

Stock daily return prediction using expanded features and feature selection

Hakan GÜNDÜZ*, Zehra ÇATALTEPE, Yusuf YASLAN

Department of Computer Engineering, Faculty of Computer and Informatics Engineering,
İstanbul Technical University, İstanbul, Turkey

Received: 22.04.2017

Accepted/Published Online: 03.10.2017

Final Version: 03.12.2017

Abstract: Stock market prediction is a very noisy problem and the use of any additional information to increase accuracy is necessary. In this paper, for the stock daily return prediction problem, the set of features is expanded to include indicators not only for the stock to be predicted itself but also a set of other stocks and currencies. Afterwards, different feature selection and classification methods are utilized for prediction. The daily close returns of the 3 most traded stocks (GARAN, THYAO, and ISCTR) in Borsa İstanbul (BIST) are predicted using indicators computed on those stocks, indicators for all the other stocks listed in the BIST100 index, and indicators on the dollar-gold prices. Twenty-five different indicators on daily stock prices are computed to form feature vectors for each trading day. These feature vectors are assigned class labels according to the daily close returns. Expanding the feature space with BIST100 stocks features results in a high dimensional feature space, with possibly noisy or irrelevant features. Therefore, feature selection methods are utilized to select the most informative features. In order to determine relevance scores of features, fast filter-based methods, gain ratio and relief, are used. Experiments are performed based on individual stock features, dollar-gold features (DG), BIST100 stock features (BIST100), and a combination of BIST100 and DG with and without feature selection. Using the gain ratio feature selection with a gradient boosting machine (GBM), the movements of GARAN stock were predicted with an accuracy of 0.599 and an F-measure of 0.614. For THYAO, the relief feature selection with the GBM gave an accuracy of 0.558, and for ISCTR, the gain ratio feature selection with logistic regression achieved an accuracy of 0.581. It was found that using BIST100 stock features boosts classification results for all stocks in terms of accuracy.

Key words: Stock market prediction, feature selection, Borsa İstanbul (BIST), feature combination, feature expansion

1. Introduction

Financial markets, especially emerging markets, are influenced by several factors such as global conditions, political situations, economic indicators (inflation rate, unemployment rate, etc.), company policies, and trader expectations. Relationships between these factors make the nature of stock markets complex, noisy, nonlinear, and dynamic [1]. All these factors make the prediction of the stock prices/directions a challenging task. It is important for traders and investors to make an accurate decision about their investment and predicting stock performance plays a key role in their strategies. There are several methods in stock market prediction and most of them use numerical and structured data such as technical indicators [2]. In the literature, technical analysis and fundamental analysis are generally used to predict the future stock price. The first one uses past stock prices to predict future prices (e.g., past daily, weekly, monthly stock prices), whereas the second one

*Correspondence: hakangunduz@itu.edu.tr

takes advantage of using data about the structure of the economy (e.g., inflation rates, exchange rates, interest rates, unemployment percentage). In recent financial studies, extraction of relevant information from financial data is performed with the help of machine learning algorithms. While artificial neural networks (ANNs) and support vector machines (SVMs) are the most commonly used algorithms, naive Bayes, logistic regression, and K-nearest neighbors are also still used due to their robustness and simplicity and also the ease of explanation of the decisions made by them [3–6]. Although ensemble methods have superior performance in tasks such as image processing, bioinformatics, natural language processing, and analytics, most studies in the financial domain do not utilize them. The proven performance of ensemble learning in other domains encourages researchers to use it in stock market prediction [7,8]. Especially in the presence of noisy or irrelevant features, feature selection methods help to improve classification accuracy and they have been utilized in a number of studies [4].

In this study, we aim to predict the daily stock price direction of the three most traded stocks, GARAN, THYAO, and ISCTR, in Borsa İstanbul (BIST). In order to do this, we use different types of classification algorithms such as the gradient boosting machine (GBM) and logistic regression (LR). Prices of individual stocks, prices of other BIST100 stocks, and gold and dollar prices are used as inputs into our prediction framework. Thus, the total number of features is 5860 for the expanded feature space. This would yield a high dimensional feature vector for each stock and, in order to handle the curse of dimensionality problem, we use feature selection methods, gain ratio and relief, and conduct experiments with different numbers of features. Experimental results are evaluated based on two performance measures: accuracy and F-measure.

The main contributions of our study can be summarized as follows. First of all, we use a feature expansion mechanism that allows us to use not only the indicators for the instrument but all other available instruments. To the best of our knowledge, this is the first study that uses features computed on all other stocks to predict the movement of a particular stock in BIST. Although previously individual stock features have been used for the prediction of a stock in numerous studies [9,10] and correlations between stock prices have been investigated [11], to the best of our knowledge, features of other stocks have not yet been used for prediction in the Turkish markets. Our second contribution is the use of different types of feature selection and classification methods together. The classification methods that we use include an ensemble-based method, GBM and LR. We compare performances of feature selection and classification methods using not only accuracy but also F-measure. Especially when data contain imbalanced classes, accuracy is not an appropriate performance measure. In this situation, classification performance is evaluated for each class using the F-measure and overall evaluation is performed by taking the average of the class-based F-measure (macroaveraged) rates [4].

The remainder of this paper is organized as follows: in the next section, we give an overview of the related work. In Section 3, we describe the dataset used in this study and provide information about the feature selection and the classification methods used. Section 4 gives details of the experimental results. Section 5 concludes the paper.

2. Background

In this section, we review previous studies on stock market prediction using machine learning and we also cover the related work on ensemble learning in the finance domain. In addition, we go through the BIST-related studies.

2.1. Machine learning for stock market prediction

Several machine learning algorithms have been employed for stock market prediction. Traditional algorithms such as decision tree (DT) [12], LR [13], naive Bayes (NB) [3,4], ANNs [14,15], and SVMs [16] have been shown to be effective for financial forecasting.

In the work of Martinez et al. [14], technical indicators were used for predicting daily maximum and minimum stock prices by training a multilayer perceptron (MLP) classifier. The classifier was used to decide on the most important indicators for prediction. Kayal [15] proposed a system for forecasting the foreign exchange market (FOREX) by using basic technical indicators, such as simple and exponential moving averages, RSI, and the standard deviation from several different periods. He used the MLP for the forecasting process and compared the classifier performance with random selection. Marković et al. [17] predicted the trend of the Belgrade stock exchange BELEX15 index with a SVM. Feature selection was used on technical and macroeconomic indicators and the best classification accuracy was obtained with the selected features. Weng et al. [18] combined stock market data with crowdsourced data obtained from Wikipedia and Google News to predict the daily stock movements. Besides the stock prices and technical indicators, they used Wikipedia page views, counts of published Google News articles, and technical indicators generated from both providers' data as features. Discriminative features were selected with the recursive feature elimination method and DT, ANN, and SVM classifiers were trained. The results show that the combination of market data and online sources increased the classification performance significantly and the best performance was obtained with the combination of market data, technical indicators, Wikipedia traffic, Google News counts, and generated features. Barak et al. [19] proposed a new approach to predict the stocks return and risks using 44 financial ratios and macroeconomic indicators. They developed a hybrid feature selection algorithm based on filter methods and function-based hierarchical clustering. They computed 12 different feature relevance scores for each feature and used the hierarchical divisive method for clustering features to find the most relevant features in risk and return prediction. They chose the features that fell in the first and second clusters and trained several DT-based models. The results showed that their proposed hybrid model finds relevant indicators for risk and return prediction and increases the classification performance. In [20], the daily direction of the S&P 500 Index ETF (SPY) return was predicted with 60 financial and economic indicators. Three feature reduction techniques, principal component analysis (PCA), fuzzy robust PCA, and kernel-based KPCA, were applied to the preprocessed dataset. Each reduction method generated 12 reduced datasets with different numbers of components and ANNs were trained with 36 reduced datasets. The results showed that combining the ANN with PCA results in a slightly higher accuracy rate than the other two combinations.

2.2. Ensemble learning for stock market prediction

Ensemble learning models have been employed instead of a single classifier in a number of stock market studies. The study by Patel et al. predicted the direction of the movement of Indian stock markets [8]. Their study compared classification performances of ANN, SVM, random forest (RF), and NB algorithms. They used open, high, low, and close prices of stocks to compute ten technical indicators with different numerical representations. Ten years of historical data from 2003 to 2012 were used in experiments and RF outperformed the other three prediction methods for overall performance. Cavalcante and Oliveira [21] compared the use of extreme learning machine and online sequential-extreme learning machine ensemble methods to make an intelligent trading system for generating automatic buy/sell signals. Ballings et al. [7] used ensemble methods to predict stock price movements. They evaluated the performances of ensemble models such as RF, Adaboost, and kernel

factory and they compared them with single classifier models ANN, LR, SVM, and K-nearest neighbor. The classification performance showed that ensemble methods outperformed the single classifier and RF provided the leading performance, followed by the SVM. Barak et al. [22] determined the number of classifiers to be fused by using clustering on the data samples and then combined the DT-based models, SVM, and ANN classifiers with different fusion methods. The fusion process increased the classification accuracy by about 3%. Dey et al. [23] predicted the trend of the stock market using extreme gradient boosting (XGBoost). Their proposed model was successful in long-term trend prediction and was also superior to traditional machine learning algorithms. Sosvilla-Rivero and Rodríguez [24] used a gradient boosting-based classification technique to inspect causality among the three important stock indices, S&P 500, FTSE 100, and Nikkei 225, during periods of unstable price changes. They found that the S&P 500 index contains valuable information to improve the prediction of returns for both itself and the others.

2.3. Borsa İstanbul studies

Several studies have utilized historical index prices on BIST as features for prediction. Bildirici and Ozgur [25] predicted the BIST100 index daily returns using different autoregressive models with a neural network. In order to predict the BIST direction, Kara et al. [26] used technical indicators as features. They used a SVM and ANN for classification. Both methods were quite successful in predicting the direction of the BIST100 Index, but the ANN showed a slightly better performance than the SVM. In our previous work [4], we proposed a prediction method that combines the analysis of news articles and stock prices to predict future market movements. We devised a feature selection method called balanced mutual information (BMI) to determine the more relevant features for prediction of the BIST100 index direction. A NB algorithm was used for prediction, using news articles and without incorporating any technical indicator. However, in this paper, we aim to analyze the effect of features obtained from other stocks in the BIST100 with and without feature selection and try to improve the classification accuracy of individual stock directions using ensemble models.

3. Materials and methods

3.1. Dataset

In this study we use the daily stock prices of the companies listed in the BIST100 index as our experimental dataset. Each stock has daily open, close, high, and low prices and volume for 931 trading days in between the years of 2011 and 2016. The dataset was partitioned into two parts, a training set (from January 2011 to December 2014) and testing set (from January 2015 to December 2015). Twenty-five technical indicators with different time lags (resulting in a total of 58 features) were used to compute the input features. We selected the technical indicators by considering previous works [15,17,27]. Details of the technical indicators we used are summarized in Table 1.

In order to be able to extract as many relevant features as possible, several types of technical indicators were computed. First, open, high, low, and close prices of examined stock were taken as features. Smoothing methods like MA, EMA, and TEMA are indicators that identify potential changes in price information. RDP and MOM indicators are effective at finding trends in series. Oscillator indicators such as K%, D% (moving average of K%), WILLR, BIAS, RSI, CCI, and PPO are good at identifying market situations like oversold or overbought. In order to evaluate the volatility of a stock, the ATR indicator can be applied to prices. Higher values of ATR show strong bidirectional movement that can be used in stop-loss decision. MACD and ULTOSC are indicators needed to be used together for providing reliable signals about the momentum of the market.

Table 1. Used technical indicators.

Indicator	Explanation	Indicator	Explanation
OP	Open price	MOM(x)	Momentum measures change in stock price over last x days
HI	High price	MACD(x,y)	x days moving average convergence and divergence
LO	Low price	TEMA(x)	Triple exponential moving average
CL	Close price	PPO(x,y)	Percentage price oscillator
ROC(x)	Rate of change	CCI(x)	Commodity channel index
ROCP(x)	Rate of change pct.	WILLR(x)	Larry William's R%
K%(x)	Stochastic K%	RSI(x)	Relative strength index
D% (x)	D% is the moving average of K%	ULTOSC(x,y,z)	Ultimate oscillator
BIAS(x)	x-days bias	ATR(x)	Average true range
MA(x)	x-days moving average	MEDPRICE(x)	Median price
EMA(x)	x-days exponential moving average	MIDPRICE(x)	Medium price
signL(x,y)	A signal line is also known as a trigger line	HH(x)	Highest price
		LL(x)	Lowest price

LL and HH show the support and resistance levels, respectively. Those indicators can indicate near-low and near-high price levels according to the trend of the market [27].

In order to perform our experiments, we selected the three most traded stocks during 2011 and 2016, named “GARANTI BANK (GARAN)”, “IS BANK (ISCTR)”, and “TURKISH AIRLINES (THYAO)”, from BIST100 and computed the selected indicators. After the indicator computation, we had a dataset with 931 instances and 58 features. Since we wanted to exploit the relationships between different BIST100 stocks [11], we wanted to utilize their features in the prediction process. We computed all technical indicators for all BIST100 stocks. Besides the features of all stocks, we also added some extra features using gold and dollar daily close prices for the same time interval. We computed smoothing (SMA, EMA, TEMA) and momentum (MOM, ROCP, ROCR100, RSI, PPO) indicators. Finally, we obtained 30 features for each commodity. After forming the feature vectors for each stock and each trading day, these vectors were grouped as individual stock features, dollar-gold (DG) features, BIST100 features, and the combination of BIST100 and DG.

Class labels indicate the movements of the stock prices and they are based on close prices of the stocks. Let $c(t)$ and $c(t-1)$ denote the close price for a stock on day t and the previous trading day $t-1$. The class label for the t th day, $r(t)$, is computed as:

$$r(t) = \begin{cases} 1, & c(t) > c(t-1) \\ -1, & \text{Otherwise} \end{cases} \quad (1)$$

After class labels are computed, they are assigned to their respective feature vectors.

3.2. Feature selection

Feature selection algorithms aim to remove noninformative, noisy, or redundant features from high dimensional data. Removal of such features usually results in not only better classifier accuracy but also faster classification and machine learning models that are easier to interpret. In our study, two feature selection methods, gain ratio and relief, were applied to our data for feature selection.

3.3. Gain ratio

Gain ratio is a filter-based (i.e. does not require classifier training) feature selection that is a modification of the information gain method. Gain ratio was first used in decision tree C4.5 by Quinlan [28]. In information gain selection, a feature that takes a large number of unique values usually has a higher value of relevance because such a feature has a better chance of correct classification. At the limit, we would be able to achieve perfect classification (and no generalization) using a feature that takes as many different values as the number of instances. Therefore, selection of this nonrelevant feature can cause overfitting in prediction. Gain ratio considers the number of values of a feature while doing selection. This problem is solved by normalizing the information gain value with the entropy of feature values. Information gain and *gain ratio* are defined as follows:

$$InfoGain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

$$GainRatio(S, A) \equiv \frac{InfoGain(S, A)}{SplitInformation(S, A)} \quad (3)$$

$$SplitInformation(S, A) \equiv - \sum_{i=1}^c \frac{|S_i|}{|S|} \log \frac{|S_i|}{|S|} \quad (4)$$

$$Entropy(S) = -p_{+1} \log(p_{+1}) - p_{-1} \log(p_{-1}) \quad (5)$$

In these equations, S shows all instances in the data. p_{-1} and p_{+1} are the proportions of -1 class instances and $+1$ class instances in S , respectively. S_i is a subset of S for which feature A has value v_i . When two features are compared, the one with the higher gain ratio score is taken to be more relevant in gain ratio feature selection.

3.3.1. Relief

The relief algorithm was proposed by Kira and Rendell [29] as an effective and basic method to evaluate relevance of features. The algorithm outputs the vector of relevance scores (weights) for features. The weight of a feature is updated in the relief algorithm iteratively as follows: a random instance is selected from the training data, and the nearest neighbor instance that is in the same class with the selected and the nearest neighbor instance that belongs to a different class are found. The feature values of these instances are compared and the weight of a feature is modified. If a change in a feature value results in a change in class value, this feature probably is an important feature for class discrimination and its weight will increase. On the other hand, if a change in feature value does not result in a change in class value, its weight will decrease.

This weight updating procedure is repeated for a set of random instances selected from the training data or for all instances in training data. In the last step, weight updates are averaged and the final weights are computed in the interval $[-1: +1]$.

3.4. Classifiers

We explain the details of the classifiers used in this subsection.

3.4.1. Logistic regression

The first classifier we used was LR. In our study we used LR to evaluate the relationship between binary class labels (-1 or +1) and multiple features (technical indicators). The LR model gave us the probability of the following trading day being decided as “Up”: +1 or “Down”: -1. In our case, we determined the threshold value as 0.5 and we assigned the class label as “Up” if the probability exceeded the threshold. LR estimates the probability of the output as follows:

$$p(y | x_1, \dots, x_n) = \frac{e^{w_0 + w_1 x_1 + w_2 x_2 + \dots + w_n x_n}}{1 + e^{w_0 + w_1 x_1 + w_2 x_2 + \dots + w_n x_n}} \quad (6)$$

In this equation, y is the class label and $p(y | x_1, \dots, x_n)$ is the prediction probability of the movement direction of a trading day based on n technical indicators (x_1, \dots, x_n) . The maximum likelihood method is used to optimize the model parameters (w_1, \dots, w_n) [30].

3.4.2. Gradient boosting machine

Boosting is a prediction method that combines a set of weak learners to build a single strong learner. The boosting method has been applied to many applications due to its successful performances [31,32].

When gradient boosting (GB) [33] is used, models are built in an additive manner by training weak learners. As a weak learner, LR, DT, and regression tree approaches can be chosen. The GB model is represented as follows:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in F \quad (7)$$

Here, K is the number of weak learners and f_k is a function (e.g., decision tree or a linear function) in the functional space F . x_i is an instance from the dataset, and \hat{y}_i is the prediction of a class label y_i . GB models are fitted iteratively. At each stage, a learner (such as a new decision tree) is fitted to recover the misclassifications of the existing model. Function fit is performed by minimizing the objective function with a gradient descent algorithm. Further details of the algorithm can be found in [34].

4. Results

In our experiments, the classifiers we used had different parameters that could take different values. In order to make sure the test performance values we computed were actual, we did not use the test set to decide on which parameter values to use. Instead, we partitioned the dataset into a training and test set and used cross-validation within the training set. We used a grid search procedure with 5-fold cross-validation. The parameter set with the best average cross-validation performance was then used for performing training on all the training data.

We first performed experiments using only the individual stock indicator features. Each stock had 58 features with 662 training and 267 test instances. The classification performances were evaluated using accuracy and macroaveraged F-measure metrics. The results are shown in Table 2. In this table, the LR classifier shows slightly better performance than GBM for GARAN and THYAO stocks in terms of accuracy. However, the ISCTR stock has a higher accuracy rate with the GBM classifier. For all predictions, GBM had better F-measure values than LR.

After computing the results with stock features, we used features computed with the dollar-gold currencies in the classification process. Due to having only close prices of these currencies, we computed technical indicators

Table 2. Classification results without feature selection.

Stock	Feature type	#Feats	GBM		LR	
			Acc	FM	Acc	FM
ISCTR	Stock	58	0.509	0.589	0.542	0.429
	Dollar-Gold	60	0.543	0.591	0.521	0.424
	BIST100	5800	0.479	0.512	0.512	0.438
	BIST100+DG	5860	0.513	0.549	0.505	0.452
THYAO	Stock	58	0.491	0.438	0.527	0.441
	Dollar-Gold	60	0.483	0.384	0.522	0.467
	BIST100	5800	0.506	0.459	0.521	0.487
	BIST100+DG	5860	0.509	0.433	0.519	0.497
GARAN	Stock	58	0.506	0.522	0.523	0.440
	Dollar-Gold	60	0.524	0.490	0.532	0.458
	BIST100	5800	0.521	0.402	0.529	0.409
	BIST100+DG	5860	0.513	0.449	0.519	0.445

that only use close prices. In this dataset, we had 60 features (30 features for dollar currency, 30 features for gold currency). In Table 2, the results of this classification as dollar-gold (DG) are shown. Compared to individual features of stocks, in DG we got better accuracy rates for the ISCTR and GARAN stocks. In order to examine the effects of other stock prices on each stock, we decided to use features of stocks listed in the BIST100 index. We computed 58 individual features for each BIST100 stock and we expanded our feature set size to 5800. We showed the classification performances of these features as BIST100 in Table 2. When we compared the results of BIST100 features with individual stock features, the accuracy rates of GARAN and THYAO were increased. Feature expanding continued with DG features and adding them to BIST100 features resulted in a feature space with 5860 dimension. We referred to these features as BIST 100+DG features in our classification process. Although the classification performance of BIST100+DG features was not so successful in prediction, it boosted the ISCTR accuracy rate. According to Table 2, in general LR resulted in better accuracy than GBM. In terms of F-measure, ISCTR and GARAN had better F-measure values when GBM was used.

Due to the high dimensionality of BIST100 and BIST100+DG data and since the number of instances was respectively smaller, a feature selection method was needed to select relevant features. In our study, we used filter-based feature selection methods, gain ratio and relief. In order to compare performances of different feature selection methods, we used each feature selection method to select the most relevant features from all of the training data. For individual stock and DG data, we selected 10, 20, 30, 40, and 50 features from 58 and 60 features, respectively. For BIST100 data, we selected 500 features from 5800 according to their relevance scores. We started to train the classifier with the most relevant 25 features and expanded them by 25 in each experiment. We repeated the same procedure for BIST100+DG data in addition to adding DG features in each feature expansion. Using the features selected by each feature selection method, we trained GBM and LR classifiers on training data and evaluated their performance on test data. In Table 3, the results obtained by gain ratio and relief feature selection for each stock are shown.

For the ISCTR stock, the gain ratio-based feature selection method achieved an accuracy value of 0.581 with an F-measure value of 0.545 using only 250 features (BIST100+DG) with the LR classifier. On the other hand, relief feature selection achieved an accuracy value of 0.569 with an F-measure value of 0.615 using 200 features (BIST100) with the GBM classifier. In the prediction of the THYAO stock, gain ratio feature selection showed a similar performance to relief in terms of accuracy value. While gain ratio achieved an accuracy value

Table 3. Classification results using gain ratio and relief (THYAO, ISCTR, GARAN).

		Feature type	Gain ratio			Relief		
			#Feats	Acc.	FM	#Feats	Acc.	FM
THYAO	GBM	Stock	50	0.543	0.509	10	0.494	0.639
		Dollar-Gold	20	0.539	0.592	10	0.543	0.629
		BIST100	175	0.554	0.481	25	0.558	0.648
		BIST100+DG	75	0.551	0.614	500	0.524	0.477
	LR	Stock	10	0.528	0.571	10	0.539	0.559
		Dollar-Gold	30	0.506	0.571	40	0.517	0.498
		BIST100	225	0.528	0.540	450	0.539	0.586
		BIST100+DG	375	0.528	0.484	450	0.551	0.574
ISCTR	GBM	Stock	50	0.517	0.569	10	0.554	0.577
		Dollar-Gold	40	0.547	0.575	50	0.551	0.589
		BIST100	375	0.539	0.565	200	0.569	0.615
		BIST100+DG	425	0.558	0.640	75	0.558	0.599
	LR	Stock	20	0.532	0.593	40	0.528	0.297
		Dollar-Gold	20	0.547	0.619	50	0.558	0.250
		BIST100	25	0.577	0.619	225	0.528	0.487
		BIST100+DG	250	0.581	0.548	125	0.524	0.525
GARAN	GBM	Stock	50	0.517	0.574	30	0.506	0.400
		Dollar-Gold	50	0.539	0.514	60	0.524	0.490
		BIST100	100	0.599	0.614	475	0.566	0.577
		BIST100+DG	75	0.577	0.598	100	0.532	0.582
	LR	Stock	40	0.554	0.568	40	0.547	0.631
		Dollar-Gold	30	0.566	0.449	40	0.517	0.617
		BIST100	250	0.573	0.640	125	0.532	0.590
		BIST100+DG	150	0.566	0.506	125	0.558	0.585

of 0.551 (F-measure value of 0.574) with 75 features (BIST100+DG), relief had an accuracy value of 0.558 (F-measure value of 0.648) with 25 features. Both classification results were obtained with the GBM classifier. Using DG features in classification also resulted in an accuracy value of 0.543. For GARAN prediction, the best accuracy value was achieved in the gain ratio method with the GBM classifier. Using only 100 features (BIST100) results in an accuracy value of 0.599 and an F-measure value of 0.614. In relief selection, an accuracy value of 0.566 was achieved with 475 features (BIST100) with the LR classifier.

To sum up the results presented above, in general, feature selection increased the prediction performance of all stocks and different feature types.

5. Discussion

In this study, we predicted the daily movements of three stocks in Borsa İstanbul using different types of technical indicators as features. Class labels were assigned using each stock's close price and the movements in stock prices were indicated by these labels. Using two types of feature selection methods, relevant features

out of the thousands of features (5860 features) were selected. A GBM and a LR classifier were trained and performance was measured in terms of accuracy and F-measure metrics.

Experiments were performed based on different types of feature sets (individual stock features, DG features, BIST100 stock features, and the combination of BIST100+DG). In order to perform feature selection, the relevance score of each feature with the class labels was computed with gain ratio and relief methods. Features were selected according to their relevance values, starting with the highest ones. Using the gain ratio feature selection with the GBM achieved a better accuracy value (0.599) than all feature selection-classifier combinations. The movements of THYAO were successfully predicted using the relief selection-GBM classifier pair with an accuracy of 0.558. In ISCTR prediction, the LR classifier with gain ratio selection achieved an accuracy rate of 0.581.

When the classification results were examined, it was found that feature selection improved the classification results according to results obtained without selection. We can conclude that using BIST100 stock features in prediction boosts the classification performance for all stocks in terms of accuracy.

The aspects of our work that are different from the recent stock market studies can be summarized as follows:

- The structure of our input data differs from the relevant stock market studies. Recent studies used technical indicators [8,18], financial ratios and company reports [19], and commodity prices and exchange rates [20] to predict the direction of stock markets. Fundamental information such as equity ratios, stock book values, and debt-to-asset ratios can also be used to estimate risk and return ratios [22]. Unlike these studies, in order to exploit the relationships between the stocks, we used the features of all stocks in the BIST100 together with individual stock features to predict the daily direction of the stocks. Using BIST100 stock features in prediction gives better performance than the individual stock features for all stocks.
- Another aspect that distinguishes our study from the other studies are the performance evaluation metrics. In recent studies, it has been observed that accuracy is often used as an evaluation metric [35]. Along with the accuracy metric, there are also studies using sensitivity, specificity, and area under the curve metrics [9,19]. Accuracy can be misleading if the datasets have an imbalanced class distribution. In our study, the F-measure metric was used in addition to accuracy to do reliable performance evaluation. Classification results were evaluated for each class using the F-measure and overall evaluation was performed by taking the average of class-based F-measure values.
- Since stock market data contain large amounts of noise, this should be taken into account when selecting the classification methods. SVM, ANN, and DT classifiers have been used frequently in recent stock market studies [18,19,22]. Due to the relatively small number of data instances in stock market data, these models easily overfit and their variances tend to increase. In order to address these problems in this study, an ensemble learning approach was introduced. Reduction in model variance and enhancement in generalization make ensemble learning more resistant to noise [36]. The use of ensemble learning in stock market studies has gained popularity in recent years. The survey study by Jadhav et al. [35] showed that the best classification result in 20 of the 44 examined stock market prediction studies was obtained by ensemble learning or hybrid methods. The success of the ensemble learning approach in stock market prediction has led us to use the GBM classifier. The GBM's success in different time series classification problems and its robustness to noisy data are the key reasons for our choice [37].

- As the use of other stock features in the classification extends our feature space, dimensionality reduction is required. In recent prediction studies, dimensionality reduction is performed by a filter approach that evaluates the relation between each feature and class labels [9,19]. In our study, as the number of dimensions increased up to 5660, filter-based and computationally efficient methods, relief and gain ratio, were chosen. Relief has robustness against incomplete and noisy data while gain ratio is successful against features with different numbers of values. The combination of these selection methods with different types of learners improves the classification results according to results obtained without selection.

In the future we plan to do further analysis of these dependencies and produce stock networks based not only on the labels but also on the features. We believe that this could allow us to make more causal dependency projections between classes.

References

- [1] Abu-Mostafa YS, Atiya AF. Introduction to financial forecasting. *Appl Intell* 1996; 6: 205-213.
- [2] Marcek D. Stock price forecasting: statistical, classical and fuzzy neural network approach. In: MDAI; 2004. pp. 41-48.
- [3] Atsalakis GS, Valavanis KP. Surveying stock market forecasting techniques—Part II: Soft computing methods. *Expert Syst Appl* 2009; 36: 5932-5941.
- [4] Gunduz H, Cataltepe Z. Borsa Istanbul (BIST) daily prediction using financial news and balanced feature selection. *Expert Syst Appl* 2015; 42: 9001-9011.
- [5] Guresen E, Kayakutlu G, Daim TU. Using artificial neural network models in stock market index prediction. *Expert Syst Appl* 2011; 38: 10389-10397.
- [6] Alkhatib K, Najadat H, Hmeidi I, Shatnawi MKA. Stock price prediction using k-nearest neighbor (kNN) algorithm. *International Journal of Business, Humanities and Technology* 2013; 3: 32-44.
- [7] Ballings M, Van den Poel D, Hespeels N, Gryp R. Evaluating multiple classifiers for stock price direction prediction. *Expert Syst Appl* 2015; 42: 7046-7056.
- [8] Patel J, Shah S, Thakkar P, Kotecha K. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Syst Appl* 2015; 42: 259-268.
- [9] Pehlivanli AC, Asikgil B, Gulay G. Indicator selection with committee decision of filter methods for stock market price trend in ISE. *Appl Soft Comput* 2016; 49: 792-800.
- [10] Khaidem L, Snehanshu S, Sudeepa RD. Predicting the direction of stock market prices using random forest. *arXiv* 2016; 1605.00003.
- [11] Ergur B, Cataltepe Z. Relational stock market network analysis. In: 21st Signal Processing and Communications Applications Conference (SIU); 2013.
- [12] Basti E, Kuzey C, Delen D. Analyzing initial public offerings' short-term performance using decision trees and SVMs. *Decis Support Syst* 2015; 73: 15-27.
- [13] Dutta A, Bandopadhyay G, Sengupta S. Prediction of stock performance in Indian stock market using logistic regression. *International Journal of Business and Information* 2015; 7: 105-136.
- [14] Martinez LC, da Hora DN, Palotti JRDM, Meira W, Pappa GL. From an artificial neural network to a stock market day-trading system: a case study on the BM&F BOVESPA. In: *IJCNN*; 2009. pp. 2006-2013.
- [15] Kayal A. A neural networks filtering mechanism for foreign exchange trading signals. In: *Intelligent Computing and Intelligent Systems*; 2010.
- [16] Kim K. Financial time series forecasting using support vector machines. *Neurocomputing* 2003; 55: 307-319.

- [17] Marković IP, Stojanović MB, Stanković JZ, Božić MM. Stock market trend prediction using support vector machines. *Facta Universitatis Series Automatic Control and Robotics* 2014; 13: 147-158.
- [18] Weng B, Ahmed MA, Megahed FM. Stock market one-day ahead movement prediction using disparate data sources. *Expert Syst Appl* 2017; 79: 153-163.
- [19] Barak S, Modarres M. Developing an approach to evaluate stocks by forecasting effective features with data mining methods. *Expert Syst Appl* 2015; 42: 1325-1339.
- [20] Zhong X, Enke D. Forecasting daily stock market return using dimensionality reduction. *Expert Syst Appl* 2017; 67: 126-139.
- [21] Cavalcante RC, Oliveira ALI. An autonomous trader agent for the stock market based on online sequential extreme learning machine ensemble. In: *IJCNN*; 2014.
- [22] Barak S, Arjmand A, Ortobelli S. Fusion of multiple diverse predictors in stock market. *Inform Sciences* 2017; 36: 90-102.
- [23] Dey S, Kumar Y, Saha S, Basak S. Forecasting to classification: predicting the direction of stock market price using Xtreme Gradient Boosting. Working paper. DOI: 10.13140/RG.2.2.15294.48968.
- [24] Sosvilla-Rivero S, Rodríguez PN. Linkages in international stock markets: evidence from a classification procedure. *Appl Econ* 2010; 42: 2081-2089.
- [25] Bildirici M, Ersin ÖÖ. Improving forecasts of GARCH family models with the artificial neural networks: an application to the daily returns in Istanbul Stock Exchange. *Expert Syst Appl* 2009; 36: 7355-7362.
- [26] Kara Y, Boyacioglu MA, Baykan ÖK. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Syst Appl* 2011; 38: 5311-5319.
- [27] Kumar D, Meghwani SS, Thakur M. Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets. *Journal of Computational Science* 2016; 17: 1-13.
- [28] Quinlan JR. Improved use of continuous attributes in C4.5. *J Artif Intell Res* 1996; 4: 77-90.
- [29] Kira K, Rendell LA. The feature selection problem: traditional methods and a new algorithm. In: *AAAI*; 1992. pp. 129-134.
- [30] Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression* (Vol. 398). New York, NY, USA: John Wiley & Sons, 2013.
- [31] Freund Y, Schapire RE. Experiments with a new boosting algorithm. In: *ICML* (Vol. 96); 1996. pp. 148-156.
- [32] Schapire RE, Freund Y. *Boosting: Foundations and Algorithms*. Cambridge, MA, USA: MIT Press, 2012.
- [33] Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2001; 29: 1189-1232.
- [34] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2016. pp. 785-794.
- [35] Jadhav S, He H, Jenkins KW. An academic review: applications of data mining techniques in finance industry. *International Journal of Soft Computing and Artificial Intelligence* 2017; 4: 79-95.
- [36] Fersini E, Messina E, Pozzi FA. Sentiment analysis: Bayesian ensemble learning. *Decis Support Syst* 2014; 68: 26-38.
- [37] Nagy GI, Barta G, Kazi S, Borbély G, Simon G. GEFCom2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach. *Int J Forecasting* 2016; 32: 1087-1093.