURO)
> accuracy(naive_eur, realEURO)
> accuracy(rwf_gbp, realGBP)
> accuracy(naive_gbp, realGBP)
```

Appendix 4: R Codes for Currency Prediction with Sentiment Analysis

```
> install.packages("stringr")
> library(stringr)
> install.packages("stringi")
> library(stringi)
> install.packages("tm")
> library(tm)
> install.packages("tidytext")
> library(tidytext)
> install.packages("dplyr")
> library(dplyr)
> install.packages("ggplot2")
> library(ggplot2)
> install.packages("wordcloud")
> library(wordcloud)
> install.packages("wordcloud2")
> library(wordcloud2)
>
> dolar=read.csv("Desktop/Final Data/Revize/dolarTW.csv")
> head(dolar)
> dolar.df <-as.data.frame(dolar)
> stp=read.csv("Desktop/stopwords", header=FALSE)
>
> dolar.clean <- dolar.df
>
> head(dolar.df)

> dolar.clean$text <- stri_enc_toutf8(dolar.clean$text)
> dolar.clean$text <- ifelse(str_sub(dolar.clean$text,1,2) == "RT",
                             substring(dolar.clean$text,3),
                             dolar.clean$text)
> dolar.clean$text <- str_replace_all(dolar.clean$text, "http[^[:space:]]*", "")
> dolar.clean$text <- str_replace_all(dolar.clean$text, "the", "")
> dolar.clean$text <- str_replace_all(dolar.clean$text, "#\\S+", "")
> dolar.clean$text <- str_replace_all(dolar.clean$text, "@\\S+", "")
> dolar.clean$text <- str_replace_all(dolar.clean$text, "[[:punct:][:blank:]]+", " ")
> dolar.clean$text <- str_to_lower(dolar.clean$text, "tr")
> dolar.clean$text <- removeNumbers(dolar.clean$text)
> dolar.clean$text <- str_replace_all(dolar.clean$text, "[<].*[>]", "")
> dolar.clean$text <- gsub("\uFFFD", "", dolar.clean$text, fixed =  TRUE)
> dolar.clean$text <- gsub("\n", "", dolar.clean$text, fixed =  TRUE)
> dolar.clean$text <- str_replace_all(dolar.clean$text, "[^[:alnum:]]"," " )
```

```
> US_tidy_tweets <- dolar.clean %>% select(text) %>%
> mutate(linenumber = row_number()) %>% unnest_tokens(word, text)
> US_tidy_tweets
> names(stp)[1] <- "word"
> View(stp)
> US_tidy_tweets <- US_tidy_tweets %>% anti_join(stp)
>
> US_tidy_tweets %>%
> count(word, sort = TRUE) %>%
> filter(n > 3150) %>%
> mutate(word = reorder(word, n)) %>%
> ggplot(aes(word, n)) +
> geom_col() +
> xlab(NULL) +
> coord_flip() + theme_linedraw() +
> ggtitle("USD/TRY Tweets most common words")
>
> US_tidy_tweets %>%
> count(word, sort = TRUE)
>
> US_tidy_tweets %>%
> count(word) %>%
> with(wordcloud(word, n, max.words = 250, random.order = FALSE, color=1:100))
>
> View(sentiments)
> count(sentiments)
> write.csv(stopwords, "Desktop/Final Data/stop.csv")
> sentiment <- read.csv("Desktop/Final Data/sentiments.csv")
> duygu <- read.csv("Desktop/Final Data/duygu.csv")
> count(sentiment)
>
> bing_tidytwt <- US_tidy_tweets %>% inner_join(duygu) %>%
  count(word,sentiment,sort = TRUE) %>% ungroup()
> bing_tidytwt %>% group_by(sentiment) %>% top_n(100) %>% ungroup() %>%
  mutate(word = reorder(word, n)) %>% ggplot(aes(word, n, fill=sentiment)) +
  geom_col(show.legend = FALSE) + facet_wrap(~sentiments, scales="free_y") +
  labs(title="usd/try", y = "sentiment freq", x = NULL) + coord_flip() + theme_bw()
> svm_usd <- svm(y ~ x,data=USD, type="eps-
  regression",kernel="radial",cost=10000, gamma=10)
> e_pre <- 1:620
> usd_pred <- predict(svm_usd, usd_sentiment_scores,
  newdata=data.frame(x=e_pre))
> View(usd_pred)
```

```
>
> euro_tw=read.csv("Desktop/Final Data/Revize/euroTweet-TR.csv")
> View(euro_tw)
> head(euro_tw)
> euro.df <-as.data.frame(euro_tw)
>
> euro.clean <- euro.df
> count(euro_tw)
>
> head(euro.df)
>
> euro.clean$text <- stri_enc_toutf8(euro.clean$text)
> euro.clean$text <- ifelse(str_sub(euro.clean$text,1,2) == "RT",
                  ▪ substring(euro.clean$text,3),
                  ▪ euro.clean$text)
> euro.clean$text <- str_replace_all(euro.clean$text, "http[^[:space:]]*", "")
> euro.clean$text <- str_replace_all(euro.clean$text, "the", "")
> euro.clean$text <- str_replace_all(euro.clean$text, "#\\S+", "")
> euro.clean$text <- str_replace_all(euro.clean$text, "@\\S+", "")
> euro.clean$text <- str_replace_all(euro.clean$text, "[[:punct:][:blank:]]+", " ")
> euro.clean$text <- str_to_lower(euro.clean$text, "tr")
> euro.clean$text <- removeNumbers(euro.clean$text)
> euro.clean$text <- str_replace_all(euro.clean$text, "[<].*[>]", "")
> euro.clean$text <- gsub("\uFFFD", "", euro.clean$text, fixed =  TRUE)
> euro.clean$text <- gsub("\n", "", euro.clean$text, fixed =  TRUE)
> euro.clean$text <- str_replace_all(euro.clean$text, "[^[:alnum:]]"," " )
> EUR_tidy_tweets <- euro.clean %>% select(text) %>%
> mutate(linenumber = row_number()) %>% unnest_tokens(word, text)
> EUR_tidy_tweets
> names(stp)[1] <- "word"
> EUR_tidy_tweets <- EUR_tidy_tweets %>% anti_join(stp)
>
> EUR_tidy_tweets %>%
> count(word, sort = TRUE) %>%
> filter(n > 1500) %>%
> mutate(word = reorder(word, n)) %>%
> ggplot(aes(word, n)) +
> geom_col() +
> xlab(NULL) +
> coord_flip() + theme_minimal() +
> ggtitle("EUR/TRY Tweets most common words")
>
> EUR_tidy_tweets %>%
> count(word, sort = TRUE)
```

```
>
>   EUR_tidy_tweets %>%
>   count(word) %>%
>   with(wordcloud(word, n, max.words = 250, color=1:250))
>
>   # sentiment #
>   duygu <- read.csv("Desktop/duygu_2.csv")
>   count(duygu)
>
>   eu_bing_tidytwt <- EUR_tidy_tweets %>% inner_join(duygu) %>%
    count(word,sentiment,sort = TRUE) %>% ungroup()
>   eu_bing_tidytwt %>% group_by(sentiment) %>% top_n(100) %>% ungroup()
    %>% mutate(word = reorder(word, n)) %>% ggplot(aes(word, n, fill=sentiment)) +
    geom_col(show.legend = FALSE) + facet_wrap(~sentiment, scales="free_y") +
    labs(title="EUR/TRY", y = "sentiment freq", x = NULL) + coord_flip() +
    theme_bw()
>
>   tidy_tweets_20 <- euro.clean %>% select(text, created_at) %>%
>   mutate(linenumber = row_number()) %>% unnest_tokens(word, text)
>
>   date_df <- semi_join(tidy_tweets_20, duygu)
>   set.seed(200)
>
>   date_20 <- semi_join(duygu, tidy_tweets_20)
>   duyguplot <- date_df
>
>   duyguplot %<>%
>   mutate(
>   created = created_at %>%
>   str_remove_all(pattern = '\\+0000') %>%
>   parse_date_time(orders = '%y-%m-%d %H%M%S')
>   )
>
>   duyguplot %<>%
>   mutate(Created_At_Round = created%>% round(units= 'days') %>%
    as.POSIXct())
>
>   duyguplot %>% pull(created) %>% min()
>   duyguplot %>% pull(created) %>% max()
>
>   plt <- duyguplot %>%
>   dplyr::count(Created_At_Round) %>%
>   ggplot(mapping = aes(x=Created_At_Round, y=n)) +
>   theme_light() +
```

```
>   geom_line() +
>   xlab(label = 'Date') +
>   ylab(label = NULL) +
>   ggtitle(label = 'EUR/TRY Sentiments Words per Date')
>
>   plt %>% ggplotly()
>
>   df_r <- data.frame(x = date_df$created_at, y = date_df$word)
>   ggplot(df_r, aes(x,y)) + geom_line()
>
>   europlot <- euro.clean
>
>   europlot %<>%
>   mutate(
>   created = created_at %>%
>   str_remove_all(pattern = '\\+0000') %>%
>   parse_date_time(orders = '%y-%m-%d %H%M%S')
>   )
>
>   europlot %<>%
>   mutate(Created_At_Round = created%>% round(units= 'days') %>%
    as.POSIXct())
>
>   europlot %>% pull(created) %>% min()
>   europlot %>% pull(created) %>% max()
>
>   plt <- europlot %>%
>   dplyr::count(Created_At_Round) %>%
>   ggplot(mapping = aes(x=Created_At_Round, y=n)) +
>   theme_light() +
>   geom_line() +
>   xlab(label = 'Date') +
>   ylab(label = NULL) +
>   ggtitle(label = 'EUR/TRY Tweets per Date')
>
>   plt %>% ggplotly()
>
>   pos_sents <- read.csv("Desktop/pos_sents.csv")
>   date_df_pos <- semi_join(tidy_tweets_20, pos_sents)
>
>   date_df_pos %<>%
>   mutate(
>   created = created_at %>%
>   str_remove_all(pattern = '\\+0000') %>%
```

```
>   parse_date_time(orders = '%y-%m-%d %H%M%S')
>   )
>
>   date_df_pos %<>%
>   mutate(Created_At_Round = created%>% round(units= 'days') %>%
    as.POSIXct())
>
>   date_df_pos %>% pull(created) %>% min()
>   date_df_pos %>% pull(created) %>% max(6000)
>
>   plt <- date_df_pos %>%
>   dplyr::count(Created_At_Round) %>%
>   ggplot(mapping = aes(x=Created_At_Round, y=n,)) +
>   theme_light() +
>   geom_line() +
>   xlab(label = 'Date') +
>   ylab(label = NULL) +
>   ggtitle(label = 'EUR/TRY Positive Words per Date')
>
>   plt %>% ggplotly()
>
>   neg_sents <- read.csv("Desktop/neg_sents.csv")
>   date_df_neg <- semi_join(tidy_tweets_20, neg_sents)
>
>   date_df_neg %<>%
>   mutate(
>   created = created_at %>%
>   str_remove_all(pattern = '\\+0000') %>%
>   parse_date_time(orders = '%y-%m-%d %H%M%S')
>   )
>
>   date_df_neg %<>%
>   mutate(Created_At_Round = created%>% round(units= 'days') %>%
    as.POSIXct())
>
>   date_df_neg %>% pull(created) %>% min()
>   date_df_neg %>% pull(created) %>% max()
>
>   plt <- date_df_neg %>%
>   dplyr::count(Created_At_Round) %>%
>   ggplot(mapping = aes(x=Created_At_Round, y=n)) +
>   theme_light() +
>   geom_line() +
>   xlab(label = 'Date') +
```

```
>   ylab(label = NULL) +
>   ggtitle(label = 'EUR/TRY Negative Words per Date')
>
>   plt %>% ggplotly()
>
>   svm_eu <- svm(y ~ x,data=EU, type="eps-regression",kernel="radial",cost=10000,
    gamma=10)
>   e_pre <- 1:620
>   eur_pred <- predict(svm_eur, eur_sentiment_scores, newdata=data.frame(x=e_pre))
>   View(eur_pred)


>   strln=read.csv("Desktop//Final Data/Revize/strln.csv")
>   count(strln)
>   head(strln)
>   strln.df <-as.data.frame(strln)
>   stp=read.csv("Desktop/stopwords", header=FALSE)
>
>   strln.clean <- strln.df
>
>   head(strln.df)
>
>   strln.clean$text <- stri_enc_toutf8(strln.clean$text)
>   strln.clean$text <- ifelse(str_sub(strln.clean$text,1,2) == "RT",
                ▪ substring(strln.clean$text,3),
                ▪ strln.clean$text)
>   strln.clean$text <- str_replace_all(strln.clean$text, "http[^[:space:]]*", "")
>   strln.clean$text <- str_replace_all(strln.clean$text, "the", "")
>   strln.clean$text <- str_replace_all(strln.clean$text, "#\\S+", "")
>   strln.clean$text <- str_replace_all(strln.clean$text, "@\\S+", "")
>   strln.clean$text <- str_replace_all(strln.clean$text, "[[:punct:][:blank:]]+", " ")
>   strln.clean$text <- str_to_lower(strln.clean$text, "tr")
>   strln.clean$text <- removeNumbers(strln.clean$text)
>   strln.clean$text <- str_replace_all(strln.clean$text, "[<].*[>]", "")
>   strln.clean$text <- gsub("\uFFFD", "", strln.clean$text, fixed =  TRUE)
>   strln.clean$text <- gsub("\n", "", strln.clean$text, fixed =  TRUE)
>   strln.clean$text <- str_replace_all(strln.clean$text, "[^[:alnum:]]"," " )
>   ST_tidy_tweets <- strln.clean %>% select(text) %>%
>   mutate(linenumber = row_number()) %>% unnest_tokens(word, text)
>   ST_tidy_tweets
>   names(stp)[1] <- "word"
>   View(stp)
>   ST_tidy_tweets <- ST_tidy_tweets %>% anti_join(stp)
```

```
>  ST_tidy_tweets %>%
>  count(word, sort = TRUE) %>%
>  filter(n > 200) %>%
>  mutate(word = reorder(word, n)) %>%
>  ggplot(aes(word, n)) +
>  geom_col() +
>  xlab(NULL) +
>  coord_flip() + theme_minimal() +
>  ggtitle("GBP/TRY Tweets most common words")
>
>  ST_tidy_tweets %>%
>  count(word, sort = TRUE)
>
>  ST_tidy_tweets %>%
>  count(word) %>%
>  with(wordcloud(word, n, max.words = 250, random.order = FALSE, colors =
   1:100))
>
>  count(ST_tidy_tweets)
>
>  duygu <- read.csv("Desktop/duygu_2.csv")
>  count(duygu)
>
>  st_bing_tidytwt <- ST_tidy_tweets %>% inner_join(duygu) %>%
   count(word,sentiment,sort = TRUE) %>% ungroup()
>  st_bing_tidytwt %>% group_by(sentiment) %>% top_n(100) %>% ungroup()
   %>% mutate(word = reorder(word, n)) %>% ggplot(aes(word, n, fill=sentiment)) +
   geom_col(show.legend = FALSE) + facet_wrap(~sentiment, scales="free_y") +
   labs(title="GBP/TRY", y = "sentiment freq", x = NULL) + coord_flip() +
   theme_bw()
>
>  tidy_tweets_10 <- strln.clean %>% select(text, created_at) %>%
>  mutate(linenumber = row_number()) %>% unnest_tokens(word, text)
>
>  date_df <- semi_join(tidy_tweets_10, duygu)
>  set.seed(200)
>
>  date_2 <- semi_join(duygu, tidy_tweets_10)
>
>  duyguplot <- date_df
>
>  duyguplot %<>%
>  mutate(
>  created = created_at %>%
```

```
>   str_remove_all(pattern = '\\+0000') %>%
>   parse_date_time(orders = '%y-%m-%d %H%M%S')
>   )


>   duyguplot %<>%
>   mutate(Created_At_Round = created%>% round(units= 'days') %>%
    as.POSIXct())
>
>   duyguplot %>% pull(created) %>% min()
>   duyguplot %>% pull(created) %>% max()
>
>   plt <- duyguplot %>%
>   dplyr::count(Created_At_Round) %>%
>   ggplot(mapping = aes(x=Created_At_Round, y=n)) +
>   theme_light() +
>   geom_line() +
>   xlab(label = 'Date') +
>   ylab(label = NULL) +
>   ggtitle(label = 'GBP/TRY Sentiments Words per Date')
>
>   plt %>% ggplotly()
>
>   df_r <- data.frame(x = date_df$created_at, y = date_df$word)
>   ggplot(df_r, aes(x,y)) + geom_line()
>   strlnplot <- strln.clean
>   strlnplot %<>%
>   mutate(
>   created = created_at %>%
>   str_remove_all(pattern = '\\+0000') %>%
>   parse_date_time(orders = '%y-%m-%d %H%M%S')
>   )
>   strlnplot %<>%
>   mutate(Created_At_Round = created%>% round(units= 'days') %>%
    as.POSIXct())
>   strlnplot %>% pull(created) %>% min()
>   strlnplot %>% pull(created) %>% max()
>
>   plt <- strlnplot %>%
>   dplyr::count(Created_At_Round) %>%
>   ggplot(mapping = aes(x=Created_At_Round, y=n)) +
>   theme_light() +
>   geom_line() +
>   xlab(label = 'Date') +
```

```
>   ylab(label = NULL) +
>   ggtitle(label = 'GBP/TRY Tweets per Date')
>
>   plt %>% ggplotly()
>
>   pos_sents <- read.csv("Desktop/pos_sents.csv")
>   date_df_pos <- semi_join(tidy_tweets_10, pos_sents)
>
>   date_df_pos %<>%
>   mutate(
>   created = created_at %>%
>   str_remove_all(pattern = '\\+0000') %>%
>   parse_date_time(orders = '%y-%m-%d %H%M%S')
>   )
>
>   date_df_pos %<>%
>   mutate(Created_At_Round = created%>% round(units= 'days') %>%
    as.POSIXct())
>
>   date_df_pos %>% pull(created) %>% min()
>   date_df_pos %>% pull(created) %>% max(6000)
>
>   plt <- date_df_pos %>%
>   dplyr::count(Created_At_Round) %>%
>   ggplot(mapping = aes(x=Created_At_Round, y=n,)) +
>   theme_light() +
>   geom_line() +
>   xlab(label = 'Date') +
>   ylab(label = NULL) +
>   ggtitle(label = 'GBP Positive Words per Date')
>
>   plt %>% ggplotly()
>
>   neg_sents <- read.csv("Desktop/neg_sents.csv")
>   date_df_neg <- semi_join(tidy_tweets_10, neg_sents)
>
>   date_df_neg %<>%
>   mutate(
>   created = created_at %>%
>   str_remove_all(pattern = '\\+0000') %>%
>   parse_date_time(orders = '%y-%m-%d %H%M%S')
>   )
>
>   date_df_neg %<>%
```

```
>   mutate(Created_At_Round = created%>% round(units= 'days') %>%
    as.POSIXct())
>
>   date_df_neg %>% pull(created) %>% min()
>   date_df_neg %>% pull(created) %>% max()
>
>   plt <- date_df_neg %>%
>   dplyr::count(Created_At_Round) %>%
>   ggplot(mapping = aes(x=Created_At_Round, y=n)) +
>   theme_light() +
>   geom_line() +
>   xlab(label = 'Date') +
>   ylab(label = NULL) +
>   ggtitle(label = 'GBP/TRY Negative Words per Date')
>
>   plt %>% ggplotly()
>
>   svm_gbp <- svm(y ~ x,data=ST, type="eps-
    regression",kernel="radial",cost=10000, gamma=10)
>   e_pre <- 1:620
>   gbp_pred <- predict(svm_gbp, gbp_sentiment_scores,
    newdata=data.frame(x=e_pre))
>   View(gbp_pred)
>
```