For citation / Atıf için:
AKAY, Ö., & ALTINDAĞ İ, (2025). Comparison of decision tree algorithms in predicting consumer confidence index. *Uluslararası Sosyal Bilimler ve Eğitim Dergisi – USBED 7*(12), 255–272. https://doi.org/10.5281/zenodo.15049637 , https://dergipark.org.tr/tr/pub/usbed

Comparison of decision tree algorithms in predicting consumer confidence index

Tüketici güven endeksi tahmininde karar ağacı algoritmalarının karşılaştırılması

Özlem AKAY[1]
Doç. Dr.; Gaziantep İslam Bilim ve Teknoloji Üniversitesi, Tıp Fakültesi, Biyoistatistik, 27000, Gaziantep, Türkiye
E-mail: ozlem.akay@gibtu.edu.tr ORCID: 0000-0002-9539-7252

İlkay ALTINDAĞ
Doç. Dr.; Necmettin Erbakan Üniversitesi, Uygulamalı Bilimler Fakültesi, Finans ve Bankacılık, 42000, Konya, Türkiye
E-mail: ialtindağ@erbakan.edu.tr ORCID: 0000-0001-5359-8964

Makale Türü / Article Type:	Araştırma Makalesi / Research Article
Gönderilme Tarihi / Submission Date:	31/01/2025
Revizyon Tarihleri / Revision Dates:	14/03/2025 (Minor r.)
Kabul Tarihi / Accepted Date:	19/03/2025

Etik Beyan / Ethics Statement
✓ Ethical approval was not obtained for this study. The author(s) declare(s) that the study is not subject to ethics committee approval.

Araştırmacıların çalışmaya katkısı / Researchers' contribution to the study
1. Author contribution: Wrote the article, collected the data, and analyzed/reported the results (50%).
2. Author contribution: Wrote the article, collected the data, and analyzed/reported the results (50%).

Çıkar çatışması / Conflict of interest
The author(s) declare(s) that there is no potential conflict of interest in this study.

Benzerlik / Similarity
This study was scanned using the iThenticate program. The final similarity rate is 4 %.

[1] Sorumlu yazar / Corresponding author

Comparison of decision tree algorithms in predicting consumer confidence index

Abstract

The economic conditions and future expectations of individuals in a country can influence their spending and/or saving behaviors. The reflections of these behavioral patterns on the economy can be measured through the consumer confidence index. The aim of this study is to determine the most suitable algorithm for predicting the consumer confidence index by comparing various decision tree algorithms. Independent variables such as unemployment rate, BIST100 index, housing price index, exchange rate, and consumer price index, which are thought to impact the consumer confidence index, were used in the study. In the analyses, monthly data for the period of 01.2014-08.2024 were used, and 70% of the data was separated for training and 30% for testing. Decision tree-based algorithms, including Random Forest, XGBoost, LightGBM, and CatBoost, were applied to these data to develop predictive models. MSE, RMSE, MAE and MAPE error criteria were used to evaluate the performance of the algorithms. Based on the MSE and RMSE results, the RF algorithm, and based on the MAE and MAPE results, the XGBoost algorithm have both been determined as suitable decision tree algorithms for CCI prediction.

Keywords: Consumer confidence index, CatBoost, LightGBM, Random forest, XGBoost.

EXTENDED ABSTRACT

Introduction

This study aims to examine the effectiveness of decision tree algorithms in predicting the Consumer Confidence Index (CCI). The CCI is a critical economic indicator that reflects consumers' perceptions of current and future economic conditions, providing valuable insights into key economic indicators such as economic growth, consumption trends, and recessions. Therefore, accurately forecasting the CCI is of significant importance for policymakers, market analysts, and businesses.

The main objective of this study is to compare the performance of four different decision tree algorithms (Random Forest, XGBoost, LightGBM, and CatBoost) in predicting the CCI and determine which algorithm achieves the highest accuracy. In this context, the study seeks to address the following questions: Which decision tree algorithm provides the highest accuracy in predicting the CCI? What are the capabilities of decision tree algorithms in analyzing complex time series data like the CCI? How can the applicability of machine learning techniques in forecasting economic indicators be enhanced?

This study contributes significantly to the literature by examining the use of machine learning models in forecasting economic indicators, in addition to traditional statistical methods. Specifically, the ability of decision tree algorithms to capture hidden patterns in complex datasets makes them a powerful tool for forecasting a dynamic and multidimensional index like the CCI. Moreover, this research has the potential to guide more effective strategic planning and decision-making processes for policy makers and the business world, and to help develop more accurate forecasting models. The limitations encountered during the study may include challenges related to the size and quality of the dataset, the computational resources required for hyperparameter optimization of algorithms, and the difficulties in modeling seasonalities and trends inherent in time series data. Additionally, the impact of external shocks and unforeseen events on economic indicators can complicate the modeling process.

Conceptual and Theoretical Framework

The theoretical framework of this study is centered on understanding consumer behavior and exploring the applicability of machine learning algorithms in forecasting economic indicators. The core theoretical foundation is based on the notion that individuals' perceptions and expectations regarding economic conditions significantly influence their spending and/or saving behaviors. The CCI serves as a critical tool in measuring the reflections of these behavioral patterns on the economy, providing valuable insights into economic activities and trends.

In this context, the study examines the effectiveness of decision tree algorithms in modeling complex time series data, particularly in the prediction of the CCI. Decision trees are known for their ability to learn intricate patterns and relationships in data. The study specifically focuses on four decision tree-based algorithms Random Forest, XGBoost, LightGBM, and CatBoost which are applied to forecast the CCI.

Additionally, the theoretical framework includes a review of the literature on the integration of economic indicators with machine learning techniques. The role of machine learning in predicting dynamic economic indicators like the CCI is explored in relation to prior studies in the field of economics and finance. The study also discusses the potential relationships between independent variables (such as unemployment rate, BIST100 index, housing price index, exchange rate, and consumer price index) and the CCI, outlining the relevance of these variables in forecasting economic trends.

Method

In this study, four different decision tree algorithms (Random Forest, XGBoost, LightGBM, and CatBoost) were employed to predict the CCI. Decision tree algorithms are effective methods for learning complex relationships within datasets and making predictions. The performance of each algorithm was evaluated using error metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

The independent variables thought to influence the Consumer Confidence Index were selected based on findings from previous research. These variables include: Unemployment Rate (%), BIST100 Index, Housing Price Index, Exchange Rate (CPI-based real effective exchange rate), Consumer Price Index (CPI). These independent variables were derived from monthly data covering the period from January 2014 to August 2024, obtained from the Turkish Central Bank's EVDS (Electronic Data Distribution System) database.

The dataset was split into training and testing sets. Seventy percent of the data was used for training the model, while thirty percent was reserved for testing the model's accuracy. This division is commonly used to ensure the generalizability of the model and avoid overfitting. The training data was used to model the decision tree-based algorithms, and the predictive accuracy of these models was evaluated using the test data. All analyses were conducted using the R programming language (version 4.3.1; R Foundation for Statistical Computing, Istanbul, Turkey).

The results were assessed by comparing the performance of each algorithm using various error metrics. These metrics included MSE, RMSE, MAE, and MAPE. This allowed for an objective comparison of the algorithms' accuracy and identification of the most effective algorithm.

Findings

The findings of the study indicate that Random Forest achieves the best performance concerning MSE and RMSE, whereas XGBoost is the most appropriate model according to MAE and MAPE. These findings suggest that both models are suitable for CCI prediction.

Conclusion, Discussion, and Recommendations

The analysis revealed that the Random Forest (RF) algorithm demonstrated superior performance in predicting the CCI, achieving the lowest error metrics compared to other models. The XGBoost algorithm ranked second, exhibiting similar error rates but with a relatively lower MAE. In contrast, the CatBoost algorithm showed the least effective forecasting performance. Variable importance analysis indicated that key economic indicators, particularly the housing price index (HPI) and the exchange rate (ER), significantly influence the CCI, highlighting their pivotal role in shaping perceptions of economic uncertainty. Additionally, the CPI emerged as an important variable, especially for the XGBoost and LightGBM algorithms. Notably, the unemployment rate (UR) was prioritized as the most significant variable by the CatBoost algorithm, suggesting differing approaches to variable ranking across models.

These findings align with previous studies, which emphasize the significant impact of economic indicators such as housing prices, exchange rates, and unemployment rates on consumer confidence. The observed variations in variable prioritization across algorithms further contribute to the literature on the application of machine learning techniques in economic forecasting, offering insights into the adaptability and specific advantages of different algorithms.

Future research could assess the generalizability of these results by incorporating other machine learning algorithms, such as deep learning models, to explore potential improvements in prediction accuracy. Additionally,

extending the analysis to include other relevant economic variables and periodic effects (e.g., seasonality) may further refine the models and enhance their forecasting capabilities.

INTRODUCTION

Economic indicators are crucial for analyzing the economic situation of countries and predicting future trends. The Consumer Confidence Index (CCI) reflects consumers' perceptions of the current and future economic conditions, serving as a critical information source for policymakers and economic forecasters. This index plays a significant role in both public policies and strategic decision-making in the business world (Islam & Mumtaz, 2016; Shayaa et al., 2018).

The CCI has a direct impact on economic growth, as it measures the degree of economic optimism or pessimism of consumers. Positive consumer confidence can lead to an increase in spending and therefore economic growth, while negative confidence can lead to a decrease in economic growth (Mazurek & Mielcová, 2017). Consumption behavior is of great importance in terms of macroeconomic modeling and policy making. The CCI includes consumers' retrospective assessments of both personal and national economic conditions and their expectations for the future. Research conducted in developed countries shows that economic assessments usually focus on past periods. However, it is stated that participant expectations should be included in the focus of the analysis during periods of economic instability. (Grzywińska-Rąpca & Ptak-Chmielewska, 2023). Considering the influence of psychological factors on economic behavior, the CCI is a significant tool for measuring both individual and societal economic sentiment (Tjandrasa & Dewi, 2022). CCI is created by answers to questions related to economic factors such as interest rates, employment opportunities and price stability (Wang et al., 2019). The CCI includes the current economic situation index and the consumer expectations index and has a scale ranging from 0 to 200. CCI values above 100 indicate an optimistic situation, while values below 100 indicate a pessimistic situation (Karagöz & Aktaş, 2015). The predictive power of the CCI has attracted significant attention in both academic literature and practical contexts. Consumer confidence plays a guiding role in strategic planning for governments and businesses, as it reflects individuals' financial situations, spending intentions and their views on general economic conditions. If low consumer confidence is detected, businesses turn to more value-oriented products and services, while governments may try to increase confidence with economic stimulus packages (Shayaa et al., 2018). As a result, CCI is a valid economic indicator that provides information about economic growth, recession and consumption trends and is an indispensable tool for policy makers, market analysts and researchers (Çelik, 2010; Su et al., 2023).

Predicting the Consumer Confidence Index (CCI) is invaluable for economic decision-makers to adapt more quickly to market conditions and develop effective measures. To improve the accuracy of forecasting processes, various methods need to be employed. In addition to traditional statistical methods, machine learning and artificial intelligence-based models are increasingly preferred for analyzing complex data such as the CCI. This study examines the effectiveness of decision tree algorithms in predicting the CCI.

Decision tree algorithms are powerful techniques widely used in data mining and machine learning for classification and regression analysis. These algorithms learn the relationships within a dataset and represent the decision-making process through a visualizable tree structure. Decision trees excel in analyzing time series data like the Consumer Confidence Index (CCI) due to their ability to capture hidden patterns in complex datasets.

The aim of this study is to compare the performance of four different decision tree algorithms (Random Forest, XGBoost, LightGBM, and CatBoost) in forecasting the CCI and to identify the algorithm that yields the highest accuracy. In this context, the forecasting performance of decision tree-based models is evaluated using metrics such as accuracy and error rates. The study aims to contribute to more effective and accurate forecasting of the CCI and to add to the literature on the applicability of machine learning techniques to forecasting economic indicators.

LITERATURE REVIEW

The Consumer Confidence Index (CCI) is a critical indicator that measures consumers' perceptions and expectations about the economic situation. Frequently used in economic analyses, this index serves as a valuable tool for forecasting potential shifts in economic trends and providing insights into policy perspectives. This literature review compiles academic studies conducted on CCI.

Karagöz and Aktaş (2015) analyzed CCI data using the one-way repeated measures analysis method. Their study identified that the dynamics of CCI are linked to structural movements in the economy.

Mazurek and Mielcová (2017) examined the statistical relationship between CCI and GDP to determine whether CCI serves as a reliable predictor of economic growth or recessions. The results from Granger causality tests embedded in VAR models indicated that, for U.S. data, CCI could be considered a valid predictor of economic growth. However, they noted that short-term forecasts might deviate from long-term trends.

Canöz (2018) investigated the relationship between the CCI announced in Turkey and the Borsa Istanbul 100 Index. Using the Toda-Yamamoto causality test, the findings revealed a unidirectional causal effect of stock returns on consumer confidence.

Wang et al. (2019) applied large-scale data from microblogging platforms to the CCI calculation process. By leveraging a user-defined CCI dictionary derived from the frequency of target words, they tested the predictive capability of this method for the following month's CCI and validated its efficacy using another set of microblog data.

Akkuş and Zeren (2019) explored the relationship between Turkey's Katılım-30 Islamic stock index and investor sentiment, represented by the CCI. Their findings indicated no causal relationship between the two indices. However, cointegration analysis results suggested that in the presence of positive shocks, these indices exhibit a compatible structure in the long term.

Münyas (2019) examined the relationships between Borsa Istanbul stock indices and various confidence indices. Results from the Quantile Regression Model showed statistically significant relationships between stock indices and the Economic Confidence Index (ECI), CCI, and the Real Sector Confidence Index (RSCI).

Durgun (2019) analyzed the relationship between CCI, RSCI, and certain macroeconomic variables using the VAR model. Empirical results indicated that both CCI and RSCI are influenced by and also impact these macroeconomic variables.

Ohmura (2020) proposed extending and enhancing the use of Japan's CCI data through an alternative index. By employing time series clustering analysis (TSCA) with dynamic time warping (DTW) distances, the study cross-referenced and validated the reliability of traditional indices.

Qiu (2020) introduced a new MIDAS (Mixed Data Sampling) approach by integrating regression tree-based algorithms into the MIDAS framework. An out-of-sample forecasting study for CCI revealed that the proposed method utilized past sentiment data more comprehensively, significantly improving forecast accuracy and highlighting the role of social media in influencing consumer confidence.

Beşiktaşlı and Cihangir (2020) investigated the direction and relationship between CCI, financial markets, and general economic indicators through Cointegration and Granger Causality Tests. Their findings showed a long-term relationship between CCI and money markets and general macroeconomic indicators, but no long-term relationship with capital market variables.

Caleiro (2021) analyzed the relationship between consumer confidence levels and unemployment rates using learning models called regression and classification trees. Classification trees demonstrated that the distinction between low and high consumer confidence values could be made using a sufficient threshold for the unemployment rate. Regression trees indicated an inverse proportionality between consumer confidence levels and unemployment rates.

Tjandrasa and Dewi (2022) aimed to explore a new variable affecting Indonesia's CCI and explain the effects among variables using secondary data. Results suggested that Indonesia's CCI is influenced by inflation rates, unemployment rates, exchange rates, and corruption control conditions.

Şeyranlıoğlu (2023) investigated the relationships between CCI and real returns on various financial investment instruments. Results from Hacker and Hatemi-J's (2012) bootstrap causality test showed that CCI could not be used as a leading indicator for predicting investment instrument returns.

Nguyen et al. (2023) evaluated two multivariate classical econometric models (MLR and ARDL) and two machine learning models (SVR and MARS) for CCI forecasting in the U.S. Results revealed that MARS, SVR, ARDL, and MLR models achieved high accuracy parameters, outperforming many existing baseline models in CCI prediction.

Han et al. (2023) developed a conceptual framework examining the relationship between CCI and web search keywords. The study used six machine learning and deep learning models BP neural network, convolutional neural network, support vector regression, random forest, ELMAN neural network, and extreme learning machine to predict CCI. Results demonstrated that machine learning models provided superior predictive performance for CCI.

Lin et al. (2024) presented an innovative framework integrating social network analysis (SNA) with machine learning (ML) models to forecast China's CCI. The proposed model addressed the limitations of traditional econometric and ML methods by incorporating complex dependencies among economic variables, particularly enhancing prediction accuracy during periods of economic volatility.

Vitkauskaitė (2024) evaluated the efficiency and reliability of consumer confidence indicators derived from social media and administrative data. SARIMAX, VECM, Random Forest, and XGBoost models were employed for CCI forecasting. XGBoost achieved the highest accuracy, with SARIMAX showing comparable performance, while Random Forest and VECM demonstrated relatively lower accuracy.

MATERIALS and METHODS

In this study, decision tree algorithms, including Random Forest, XGBoost, LightGBM, and CatBoost, were employed to predict the consumer confidence index. Numerous studies in the literature have examined the determinants of the consumer confidence index, with each utilizing different variables. Considering the findings from previous research, variables thought to influence the consumer confidence index were selected as follows: unemployment rate (%), BIST100 index (based on XU100 closing prices), housing price index, exchange rate (CPI-based real effective), and consumer price index. These variables consist of monthly data covering the period from January 2014 to August 2024, obtained from the Turkish Central Bank's EVDS database. For all algorithms, 70% of the dataset was used for training, and 30% was used for testing. The analyses were conducted using the R (version 4.3.1; R Foundation for Statistical Computing, Istanbul, Turkey) programming language.

Random Forest

Random Forest (RF) was proposed by Breiman in 2001 as a classification and regression method. This algorithm is robustness and flexibility in modeling the input–output functional relationship appropriately **(**Adusumilli et al., 2013)**.** This algorithm combines multiple decision trees with the same distribution to create a forest, which is then utilized for training and making predictions on the sample dataset (Kuhn and Johnson 2013). A decision tree is a non-parametric supervised learning approach that extracts decision rules from a dataset consisting of features and labels. It utilizes the tree's structure to represent these rules, effectively addressing classification and regression tasks (Zhang et al, 2021). When RF is provided with an input vector (X), comprising the values of various evidential features from a specific training area, it constructs K regression trees and computes the average of their outputs (Rodriguez-Galiano et al., 2015).

Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) was proposed by Chen and Guestrin (2016) as an alternative method to estimate a response variable based on certain covariates **(**Pesantez-Narvaez et al., 2019)**.** Itis a highly efficient gradient-boosting machine learning technique that outperforms many other tree-based algorithms. It is an ensemble method, which combines the predictions of multiple base learners and aggregates the same to generate a final result. It is a learning method that enhances the performance of three models through boosting (Pramanik, et al., 2024). As the tree structure, *f(x)*, the final prediction was calculated by summing up the scores across all leaves and this can be expressed as $\hat{y}_i = \sum_{j=1}^{N} f_j(x_i)$ (Huang et al., 2020). There are *N* trees in the model (Ge et al., 2022). The following regularized objective is minimize to learn the set of functions used in the model.

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \gamma T + \frac{1}{2}\lambda \|w\|^2$$

where l is a differentiable convex loss function that measures the difference between the prediction $\hat{y}_i$ and the target y_i (Chen and Guestrin, 2016). γ refers to the user-definable penalty which meant to encourage pruning; T denotes the number of terminal nodes or leaves in an individual tree structure; λ is a regularization term that reduce the prediction's insensitivity to individual observation; $\|w\|$ represents the leaf weights which also considered as the output value of the leaf (Mariadass et al., 2022). To explore the detailed theory behind XGBoost, the original paper by Chen and Guestrin (2016) is recommended.

Light Gradient Boosted Machine Learning Algorithms (LigthGBM)

Light Gradient Boosted Machine Learning Algorithms **(**LightGBM), an advanced version of the Gradient Boosting Decision Tree (GBDT), was introduced in 2017. GBDT is an ensemble algorithm that builds a sequence of models, where each is a linear combination of subsets (Mpofu et al. ,2023). GBDT is widely utilized across numerous fields and applications, including click-through rate prediction, search ranking systems, and multi-class classification problems (Huang, 2023).

LightGBM is a fast, scalable, and high-performance gradient boosting decision tree algorithm that leverages the Histogram technique. By organizing large datasets into histograms, it reduces memory usage and computational overhead. Unlike traditional methods, LightGBM employs a leaf-wise tree growth strategy, which splits trees based on leaves rather than levels or depth. This approach allows it to identify critical points and halt calculations more effectively. As a result, the leaf-wise method minimizes loss more efficiently than level-wise methods, leading to improved accuracy. However, a downside of the leaf-wise approach is its tendency to create deeper trees, increasing the risk of overfitting. To mitigate this, LightGBM incorporates a maximum depth limit, striking a balance between high efficiency and overfitting prevention (Zhu et al., 2022). LightGBM utilizes two key techniques, GOSS and EFB, to enhance efficiency by reducing the number of data samples and features, respectively. The GOSS method prioritizes instances with high gradients while randomly sampling those with smaller gradients. On the other hand, EFB combines mutually exclusive features, effectively lowering feature dimensionality without any loss of information (Li et al., 2022). These methods aim to simplify the process of constructing histograms (Ponsam et al., 2021).

CatBoost

Catboost based on the gradient boosting decision tree (GBDT) proposed in 2017 is a non-linear regression algorithm that uses ensemble learning (Joo et al, 2023). CatBoost has the ability to transform categorical features into numerical ones during the training process. A highly effective approach for managing categorical features with minimal information loss is the target statistic, which estimates the expected target value for each category. CatBoost improves this by using a more efficient and practical method called ordered target statistics. Specifically, it randomly permutes the dataset and computes the average label value for each instance, considering only the categories that appear before the current one in the permutation (Yang et al., 2021). CatBoost addresses issues like gradient bias and prediction shift, offering severl advantages. It integrates an innovative algorithm that automatically treats categorical features as numerical values, utilizes a combination of categorical features to exploit relationships between them, thus enhancing feature dimensions, and employs a perfectly balanced tree structure to reduce overfitting, ultimately improving both the accuracy and generalizability of the model (Luo et al., 2021).

FINDINGS

Descriptive statistics of the variables used in the study are given in Table 1.

Table 1. Descriptive statistics of the variables

Variables	Mean	Standard deviation	Median	Minimum	Maximum
CCI	83.811016	7.573573	82.835	63.410	97.370
UR	11.027344	1.665430	10.600	7.700	15.100
HPI	31.086250	38.923002	11.360	6.670	146.400
BIST	2250.389922	2615.055345	1047.815	618.100	10647.910
ER	77.227031	18.491202	75.635	47.610	108.600
CPI	642.428750	558.604512	411.575	233.540	2453.340

Table 1 provides descriptive statistics for the variables covering the period from January 2014 to August 2024, summarized as mean, standard deviation, median, minimum, and maximum values. According to Table 1, the consumer confidence index (CCI) ranged between 63.410 and 97.370, with a mean value of 83.811 during the analyzed period. The unemployment rate (UR) varied between 7.70% and 15.10%, with an average of 11.027%. The housing price index(HPI) had a mean value of 31.086, with the highest and lowest values being 146.4 and 6.67, respectively. The BIST100 index (BIST) recorded a minimum value of 618.1 and a maximum value of 10,647.91. The average exchange rate (ER) (CPI-based real effective) was 77.227. This variable ranged from a minimum of 47.610 to a maximum of 108.600.The consumer price index ranged from a minimum of 233.54 to a maximum of 2453.34. Performance metrics for the test dataset were calculated and compared for the algorithms used to predict the consumer confidence index. To evaluate the predictive performance of the models, values for Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) were computed. The results are summarized in Table 2.

Table 2. Performance metrics

	Random Forest	XGBoost	LightGBM	CatBoost
MSE	11.18624	11.47668	17.59433	32.82665
RMSE	3.344584	3.387725	4.194559	5.729455
MAE	2.455291	2.351369	2.743695	4.561562
MAPE	3.10142	3.007493	3.54959	5.704303

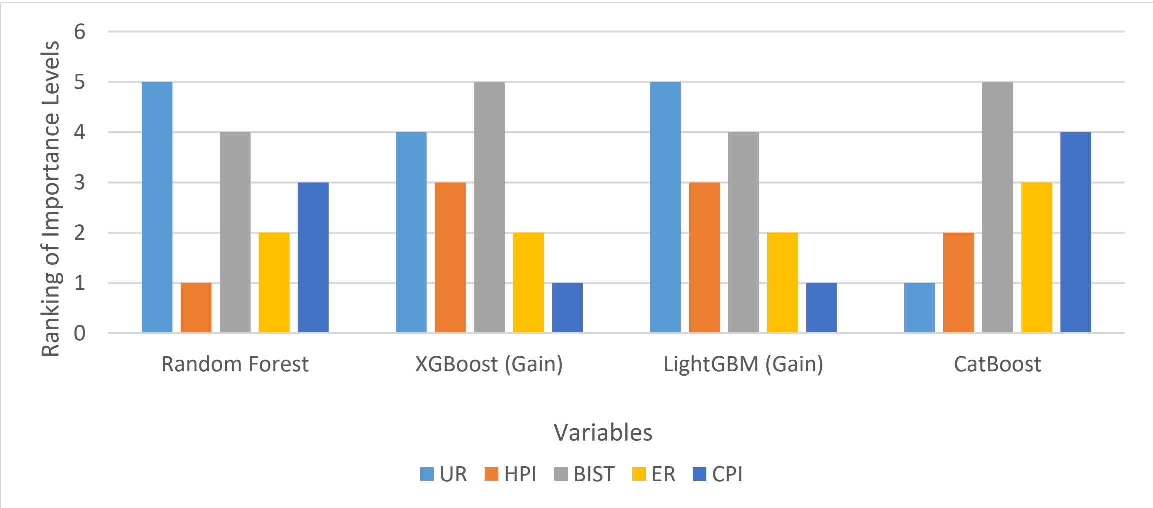
When examining the performance metrics of the algorithms, the Random Forest algorithm exhibited the lowest MSE and RMSE values, whereas the CatBoost algorithm had the highest MSE and RMSE values. Regarding MAE and MAPE, the XGBoost algorithm demonstrated the smallest values, while CatBoost again showed the largest. Based on these findings, it can be concluded that the Random Forest algorithm performs best in terms of MSE and RMSE, whereas XGBoost is the most suitable algorithm according to MAE and MAPE.

The factor importance levels of the variables influencing the consumer confidence index for each algorithm are presented in Table 3. The ranking of variable importance for each algorithm is visually depicted in Figure 1.

Table 3. Importance levels of independent variables

	Random Forest	XGBoost (Gain)	LightGBM (Gain)	CatBoost
UR	386.0468	0.03579791	0.03126459	30.17413
HPI	1163.8851	0.18650420	0.07532212	23.77840
BIST	791.1017	0.03173478	0.04286374	11.26253
ER	1136.4046	0.23860450	0.13035099	20.60833
CPI	1014.2411	0.50735861	0.72019856	14.17660

Figure 1. Graph of factor importance levels of variables affecting the consumer confidence index of algorithms



Based on Table 3 and Figure 1, the most influential variable affecting the consumer confidence index for the Random Forest algorithm is the housing price index (HPI: 1163.8851), followed by the exchange rate (ER: 1136.4046). The other variables, in descending order of importance, are the consumer price index (CPI: 1014.2411) and the BIST100 index (BIST: 791.1017). For the XGBoost and LightGBM algorithms, the top three factors, in the same order of importance, are CPI, ER, and HPI. For the XGBoost algorithm, the fourth and fifth most influential variables are the unemployment rate (UR) and the BIST100 index (BIST), respectively. In contrast, for the LightGBM algorithm, the fourth and fifth variables are BIST100 and UR, respectively. In the CatBoost algorithm, the variable importance rankings are as follows: UR (30.174), HPI (23.778), ER (20.608), CPI (14.176), and BIST100 (11.262).

Figure 2. Consumer Confidence Index Prediction Results Using Decision Tree Algorithms

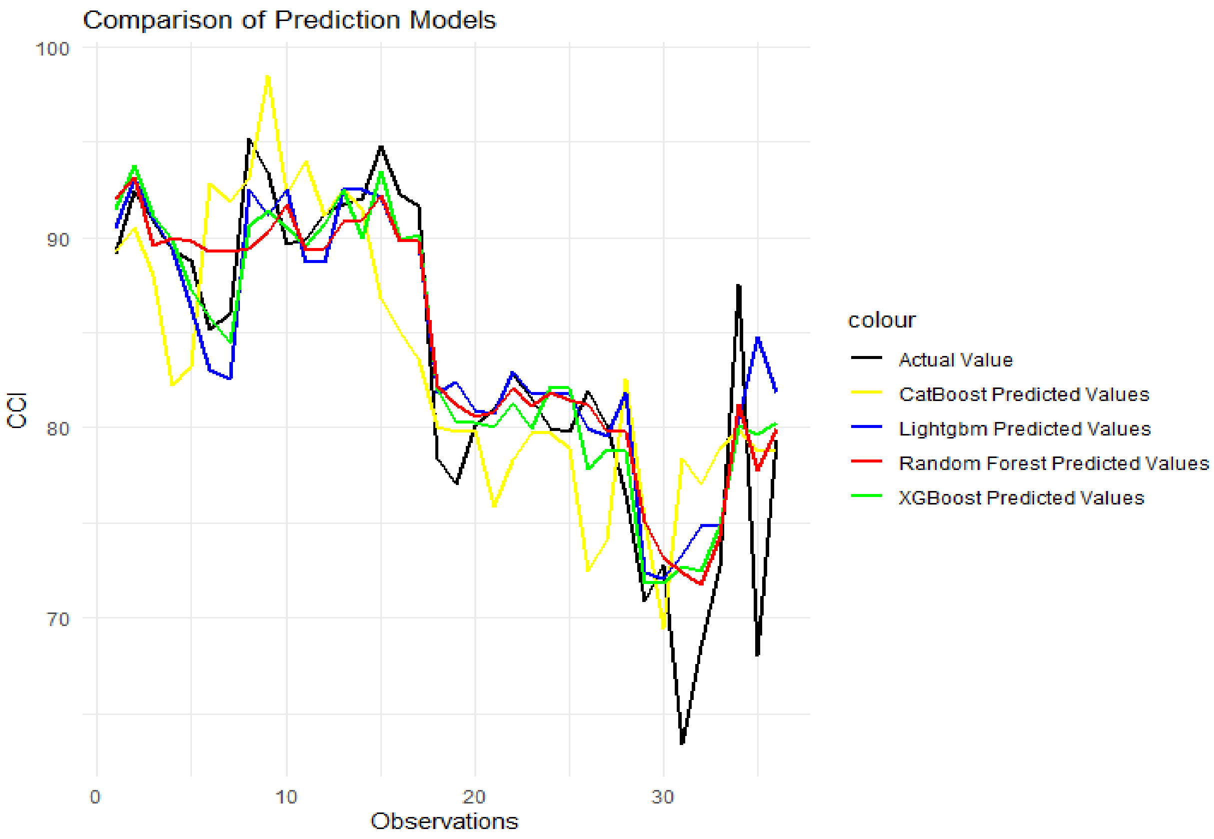


Figure 2 illustrates the prediction results of the consumer confidence index obtained using the Random Forest, XGBoost, LightGBM, and CatBoost algorithms, along with the actual consumer confidence index values. The black line represents the target (actual) values, while the red, green, blue, and yellow lines correspond to the predicted values from the Random Forest, XGBoost, LightGBM, and CatBoost algorithms, respectively. As evident from the figure, predictions generated using the Random Forest and XGBoost algorithms align more accurately with the data points. Among the algorithms employed, the Random Forest and XGBoost algorithms are the more successful in predicting the consumer confidence index.

CONCLUSION

The Consumer Confidence Index (CCI) reflects consumers' perceptions of the current and future economic conditions, playing a critical role in forecasting spending behavior and overall economic activity (Lin et al., 2024). Accurate prediction of the CCI provides early signals about economic growth, recessions, and consumption trends, enabling both public policies and private sector strategies to be shaped more effectively (Jimenez, 2013). CCI forecasts assist market analysts and researchers in evaluating economic uncertainties while serving as a vital guide for long-term strategic planning. Therefore, predicting the CCI is an essential tool for supporting economic and financial stability.

This study compares the performance of four different decision tree algorithms Random Forest, XGBoost, LightGBM, and CatBoost in predicting the Consumer Confidence Index (CCI). Based on macroeconomic data from January 2014 to August 2024, the study examines the impact of five independent variables unemployment rate (UR), housing price index (HPI), BIST100 index (BIST), exchange rate (ER), and consumer price index (CPI) on CCI. MSE, RMSE, MAE, and MAPE measures were used to evaluate the effectiveness of the estimation models in the study.

According to the calculated MSE and RMSE results, the RF algorithm, and according to the MAE and MAPE results, the XGBoost algorithm, are both identified as suitable decision tree algorithms for CCI prediction. In contrast, the CatBoost algorithm showed weaker forecasting performance.

Variable importance analysis indicated that economic indicators such as HPI and ER have a significant impact on the CCI. This finding highlights the decisive role of the HPI and ER in the perception of economic uncertainty. The CPI also emerged as an important variable, particularly for XGBoost and LightGBM algorithms. The fact that the UR is ranked first in the CatBoost algorithm shows that the algorithm makes a different variable prioritization. These differences emphasize the diversity in the approaches of the algorithms to evaluate the relationships of variables.

Analyzing the usability of decision tree algorithms in predicting the CCI and the influence of key economic indicators on the index, this study shows that the high importance of HPI and ER affects the economic uncertainty on consumer confidence. The findings provide important clues for policy makers and the financial sector in understanding consumer behavior and predicting future trends in economic indicators. Future studies can evaluate the generalizability of these findings by examining the performance of other machine learning algorithms. Furthermore, model performance can be investigated in more detail by including different periodic effects and variables.

REFERENCES

Adusumilli, S., Bhatt, D., Wang, H., Bhattacharya, P., & Devabhaktuni, V. (2013). A low-cost INS/GPS integration methodology based on random forest regression. *Expert Systems with Applications, 40*(11), 4653-4659.

Akkuş, H. T., & Zeren, F. (2019). Tüketici güven endeksi ve Katılım-30 İslami hisse senedi endeksi arasındaki saklı ilişkinin araştırılması: Türkiye örneği. *Third Sector Social Economic Review, 54*(1), 53-70.

Beşiktaşlı, D. K., & Cihangir, Ç. K. (2020). Tüketici güven endeksinin finansal piyasalara ve makroekonomik yapıya etkisi. *Finans Ekonomi ve Sosyal Araştırmalar Dergisi, 5*(1), 54-67.

Breiman, L. (2001). Random forests. *Machine learning, 45*, 5-32.

Caleiro, A. B. (2021). Learning to classify the consumer confidence indicator (in Portugal). *Economies, 9*(3), 1-12.

Canöz, İ. (2018). Borsa İstanbul 100 endeksi ile tüketici güven endeksleri arasındaki nedensellik ilişkisi: Türkiye örneği. *Fiscaoeconomia, 2*(1), 136-153.

Çelik, S. (2010). An unconventional analysis of consumer confidence index for the Turkish economy. *International Journal of Economics and Finance Studies, 2*(1), 121-129.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

Durgun, A. (2019). Türkiye'de tüketici ve reel kesim güven endeksi ile seçilmiş makro değişkenler arasındaki ilişkiler: 2010-2018. *Journal of Management and Economics Research, 17*(1), 314-332.

Ge, J., Zhao, L., Yu, Z., Liu, H., Zhang, L., Gong, X., & Sun, H. (2022). Prediction of greenhouse tomato crop evapotranspiration using XGBoost machine learning model. *Plants, 11*(15), 1-17.

Grzywińska-Rąpca, M., & Ptak-Chmielewska, A. (2023). Backward assessments or expectations: What determines the consumer confidence index more strongly? Panel model based on the CCI of European countries. Wiadomości Statystyczne. *The Polish Statistician, 68*(2), 1-15.

Han, H., Li, Z., & Li, Z. (2023). Using machine learning methods to predict consumer confidence from search engine data. *Sustainability, 15*(4), 1-12.

Huang, J. C., Tsai, Y. C., Wu, P. Y., Lien, Y. H., Chien, C. Y., Kuo, C. F., ... & Kuo, C. H. (2023). Predictive modeling of blood pressure during hemodialysis: A comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method. *Computer Methods and Programs in Biomedicine, 195*, 1-6.

Huang, S. (2023, December). Price Prediction and Analysis Based on the LightGBM Regression Model. In *2023 4th International Conference on Computer, Big Data and Artificial Intelligence (ICCBD+ AI)* (pp. 580-584). IEEE.

Islam, T. U., & Mumtaz, M. N. (2016). Consumer confidence index and economic growth: An empirical analysis of EU countries. *EuroEconomica, 35*(2), 17-22.

Jimenez, B. S. (2013). Strategic planning and the fiscal performance of city governments during the Great Recession. *The American Review of Public Administration, 43*(5), 581-601.

Joo, C., Park, H., Lim, J., Cho, H., & Kim, J. (2023). Machine learning-based heat deflection temperature prediction and effect analysis in polypropylene composites using catboost and shapley additive explanations. *Engineering Applications of Artificial Intelligence, 126*, 1-12.

Karagöz, D., & Aktaş, S. (2015). Evaluation of consumer confidence index of Central Bank of Turkey consumer tendency survey. *TOJSAT, 5*(3), 31-36.

Kuhn, M., and K. Johnson. 2013. Applied Predictive Modeling.

Li, B., Chen, G., Si, Y., Zhou, X., Li, P., Li, P., & Fadiji, T. (2022). GNSS/INS integration based on machine learning LightGBM model for vehicle navigation. *Applied Sciences, 12*(11), 1-12.

Lin, Y. C., Sung, B., & Park, S. D. (2024). Integrated systematic framework for forecasting China’s consumer confidence: A machine learning approach. *Systems*, *12*(11), 1-33.

Luo, M., Wang, Y., Xie, Y., Zhou, L., Qiao, J., Qiu, S., & Sun, Y. (2021). Combination of feature selection and catboost for prediction: The first application to the estimation of aboveground biomass. *Forests, 12*(2), 1-21.

Mariadass, D. A., Moung, E. G., Sufian, M. M., & Farzamnia, A. (2022, November). EXtreme gradient boosting (XGBoost) regressor and shapley additive explanation for crop yield prediction in agriculture. In *2022 12th International Conference on Computer and Knowledge Engineering (ICCKE)* (pp. 219-224). IEEE.

Mazurek, J., & Mielcová, E. (2017). Is consumer confidence index a suitable predictor of future economic growth? An evidence from the USA. *E & M Ekonomie A Management, 20*(2), 30–45.

Mpofu, K., Adenuga, O. T., Popoola, O. M., & Mathebula, A. (2023). LightGBM and SVM algorithms for predicting synthetic load profiles using a non-intrusive approach. https://doi:10.20944/preprints202308.0257.v1.

Münyas, T. (2019). Borsa İstanbul Endeksleri ile Güven Endeksleri arasindaki ilişkinin araştirilmasi üzerine bir inceleme. *TESAM Akademi Dergisi, 6,* 299-320.

Nguyen, T. T., Nguyen, H. G., Lee, J. Y., Wang, Y. L., & Tsai, C. S. (2023). The consumer price index prediction using machine learning approaches: Evidence from the United States. *Heliyon, 9*(10), 1-17.

Ohmura, H. (2020). A new measurement for Japanese Consumer Confidence Index. *Economics Bulletin, 40*(2), 1557-1569.

Pesantez-Narvaez, J., Guillen, M., & Alcañiz, M. (2019). Predicting motor insurance claims using telematics data—XGBoost versus logistic regression. *Risks, 7*(2), 1-16.

Ponsam, J. G., Gracia, S. J. B., Geetha, G., Karpaselvi, S., & Nimala, K. (2021, December). Credit risk analysis using LightGBM and a comparative study of popular algorithms. In *2021 4th International Conference on Computing and Communications Technologies (ICCCT)* (pp. 634-641). IEEE.

Pramanik, P., Jana, R. K., & Ghosh, I. (2024). AI readiness enablers in developed and developing economies: Findings from the XGBoost regression and explainable AI framework. *Technological Forecasting and Social Change, 205*, 1-18.

Qiu, Y. (2020). Forecasting the Consumer Confidence Index with tree-based MIDAS regressions. *Economic Modelling, 91*, 247-256.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. O. G. R. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews, 71*, 804-818.

Şeyranlıoğlu, O. (2023). Tüketici Güven Endeksi ile finansal yatırım araçlarının reel getirileri arasındaki nedensellik ilişkilerinin değerlendirilmesi: Türkiye örneği. *Karadeniz Sosyal Bilimler Dergisi, 15*(29), 572-593.

Shayaa, S., Ainin, S., Jaafar, N. I., Zakaria, S. B., Phoong, S. W., Yeong, W. C., ... & Zahid Piprani, A. (2018). Linking consumer confidence index and social media sentiment analysis. *Cogent Business & Management, 5*(1), 1-12.

Su, C. W., Meng, X. L., Tao, R., & Umar, M. (2023). Chinese consumer confidence: A catalyst for the outbound tourism expenditure? *Tourism Economics, 29*(3), 696-717.

Tjandrasa, B. B., & Dewi, V. I. (2022). Determinants of Consumer Confidence Index to predict the economy in Indonesia. *Australasian Accounting, Business and Finance Journal, 16*(4), 3-13.

Vitkauskaitė, A. (2024). Evaluation of consumer confidence indicators using social media and administrative data (Doctoral dissertation, Vilniaus universitetas.).

Wang, P., Li, X., Zhan, X., Zhang, Y., Yan, Y., & Meng, W. (2019). Building consumer confidence index based on social media big data. *Human Behavior and Emerging Technologies, 1*(3), 261-268.

Yang, X., & Chen, Z. (2021, April). A hybrid short-term load forecasting model based on catboost and lstm. In *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)* (pp. 328-332). IEEE.

Zhang, W., Wu, C., Li, Y., Wang, L., & Samui, P. (2021). Assessment of pile drivability using random forest regression and multivariate adaptive regression splines. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 15*(1), 27-40.

Zhu, J., Su, Y., Liu, Z., Liu, B., Sun, Y., Gao, W., & Fu, Y. (2022). Real-time biomechanical modelling of the liver using LightGBM model. *The International Journal of Medical Robotics and Computer Assisted Surgery, 18*(6), 1-11.