BMIJ
ISSN: 2148-2586

bmij (2022) 10 (4):1545-1564
doi: https://doi.org/10.15295/bmij.v10i4.2162
Research Article

Data mining application with machine learning algorithms to manage interest rate risk

Faiz oranı riskini yönetmek için makine öğrenimi algoritmaları ile veri madenciliği uygulaması

Enes Koçoğlu[1]

Filiz Ersöz[2]

[1] Branch Manager, Türkiye Emlak Katılım Bankası A.Ş, Trabzon, Türkiye, eneskocoglu@hotmail.com

ORCID: 0000-0002-7137-5323

[2] Prof. Dr., Department of Industrial Engineering, Karabük University, Karabük, Türkiye, fersoz@karabuk.edu.tr

ORCID: 0000-0002-4964-8487

Corresponding Author:

Enes Koçoğlu,

Türkiye Emlak Katılım Bankası A.Ş, Trabzon, Türkiye, eneskocoglu@hotmail.com

Submitted: 3/11/2022

Revised: 19/12/2022

Accepted: 23/12/2022

Online Published: 25/12/2022

Citation: Koçoğlu, E., & Ersöz, F., Data mining application with machine learning algorithms to manage interest rate risk, bmij (2022) 10 (4): 1545-1564, doi: https://doi.org/10.15295/bmij.v10i4.2162

Abstract

In trade, the risks taken may increase the expected income; however, they may also cause large amounts of losses as well. Banks transfer the capital and the deposits they collect from their clients to the individuals or institutions in need of profit, taking certain risks into account. One of the important risks taken in this process of capital transfer is the market's change in interest or profit share rates. If the bank transfers the deposit collected with a certain commitment to the market at a lower rate, it will make a loss. Models for predicting future interest or profit share rates gain importance for preventing this situation. The aim of this study is to determine which variables will be taken into account for the loan interest rate that banks will offer to their customers during the lending process, and to create a machine learning model that can predict the loan interest rate that the bank will offer to its customers by using these variables. Multiple Linear Regression analysis was performed to demonstrate the relationship between the variables selected based on the literature review, expert opinions, and the interest rate. In order to facilitate decision-makers in practice, Random Forests, Decision Trees, K-Nearest Neighbours (KNN), Artificial Neural Networks (ANN), and Support Vector Machine (SVM) algorithms from machine learning algorithms were compared by using the prediction model. Accuracy Rate, Cohen's Kappa, Precision, Sensitivity, and F-Measure measurements were used to compare the algorithms used in the study. According to the analysis results, it was observed that the Random Forest algorithm was more successful on the first model consisting of weekly data. The Decision Tree algorithm succeeded more on the second model consisting of monthly data prediction performance. In the model consisting of weekly data, USD Selling Price, Stock Index (BIST100), and Central Bank Gold Reserve from the Multiple Linear Regression variables were found significant in affecting the interest rate.

Keywords: Data Mining, Decision Tree. Machine Learning, KNN, ANN, Random Forest, SVM

Jel Codes: C45, C51, C53

Öz

Ticarette alınan riskler beklenen getiriyi artırabilir; ancak büyük miktarlarda kayıplara da neden olabilirler. Bankacılıkta alınan önemli risklerden biri de piyasadaki faiz veya kâr payı oranlarının değişmesidir. Banka belli bir taahhütle topladığı mevduatı piyasaya daha düşük bir oranda kredi olarak kullandırısa zarar etmiş olur. Bu çalışmada, bankaların kredi verme sürecinde müşterilerine teklif edeceği kredi faiz oranı için hangi değişkenlerin dikkate alınacağını belirlemek ve bu değişkenleri kullanarak bankanın müşterisine teklif edeceği kredi faiz oranını tahmin edebilecek birmakine öğrenimi modeli oluşturulması amaçlanmıştır. Karar vericilere uygulamada kolaylık sağlamak amacıyla makine öğrenmesi algoritmalarından Random Forest, Decision Tree, K-En yakın Komşular (KNN), Yapay Sinir Ağları (YSA) ve Destek Vektör Makinesi (SVM) algoritmaları tahmin modeli kullanılarak karşılaştırılmıştır. Analiz sonuçlarına göre, tahmin performansı açısından Random Forest algoritmasının haftalık verilerden oluşan birinci modelde, Karar Ağacı algoritmasının aylık verilerden oluşan ikinci modelde daha başarılı olduğu görülmüştür. Haftalık verilerden oluşan modelde Çoklu Doğrusal Regresyon'dan ABD Doları Satış Fiyatı, Hisse Senedi Endeksi (BİST100) ve Merkez Bankası Altın Rezervi değişkenlerinin faiz oranını etkilemede anlamlı olduğu görüldü. Aylık verilerin Çoklu Doğrusal Regresyon analizi sonuçlarına göre, faiz oranını etkileyen en önemli değişken sırasıyla Enflasyon Oranı ve İşsizlik Oranı olmuştur. Tüketici Fiyat Endeksi, Sanayi Üretim Endeksi ve İstihdam Oranı değişkenlerinin faiz oranını etkileme açısından anlamsız olduğu görülmüştür.

Anahtar Kelimeler: Veri Madenciliği, Karar Ağaçları, Makine Öğrenimi, KNN, YSA, SVM

JEL Kodları: C45, C51, C53

Introduction

The success of decisions depends on external environmental factors (Tekin, 2016). Knowing the conditions and constraints of the environment in which the decision is made is necessary. In case the environment in which the decision is made cannot be defined correctly, the decisions taken cannot be classified systematically and made into data to be used when needed. Classification of the environments in which the decision is made reduces the cost of accessing information when a decision output is needed in a similar environment in the future by focusing on the decision outputs regarding the appropriate environment. Experiences gained, and scientific studies constitute the basis to be contributed by future generations.

When making decisions in uncertain environments, there is a possibility that the decision will be affected in the future. Therefore, the decision-maker should make the right choice by considering different possibilities. The "Behavioral Approach" developed by Kahneman and Tversky investigated the decisions taken in uncertain environments and observed that the fear of losing is more dominant than the desire to win in individuals. The results of this research showed that; numbers are not always important in decision making. As a result of the experiments, they revealed that the pain experienced by individuals in the event of losing was twice the satisfaction they experienced in the event of winning (Şentürk & Fındık, 2014). In determining banks' interest rates, decision-makers may adopt a behavioural approach due to the fear of losing when many variables must be considered, probabilities must be calculated, and there is intensive uncertainty.

Developing countries aimed to target interest rates rather than monetary aggregates due to the instability in money demand after the 1970s. Instead of increasing the monetary resource, countries attempted to create the required resource by managing the interest rate correctly. Some of the countries adopted inflation-oriented policies for managing the interest rate. The common objective of the two goals was to ensure price stability of monetary policy, which was achieved (Mishkin, 2000). As a result of historical economic processes, central banks have had to make the right decisions regarding the interest rate to ensure market stability in the country. Except for the central bank, all banks that provide services must manage the interest risk.

Data mining applications allow the decision maker to assess the risk of a process better so that they can determine optimal operational decisions with the overall risk condition and react promptly to economic and environmental changes(Chen, Wang, Yang, Ng & Cheng, 2022; Ersöz, 2019; Koçoğlu & Ersöz, 2021).

In times of financial crisis, uncertainties affect the accuracy of decisions made by professional analysts or decision-makers of banks regarding the prediction of interest rates. The continuity of the relationships between financial instruments in the long term should be considered along with their directions if they are to continue. The high level of uncertainty may make it difficult for decision-makers to take similar decisions. Measurement of uncertainty may be important in terms of the early signals from the market that decision-makers should perceive (Kunze, Wegener, Bizer & Spiwoks, 2017). When introducing a loan product to the market, incorrect prediction of the future interest rate risks causing major losses for banks.

The problem of predicting the future direction of the interest rate is one of the most important problems in the field of finance. The correlation of interest rates with many other economic variables creates non-linear cycles. As a result, it is difficult to predict the price, exchange rate, time, direction, etc., in financial markets, which have a dynamic nature. The reason for this difficulty is the presence of mixed, complex and non-linear time series. The reason why linear time series cannot be formed is that financial markets are affected by politics, investors, and the overall economy(Bezerra & Albuquerque, 2017; Diaz, Theodoulidis & Dupouy, 2016; Göçken, Özçalıcı & Boru, 2016; Henrique, Sobreiro & Kimura, 2019; Tay & Cao, 2001; Zhang, Lin & Shang, 2017; Zhong & Enke, 2017).

Thanks to their market power, banks have a 1.3% lower cost advantage when they receive deposits with interest or dividend payments at the end of a certain maturity. Increasing the market power is observed to affect the cost of deposits collected. In case the collected deposits are distributed as loans, the banks' profit margin decreases by an average of 40 basis points as a result of the 1% decrease in the interest rates. Banks' profits constantly fall in markets with permanently low-interest rates (Whited, Wu & Xiao, 2021). Therefore, the ability of the banks to predict the interest rates for granting loans is important in managing profitability.

Xiao-Lin Li et al. (2021) studied how interest rates affected the bill market, bond market, and banks, considering the variability caused by the non-linear movement of interest rates within the monetary policy in China. According to the results, changes in monetary policy were observed to negatively affect the market interest rates applied in other banks and the positive interbank interest rates. These findings proved that lowering interest rates in China would not have the desired effects on the real economy (Li, Si & Ge, 2021). This finding indicates that it is necessary to focus on effectively managing interest rates for the market rather than reducing or increasing them. In addition, possible predictions at certain maturity ranges are required for effective interest management.

According to the studies in the literature, ANN and SVM techniques have been considered common and useful tools for modelling the complex relationships between interest rates and relevant variables (Diaz et al., 2016). Henrique et al. (2019) reviewed 57 studies on machine learning techniques used to predict financial markets. According to the results, they observed that SVM, ANN, and Random Forest techniques were the most frequently used algorithms for prediction (Henrique et al., 2019).

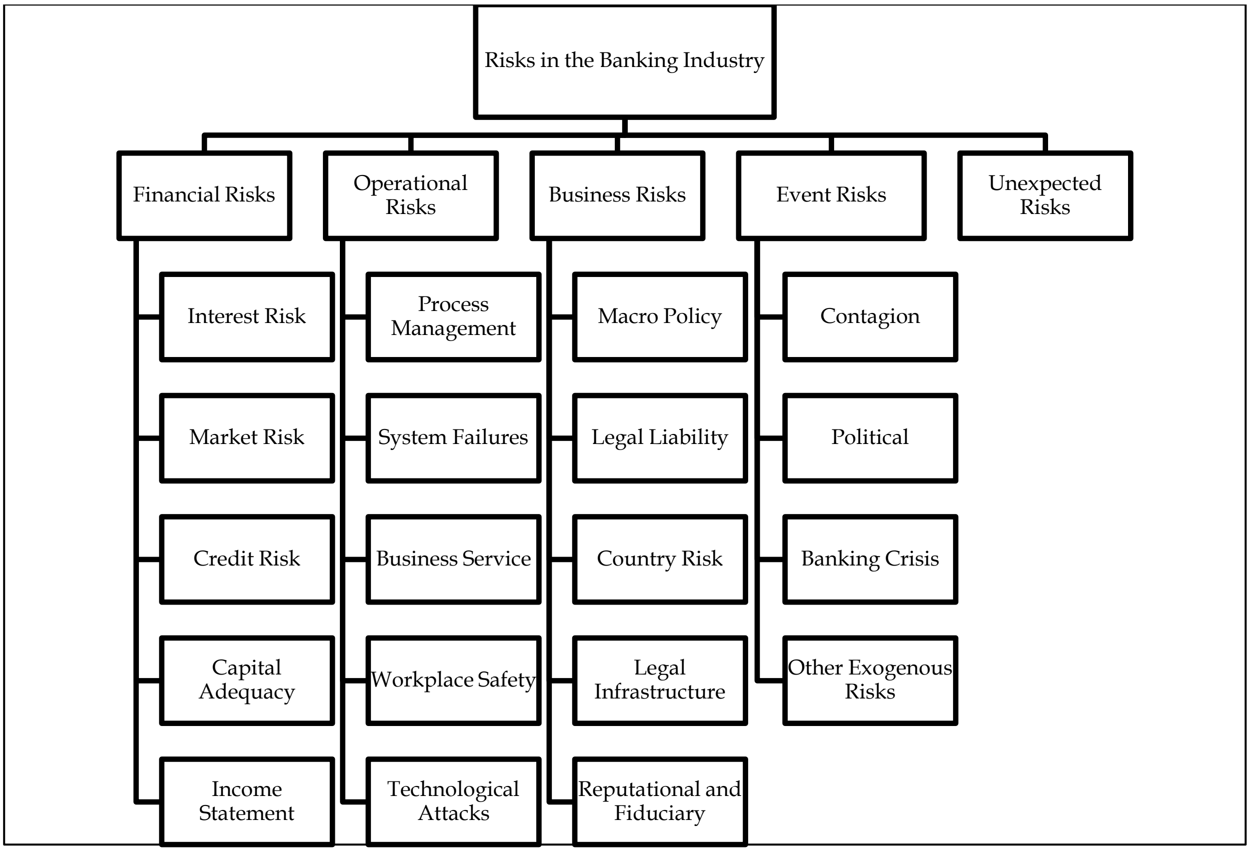

Figure 1: Banking Risk Exposure

Source: (Greuning & Iqbal, 2008).

The banking sector risks seen in Figure 1 are significant enough to endanger the bank's operations when realized. Risks must be managed to prevent these risks that may occur. Generally, the first thing that comes to mind is the credit dilemma when considering risk in the banking sector. Risk management is more important than credit swamps in the banking sector. For example, incorrect calculation of interest rates at the stage of granting a loan may cause operating losses that will damage the long-term balance sheets of banks. Therefore, interest rate risk is important in banking.

In this study, the variables affecting the interest rate risk, one of the financial risks, will be examined. Although the risks in the banking sector are high, the importance of interest risk is very high among the risks. It may be a problem for the bank to continue its activities without correctly managing the interest rate. For this reason, banks that can predict the interest rate risk correctly can earn more profit from their activities. On the other hand, banks that cannot manage interest rates correctly may suffer operating losses.

This study aimed to present a model that can produce predictions for the loan interest rate or profit share rate to be offered to the client regarding the granting of loans, which is one of the main duties of banks, by determining which variables to consider and which machine learning methods to use. Multiple Linear Regression analysis was performed to demonstrate the relationship between the variables selected based on the literature review, expert opinions, and the interest rate. In addition, random Forest, SVM, Decision Tree, KNN and ANN algorithms from machine learning algorithms were compared over the prediction model to increase its applicability in practice and facilitate decision makers work. Accuracy Rate, Cohen's Kappa, Precision, Sensitivity, and F-Measure measurements were used for performances used in the study.

Literature

The studies in the literature about the variables affecting interest rates and the algorithms used for prediction in the world are presented in the following paragraphs. The variables of this study were selected after the literature review.

Oh & Han (2000) made predictions for interest rates using the neural network models by determining the points of change formed by the government's interventions, obtaining significant intervals divided from the points of change, and predicting the interest rates by taking these intervals into account. A backpropagation neural network algorithm was used to detect the points of change and make predictions after detection. According to the results of the analysis, it was observed that determining the point of change was an important tool for improving performance. Furthermore, this model has been proven to significantly increase the predictability of interest rates (Oh & Han, 2000).

Created a prediction model for predicting inflation in India by using the data of the central bank between 2000 and 2012 in the ANN model. The model's accuracy was found to be satisfactory compared to the current predictions (Thakur, Bhattacharya & Mondal, 2016).

Gupta & Kotze (2017) used the Bayes Vector Autoregression model to investigate the effect of oil prices on the prediction of interest rates in South Africa based on the data, including CPI, exchange rates, oil prices, and interest rates between the years 1979 and 1997. According to the results, the model was observed to provide successful predictions when making long-term decisions and tend to outperform traditional approaches (Gupta & Kotzé, 2015). Çolak & Karahan (2017) examined the causality relationship between the exchange rate and interest rate from 2002 to 2016 using the ARDL model. The results showed a negative relationship between exchange rate and interest rates in the long run (Çolak & Karahan, 2017).

Karahan & Gürbüz (2017) analysed the relationship between consumer loans and inflation for 2002 and 2016 using the Johansen Cointegration test. According to the results, inflation was observed to have a stronger effect on individual bank loans. This reveals the necessity of reducing inflation to reduce the pricing of bank loans (Karahan & Gürbüz, 2016).

Assefa et al. (2017) investigated the relationship between interest rates and stock returns between 1999 and 2013 in 40 countries. The results demonstrated the significant effects of interest rates on stock returns in developed economies. According to the results, only the global market affected stock returns in developing countries (Assefa, Esqueda & Varella, 2017).

The relationship between interest rates and inflation was brought to the literature by Irving Fisher in 1930. According to Fisher, the nominal interest rate refers to the sum of the expected real interest rate and the expected inflation rate. According to the hypothesis of Fisher, inflation was the main determinant of interest rates; when the inflation rate increased by one per cent, the interest rate increased at the same rate (Bal, Erdoğan & Palandökenlier, 2019).

Güler & Özçalık (2018) performed a statistical analysis, the Granger causality test, and Vector Auto Regression analysis to reveal the relationships between BIST 100 index, USD index, mean CBRT interest rates, and dollar/TL exchange rate between 2016 and 2018. According to the results, all variables were observed to be related to each other (Güler & Özçalık, 2018). Erkişi (2018) analysed the data consisting of 38 observations between 1980 and 2017 with the VECM prediction algorithm to reveal the long-term effects of interest rates, money supply, and exchange rates on each other in Turkey. According to the prediction results, there was a positive and significant relationship between the exchange and interest rates. In addition, a significant and positive relationship was observed between the exchange rate and the money supply (Erkişi, 2018).

Kim & Shi (2018) used an ordinal profit model with quarterly frequency data from 1987-2013 to predict the two main interest rates in China: the benchmark interest rate and the lending rate. The linear Taylor rule was used as the algorithm for prediction. According to the results, it was observed that the exchange rate change was insignificant, and inflation was important in the central bank's decisions regarding the interest rate (Kim & Shi, 2018).

Güriş & İçen (2019) analysed the weekly data between 2009 and 2018 using the Harvey - Leybourne tests to determine the relationship between exchange rate, risk premium, and deposit interest rates. As a result, it was observed that the exchange rate affected the interest rates with a one-way non-linear relationship, and the CDS premiums affected the exchange rate with a one-way non-linear relationship (Güriş & İçen, 2019).

İşcan & Kaygısız (2019) analysed the data between 2009 and 2017 using the Vector Autoregression (VAR) model to determine the relationship between inflation, exchange rate and interest rate. In addition, the causality relationship between the variables was investigated. According to the results, the exchange rate was identified as the cause of inflation and interest, and inflation was identified as the cause of interest. Furthermore, according to variance decomposition tests, the exchange rate was considered the most important factor in the interest rate and inflation (İscan & Kaygısız, 2019).

Telçeken & Değirmen (2019) evaluated the relationship between loan rates and inflation within the framework of the Fisher Hypothesis. During the evaluation, the data from the 2002-2018 periods were used in the Granger Causality analysis. According to the results, there was a one-way causality relationship between the producer price index to the commercial loan interest rates and the consumer price index to the individual loan rates. In addition, according to the ARDL test performed, there was a long-term relationship between PPI and Commercial loan rates; however, there was no causality relationship between CPI and Personal loan rates (Telçeken & Değirmen, 2019).

Aksu & Emsen (2019) analysed nominal interest rates, logarithms of nominal exchange rates, and CPI rates between 2003 and 2017 using ARDL analysis. The results demonstrated that while inflation was affected by short-term nominal exchange rate changes, it was not affected by interest rates. In addition, interest rates were significantly affected by exchange rate changes, and exchange rate changes were strongly affected by interest rates(Aksu & Emsen, 2019).

Yenice & Yenisu (2019) used the data between 2003 and 2018 in the cointegration analysis with the ARDL bounds testing approach to examine the effect of exchange rates on inflation and interest rates. The Toda-Yamamoto test was performed to determine the causality relationship between the variables. According to the results, the exchange rate was seen as the cause of inflation and interest rate in one direction. In addition, there was a cointegration relationship between the interest and inflation rates. This finding confirmed the presence of the Fisher Hypothesis in Turkey (Yenice & Yenisu, 2019). Tursoy (2019) examined the relationship between interest rates and stock prices in Turkey using cointegration analysis, vector autoregression analysis and ARDL test. According to the results, interest rates and stock prices were observed to be significantly correlated with each other (Tursoy, 2019).

Apergis et al. (2019) used Bayesian Markov-Switching VECM models to investigate the effect of changes in interest rates on gold prices. According to the analysis results, a significant positive relationship was observed between gold prices and real interest rates. Furthermore, the actual output of the study was consistent across all G7 countries (Apergis et al., 2019). Yong & Dingming (2019) used the Bayesian vector autoregression model to investigate the effect of the news on the spending decisions of the US Government on interest rates. According to the results, it was observed that the increase in government expenditures led to a significant increase in interest rates both in the short and long term (Yong & Dingming, 2019).

Maehashi & Shintani (2020) analysed data from 219 variables formed between 1973 and 2018 with machine learning and factor models to predict the macroeconomic future in Japan. It was observed that factor models and machine learning algorithms outperformed traditional time series (AR) models in many cases; particularly, machine learning algorithms made better predictions in the long run. The success of machine learning stemmed from the generalization approach based on regression trees. As a result, factor models and the machine learning approach were suitable for macroeconomic forecasts such as the interest rate (Maehashi & Shintani, 2020).

Yıldırım & Sarı (2020) used ARDL and NARDL models to examine the effect of the inflation rate and exchange rate change on the interest rate of the Turkish economy for the period from January 2004 to April 2020. According to the results of the ARDL test, the integration effect of the inflation rate and the exchange rate was not observed in the determination of the interest rate. On the other hand, there was a relationship between the series in the NARDL model. According to the results, there was a relationship between the inflation rate and the interest rate in the long term. It was observed that the exchange rate was not effective in determining the interest rate in the long term (Yıldırım & Sarı, 2020).

Akın & Dağlıoğlu (2021) examined the consumption phenomenon under consumption, uncertainty, and debt constraint. They analyzed the data between 2004-2019 to determine the relationship between the USD exchange rate, the share of household consumption expenditures in income, and the consumer confidence index. According to the results, there was a negative relationship between household loans, the USD exchange rate, and consumption expenditures. On the other hand, a positive relationship was found between the consumer confidence index, per capita income, and USD exchange rates. According to the causality analysis, consumption expenditures were causally affected by the USD exchange rate, per capita income, household loans, and consumer confidence index both in the short and long term (Akın & Dağlıoğlu, 2021).

Garg & Prabheesh (2021) used daily interest rates and exchange rates data from January 31, 2020, to June 30, 2020, to reveal how the difference in interest rates would cause the exchange rate to change during the COVID-19 period. The ARLD algorithm was used for prediction and accuracy checks to observe the causality between the interest and exchange rates. According to the results, it was revealed that changes in exchange rates could be predicted by considering the differences in interest rates (Garg & Prabheesh, 2021). This finding raised the question of whether interest rates could be predicted using exchange rates. Salisu & Vo (2021) examined the behaviour of exchange rates and stock returns in low and high-interest rate environments. The study used inflation, stock, exchange rate, CPI, and oil prices as variables. The data used in the research were mostly obtained as weekly data from the website of the Federal Reserve (FRED). The Panel Autoregressive Distributed Latency (PARDL) model was used in the analysis. According to the results, the effects of low and high-interest rates were different in the short and long term. It was observed that the negative effect on stocks was higher in the short term in high-interest environments, and the negative effect on stocks was higher in the short term in low-interest environments (Salisu & Vinh, 2021).

Stolyarov & Tesar (2021) investigated how global factors affected the prediction of interest rates in the USA. Labour productivity, changing demographics, stock prices, inflation rate, annual economic growth, and bond prices were the variables used in the analysis. According to the results, the predictions derived from the data in which global factors were considered were more successful than those from the data consisting of local factors. In addition, the results indicated that the factors in the European and Asian markets strongly affected the USA's long-term interest rates (Stolyarov & Tesar, 2019).

Methodology

First, the data and variables used in the research were defined in this part. Secondly, the algorithms used in the research were examined according to the literature.

In this article, detailed information about the methods is given, and information about how the methods are used in practice (For example, how do the authors configure the ANN method? What is the number of hidden layers? What is the learning rate?) is not given. In this article, the purpose of working in this way in the application part is to conclude by using the machine learning package program for those who will apply it faster and in practice using the Knime machine learning package program.

Data and variables of the study

In this study, weekly and monthly data of the variables, which were thought to affect the interest rates in Turkey between the dates of 2014 - 2021, were used for analysis. The study consisted of open-source data published on tcmb.gov.tr and investing.com websites.

Variables; Loan Interest Rate, Gold Selling Price, Government Domestic Debt Securities, Foreign Debt Principal and Interest, USD Selling Price, Euro Selling Price, Central Bank Gold Reserve, Industrial Production Index, Unemployment Rate (%), Consumer price index (%), Stock Index (BIST100), Total Credit Volume, Employment Rate (%), Central Bank Gross Currency Reserve, Inflation Rate (%).

Machine learning algorithms used in the study

Multiple Linear Regression analysis was performed to show the relationship between the variables selected based on the literature review, expert opinions, and the interest rate. In addition, random Forest, SVM, Decision Tree, KNN and ANN algorithms were compared over the prediction model to increase its applicability in practice and facilitate decision makers work. Accuracy Rate, Cohen's Kappa, Precision, Sensitivity, and F-Measure measurements were used for performances used in the study.

It was observed that six different parameters were used in the literature to evaluate the performances of different prediction algorithms. These parameters were as follows (Imran, Badrudduza & Rifat, 2019):

- $Precision = \frac{TP}{TP+FP}$ (1)
- $Specificity = \frac{TN}{TN+FP}$ (2)
- $Sensitivity\ (recall) = \frac{TP}{TP+FN}$ (3)

- $F\ measure = 2 \times \frac{(recall\ x\ precision)}{(recall + precision)}$ (4)

- $Cohen's\ Kappa = \frac{P_0 - P_e}{1 - P_e}$ (5)

$P_0 =$ Relative observed agreement among accuracy.

$P_e =$ Hypothetical probability of chance agreement

- $Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$ (6)

TP : True Positives (Positive tuples correctly labelled by the classifier.)

FP: False Positives (Negative tuples incorrectly labelled as positive by classifier)

TN : True Negatives (Negative tuples correctly labelled by the classifier.)

FN: False Negatives (Positive tuples incorrectly labelled as negative by the classifier.)

The F measure value is the harmonic mean of the Precision and Sensitivity values. The factors used to reveal the advantages or weaknesses of the algorithms over each other are important. In this study, valuation factors frequently seen in the literature were preferred. The effects of factors on prediction abilities constitute the distinguishing features of algorithms.

Support vector machine algorithm

Following the study of Lerner & Vapnik in 1963 and the study of Chervonenkis & Vapnik in 1964, it was used in a text, speech, and time series prediction of Vapnik & Cortes in 1995 (Cortes & Vapnık, 1995). As displayed in Figure 1, SVM finds a plane by dividing it into two areas with the maximum distance between the data over the hyperplane with the dispersed form of the data. The distance between the hyperplane and any data is called the margin. The data closest to the hyperplane on the two sides separated by the hyperplane are called support vectors (Nakagawa, Hochin, Nomiya, Nakanishi & Shoji, 2021). The structure of the SVM is displayed in Figure 2.

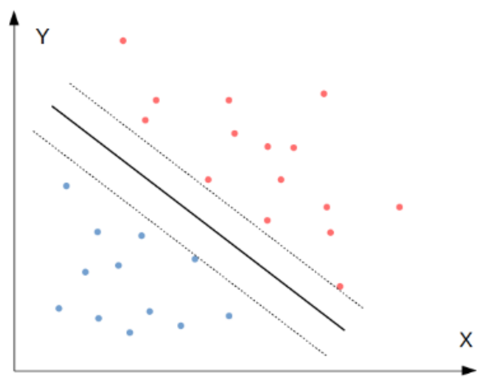

Figure 2: Support Vector Machine (SVM)

Each of the variables considered to affect the dependent variable in a data set does not equally affect the dependent variable. Therefore, the relationship effect of some variables may be more or less than those of other variables. Both high computational cost and performance decrease can be observed in the analysis made with the assumption that all variables are equally effective. For this reason, performance can be increased by feature selection or feature extraction algorithms. The feature extraction approach provides the optimal transformation with a lower dimension, which can represent the master data most effectively. The feature selection approach provides a transition from the original set containing the full features to the most representative subset (Guo, Zhang & Tang, 2021).

Artificial neural network algorithm

ANN is a classifying data mining algorithm that can analyse and make sense of algorithms developed to make sense of the mixed data that are considered meaningless. ANN learns by exemplifying, similar to human beings. It makes inferences according to the old learning when new data is entered as a result of learning. ANNs operate the process of information transfer and learning using connection tools such as neurons in human brains (Elmas, 2021). The structure of the ANN is presented in Figure 3.

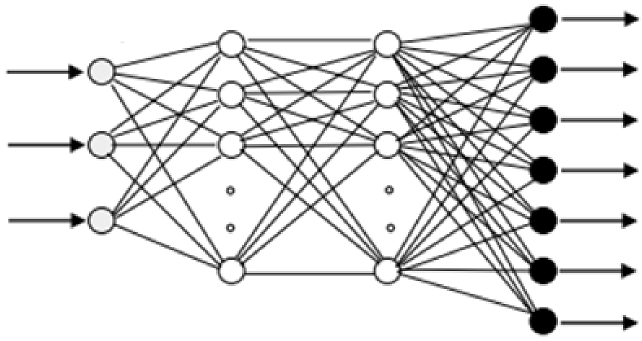

Figure 3: General Structure of ANN (Atalay & Çelik, 2017).

ANN is a data mining algorithm that is widely used in the literature in stock price determination, stock return determination, index return determination, credit scoring, and predictions about the economy, economics, finance and exchange rates (Aksoy, 2021; Altaş & Gülpınar, 2012; Çınaroğlu & Avcı, 2020; Ersöz, 2019b; Fılız, Karaboğa & Akogul, 2017; Karakul, 2020; Kavcıoğlu, 2019; Taş, Gülüm & Tulum, 2021; Yiğiter, Sari & Başakin, 2017).

ANN is a single-layer, easy-to-apply model used in the solution of models with only linear variables (Çavdar & Aydın, 2018). As a result of comparing the ANN and SVM algorithms in solving problems such as credit rating prediction and stock market index prediction, the two algorithms were equally successful (Huang, Chen, Hsu, Chen & Wu, 2004; Kara, Acar Boyacioglu & Baykan, 2011). The results were observed to be very close to the real values in the stock pricing prediction made using the ANN algorithm (Çınaroğlu & Avcı, 2020). Models were created to predict the variables affecting the Borsa İstanbul (abbreviated as BIST-100) index. According to the results, it was demonstrated that the ANN algorithm could be used in finance (Karakul, 2020).

K-Nearest Neighbours algorithm

The K-Nearest Neighbours (KNN) algorithm was first introduced in the early 1950s and started to be widely used in the 1960s. The KNN algorithm, widely used in machine learning, can be used for *Multiple Linear Regression* and classification as one of the supervised learning algorithms. The structure of the KNN is presented in Figure 4.

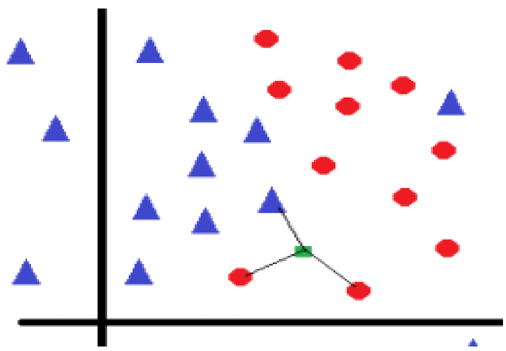

Figure 4: KNN Algorithm image (Çavuşoğlu & Kaçar, 2019)

The unclassified data, whose properties are not defined in the KNN algorithm, are compared with the trained data and classified depending on their distance. The data are classified when the calculated distances are classified. In the KNN algorithm, the distance of the new data to the existing data is calculated one by one for each new data included in the data set and assigned to the class with the most appropriate distance. There is a long classification process in the KNN algorithm, which repeats the search for each new data (Dilki & Başar, 2020; Hu, Huang, Ke & Tsai, 2016; Khan, Ding & Perrizo, 2015). The formula used to measure the distances between the data in the KNN algorithm is as follows (Dolgun, Özdemir & Oğuz, 2009; Kılınç, Borandağ, Yücalar, Tunalı, Şimşek & Özçift, 2016):

$$d_{(i,j)} = \sqrt{\sum_{k=1}^{p} (X_{ik} - X_{jk})^2} \tag{7}$$

In the KNN algorithm, the k value represents the number of other close-range data to be considered. Optimization is required to determine the K value. An increase in the K value increases the fit to the test data. In *Multiple Linear Regression* problems based on prediction, the dependent variable is calculated by taking the arithmetic mean of the independent variables as much as the optimum k value of the independent variables (Altunkaynak, Başakın & Kartal, 2020).

Decision tree algorithm

Decision trees are frequently used in classification problems due to their ease of use. The main reason for its widespread use is that the rules that make up the decision tree can be defined in simple terms. Furthermore, decision trees are more advantageous for decision-makers when evaluating the results than other data mining algorithms in terms of their ability to interpret (Chien & Chen, 2008; Koçak, 2020).

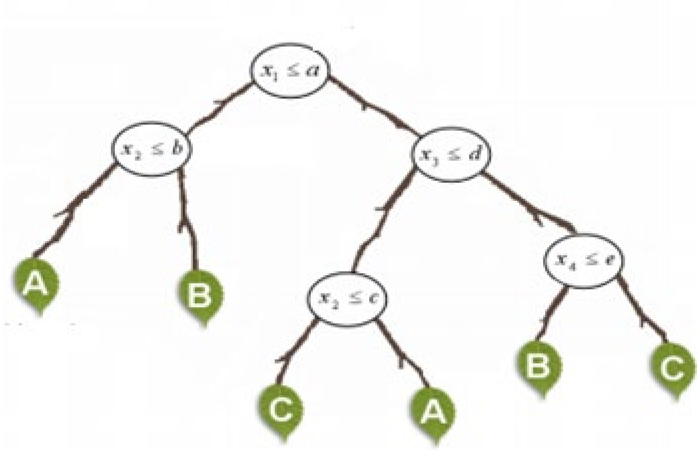

Figure 5: Decision Tree (Kavzoğlu & Çölkesen, 2010)

The Decision Tree structure consists of leaves, branches and nodes, as shown in Figure 5. Nodes define the attributes of the related problem. Flow elements between roots and leaves are defined as branches. The data enters through the root node. Internal nodes (branches) and leaves (end nodes) are the main tools for creating the structure of the decision. The Decision Tree algorithm starts working with the data transferred to the tree via the root node. It ends with asking questions about the problem through branches and leaves and making a decision based on the results obtained (Pal & Mather, 2003).

Random forest algorithm

The Random Forest Algorithm is a decision tree approach created based on the random preference of supervised learning-based variables within machine learning algorithms. It was proposed as an alternative to the "Boosting" algorithm, which was introduced by Breiman (2001), and Amit & Geman (1997) (Breiman, 2001; Cutler, Cutler & Stevens, 2012). The Random Forest algorithm is a regression analysis approach based on the decision tree. As the sub-algorithm of the decision trees, the node is divided into branches by choosing the best of the randomly selected variables in the selected node, unlike the decision trees of the Random Forest algorithm. Decision trees are created by choosing a random variable (Akar & Güngör, 2012; Breiman, 2001; Nuray, Gençdal & Arama, 2021).

The proposed model

Figure 5 below shows the model proposed in this study. Model consists of “Business Understanding”, “Data Understanding”, “Data Preparation”, “Modelling”, “Evaluation” and “Deployment” stages.

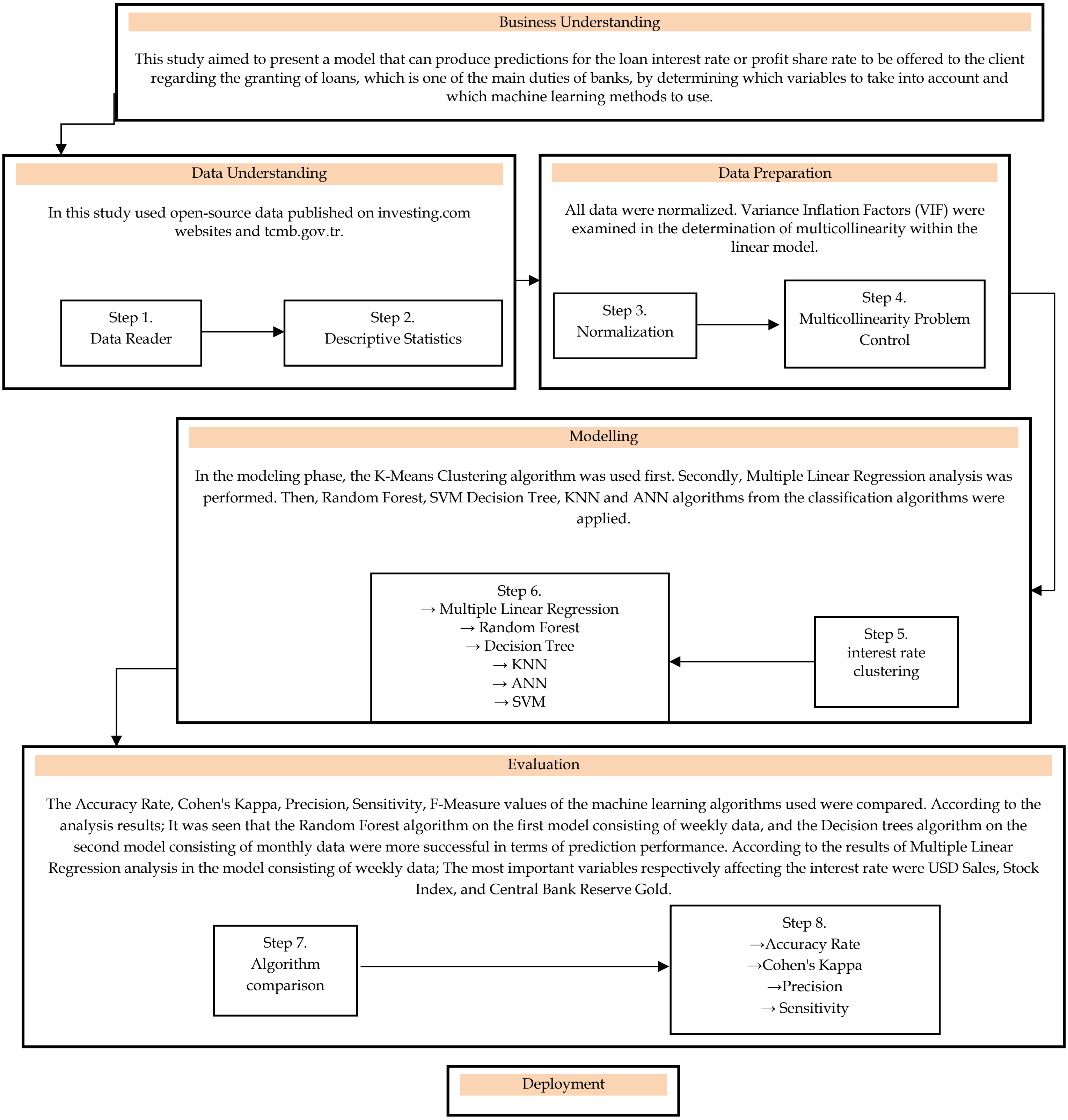

Figure 5: Proposed Model and Methodology

Results

Descriptive statistics

The aim of this study is to determine which variables will be taken into account for the loan interest rate that banks will offer to their customers during the lending process, and to create a machine learning model that can predict the loan interest rate that the bank will offer to its customers by using these variables. All analyses in this article were made with the KNIME Analytics Platform. The results of the descriptive statistics regarding the weekly data used in the study are presented in Table 4.

Table 4: Descriptive Statistics of Weekly Data

	Interest Rate (%)	Government Domestic Debt Securities (TL)	Foreign Debt Principal and Interest (USD)	Total Credit Volume (TL)	USD Selling Price (TL)	Euro Selling Price (TL)	Central Bank Gross Currency (USD)	Central Bank Gold Reserve (USD)	Stock Index	Gold Selling Price (TL)
Minimum	9.17	408 544.40	10.02	1 108 641 251	2.16	2.67	14 051	40 373	548.10	83.35
Maximum	35.85	1 117 300.60	3158.8	3 568 207 079	8.42	9.85	45 229	114 030	1207.02	534.98
Mean	17.04	602 579.80	264.7	2 081 835 514	4.44	5.07	23 562	84 340	830.97	204.97
S.Deviation	5.32	204 667.00	415.2	672 340 423	1.69	1.98	7 644	18 864	139.08	113.21

The descriptive statistics of weekly data are presented in Table 4. It was observed value was 17.04 ± 5.32 for the Interest rate. It was 602579.8 ± 204667.0 for GDDS, 264.06 ± 415.22 for foreign debt, 2.081.835.514 ± 672.340.423 for Total credit volume, 4.44 ± 1.69 for USD selling price; 5.07 ± 1.98 for Euro selling price, 23562 ± 7644 for gold reserve, 84340 ± 18864 for foreign currency reserve, 830.97 ± 139.08 for BIST 100, and 204.97 ± 113.21 for gold selling price.

Table 5: Descriptive Statistics of Monthly Data

	Interest Rate (%)	CPI	IPI	Employment Rate (%)	Unemployment Rate (%)	Inflation Rate (%)
Minimum	9.74	233.54	78.22	40.30	8.40	6.57
Maximum	34.48	498.58	134.95	48.50	15.10	25.24
Mean	17.19	333.18	107.20	45.70	11.40	11.30
S. Deviation	5.23	78.75	12.79	1.70	1.60	4.25

The descriptive statistics of monthly data are presented in Table 5. It was observed that the value for the Interest rate was 17.19 ± 5.23, it was 333.18 ± 78.75 for CPI, 107.20 ± 12.79 for Industrial production index, 45.7 ± 1.7 for the Employment rate, 11.4 ± 1.6 for the Unemployment rate, 11.30 ± 4.25 for the Inflation rate.

In the research, Multiple Linear Regression, Random Forest, Decision Tree, KNN, ANN and SVM and algorithms were used to determine the variables affecting the interest rate. Random Forest, Decision Tree, KNN, ANN and SVM classifier algorithm results are given only in the model comparison Table 16. In the study, Multiple Linear Regression analysis was performed to determine the variables affecting the interest rate. In addition, Variance Inflation Factors (VIF) were examined to determine multicollinearity within the linear model. The diagonal elements of the inverse of the correlation matrix for the independent variables are called the VIFs. VIF is calculated to determine the degree of relationship of an independent variable with other independent variables. If VIF is greater than or equal to 10, there is a multicollinearity problem. When the VIF values were examined during the analysis, there was a multicollinearity problem in the central bank gross currency rate, gold selling price, total credit value, and government domestic debt securities variables; therefore, these variables were excluded from the model with weekly data. VIF values of the USD selling price, foreign debt principal and interest, central bank gold reserve, and stock index (BIST100) variables included in the model are presented in Table 6 below.

Table 6: VIF values of the Weekly Dataset

Variable	VIF Value
Foreign Debt Principal and Interest	1.04
Stock Index (BIST100)	1.90
Central Bank Gold Reserve	2.77
USD Selling Price	3.99

According to the outputs of the model summarized in Table 7, the rate of the explanation of the adjusted model by the weekly data was found as (R^2) 0.399. Therefore, based on the R^2 value obtained, it can be argued that the model's explanation rate by the variables used in the study was not high.

Table 7: Summary of Model 1

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.637	.406	.399	.6346	.406	58.007	4	339	.000	.427*

$p<0.05$

According to the results of the *Multiple Linear Regression* analysis belonging to the model formed with the weekly data displayed in Table 8, *the total foreign debt principal and interest variable were insignificant. On the other hand, USD selling price, stock index, and central bank gold reserve* variables were found to be significant in affecting the interest rate.

Table 8: Results of the Multiple Linear Regression Analysis for Weekly Data

Variable	**Coeff.**	**Std. Err.**	**t-value**	**p**
USD Selling Price	1.243	0.084	14.721	0.00*
Stock Index (BIST100)	-1.030	0.078	-13.281	0.00*
Central Bank Gold Reserve	0.496	0.075	6.647	0.00*
Foreign Debt Principal and Interest	-0.014	0.088	-0.162	0.87

*$p<0.05$

In the second model consisting of monthly data, Multiple Linear Regression analysis was performed to determine the variables affecting the interest rate. When the VIF values were examined during the analysis, it was observed that there was no multicollinearity problem between any of the variables. *The VIF values of the industrial production index, inflation rate, employment rate, unemployment rate, and CPI variables* included in the model are presented in Table 9 below.

Table 9: VIF Values of the Monthly Dataset

Variable	**VIF Value**
Industrial Production Index	2.14
Inflation Rate (%)	2.25
Employment rate (%)	2.34
Unemployment rate (%)	3.54
CPI	4.47

According to the outputs of the model summarized in Table 10, the rate of the explanation of the adjusted model by the monthly data was found as (R^2) 0.301. The Durbin-Watson and significance values indicated that the model was acceptable. The R^2 value obtained indicated that the monthly data had a lower representation ability than the weekly data.

Table 10: Summary of Model 2

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
2	0.586	0.344	0.301	4.3726026	0.344	8.063	5	77	0.000	0.319*

$p<0.05$

According to the results of the Multiple Linear Regression analysis presented in *Table 11*, the *Inflation rate and Employment Rate (%)* variables were found to be effective in terms of affecting the interest rate ($p=0.00<0.05$). The most important variables affecting the interest rate were “Inflation rate” and “Unemployment rate”. On the other hand, the *Consumer price index, Industrial production index, and Employment rate variables* were found to be insignificant in terms of affecting the interest rate.

Table 11: Results of the Multiple Linear Regression Analysis for Monthly Data

Variable	Coeff.	Std. Err.	t-value	p
Inflation Rate (%)	0.385	0.121	3.179	0.00*
Unemployment Rate (%)	-0.523	0.138	-3.796	0.00*
CPI	0.235	0.136	1.721	0.08
Industrial Production Index	0.169	0.125	1.349	0,18
Employment Rate (%)	-0.059	0.141	-0.420	0.67

*p<0.05

K-Means clustering algorithm results

An unsupervised learning approach was used since all of the data used in this study consisted of numerical values. When preparing the data, the interest rates were first classified as the dependent variable using the K-Means module in the Knime software, and the interest rates were labelled. Since the variables are defined with different measurement units, the normalization process was performed in the Knime software in the second stage. After the normalization process, the training and test data were created by dividing the data set as 80% and 20%.

Since the dependent variable is not defined and labelled in the data set used in this study, it was necessary to label the interest situation at a certain time. Therefore, at this stage of the study, weekly and monthly data sets were clustered with the K-Means algorithm to be able to analyse the data. They were clustered into three (High, low and normal interest) and four (low, normal, high, and very high interest) clusters, as presented in Tables 12 and 14 below.

The data in Table 12 were obtained using the KNIME Analytics Platform.

Table 12: Classification Table of Weekly Data

Name of the Class	Interest Rate (%)	Foreign Debt Principal and Interest (USD)	USD Selling Price (TL)	Central Bank Gross Currency (USD)	Stock Index
High Interest	20.822	252 392	5.205	78 550.188	777.806
Low Interest	14.636	210 136	2.963	99 563.374	914.763
Normal Interest	14.753	479 571	7.335	48 684.979	693.759

The clusters were determined by applying hierarchical clustering on IBM SPSS Statistics software, and the weekly data were classified into three clusters: high interest, low interest, and normal interest, as presented above in Table 12. The reason why the classification was made with three clusters was due to the success of clustering. As presented in Table 13, the clustering success was 98%. The high clustering success is important in increasing the data's ability to be analysed and the commentators' ability to predict. Furthermore, the high classification success observed in this study confirms the robustness of the data and its ability to represent reality.

Table 13: Clustering Success Rate of Weekly Data

Cluster Number of Case		1	2	3	Total
%	1	97.9	0.0	2.1	100.0
	2	0.0	100.0	0.0	100.0
	3	0.0	3.5	96.5	100.0
98,0% of original grouped cases were correctly classified.					

The high number of classifications in monthly data indicates market variability. Excessive changes in a small amount of data reduce the predictability of the markets as they increase the differentiation of the data. For these reasons, the clustering success was observed high in four classes in the model in which monthly data were examined.

Table 14: Classification Table of Monthly Data

Name of the Class	Interest Rate (%)	CPI	IPI	Employment Rate (%)	Unemployment Rate (%)	Inflation Rate (%)
Low Interest	14.07	248.83	95.13	45.49	10.04	8.27
Normal Interest	15.25	281.04	103.96	45.91	11.11	8.07
High Interest	17.13	434.46	113.80	44.62	13.25	15.17
Very High Interest	22.77	328.88	114.44	47.42	10.38	12.06

As presented in Table 14, the monthly data were classified into four cluster sets: very high interest, high interest, normal interest, and low interest. The reason why the classification was made with four clusters was due to the success of clustering. As presented in Table 15, the clustering success was 99%.

Table 15: Clustering Success Rate of Monthly Data

Cluster Number of Case		1	2	3	4	Total
%	1	100.0	0.0	0.0	0.0	100.0
	2	0.0	100.0	0.0	0.0	100.0
	3	0.0	0.0	100.0	0.0	100.0
	4	0.0	0.0	7.1	92.9	100.0
a. 98, 8% of original grouped cases were correctly classified.						

Comparison of model achievements of classification algorithms

The Random Forest, SVM, Decision Tree, KNN and ANN algorithms from the machine learning classification algorithms were applied to measure the relationship between the variables that were thought to be related to the bank interest rate. Precision, sensitivity, specificity, f-measure, accuracy statistics, and cohen’s kappa values were used for the prediction results of the classification models.

Table 16: Comparison of Weekly Data and the Method

Method	Class	Precision	Sensitivity	Specificity	F-measure	Accuracy Statistics	Cohen's Kappa
Random Forest	Low Interest	1.000	1.000	1.000	1.000	0.997	0.995
	Normal Interest	1.000	0.987	1.000	0.994		
	High Interest	0.98	1.000	0.997	0.99		
ANN	Low Interest	1.000	0.979	1.000	0.989	0.986	0.970
	Normal Interest	0.933	1.000	0.982	0.966		
	High Interest	1.000	1.000	1.000	1.000		
KNN	Low Interest	0.997	0.982	0.961	0.979	0.971	0.945
	Normal Interest	0.948	0.924	0.985	0.936		
	High Interest	0.98	1.000	0.997	0.99		
Decision Tree	Low Interest	0.995	1.000	0.992	0.998	0.858	0.724
	Normal Interest	0.619	0.987	0.819	0.761		
	High Interest	0.961	0.724	1.000	0.667		
SVM	Low Interest	0.847	1.000	0.526	0.917	0.841	0.561
	Normal Interest	1.000	0.267	1.000	0.421		
	High Interest	0.667	1.000	0.969	0.800		

Table 16 above shows that the accuracy rate and cohen's kappa values were ranked similarly. The Random Forest method's high value according to the accuracy rate and cohen's kappa indicated its superiority over other methods. Since cohen’s kappa value is an important reference showing the reliability of the result, it indicated that the reliability of the prediction success of the Random Forest algorithm was superior to other algorithms.

All algorithms used were found to have a high success rate regarding classification accuracy, as presented in Table 16. When the algorithms were listed from high to low according to the Accuracy rate, the sequence of the classification algorithms was Random Forest, ANN, KNN, Decision Tree, and SVM algorithms. When the algorithms were listed according to cohen's kappa values, the sequence was Random Forest, ANN, KNN, Decision Tree, and SVM algorithms. The accuracy rate and cohen's kappa values were ranked with similar results. The Random Forest method's high value according to the accuracy rate and cohen's kappa indicated its superiority over other methods. Since cohen’s kappa value is an important reference showing the reliability of the result, it indicated that the reliability of the prediction success of the Random Forest algorithm was superior to other algorithms.

Table 16 was compared; this finding demonstrated that the Random Forest algorithm was superior to other algorithms in terms of precision in difficult environments. According to F-Measure values, the Random Forest algorithm was found to be more successful compared to other algorithms in a low-interest environment, and the ANN algorithm was found to be more successful compared to other algorithms in a high-interest environment. This finding showed that the sensitivity and accuracy of algorithms differed in high-interest or low-interest environments.

Table 17: Comparison of Monthly Data and the Method

Method	Class	Precision	Sensitivity	Specificity	F-measure	Accuracy Statistics	Cohen's Kappa
Decision Tree	Low Interest	1.000	1.000	1.000	1.000	0.988	0.982
	Normal Interest	1.000	1.000	1.000	1.000		
	High Interest	0.970	1.000	0.970	1.000		
	Very High Interest	1.000	0.970	1.000	0.980		
Random Forest	Low Interest	1.000	0.833	1.000	0.909	0.976	0.964
	Normal Interest	0.923	1.000	0.986	0.960		
	High Interest	1.000	0.970	1.000	0.985		
	Very High Interest	0.970	1.000	0.980	0.985		
SVM	Low Interest	1.000	1.000	1.000	1.000	0.882	0.823
	Normal Interest	1.000	0.667	1.000	0.800		
	High Interest	0.750	1.000	0.818	0.857		
	Very High Interest	1.000	0.857	1.000	0.923		
KNN	Low Interest	0.500	0.500	0.961	0.500	0.807	0.707
	Normal Interest	0.667	0.500	0.958	0.571		
	High Interest	0.848	0.848	0.900	0.848		
	Very High Interest	0.857	0.938	0.902	0.896		
ANN	Low Interest	0.400	1.000	0.800	0.571	0.765	0.671
	Normal Interest	1.000	0.833	1.000	0.909		
	High Interest	0.857	1.000	0.909	0.923		
	Very High Interest	0.400	0.648	1.000	0.521		

According to the monthly data analysis, all algorithms used were found to have a high success rate regarding classification accuracy, as presented in Table 17. The accuracy rate and cohen's kappa values differed within the algorithms in parallel with each other. The high value of the Decision Tree algorithm, according to both the accuracy rate and cohen's kappa, indicated its superiority over other methods.

According to the decision tree analysis results, If the inflation rate is less than 19%, the interest rate is likely to be below 50% and normal for 44%. On the other hand, when the inflation rate is higher than 7%, there is an 87% normal interest probability; when it is lower than 7%, there is a 40% lower and 60% normal interest rate.

According to F-Measure values, the Decision Tree algorithm was more successful than other algorithms in all low-interest, normal-interest, high-interest, and very-high-interest environments. This finding showed that the sensitivity and accuracy of algorithms differed in high-interest or low-interest environments. This revealed the necessity of classifying the interest rates in the market according to the past periods and choosing the prediction algorithm according to the high or low-interest conditions.

Conclusion

The aim of this study is to determine which variables will be taken into account for the loan interest rate that banks will offer to their customers during the lending process, and to create a machine learning model that can predict the loan interest rate that the bank will offer to its customers by using these variables. Since the weekly and monthly data variables were different, interest rate prediction proposal models were created by establishing two models. The models were compared to the predictive classification models of Random Forest, SVM, Decision Tree, KNN and ANN algorithms from the machine learning algorithms. Classification algorithms used in the study were compared using the accuracy rate, cohen's kappa, precision, sensitivity, and F-measure measurements.

The algorithms applied to the first data set that consisted of the variables with weekly data were ranked from high to low according to their accuracy rates as the Random Forest, ANN, KNN, Decision Tree, and SVM. The success of the Random Forest algorithm was also observed to be higher in Cohen's alpha values than other algorithms. This indicated that the Random Forest algorithm was more successful than other algorithms in producing correct predictions and in the reliability of the predictions it produced in the model where weekly data were taken into account.

The success of the Decision Tree algorithm was observed to be higher than other algorithms in Cohen's Kappa values. In the data set consisting of monthly data, the accuracy success rate of the models (0.98) was lower than that of the first model consisting of weekly data (0.99). While the Random Forest algorithm was successful in the analysis performed on a greater amount of data, the accuracy success rate of the Decision Tree algorithm was higher in the analysis performed on a smaller amount of data. This finding contributes to the literature as important information for the analysts to select the algorithms.

The fact that the Decision Tree algorithm achieved 100% success in terms of F-measure values in the model with monthly data in low, normal, and very high-interest environments indicates that it can successfully prevent the fragility of the market.

According to the decision analysis results in the second model consisting of monthly data, in cases where the inflation rate is higher than 19%, the interest rate is 60% likely to be very high and 40% likely to be high. If the inflation rate is lower than 19%, the interest rate is 50%, likely low, and 44% to normal. When the inflation rate is higher than 7%, there is an 87% probability of normal interest; when it is lower than 7%, there is a low 40% probability and a 60% probability of a normal interest rate. When the inflation rate is higher than 5%, there is an 82% probability of low interest. When the inflation rate is lower than 5%, there is a 33% probability of a low-interest rate and a 66% probability of a normal interest rate.

According to the results of the Multiple Linear Regression analysis of the weekly data, the R^2 value was found as 0.399. This indicates that the model does not have a high representation ability; however, the R^2 value was interpreted as acceptable based on expert opinions since it involved a wide area related to the country's economy. The "Foreign Debt Principal and Interest Sum" variable was found to be insignificant in terms of affecting the interest rate, which was the dependent variable. Among the variables, the "USD Selling Price", "Stock Index (BIST100)', and "Central Bank Gold Reserve" were significant in affecting the interest rate. These findings indicate that decision-makers must consider the variables found to be significant when pricing a loan.

According to the results of the Multiple Linear Regression analysis of the monthly data, the Inflation and Unemployment Rate variables significantly affected the interest rate. The most important variables affecting the interest rate were "Inflation rate" and "Unemployment rate". On the other hand, the CPI, Industrial Production Index, and Employment Rate variables were insignificant regarding the interest rate. Analysis results consisting of monthly data also show that variables will be effective in determining the interest rate at the stage of granting a loan. Therefore, the outputs of the Multiple Linear Regression analysis performed in the study would contribute to the literature highly in terms of involving very important outputs for the senior management of the banks.

Credit pricing is very important for banks in developing countries. The reason why pricing is so important is that the maturity structure of short-term deposits collected, which is the source of the loan to be extended, is not compatible with the maturity structure of the loans extended. Therefore, expected future interest rates must be predictable to manage long-term loans and short-term deposits effectively. If the bank can foresee the change in the cost within 24 months while extending a 24-month loan, it can prevent major losses. Actual results in the literature have been presented in this study to demonstrate the use of machine learning algorithms in increasing the prediction capabilities of banks. This study consists of the analysis of variables and algorithms for the solution to this problem. In order to contribute to the literature and the banking sector, there is a need for studies in which macroeconomics methods are used with machine learning methods. Therefore, it is recommended to conduct further studies with machine learning algorithms to predict and manage interest costs.

Peer-review:

Externally peer-reviewed

Conflict of interests:

The authors have no conflict of interest to declare.

Grant Support:

The authors declared that this study had received no financial support.

Author Contributions:

Idea/Concept/Design: **K.E**. Data Collection and/or Processing: **K.E**. Analysis and/or Interpretation: **K.E**. Literature Review: **K.E.**, **E.F.**, Writing the Article: **K.E., E.F**. Critical Review: **AÇ., Y.**Y., Approval: **A.Ç., Y.Y., K.Ö., D.G.**

References

Akar, Ö., & Güngör, O. (2012). Rastgele Orman Algoritması Kullanılarak Çok Bantlı Görüntülerin Sınıflandırılması. *Journal of Geodesy and Geoinformation*, *1*(2), 139–146.

Akın, T., & Dağlıoğlu, C. (2021). Tüketim, Belirsizlik ve Borçlanma kısıtı Üçlemi: Türkiye Örneği. *Journal of Yaşar University*, *16*(612020), 414–433.

Aksoy, B. (2021). Pay Senedi Fiyat Yönünün Makine Öğrenmesi Yöntemleri ile Tahmini: Borsa İstanbul Örneği. *Business and Economics Research Journal*, *12*(1), 89–110.

Aksu, H., & Emsen, Ö. S. (2019). Enflasyon, Faiz ve Döviz Kuru İlişkileri: Türkiye İçin ARDL Analizleri İle Asimetrik Eş - Bütünleşme Araştırması. *İktisadi ve İdari Bilimler Dergisi*, *33*(1), 69–90.

Altaş, D., & Gülpınar, V. (2012). Karar Ağaçları ve Yapay Sinir Ağlarının Sınıflandırma Performanslarının Karşılaştırılması: Avrupa Birliği Örneği. *Trakya Üniversitesi Sosyal Bilimler Dergisi*, *14*(1), 1–22.

Altunkaynak, A., Başakın, E. E., & Kartal, E. (2020). Dalgacık K-EN Yakın Komşuluk Yöntemi İle Hava Kirliliği Tahmini. *Uludağ Üniversitesi Mühendislik Fakültesi Dergisi*, *25*(3), 1547–1556.

Apergis, N., Cooray, A., Khraief, N., & Apergis, I. (2019). Do Gold Prices Respond to Real Interest Rates? Evidence From the Bayesian Markov Switching VECM Model. *Journal of International Financial Markets, Institutions and Money*, *60*, 134–148.

Assefa, T. A., Esqueda, O. A., & Varella, A. (2017). Stock Returns and Interest Rates Around the World: A Panel Data Approach. *Journal of Economics and Business*, *89*, 20–35.

Atalay, M., & Çelik, E. (2017). Büyük Veri Analizinde Yapay Zeka ve Makine Öğrenmesi Uygulamaları. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *9*(22), 155–172.

Bal, H., Erdoğan, E., & Palandökenlier, B. (2019). Enflasyon ve Faiz Oranı Arasındaki Nedensellik İlişkisi: Seçilmiş Ülkeler İçin Ampirik Bir Analiz. *Internatıonal Conference On Eurasıan Economıes*, 17–24.

Bezerra, P. C. S., & Albuquerque, P. H. M. (2017). Volatility Forecasting Via SVR–GARCH With Mixture of Gaussian Kernels. *Computational Management Science*, *14*(2), 179–196.

Breiman, L. E. O. (2001). Random Forests. *Kluwer Academic Publishers*, 5–32.

Çavdar, Ş. Ç., & Aydın, A. D. (2018). *Finans Alanında Yapay Zeka ve Ekonometrik Uygulamalar*. Seçkin Yayıncılık.

Çavuşoğlu, Ü., & Kaçar, S. (2019). Anormal Trafik Tespiti için Veri Madenciliği Algoritmalarının Performans Analizi. *Academic Platform Journal of Engineering and Science*, *7*(2), 205–216.

Chen, R., Wang, Z., Yang, L., Ng, C. T., & Cheng, T. C. E. (2022). A Study on Operational Risk and Credit Portfolio Risk Estimation Using Data Analytics. *Decision Sciences*, *53*(1), 84–123.

Chien, C., & Chen, L. (2008). Data Mining to Improve Personnel Selection and Enhance Human Capital: A Case Study in High-Technology Industry. *Expert Systems with Applications*, *34*(1), 280–290.

Çınaroğlu, E., & Avcı, T. (2020). THY Hisse Senedi Değerinin Yapay Sinir Ağları ile Kestirimi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, *34*(1), 1–20.

Çolak, O., & Karahan, Ö. (2017). Enflasyon Hedeflemeli̇ Reji̇m Altinda Türki̇ye Ekonomi̇si̇nde Fai̇z Orani Ve Dövi̇z Kuru Ili̇şki̇si̇. *International Journal of Management Economics and Business*, *13*(ICMEB17), 983–991.

Cortes, C., & Vapnik, V. (1995). *Support-Vector Networks*. *297*, 273–297.

Cutler, A., Cutler, D., & Stevens, J. (2012). *Random Forests. In: Ensemble Machine Learning* (Springer).

Diaz, D., Theodoulidis, B., & Dupouy, C. (2016). Modelling and Forecasting Interest Rates During Stages of the Economic Cycle : A Knowledge-Discovery Approach. *Expert Systems With Applications*, *44*, 245–264.

Dilki, G., & Başar, Ö. D. (2020). İşletmelerin İflas Tahmininde K- EN Yakın Komşu Algoritması Üzerinden Uzaklık Ölçütlerinin Karşılaştırılması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, *19*(38), 224–233.

Dolgun, M. Ö., Özdemir, T. G., & Oğuz, D. (2009). Veri Madenciliğinde Yapısal Olmayan Verinin Analizi: Metin ve Web Madenciliği. *İstatistikçiler Dergisi*, *2*, 48–58.

Elmas, Ç. (2021). *Yapay Zeka Uygulamaları* (Seçkin Yay). Seçkin Yayıncılık.

Erkişi, K. (2018). Para Politikasının Döviz Kuru Üzerindeki Etkileri: Türkiye için bir Uzun Dönem Analizi. *Eurosıa Journal of Social Science & Humanities Internetional Indexed & Refereed*, *5*, 54–64.

Ersöz, F. (2019a). Data Mining and Text Mining With Big Data: Review of Differences. *International Journal of Recent Advances in Multidisciplinary Research*, *06*(01), 4391–4395.

Ersöz, F. (2019b). *Veri Madenciliği Teknikleri ve Uygulamaları*. Seçkin Yayıncılık.

Filiz, E., Karaboğa, H. A., & Akogul, S. (2017). BIST-50 Endeksi Değişim Değerlerinin Sınıflandırılmasında Makine Öğrenmesi Yöntemleri ve. *Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi*, *26*(1), 231–241.

Göçken, M., Özçalıcı, M., & Boru, A. (2016). Integrating Metaheuristics and Artificial Neural Networks for Improved Stock Price Prediction. *Expert Systems With Applications*, *44*, 320–331.

Greuning, H. van, & Iqbal, Z. (2008). *Risk Analysis for Islamic Banks.* World Bank.

Güler, S., & Özçalık, M. (2018). Hisse Getirisi, Faiz Oranı ve Dolar Kuru İlişkisi: BIST'te Bir Uygulama. *Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, *16*(4), 291–306.

Guo, Y., Zhang, Z., & Tang, F. (2021). Feature Selection with Kernelized Multi-Class Support Vector Machine. *Pattern Recognition*, *117*, 107988.

Gupta, R., & Kotzé, K. (2015). The Role of Oil Prices in the Forecasts of South African Interest Rates : A Bayesian Approach. *University of Pretoria Department of Economics Working Paper Series*, *31*, 1–16.

Güriş, B., & İçen, H. (2019). Türkiye 'de Döviz Kuru Faiz ve Risk Arasındaki İlişki: Doğrusal Olmayan Nedensellik Analizi. *2nd International Congress On New Horizons In Education And Social Sciences (ICES-2019) Proceedings*, 592–597.

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature Review: Machine Learning Techniques Applied to Financial Market Prediction. *Expert Systems with Applications*, *124*, 226–251.

Hu, L. Y., Huang, M. W., Ke, S. W., & Tsai, C. F. (2016). The Distance Function Effect on K‑Nearest Neighbor Classification For Medical Datasets. *SpringerPlus*, *5*(1304), 1–9.

Huang, Z., Chen, H., Hsu, C., Chen, W., & Wu, S. (2004). Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study. *Decision Support Systems*, *37*(4), 543–558.

Imran, A. Al, Badrudduza, A. S. ., & Rifat, R. I. (2019). Educational Performance Analytics of Undergraduate Business Students. *International Journal of Modern Education and Computer Science*, *11*(7), 44–53.

İscan, H., & Kaygısız, A. D. (2019). Türkiye'de Döviz Kuru, Enflasyon ve Faiz Oranı İlişkisi: 2009-2017

Uygulaması. *Iğdır Üniverstesi Sosyal Bilimler Dergisi*, *17*(1), 581–604.

Kara, Y., Acar Boyacioglu, M., & Baykan, Ö. K. (2011). Predicting Direction of Stock Price Index Movement Using Artificial Neural Networks and Support Vector Machines: The Sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, *38*(5), 5311–5319.

Karahan, Ö., & Gürbüz, Y. E. (2016). Türkiye'de Bi̇reysel Banka Kredi̇leri̇ ve Enflasyon İli̇şki̇si̇. *Uluslararası Yönetim İktisat ve İşletme Dergisi*, *ICMEB17 Öz*, 410–417.

Karakul, A. K. (2020). Yapay Sinir Ağları ile BIST-100 Endeks Değerinin Tahmin Edilmesi. *Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *7*(2), 497–509.

Kavcıoğlu, Ş. (2019). Kurumsal Kredi Skorlamasında Klasik Yöntemlerle Yapay Sinir Ağı Karşılaştırması. *İstanbul İktisat Dergisi*, *69*(2), 207–245.

Kavzoğlu, T., & Çölkesen, İ. (2010). Karar Ağaçları ile Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği. *Harita Teknolojileri Elektronik Dergisi*, *2*(1), 36–45.

Khan, M., Ding, Q., & Perrizo, W. (2015). K-Nearest Neighbor Classification on Spatial Data Streams Using P-Trees. *Proceedings of the 6th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*.

Kim, H., & Shi, W. (2018). The Determinants of the Benchmark Interest Rates in China. *Journal of Policy Modeling*, *40*(2), 395–417.

Kılınç, D., Borandağ, E., Yücalar, F., Tunalı, V., Şimşek, M., & Özçift, A. (2016). KNN Algoritması ve R Dili ile Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi. *Marmara Fen Bilimleri Dergisi*, *3*, 89–94.

Koçak, H. (2020). Çalışanların Örgütsel Bağlılıklarının CART Karar Ağacı Algoritması ile Belirlenmesi. *Ulusal İktisadi ve İdari Bilimler Dergisi*, *6*, 0–3.

Koçoğlu, E., & Ersöz, F. (2021). Data Mining Application for Decision Optimization at Risk. *International Journal of 3D Printing Technologıes and Digital Industry*, *5*(2), 195–209.

Kunze, F., Wegener, C., Bizer, K., & Spiwoks, M. (2017). Forecasting European Interest Rates in Times of Financial Crisis – What Insights Do We Get From International Survey Forecasts ? *Journal of International Financial Markets, Institutions & Money*, *48*(C), 192–205.

Li, X., Si, D., & Ge, X. (2021). China's Interest Rate Pass-Through After the Interest Rate Liberalization: Evidence from a Nonlinear Autoregressive Distributed Lag Model. *International Review of Economics and Finance*, *73*, 257–274.

Maehashi, K., & Shintani, M. (2020). Macroeconomic Forecasting Using Factor Models and Machine Learning: An Application to Japan. *Journal of The Japanese and International Economies*, *58*, 101104.

Nakagawa, S., Hochin, T., Nomiya, H., Nakanishi, H., & Shoji, M. (2021). Prediction of Unusual Plasma Discharge by Using Support Vector Machine. *Fusion Engineering and Design*, *167*, 112360.

Nuray, S. E., Gençdal, H. B., & Arama, Z. A. (2021). Zeminlerin Kıvam ve Kompaksiyon Özelliklerinin Tahmininde Rastgele Orman Regresyonu Yönteminin Uygulanabilirliği. *Mühendislik Bilimleri ve Tasarım Dergisi*, *9*(1), 265–281.

Oh, K. J., & Han, I. (2000). Using Change-Point Detection to Support Artificial Neural Networks for Interest Rates Forecasting. *Expert Systems with Applications*, *19*(2), 105–115.

Pal, M., & Mather, P. M. (2003). An Assessment of The Effectiveness of Decision Tree Methods For Land Cover Classification. *Remote Sensing of Environment*, *86*(4), 554–565.

Salisu, A. A., & Vinh, X. (2021). The Behavior of Exchange Rate and Stock Returns in High and Low Interest Rate Environments. *International Review of Economics and Finance*, *74*, 138–149.

Şentürk, F., & Fındık, H. (2014). Rasyonel Karar Alan Ekonomi Bi̇ri̇mi̇n Ri̇sk Altında Verdiği Kararlara Davranışsal Yaklaşım: Kahneman-Tversky Beklenti Teori̇si̇ Perspekti̇fi̇nden Eleştirel bi̇r Bakış. In *Öneri Dergisi* (Vol. 11, Issue 42).

Stolyarov, D., & Tesar, L. L. (2019). Interest Rate Trends in a Global Context. *Michigan Retirement Research Center Research Paper*, *402*, 1–60.

Taş, A. İ., Gülüm, P., & Tulum, G. (2021). Finansal Piyasalarda Hisse Fiyatlarının Derin Öğrenme ve Yapay Sinir Ağı Yöntemleri ile Tahmin Edilmesi; S&P 500 Endeksi Örneği. *Düzce Üniversitesi Bilim* *ve Teknoloji Dergisi*, *9*, 446–460.

Tay, F. E. H., & Cao, L. (2001). Application of Support Vector Machines in Fınancial Time Series Forecasting. *Omega*, *29*, 309–317.

Tekin, B. (2016). *Firmaların Finansman Kararları : Davranışsal Perspektif*. 169–198.

Telçeken, H., & Değirmen, S. (2019). Enflasyon ve Kredi Faizleri Arasındaki Uzun Dönemli İlişkinin Fisher Hipotezi Çerçevesinde Değerlendirilmesi: Türkiye Uygulaması. *Istanbul Business Research*, *47*(2), 154–182.

Thakur, G. S. M., Bhattacharya, R., & Mondal, S. S. (2016). Artificial Neural Network Based Model for Forecasting of Inflation in India. *Fuzzy Information and Engineering*, *8*(1), 87–100.

Tursoy, T. (2019). The Interaction Between Stock Prices and Interest Rates in Turkey: Empirical Evidence from ARDL Bounds Test Cointegration. *Financial Innovation*, *5*(7).

Whited, T. M., Wu, Y., & Xiao, K. (2021). Low Interest Rates and Risk Incentives for Banks with Market Power. *Journal of Monetary Economics*, *121*(C), 155–174.

Yenice, S., & Yenisu, E. (2019). Türkiye'de Döviz Kuru, Enflasyon ve Faiz Oranlarının Etkileşimi. *Journal of Chemical Information and Modeling*, *22*(4), 1521–1546.

Yiğiter, Ş. Y., Sarı, S. S., & Başakın, E. E. (2017). Hisse Senedi Kapanış Fiyatlarının Yapay Sinir Ağları ve Bulanık Mantık Çıkarım Sistemleri ile Tahmin Edilmesii. *Kahramanmaraş Sütçü İmam Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *7*(1), 1–22.

Yıldırım, S., & Sarı, S. (2020). Döviz Kuru ve Enflasyonun Faiz Oranı Üzerindeki Etkisinin NARDL Yaklaşımı ile İncelenmesi: Türkiye Ekonomisi için Bir Örnek. *Balkan ve Yakın Doğu Sosyal Bilimler Dergisi*, *06*(02), 147–161.

Yong, C., & Dingming, L. (2019). How Does Government Spending News Affect Interest Rates ? Evidence From the United States. *Journal of Economic Dynamics and Control*, *108*, 103747.

Zhang, N., Lin, A., & Shang, P. (2017). Multidimensional k-Nearest Neighbor Model Based on EEMD For Financial Time Series Forecasting. *Physica A: Statistical Mechanics and Its Applications*, *477*, 161–173.

Zhong, X., & Enke, D. (2017). Forecasting Daily Stock Market Return Using Dimensionality Reduction. *Expert Systems With Applications*, *67*, 126–139.