International Journal of Finance and Managerial Accounting, Vol.12, No.45, Spring 2027

Designing a Model to Predict the Uncertainty of Financial Information by Selecting Features Using Machine Learning and Neural Networks

Moslem Talebvand
Department of Accounting, Borujerd Branch, Islamic Azad University, Borujerd, Iran

Mahmoud Hematfar
Department of Accounting, Borujerd Branch, Islamic Azad University, Borujerd, Iran

Nasrollah Takhtaei
Department of Accounting, Dezful Branch, Islamic Azad University, Dezful, Iran

Submit: 29/12/2024 **Accept:** 09/04/2025

ABSTRACT

This investigation presents a novel approach to designing a model for the uncertainty of financial information in companies listed on the Tehran Stock Exchange. For this purpose, the data of 114 companies listed on the Tehran Stock Exchange are extracted, and the relationship between 12 independent variables and the dependent variable of financial information uncertainty is investigated. Initially, support vector regression (SVR) is used to evaluate the significance of independent variables or select the best features, leading to the selection of five effective independent variables. Then, the artificial neural network (ANN) algorithm with two hidden layers, the random forest algorithm, and the support vector machine algorithm are implemented in Python to analyze and predict the uncertainty level based on the selected features. The analysis results indicate that among the five selected variables, “firm risk,” “ownership structure,” and “earnings smoothing” significantly affect the uncertainty of financial information. Also, based on the mean square error (MSE) criterion, the random forest and neural network algorithms showed very high accuracy in predicting financial uncertainty. This investigation emphasizes the effectiveness of advanced machine learning techniques in predicting financial information uncertainty and provides investors and researchers in the field of finance with invaluable insights. The findings contribute to the development of more accurate prediction models, facilitating risk management and strategic decision-making in a more complex financial prospect.

Keywords:

financial information uncertainty, artificial intelligence, neural network, feature selection

1. Introduction

In today's dynamic financial environment, it is critical for investors, financial analysts, and policymakers to accurately predict financial uncertainty. Financial uncertainty can significantly affect investment decisions, corporate strategies, and overall market stability. Understanding the factors contributing to this uncertainty is essential for developing effective risk management strategies and enhancing decision-making processes. Conventional techniques of evaluating financial risk are often ineffective in understanding the complexities of modern financial markets and demand more sophisticated analytical approaches.

Recent advances in machine learning (ML) and artificial intelligence (AI) have opened up new avenues to analyze financial data, enabling researchers and specialists to uncover hidden patterns and relationships in complex datasets. Machine learning techniques, especially those involving feature selection, allow for a more in-depth understanding of the factors that cause financial uncertainty. Through systematical identification and evaluation of the most important features, stakeholders can make more informed decisions based on a robust analytical basis.

This research presents a new model for predicting the uncertainty of financial data in companies listed on the Tehran Stock Exchange. For this purpose, a dataset consisting of 114 companies is used to study the correlation between twelve independent variables and the dependent variable, i.e., financial uncertainty. To ensure the relevance and efficiency of the model, support vector regression is first used to evaluate the significance of independent variables, resulting in the identification of five effective predictors. Then, advanced machine learning algorithms, including an artificial neural network with two hidden layers, a random forest algorithm, and a support vector machine algorithm, are implemented to analyze and predict financial uncertainty based on these selected features.

The research findings show that "firm risk," "ownership structure," and "earnings smoothing" are particularly effective in the development of financial uncertainty. Additionally, the results indicate that based on the mean square error evaluation criterion, both the random forest and neural network algorithms show high prediction accuracy. The present study highlights the potential of machine learning techniques to enhance the prediction of financial information uncertainty and provides investors and researchers with invaluable insights. This research aims to contribute to the development of more accurate intelligent prediction models and facilitate effective risk management and strategic decision-making in a more complicated financial perspective.

Theoretical foundations and hypothesis development

According to the definition presented by Bekaert and Harvey [1], financial information uncertainty is described as "the degree of unpredictability" of future financial returns, which can be influenced by macroeconomic factors, market conditions, and investors' sentiments. According to the authors, financial information uncertainty is defined as the unpredictability of future financial returns, which is influenced by various factors, including market fluctuations, economic conditions, and regulatory changes.

Baker et al. [2], in their research, considered financial uncertainty as "uncertainty about future economic conditions and financial markets, which can affect consumer behavior and corporate investment." Dornov et al. [3] defined financial information uncertainty as a "lack of clarity about future market conditions, which leads to increased volatility in stock prices and affects corporate investment decisions."

Marfou and Adlzadeh [4] have used three criteria, i.e., corporate idiosyncratic volatility (IVOL), dispersion in corporate earnings forecasts, and corporate earnings forecast error, to measure corporate financial information uncertainty.

In the past, some investigations have been conducted on predicting financial information uncertainty using machine learning and neural network models. Some of these studies have been carried out with a special focus on feature selection techniques. Financial information uncertainty, characterized by unpredictable fluctuations in economic factors and firm-specific risks, has been widely studied as a critical element in financial forecasting and decision-making (Lautenbacher, 2021; Bloom, 2009) [5, 6]. A few studies use machine learning models, such as support vector regression (SVR) and random forest algorithms, to improve the forecasted accuracy of financial information uncertainty by identifying the critical factors causing such uncertainty (Jin et al., 2019; Graham et al., 2005) [7, 8]. Feature selection techniques reduce model complexity while maintaining essential forecasting capabilities, thus improving model efficiency and interpretability (Tai et al., 2017) [9].

In particular, artificial neural networks (ANNs) have shown promising results in capturing nonlinear relationships between financial variables. This review highlights the influential variables in financial information uncertainty models, including firm risk, ownership structure, and earnings management practices, all of which consistently affect financial reporting outcomes under different economic conditions (Kim et al., 2021) [10]. By combining the findings of these studies, this review establishes a

foundation for the proposed model and emphasizes the relevance of machine learning-based feature selection and ANN techniques for predicting financial information uncertainty.

Research Hypotheses or Questions

Research Hypotheses:
H1: The use of machine learning and neural networks enables highly accurate prediction of financial uncertainty.
H2: Machine learning is effective in selecting important features leading to financial uncertainty.
Research Questions:
Q1: What are the most influential factors in anticipating financial information uncertainty?
Q2: How does feature selection affect the performance of machine learning models in predicting financial information uncertainty?

Research Methodology

The statistical population of the present study is the dataset that includes 114 companies listed on the Tehran Stock Exchange during the 2018-2022 period.
The research uses an inductive methodology in which research background and theoretical sources were collected through library research, papers, and the Internet. Then, by identifying the required variables, the dataset was received from the stock exchange website and organized in the form of a dataset. The considered dataset was analyzed using machine learning algorithms such as support vector machine, random forest, and neural network in the Jupyter Notebook of Python. From the utilized dataset, 12 main variables were selected based on research background, and by applying reductions and feature selection using the support vector machine algorithm, the most important variables were selected, and the uncertainty level of financial information was predicted using the random forest and neural network algorithms based on research questions and hypotheses.

Research variables and how to calculate them

a) Dependent variable:

The target variable or dependent variable in this research is the uncertainty of financial information. The uncertainty of financial information is considered to be the idiosyncratic volatility (IVOL) in accordance with the Iranian capital market data and the results of earlier research.
The corporate idiosyncratic volatility (IVOL) is obtained as the standard deviation of a company's abnormal return. The return on investment represents the benefits obtained from that investment, and investors seek investment opportunities that maximize their return on capital. The main factor that every investor pays particular attention to in their decisions is the return on investment [11].
Deviations in the prediction of stock returns are called abnormal returns. In fact, abnormal returns are the difference between the actual return and the expected return (the mathematical expectation of stock returns).

$AR_{it} = R_{it} - E(R_{it})$
$R_{it} = (P_1 + D - P_0)/P_0$
$E(R_{it}) = R_f + \beta_i(R_{mt} - R_f)$
AR_{it}= Abnormal return of stock i at time t
R_{it}= Actual return of stock i at time t
$E(R_{it})$= Expected return of stock i at time t
P_0 = Stock price at the beginning of the year t
P_1 = Stock price at the end of year t
$R_{mt} = (TEDPIX_t - TEDPIX_{t-1})/TEDPIX_t$
$TEDPIX_t$ = Total stock price index of Tehran Stock Exchange at the end of the year
$TEDPIX_{t-1}$= Total stock price index of Tehran Stock Exchange at the beginning of the year
R_{mt}= Market return
R_f= Risk-free rate of return, which is typically considered equivalent to the interest rate on bonds or bank deposits.
$\beta_i = \frac{COV\ R_{it}R_{mt}}{Var\ R_{mt}}$ *[4]*

b) Independent variables:

The set of independent variables, or as they are called in AI, the features are as follows:
1. Information Asymmetry:
The relationship between information asymmetry and financial information uncertainty is an important field of study in finance, especially in understanding market behavior and decision-making processes.
Information asymmetry arises when one party to a transaction has more or better information than the other, which may lead to an imbalance in power and decision-making, especially in financial markets where insiders may have access to critical information that investors do not have access to.
Financial information uncertainty refers to the unpredictability of the financial performance or health of a business entity. This uncertainty can arise from incomplete or ambiguous information that makes it difficult for stakeholders to evaluate the risks accurately.
Asymmetric information increases financial uncertainty for investors and other stakeholders. It becomes challenging to make informed investment decisions when they do not have complete information about a company's operations, financial health, or future prospects [12].

Information asymmetry can lead to market inefficiencies, where asset prices do not reflect their actual value due to unequal access to information. This mispricing contributes to investors' higher uncertainty about the true value of their investments [13].

In information asymmetry scenarios, wrong choices may occur, where those with less information are at a disadvantage and may make unfavorable investment choices. This situation intensifies uncertainty about the quality of available investments [14].

Recent studies have shown that the increasing complexity of financial instruments and the rapid development of technology intensify information asymmetry and uncertainty. For instance, the emergence of fintech and algorithmic trading has created new layers of complexity that can obscure information and lead to increased uncertainty for traditional investors (Zhang et al., 2023) [15].

This study uses the effective bid-ask spread model, which is the dominant measure of transaction costs in financial markets, to measure information asymmetry.

For a given transaction, the effective bid-ask spread is calculated as below:

$$S = 2D\frac{Ask-Bid}{Bid} \times 100 \quad [16]$$

where *Ask* stands for the best asking price for the purchase of the company i's stock at time t, and *Bid* is the best asking price for the sale of company i's stock at time t.

2. Accounting Conservatism

Baso's (1997) model measures earnings accounting conservatism by showing the strength of the correlation between returns and earnings when predicting future bad news (i.e., losses). If earnings, through accruals, contain information about future losses, then the relationship with returns is stronger because stock returns also incorporate information about future losses.

However, due to the conservatism principle, earnings reflect only future losses and not future profits. In accounting practices, anticipated losses are recorded in the next period, even if they have not yet been realized. For example, future bad debts are estimated and recorded before they are realized. On the other hand, any earnings or forecasted earnings for future periods are not recorded unless they are realized. Sales revenues are only recorded when the goods are delivered to the buyer. Therefore, any forecasted growth in sales is not reflected in the sales figure.

Thus, accounting profit reflects the impact of current performance and bad news (expected losses) on future performance. Stock returns reflect expectations about future losses and future profits. Therefore, returns reflect the impact of current performance and future performance, whether bad (losses) or good (profits). Accordingly, the relationship between profit and returns is stronger when there is significant bad news (losses) in the future, and it becomes weaker when there is significant good news (profits) in the future. Based on the above argument, Basu (1997) measured

conservatism using the following regression model [17]:

$$EPS_{it}/_{P_{it-1}} = \alpha + \beta_1 NEG_{it} + \beta_2 RET_{it} + \beta_3 (NEG * RET)_{it} + \varepsilon_{it}$$

where EPS_{it} stands for earnings per share in year t, P_{it-1} represents the stock price at the beginning of year t, NEG_{it} is an indicator variable that is 1 if the stock return is negative in the previous period and zero if the stock return is positive in the previous period, and RET_{it} is the annual rate of return on the company's stock.

3. Earnings Smoothing

Smoothing of reported earnings means management's efforts to deliberately reduce earnings fluctuations to an extent that is considered reasonable and acceptable based on accounting and management policies. Earnings smoothing leads to reduced uncertainty of financial information [18]. Accordingly, earnings smoothing is the ratio of the standard deviation of a company's cash flows to the standard deviation of earnings, which is calculated as follows:

$$\text{TASmoothing} = \frac{\text{std(CFO)}}{\text{std}(NIBE)}$$

The larger the result of this fraction, the more profit smoothing has taken place.

where *CFO* stands for the operating cash flows, and NIBE represents the net income before extraordinary items.

4. Market to Book (M/B) Ratio

The ratio of market value to book value (M/B) has a significant positive effect on risk reduction and stock returns. A high book-to-market ratio indicates that the market perception of the company's value is still low and can signal an excellent investment opportunity for investors [19].

This ratio is calculated by dividing the market value at the end of the period by the book value of stockholders' equity at the end of the period. Also, to calculate the market value, the company's latest stock price is multiplied by the number of authorized shares [20].

5. Auditor's reputation

The auditor's ability to reduce the premium associated with financial information uncertainty is

directly related to the capital market's perception of the auditor's reputation for quality and independence. The higher the auditor's reputation, the more confident investors are in the signal presented by a company's financial statements. This means that investors who condition the stock value on the auditor's reputation will revise the price downward when that reputation unexpectedly deteriorates. Firms suffering from financial information uncertainty benefit most from a reliable audit and, as a result, suffer the most when the credibility of past audits is lower than expected. Hence, an unexpected decline in the auditor's reputation has a more negative impact on the stocks of client firms with higher financial information uncertainty [21].

In the present study, the auditor's reputation is regarded as a binary variable. If the company's audit was conducted by the Audit Organization or the Mofid Rahbar auditing firm, the value of the variable will be one; otherwise, its value will be zero.

6. Leverage

Estimates indicate that with an increase in any form of financial information uncertainty, companies reduce their short-term leverage levels [22].

Financial leverage is the ratio of the book value of debts to the market value of the company's assets. The market value of assets is calculated by adding the market value of shareholder's equity to total debts [23]. In general, one can calculate financial leverage as follows:

$$FD = \frac{TD}{TA}$$

where TD is total debts, TA represents total assets, and FD is the financial leverage.

7. Liquidity

Uncertainty of financial information plays a mitigating role in shareholder-debtor conflicts over investment policy. In addition, as the level of information uncertainty increases, the costs of information uncertainty borne by shareholders increase significantly [24].

Joushi et al. [25] estimate liquidity as the ratio of trading volume multiplied by the closing price divided by the prices range from high to low for the entire trading day on a logarithmic scale. The authors use the price at the end of the trading period because it is the most accurate valuation of the stock at that point in time. They use the traded volume for the day, assuming that trading volume is a linear function of time. This research has used the same method to calculate liquidity.

8. Firm Size

Strecken et al. [26] studied the relationship between firm size and financial information uncertainty and showed that there is a significant relationship between managers' decision-making for a firm's growth and information uncertainty. Thus, the present study included firm size as an independent variable in the calculation of financial information uncertainty. Based on the definition of Dang et al. [27], this study employs the natural logarithm of total assets to calculate the firm size.

9. Audit Firm Rotation

According to the research work of Amjadian and Daneshian (2010) [28], audit firm rotation affects financial information. If the company's independent auditor has changed during the last year, this variable assumes a value of 1; otherwise, it has a value of 0 [29].

In this regard, note 2 of Article 10 of the Guidelines for Trusted Audit Firms of the Stock Exchange Organization requires that audit firms and audit partners of any of the above legal entities are not allowed to accept the position of independent auditor and statutory auditor of the aforementioned company again after 4 years. In addition, in the event of a partner leaving the previous firm, the partner responsible for the work in the previous 4 years cannot accept the said position as a partner in another audit firm. According to this guideline, changes in an audit firm are made in two ways: voluntary and mandatory. If the change of a company's independent auditor during the last year is voluntary, this variable assumes a value of 1, and if it is mandatory or remains unchanged, it is zero.

10. Firm Risk

A company's risk-taking indicates its desire to seek higher profits and its willingness to pay to achieve these profits [30]. Corporate risk includes actions such as taking heavy loans, allocating a high percentage of resources to projects with uncertain outcomes, and entering unknown markets [31].

Based on the CAPM model, systematic risk or sensitivity factor is determined using the following equation:

$$\beta_i = \frac{Cov(R_{it}, R_{mt})}{Var\ (R_{mt})}$$ [32]

where $Cov(R_{it}, R_{mt})$ stands for the covariance of the stock return in market return, and $Var\ (R_{mt})$ represents the variance of market return.

11. Q Tobin

According to Moradzadeh Fard et al. (2013) [20], Q Tobin is one of the variables affecting the uncertainty of financial information. The authors calculate this ratio as follows:

$$Q = \frac{BVL + MC}{BVTA}$$

where BVL stands for the book value of liabilities at the end of the period, MC represents the market

capitalization at the end of the period, and BVTA is the book value of total assets at the end of that period.

12. Ownership Structure

According to the research of Kazemi et al. (2011) [33], there is a significant relationship between the ownership structure of companies and information asymmetry. Companies with a higher percentage of their shares held by legal shareholders, especially institutional shareholders, are better candidates for investment. This finding is consistent with the comparative advantage of institutional shareholders in collecting and processing information, such that as the percentage of institutional ownership increases, the provision of information by company managers to relevant individuals in the market will increase, and, in other words, information asymmetry will decrease.

In the present study, based on the research conducted by Mohammad Azadi et al. (2015) [34], the percentage of institutional ownership (INS) is used as the variable of ownership structure, which is the number of common stocks held by investment institutions or other commercial companies.

The method of testing hypotheses

This research first examines the desired dataset based on independent and dependent variables in terms of descriptive statistics. Then, preprocessing operations, such as cleaning, removing null and invalid data, removing outliers, data normalization, etc., are performed. Then, the data is entered into the support vector machine algorithm in a standard format, and from these 12 independent variables, the most important variables are selected. Finally, the random forest and neural network algorithms are applied to the reduced dataset. The problem under investigation is of the multivariate regression type, the formula of which is as follows:

$$\begin{aligned} &Financial\ Information\ Uncertainty_{i,t} \\ &= \beta_0 + \beta_1 Information\ Asymmetry_{i,t} \\ &+ \beta_2 Accounting\ Conservatism_{i,t} \\ &+ \beta_3 Smoothing_{i,t} + \beta_4 M/B_{i,t} \\ &+ \beta_5 Auditor\ Reputation_{i,t} + \beta_6 Leverage_{i,t} \\ &+ \beta_7 Liquidity_{i,t} + \beta_8 Firm\ Size_{i,t} \\ &+ \beta_9 Audit\ firm\ rotation_{i,t} + \beta_{10} Firm\ Risk_{i,t} \\ &+ \beta_{11} Q\ Tobin_{i,t} + \beta_{12} Ownership_{i,t} + \varepsilon_{i,t} \end{aligned}$$

Data Analysis

The considered dataset has 570 records or samples, 12 columns of the independent variable, and one column of the dependent variable "financial information uncertainty," with values ranging from 0 to 1, where a value of 1 means 100% information certainty and a value of 0 means complete uncertainty about the company in question, based on the data sample.

Understanding the Problem

This research aims to anticipate the level of financial uncertainty of Iranian listed companies based on artificial intelligence. The problem is of the regression type, and in this type of problem, the aim is to find the minimum levels of criteria such as Mean Squared Error (MSE). Based on the Mean Squared Error criterion, the deviation of the predicted target variable value from that of the actual target variable is calculated, and the smaller this difference, the more accurate the model is.

Understanding the Data

For the present study, data from 114 companies listed on the Tehran Stock Exchange during the 2018-2022 period were collected in the form of a dataset. Then, based on earlier related investigations, 12 main and important variables were selected as independent variables, and the idiosyncratic volatility variable was selected as the target or dependent variable. Table 1 shows the descriptive statistics of the dataset.

Table 1. Descriptive statistics of independent variables

Independent variable / Statistical indicator	Auditor's reputation	Audit Firm Rotation	Firm size	Leverage	Accounting Conservatism	Corporate risk
Number of samples	570	570	570	570	570	570
Mean	0.256140	0.333333	14.518315	0.574462	-1.084954e-07	-0.146078
Standard deviation	0.436884	0.471405	1.532782	0.223160	2.260245e-01	0.452814
Mean	0.000000	0.000000	11.128923	0.012733	-2.861401e+00	-2.040393
First quartile	0.000000	0.000000	13.615587	0.422459	-5.694493e-02	-0.384155
Second quartile	0.000000	0.000000	14.279342	0.587243	5.760465e-03	0.180394

Independent variable / Statistical indicator	Auditor's reputation	Audit Firm Rotation	Firm size	Leverage	Accounting Conservatism	Corporate risk
Third quartile	1.000000	1.000000	15.177275	0.721807	7.736226e-02	0.075187
Maximum	1.000000	1.000000	20.108439	1.551980	8.715298e-01	1.755496
Independent variable / Statistical indicator	Market to Book Ratio	Tobin's Q ratio	Information asymmetry	Earnings Smoothing	Stock Liquidity	Ownership Structure
Number of samples	570	570	570	570	570	570
Mean	2.613979	1.629138	0.396761	1.556399	12.797213	81.238825
Standard deviation	6.555894	0.680391	0.300029	1.839379	1.904071	13.445848
Mean	-49.703903	0.630983	0.001691	0.003444	6.480045	15.300000
First quartile	1.424336	1.192349	0.171650	0.659827	11.713957	73.015000
Second quartile	2.074133	1.413997	0.343384	1.054011	12.617569	84.680000
Third quartile	3.242889	1.846530	0.552785	1.616072	13.632471	92.740000
Maximum	121.509643	4.846253	1.728705	13.324054	19.653017	99.000000

Next, we examine the statistical distribution of the data for each independent variable graphically. Figure 1 illustrates the statistical distribution of the independent variables.

Next, the data is checked to discard null and invalid values. It is evident that we have invalid values in some columns, which are fixed using the nearest neighbor replacement method.

Given that the different range of the variables, it is necessary to normalize the data. Since normalization is sensitive to outliers, the data must be checked for the presence of outliers. Table 1 clearly shows that we have outliers in some variables. Therefore, first the outliers are removed from the dataset, and then the data is normalized from zero to 1.

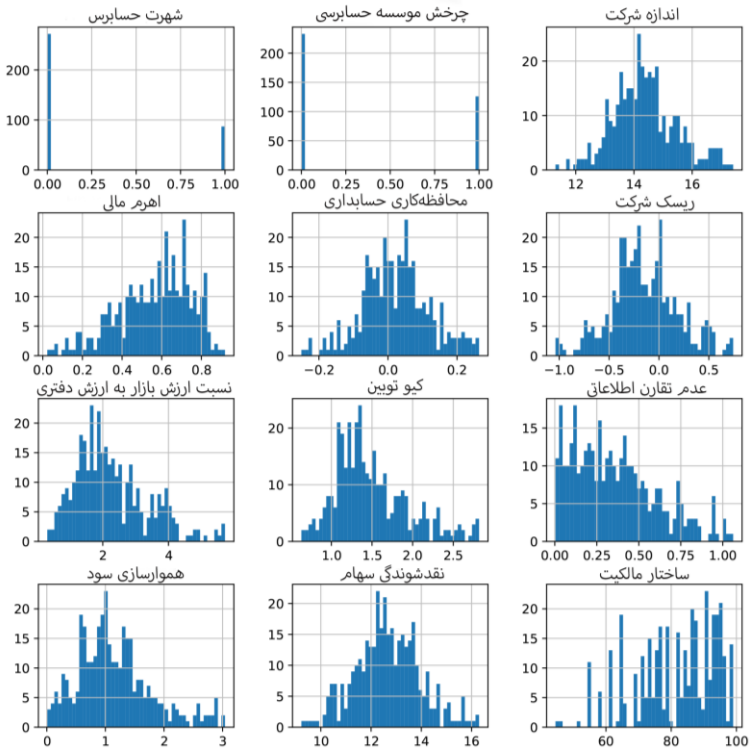

Figure 1. Statistical distribution of independent variables

Correlation measures the strength of the linear relationship between two variables. The correlation between two variables can be negative or positive and ranges between -1 and +1. Correlation values close to -1 or +1 indicate a strong linear relationship between the variables. Correlation is often used to study relationships between variables – here, independent variables or features [35]. Figure 2 depicts the correlation between features in the heatmap. As is clear from the heatmap, there is a strong correlation between some pairs of variables with a correlation coefficient close to 1. For example, the correlation coefficient between firm size and stock liquidity is 0.7, indicating a high positive correlation between the two variables. The correlation coefficient between Tobin's Q and the market-to-book ratio is also 0.84, indicating a high positive correlation between the variables.

Figure 2. Heatmap of correlations between features

Algorithms

Support Vector Machine (SVM) Algorithm

The regression problem is a generalization of the classification problem in which the model returns an output with a continuous value. In other words, a regression model estimates a multivariate function with a continuous value. SVMs solve binary classification problems by formulating them as convex optimization problems. The optimization problem involves finding the maximum margin separating the hyperplane while correctly classifying as many training points as possible. SVMs represent this optimal hyperplane with support vectors. The generalization of SVM to SVR is accomplished by introducing an e-sensitive region around the function called the e-tube. This tube reformulates the optimization problem to find a tube that best approximates the function with continuous values while balancing the model complexity and prediction error. In particular, SVR is formulated as an optimization problem by defining a loss function to minimize and find the flattest tube that contains most of the training samples. Hence, a multiobjective function is constructed from the loss function and the geometric properties of the tube [36].

Then, using appropriate numerical optimization algorithms, the convex optimization with a unique solution is solved. The hyperplane is represented in terms of support vectors with the training samples residing outside the tube boundary. In SVM, the

support vectors in SVR are the most influential samples that affect the tube shape, and it is assumed that the training and test data are independently and identically distributed, drawn from the same fixed but unknown probability distribution function [36]. Figure 3 is a representation of support vector regression in machine learning.

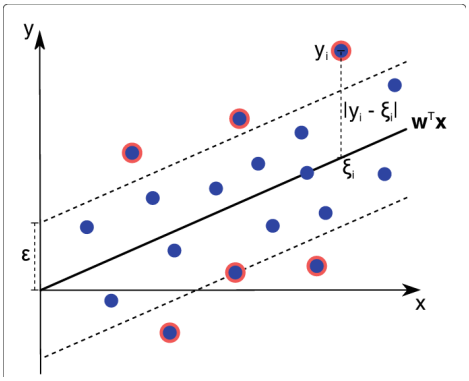

Figure 3.

Support Vector Regression (SVR). An image of an SVR regression function denoted by w T x. The insensitive tube around the function is shown as a gray tube. ξ i = w T x i is the predicted target value x i, and y i represents the actual target value. The support vectors are shown with a red border

Random Forest Algorithm

A random forest is a classifier consisting of a set of tree-structured classifiers {h(x, k), k =1, … } in which {k} are identically distributed independent random vectors, and each tree gives a single vote for the most popular class in the input x {Breiman, 2001 #42}.

The widespread popularity of the random forest algorithm is due to its user-friendly nature and adaptability, providing it with the potential to effectively solve classification and regression problems. The strength of this algorithm lies in its capability of handling complex datasets and reducing overfitting, which makes it a valuable tool for various predictive tasks in machine learning. One of the most important features of the random forest algorithm is the fact that it can handle datasets containing continuous variables, such as the regression mode, and categorical variables, such as the classification items. In this paper, the random forest is implemented on the regression problem. Figure 4 depicts the function of the random forest algorithm.

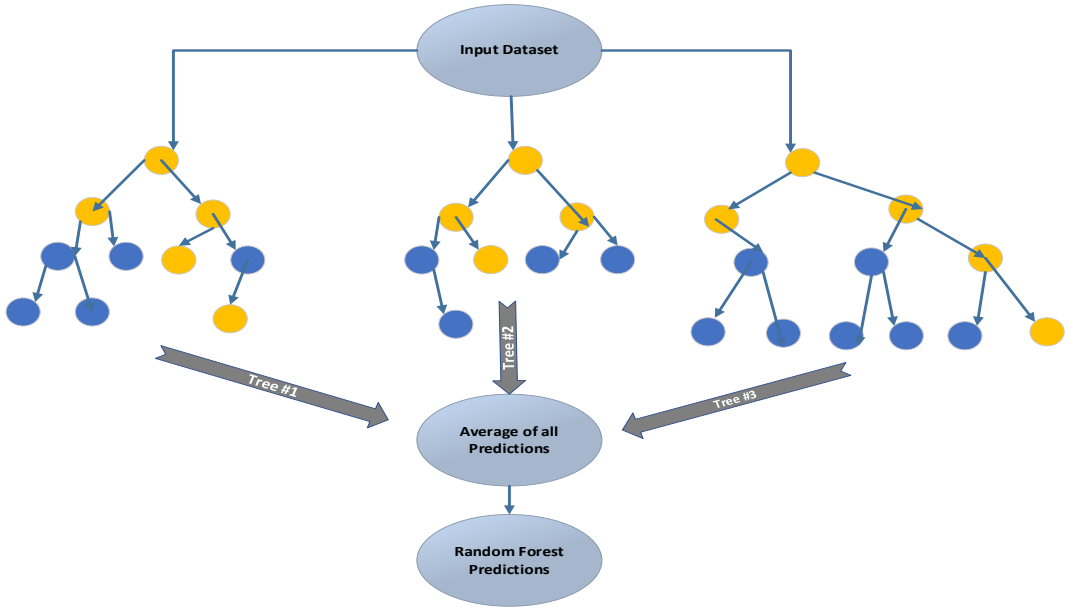

Figure 4. Schematic of the Random Forest Algorithm

Research Findings

Selecting the Best Features

To find the best features and reduce the dimensionality of the problem, a vector machine regression algorithm based on permutation feature importance is used. Permutation feature importance is a model evaluation technique that determines the contribution of each feature to the statistical performance of a fitted model for a tabular dataset. This technique is particularly useful for nonlinear or opaque estimators and involves randomly permuting the values of a single feature and observing the resulting decline in the model score {Breiman, 2001 #42}. By breaking the relationship between the feature and the target, the degree to which the model relies on the feature is determined. One of the key advantages of permutation feature importance is that it is model-agnostic, meaning that it can be applied to any fitted estimator. In addition, it can be computed multiple times with different feature permutations, which provides a further measure of the variance in the significance of the estimated features for a particular trained model. Each variable in the algorithm fitting is assigned a decimal number, and the larger this number is, the more important the variable is in finding the target variable [37]. For this purpose, after applying the support vector machine regression algorithm to the data of 12 main independent variables, five variables of firm risk, ownership structure, earnings smoothing, auditor's reputation, and audit firm rotation had the greatest impact on the target variable of information uncertainty, respectively (Fig. 5).

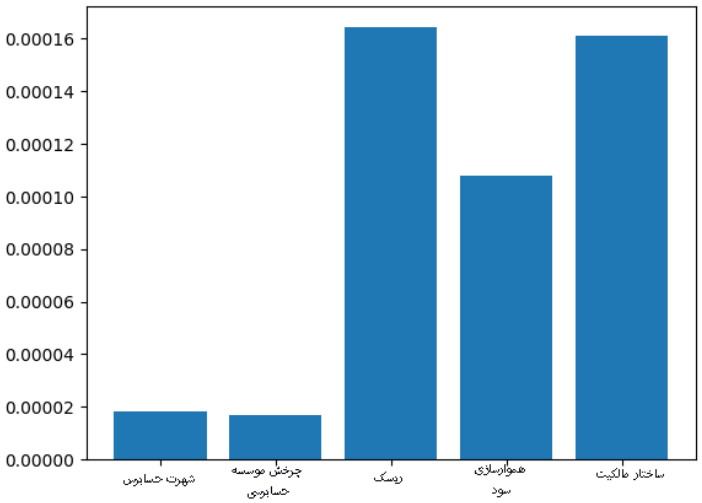

Figure 5. Selection of five important features based on their contribution to the variable of information uncertainty

Thus, the problem dimensions are reduced to five features, and the random forest and neural network algorithms are applied to these five features.

Modeling

In this stage, the random forest and neural network algorithms are trained and tested on the data. First, the data is divided into 67% training data and 33% test data. Then, the random forest and neural network algorithms are applied to these data, respectively, and the results are compared and evaluated. The evaluation criteria are the same as the regression evaluation criteria, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

Random Forest Model

The random forest model has various hyperparameters, and the optimal selection of these parameters, including the maximum depth of the tree, the number of estimators, the maximum features, the minimum samples leaf, the minimum samples split, etc., can greatly affect the performance of the algorithm. For this purpose, a greedy feature selection method called GridSearchCV is used. The output of the optimal parameters is as follows:

{'max_depth': 5, 'max_features': 5, 'min_samples_leaf': 2, 'min_samples_split': 3, 'n_estimators': 10}

RandomForestRegressor(max_depth=5, max_features=5, min_samples_leaf=2, min_samples_split=3)

By modifying the algorithm based on these parameters, the random forest algorithm has obtained

the following results on the training and testing data (Table 2).

Table 2 Results of the random forest algorithm

Criteria	Training data	Test data
MSE	0.0098	0.0176
MAE	0.0748	0.1021
RMSE	0.0098	0.1326

Neural Network Model

The neural network algorithm of the research consists of an initial layer with five input neurons, the first Dense hidden layer with 11 neurons, and a Softmax activation function, which is a normalized exponential function that transforms a vector of real numbers K into a probability distribution of K for the possible outcomes. The second Dense hidden layer has five neurons and a ReLU activation function. The rectified linear activation function (ReLU) is a piecewise linear function that outputs the input directly if it is positive; otherwise, its output is zero. It has become the default activation function for many types of neural networks because the model used by this function is easier to train and often achieves better performance. The last layer is the output layer, with a single neuron that produces a numerical prediction value of the dependent variable.

The optimization function used by the neural network is RMSprop, which is a gradient-based optimization technique. The gradients of very complex functions, such as neural networks, tend to vanish or explode as the data is propagated through the function. RMSprop uses the moving average of the squared gradients to normalize the gradient. This normalization balances the step size (momentum), reducing the step for large gradients to avoid explosion and increasing the step for small gradients to prevent vanishing {Elshamy, 2023 #43}.

The neural network has 50 epochs and a batch size of 2. The evaluation criterion is the Mean Squared Error (MSE). The learning rate of the algorithm is 0.0005.

The architecture of the neural network model is shown in the figure.

With the above settings, the constructed neural network algorithm produced the following results on the training and test data (Table 3).

Table 3. Results of the Neural Network Algorithm

Criterion	Training data	Test data
MSE	0.0157	0.0167
MAE	0.0975	0.1011
RMSE	0.1252	0.1292

Figure 6 shows the mean squared error graph of the employed neural network algorithm. The horizontal axis of the graph represents the number of epochs of applying the algorithm to the dataset, and the vertical axis represents the mean squared error as a criterion for evaluating the algorithm. It is clear from the figure that in the training phase (blue curve), as the number of epochs of the implementation of the algorithm increases, the MSE error decreases, meaning that the algorithm performs better with increasing the number of epochs. During the testing phase, we also observe the performance of the algorithm, i.e., the decrease in mean squared error by increasing the epochs of the implementation of the algorithm.

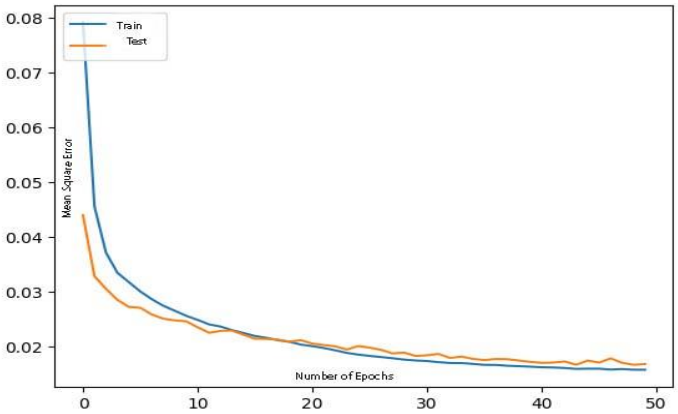

Figure 4. The graph of mean squared error of neural network algorithm for 50 epochs

Comparison of results of algorithms

The comparison of the performance of the two regression algorithms, random forest, and neural network, is presented in Table 4. The results of Table 4 show that the random forest algorithm has shown better performance in the training stage based on the three evaluation criteria of MSE, MAE, and RMSE. This means that the random forest algorithm learns better on the training data. Basically, in linear regression problems, random forest is a good choice, because in this type of problem, the essence of the problem can be better implemented using decision trees. On the other hand, the performance of the neural network is better in the testing phase because the multilayered structure of the neural network has more maneuverability on the unobserved data. In terms of duration, the neural network spent less time as well. That is, the learning speed of the neural network is higher if the neurons, layers, and network settings are selected appropriately. Thus, in total, the neural network is a better choice for the problem of predicting uncertainty in financial information.

Table 4. Comparison of the performance of the two proposed algorithms: Random Forest and Neural Network.

Algorithm	Time	Training Phase	Testing Phase	Criterion
Random Forest	15.79	0.0098	0.0176	MSE
Neural Network	12.45	0.0157	0.0167	
Random Forest		0.0748	0.1021	MAE
Neural Network		0.0975	0.1011	
Random Forest		0.0098	0.1326	RMSE
Neural Network		0.1252	0.1292	

Conclusions and Recommendations

Financial information uncertainty is a critical concept for various stakeholders, including investors, corporate managers, and researchers, which can contribute to investment decisions, risk management, market efficiency, corporate governance, regulatory compliance, and research and development.

Intelligent methods, especially those based on machine learning and artificial intelligence, can significantly enhance stakeholders' ability to anticipate financial information uncertainty. Intelligent methods can process large volumes of data from various sources (financial statements, market trends, economic indicators) more efficiently than traditional methods. Advanced algorithms can identify and prioritize relevant features that affect financial information uncertainty, enabling stakeholders to focus on the most important variables affecting their investment decisions. Machine learning models excel at recognizing patterns and complex relationships within data that may not be observable via conventional analyses. This may lead to a more accurate prediction of future financial information uncertainty based on historical trends. Intelligent methods can provide stakeholders with predictive analysis tools that provide insights into potential future financial information uncertainty and enable proactive risk management strategies.

In summary, financial information uncertainty is critical for various stakeholders because it affects investment behavior, risk management, market efficiency, and corporate governance. Intelligent methods improve the prediction of such uncertainties by using advanced data analysis techniques, enhancing accuracy, and enabling better decision-making processes.

In the present study, a financial information uncertainty prediction model was developed using machine learning and neural network techniques to identify and select key features that significantly affect financial information uncertainty. Using support vector regression, five important features were identified – firm risk, ownership structure, earnings smoothing, auditor's reputation, and audit firm rotation – that play a key role in determining the uncertainty of financial information.

Subsequently, two advanced algorithms, random forest and neural networks with two hidden layers, were used to evaluate the predictive power of these selected features. The research findings showed that the neural network model outperformed the random forest in terms of accuracy and reliability in predicting financial information uncertainty. This finding emphasizes the potential of deep learning approaches in capturing complex patterns and relationships in financial data.

This research has significant implications for financial sector practitioners and stakeholders. By understanding the factors contributing to financial information uncertainty, organizations can better manage risks, enhance decision-making processes, and improve transparency in financial reporting. In addition, this study emphasizes the importance of feature selection in building robust prediction models and shows that the scrutiny of relevant variables can lead to improved prediction capabilities.

By examining additional features, combining more diverse datasets, or using other machine learning and deep learning techniques to further enhance financial uncertainty prediction, future research can enrich the findings of the present study. The proposed model can also be applied to financial datasets in other industries.

All in all, this work contributes to the growing literature on the prediction of financial information uncertainty and reflects the efficiency of intelligent methods in addressing complex financial challenges.

Resources

G. H. Bekaert, Campbell R, "Foreign Speculators and Emerging Equity Markets," National Bureau of Economic Research Working Paper Series, vol. 6312, p. 55, 2000, doi: https://doi.org/10.1111/0022-1082.00220.

S. R. Baker, N. Bloom, and S. J. Davis" ,Measuring Economic Policy Uncertainty*," The Quarterly Journal of Economics, vol. 131, no. 4, pp. 1593-1636, 2016, doi: 10.1093/qje/qjw024.

A. Durnev, R. Morck, and B. Yeung, "Value-Enhancing Capital Budgeting and Firm-specific Stock Return Variation ",The Journal of Finance, vol. 59, no. 1, pp. 65-105, 2004, doi: https://doi.org/10.1111/j.1540-6261.2004.00627.x.

M. Marfoo and M. Adlzadeh, "Information Uncertainty and Investors' Underreaction," Empirical Accounting Research,vol. 4, no. 3, pp. 169-177 ,٢٠١٤ ,doi: 10.22051/jera.2015.1889.

S. Lautenbacher, "Subjective Uncertainty, Expectations, and Firm Behavior," ifo Institute - Leibniz Institute for Economic Research at the University of Munich, 2021. [Online]. Available: https://EconPapers.repec.org/RePEc:ces:ifowps:_349

N. Bloom, "The Impact of Uncertainty Shocks," Econometrica, vol. 77, no. 3, pp. 623-685, 2009, doi: https://doi.org/10.3982/ECTA6248.

Y. Luo and C. Zhang, "Economic policy uncertainty and stock price crash risk," Research in International Business and Finance, vol. 51, p. 101112, 2020/01/01/ 2020, doi: https://doi.org/10.1016/j.ribaf.2019.101112.

J. Yiqiang Jin, K. Kanagaretnam, Y. Liu, and G. J. Lobo, "Economic policy uncertainty and bank earnings opacity," Journal of Accounting and Public Policy, vol. 38, no. 3, pp. 199-218, 2019/05/01/ 2019, doi: https://doi.org/10.1016/j.jaccpubpol.2019.05.002.

F. E. H. Tay and L. Cao, "Application of support vector machines in financial time series forecasting," Omega, vol. 2 ,٩no. 4, pp. 309-317, 2001/08/01/ 2001, doi: https://doi.org/10.1016/S0305-0483(01)00026-3.

H. Kim and Y. Yasuda, "Economic policy uncertainty and earnings management: Evidence from Japan," Journal of Financial Stability, vol. 56, p. 100925, 2021/10 ,٢٠٢١ /٠١/doi: https://doi.org/10.1016/j.jfs.2021.100925.

Kh. Z. Rezaei F., Masoudian S. M., & Kiani S. , "Investigating the Relationship Between Market Psychological Pressure with Abnormal Stock Returns and Accumulated Abnormal Returns of Companies," Scientific Journal of Modern Research Approaches to Management and Accounting,vol.9,no. 3, p. 16, 1398, doi: https://majournal.ir/index.php/ma/article/view/250.

H. K. Baker, Kumar, S., Singh, A., ""Investor Behavior: The Role of Information Asymmetry".
",.٦٠-٤٥ ,(١)٢٣Journal of Behavioral Finance, vol. 1, no. 23, p. 16.٢٠٢٢ ,

E. F. Fama, ""Efficient Capital Markets: A Review of Theory and Empirical Work," Journal of Finance, vol. 1, no. 76, p. 33, 2021.

D. Ghosh, Kaur, R., "Adverse Selection in Financial Markets: A Review.," Journal of Financial Stability, vol ,٥٢ .p. 116, 2023.

Y. Zhang, Chen, L., Wang, J. , "Fintech and Market Dynamics: Implications for Information Asymmetry," Finance Research Letters, vol. 51, p. 8, 2023.

D. Ardia, E. Guidotti, and T. A. Kroencke, "Efficient estimation of bid–ask spreads from open, high, low, and close prices," Journal of Financial Economics, vol. 161, p. 103916, 2024/11/01/ 2024, doi: https://doi.org/10.1016/j.jfineco.2024.103916.

A. a. S. Gregoriou, Len "Does the Basu Model Really Measure the Conservatism of Earnings? ," p. 25, 2007, doi: http://dx.doi.org/10.2139/ssrn.965486.

G. Soleimany Amiri and R. Hamzi, "The effect of Income smoothing on firm's information uncertainty, stock returns and cost of equity," Accounting and Auditing Review, vol. 18, no. 64, pp. 91-112, 2011. [Online]. Available: https://acctgrev.ut.ac.ir/article_23958_0a1d580c0975860eebb80c3dff75b9c9.pdf.

B. Nugroho, "The Effect of Book to Market Ratio, Profitability, and Investment on Stock Return," International Journal of Economics and Management Studies, vol. 7, pp. 102-

107, 06/25 2020, doi: 10.14445/23939125/IJEMS-V7I6P114.

M. Moradzadeh Fard, M. Adlzadeh, M. Farajzadeh, and S. Azimi, "Information Uncertainty, Information Asymmetry, and Growth Opportunities," Empirical Studies in Financial Accounting,vol,١٠.no.39,pp.125-145,2012.[Online].

https://qjma.atu.ac.ir/article_1133_905da8f4d5172177feac9efa18814c7b.pdf.

D. M. Autore, R. S. Billingsley, and M. I. Schneller, "Information uncertainty and auditor reputation," Journal of Banking & Finance, vol. 33, no. 2, pp. 183-192, 2009/02/01/ 2009, doi: https://doi.org/10.1016/j.jbankfin.2008.07.011.

C. F. BAUM, A. STEPHAN, and O. TALAVERA, "THE EFFECTS OF UNCERTAINTY ON THE LEVERAGE OF NONFINANCIAL FIRMS," Economic Inquiry, vol. 47, no ,٢ . pp. 216-225, 2009, doi: https://doi.org/10.1111/j.1465-7295.2007.00116.x.

S. Shams, M. Yahyazadehfar, and S. Abbaszadeh, "The Study of Relationship between Financial Leverage and Information Asymmetry in Listed Firms of Tehran Stock Exchange," Accounting and Auditing Research, vol. 9, no. 34, pp. 104-121, 2017. [Online]. Available: https://www.iaaaar.com/article_98770_640d3bdaea6c0d2faae29f1e1167cb74.pdf.

H.-S. Kim, "Investment decisions and debt financing under information uncertainty," The North American Journal of Economics and Finance, vol. 52, p. 101106, 2020/04/01/ 2020, doi: https://doi.org/10.1016/j.najef.2019.101106.

N. Joshi, "A Practical Approach to Liquidity Calculation (Digest Summary)," CFA Institute Journal Review, 2024 ,doi: doi.org/10.2469/dig.v45.n2.12.

E. Sterken, R. Lensink, and P. Steen, "Uncertainty and Growth of the Firm," Small Business Economics, vol. 24, pp. 381-391, 02/01 2005, doi: 10.1007/s11187-005-7121-z.

C. Dang, Z. Li, and C. Yang, "Measuring firm size in empirical corporate finance," Journal of Banking & Finance, vol. 86, pp. 159-176, 2018/01/01/ 2018, doi: https://doi.org/10.1016/j.jbankfin.2017.09.006.

D. M. Amjadian Y., "Investigating the Effect of Audit Tenure on Information Uncertainty," Scientific Journal of Modern Research Approaches to Management and Accounting,vol. 13, no. 4, p. 17, 1399, doi: https://majournal.ir/index.php/ma/article/view/441.

A. R. M. Safarzadeh, "Representation of Financial Statements and Factors Affecting It in Different Industries" (in eng), audit knowledge, Research vol ,١٤ .no. 55, pp. 175-194, 2014. [Online]. Available: http://danesh.dmk.ir/article-1-755-fa.html.

W. Luo, Y. Wang, and X. Zhang, "Monetary policy uncertainty and firm risk-taking," Pacific-Basin Finance Journal, vol. 71, p. 101695, 2022/02/01/ 2022, doi: https://doi.org/10.1016/j.pacfin.2021.101695.

D. Lyon, G. Lumpkin, and G. Dess, "Enhancing Entrepreneurial Orientation Research: Operationalizing and Measuring a Key Strategic Decision Making Process," Journal of Management - J MANAGE, vol. 26, pp. 1055-1085, 09/01 2000, doi: 10.1177/014920630002600503.

M. Salehi, R. Hejazi, G. Talebnia, and A. Amiri, "A scheme of CAPM models considering distress risk and Firms life cycle," Financial Management Strategy, vol. 7, no. 1, pp. 95-122, 2019, doi : /١٠,٢٢٠٥١jfm.2018.20342.1668.

H. Kazemi and A. Mohammadnejad, "Investigating the Relationship between Institutional Ownership, Information Asymmetry and Financial Performance of Firms," (in Fa), Financial Accounting Knowledge Journal,vol. 1, no. 2, pp. 110-, 2012. [Online]. Available: https://www.magiran.com/paper.١٠٠٠٧٥٠/

M. Z. V. M. Esfandiar, "Institutional Ownership, Ownership Concentration, and Audit Fees," Audit Knowledge, vol. 15, no. 60, p.13, 1394. [Online].
Available: https://sid.ir/paper/516583/fa.

P. Tadikamalla and A. Durcikova, "Statistical Models (Non-Optimization Models)," in Encyclopedia of Information Systems, H. Bidgoli Ed. New York: Elsevier, 2003, pp. 227-241.

M. Awad and R. Khanna, "Support Vector Regression," 2015, pp. 67-80.

L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp ,٢٠٠١ ٠١/١٠/٢٠٠١ ,٣٢-٥ . doi: 10.1023/A:1010933404324.