

Stock Market Prediction Using Machine Learning

Nidhi Poddar¹, Chaitrali Kale², Shravani Mahabdi³, Venkatesh Puri⁴, Prof. Sagar Dhanake⁵

Students, Department of Computer Engineering^{1,2,3,4}

Assistant Professor, Department of Computer Engineering⁵

Dr. D. Y. Patil Institute of Technology, Pune, Maharashtra, India

Abstract: *Predicting how the stock market will perform is one of the most difficult things to do. There are so many factors involved in the prediction – physical factors vs. psychological, rational and irrational behavior, etc. All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy. Accurate prediction of stock market returns is a very challenging task due to volatile and non-linear nature of the financial stock markets. With the introduction of artificial intelligence and increased computational capabilities, programmed methods of prediction have proved to be more efficient in predicting stock prices. In this work, Artificial Neural Network and Random Forest techniques have been utilized for predicting the next day closing price for five companies belonging to different sectors of operation. The financial data: Open, High, Low and Close prices of stock are used for creating new variables which are used as inputs to the model. The models are evaluated using standard strategic indicators.*

Keywords: Stock market, machine learning, python, random forest, etc.

I. INTRODUCTION

The financial market is a dynamic and composite system where people can buy and sell currencies, stocks, equities and derivatives over virtual platforms supported by brokers. The stock market allows investors to own shares of public companies through trading either by exchange or over the counter markets. This market has given investors the chance of gaining money and having a prosperous life through investing small initial amounts of money, low risk compared to the risk of opening new business or the need of high salary career. Stock markets are affected by many factors causing the uncertainty and high volatility in the market. Although humans can take orders and submit them to the market, automated trading systems (ATS) that are operated by the implementation of computer programs can perform better and with higher momentum in submitting orders than any human. However, to evaluate and control the performance of ATSs, the implementation of risk strategies and safety measures applied based on human judgements are required. Many factors are incorporated and considered when developing an ATS, for instance, trading strategy to be adopted, complex mathematical functions that reflect the state of a specific stock, machine learning algorithms that enable the prediction of the future stock value, and specific news related to the stock being analysed. Time-series prediction is a common technique widely used in many real-world applications such as weather forecasting and financial market prediction. It uses the continuous data in a period of time to predict the result in the next time unit.

Many timeseries prediction algorithms have shown their effectiveness in practice. The most common algorithms now are based on Recurrent Neural Networks (RNN), as well as its special type - Long-short Term Memory (LSTM) and Gated Recurrent Unit (GRU). Stock market is a typical area that presents time-series data and many researchers study on it and proposed various models. In this project, LSTM model is used to predict the stock price.

Motivation:

The use of machine learning and artificial intelligence techniques to predict the prices of the stock is an increasing trend. The stock market prediction process is filled with uncertainty and can be influenced by multiple factors. Therefore, the stock market plays an important role in business and finance. Financial organizations and merchants have made different exclusive models to attempt and beat the market for themselves or their customers, yet once in a while has anybody accomplished reliably higher-than-normal degrees of profitability. Nevertheless, the challenge of stock forecasting is so engaging in light of the fact that the improvement of only a couple of rate focuses can build benefit by a large number of dollars for these organizations.

**Project Scope:**

In this proposed system, we focus on predicting the stock values using machine learning algorithms like Random Forest and Support Vector Machines. We proposed the system “Stock market price prediction” we have predicted the stock market price using the random forest algorithm. In this proposed system, we will be able to train the machine from the various data points from the past to make a future prediction. We will take data from the previous year stocks to train the model. We majorly will use two machine-learning libraries to solve the problem. The first one is numpy, which was used to clean and manipulate the data, and getting it into a form ready for analysis. The other is scikit, which shall be used for real analysis and prediction. The data set we will use will be from the previous year’s stock markets to be collected from the public database available online, 80 % of data will be used to train the machine and the rest 20 % to test the data. The basic approach of the supervised learning model is to learn the patterns and relationships in the data from the training set and then reproduce them for the test data.

II. DESCRIPTION OF THE PROBLEM

Problem Definition: To develop windows based model using python for stock market prediction using ML.

Objectives:

- To develop a windows application using Python.
- To study ML algorithms.
- To use random forest algorithms.
- To detect stock market price using ML.

2.1 Methodology

Machine learning:

Machine learning is a subfield of artificial intelligence (AI). The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people. Although machine learning is a field within computer science, it differs from traditional computational approaches. In traditional computing, algorithms are sets of explicitly programmed instructions used by computers to calculate or problem solve. Machine learning algorithms instead allow for computers to train on data inputs and use statistical analysis in order to output values that fall within a specific range. Because of this, machine learning facilitates computers in building models from sample data in order to automate decision-making processes based on data inputs. Machine learning is a tool for turning information into knowledge. In the past 50 years, there has been an explosion of data. This mass of data is useless unless we analyse it and find the patterns hidden within. Machine learning techniques are used to automatically find the valuable underlying patterns within complex data that we would otherwise struggle to discover. The hidden patterns and knowledge about a problem can be used to predict future events and perform all kinds of complex decision making. Most of us are unaware that we already interact with Machine Learning every single day. Every time we Google something, listen to a song or even take a photo, Machine Learning is becoming part of the engine behind it, constantly learning and improving from every interaction. It’s also behind world-changing advances like detecting cancer, creating new drugs and self-driving cars.

III. LITERATURE REVIEW

Stock Price Prediction Model Based on RBF- SVM Algorithm

This paper studies the network model based on the improved support vector machine (SVM) algorithm to realize the correct judgment of the stock price trend.

CUDA parallel computing framework for stock market prediction using K- means clustering

This proposed model uses K-means machine learning technique to models the gathered stock data and predicts the upcoming stock values.

Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms

The system predicts the closing price of stock of a company

Forecasting stock market trends using support vector regression and perceptually important points



A support vector regression model is trained on this high-level data to make trading decisions based on predicted trading signal.

Stock Market Prediction using Text-based Machine Learning

The proposed algorithm to make these predictions with a recurrent neural network (RNN)

Sentimental Analysis of News Headlines for Stock Market

The methods that we have used for comparison are K-Mean clustering, Naive Bayes, and Support Vector Machine

Sentiment Analysis for Stock Price Prediction

Investigated the impact of sentiment expressed through Stock Twits on stock price prediction

Stock Price Prognosticator using Machine Learning Techniques

Aims to predict the prices of shares

Stock Market Prediction using Supervised Machine Learning Techniques: An Overview

The study discussed how supervised machine learning techniques are applied to improve accuracy of stock market predictions.

Predicting Stock Price Using Sentimental Analysis Through Twitter Data

The sentimental analysis is used on the tweets which are obtained by using the Twitter API.

IV. SYSTEM DESIGN AND FLOW

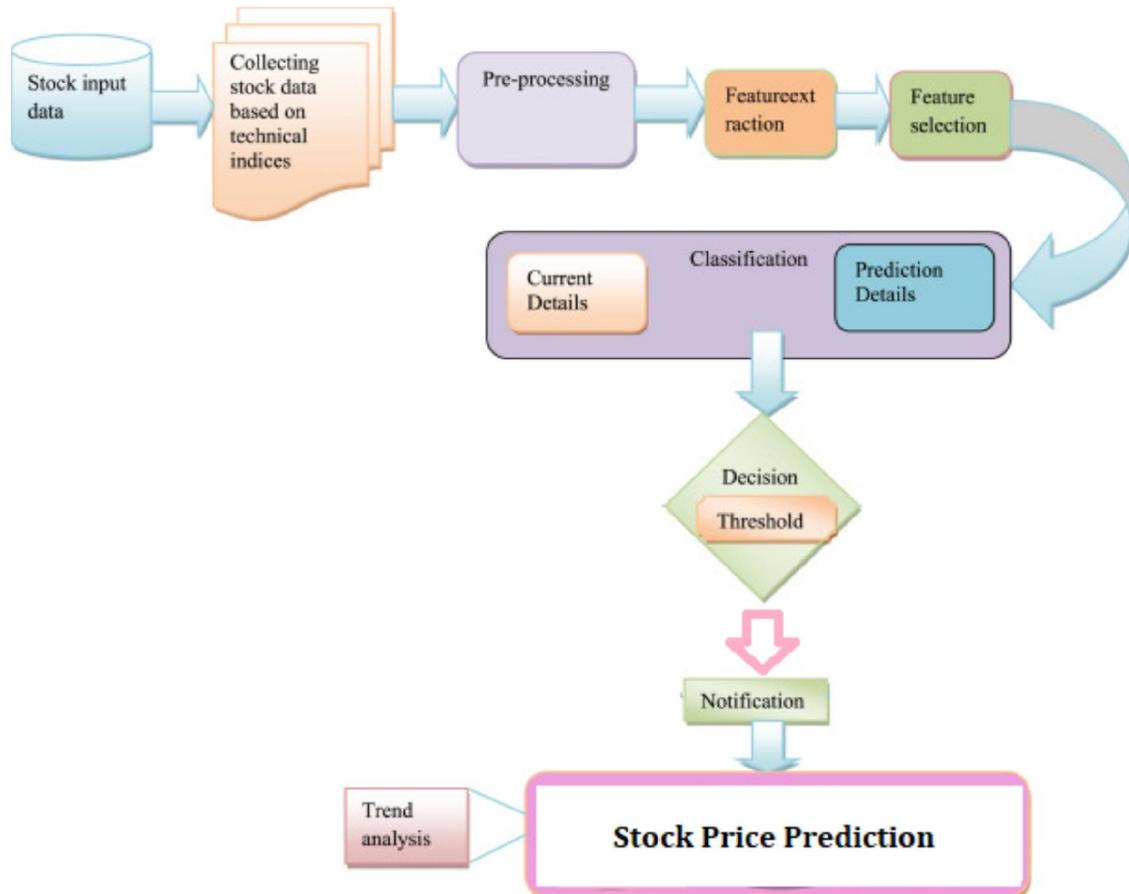


Figure: System Architecture

Stock market prediction is basically defined as trying to determine the stock value and offer a robust idea for the people to know and predict the market and the stock prices. It is generally presented using the quarterly financial ratio using the dataset. Thus, relying on a single dataset may not be sufficient for the prediction and can give a result which is inaccurate.



Hence, we are contemplating towards the study of machine learning with various datasets integration to predict the market and the stock trends.

MATHEMATICAL MODEL

Let S be Closed system defined as, $S = Ip, Op, Ss, Su, Fi, A$

To select the input from the system and perform various actions from the set of actions A so that Su state can be attained.

$S = Ip, Op, Ss, Su, Fi, A$

Where, IP1=Username, Password, Stock Name

Set of actions= $A = F1, F2, F3, F4$

Where F1= selection of random samples from a given dataset

F2= construct a decision tree for every sample

F3= Get the prediction result from every decision tree.

F4= vote for every predicted result.

S=Set of users

Ss=rest state, registration state, login state

Su- success state is successful analysis

Fi- failure state

Objects:

Input1: Ip1 = Username, Password

Input2: Ip2= Stock Name

Output1: Op1 = Decision Tree

Output2: Op2 = Voting

Output3: Op3 = Most voted prediction result as the final prediction result.

VI. PROJECT IMPLEMENTATION

Algorithm:

Random Forest:

Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

Working of Random Forest Algorithm

We can understand the working of Random Forest algorithm with the help of following

Step:

1. First, start with the selection of random samples from a given dataset.
2. Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.
3. In this step, voting will be performed for every predicted result.
4. At last, select the most voted prediction result as the final prediction result.

The following diagram will illustrate its working

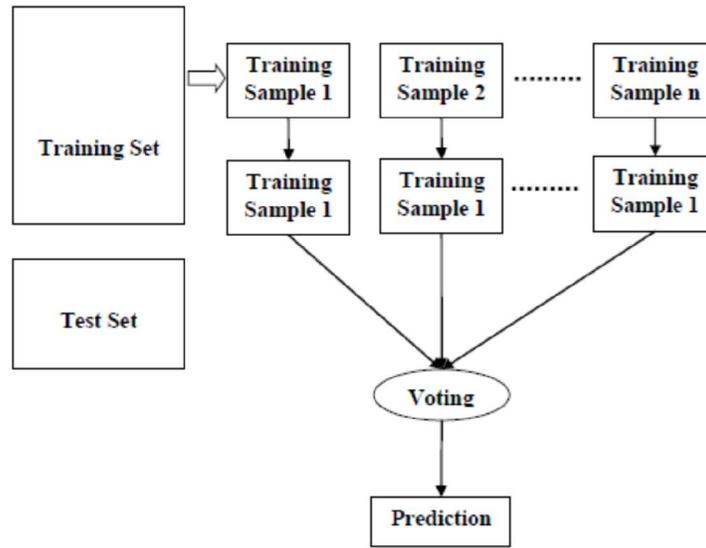


Fig: Random Forest Algorithm

VII. DISCUSSION AND RESULTS

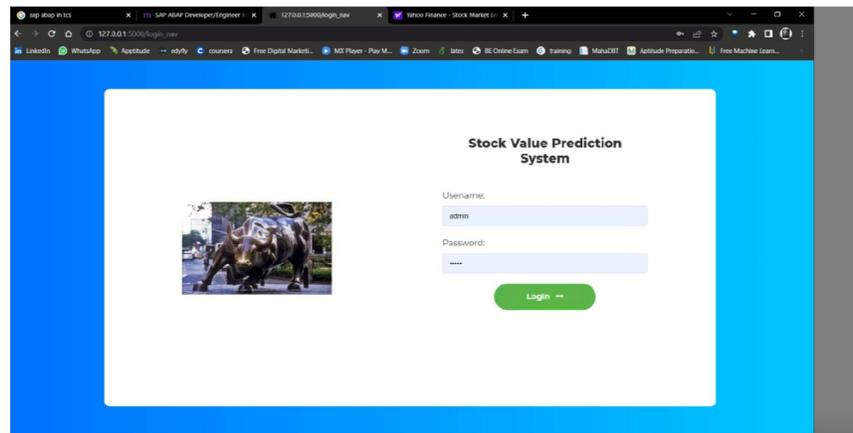


Fig. User Login

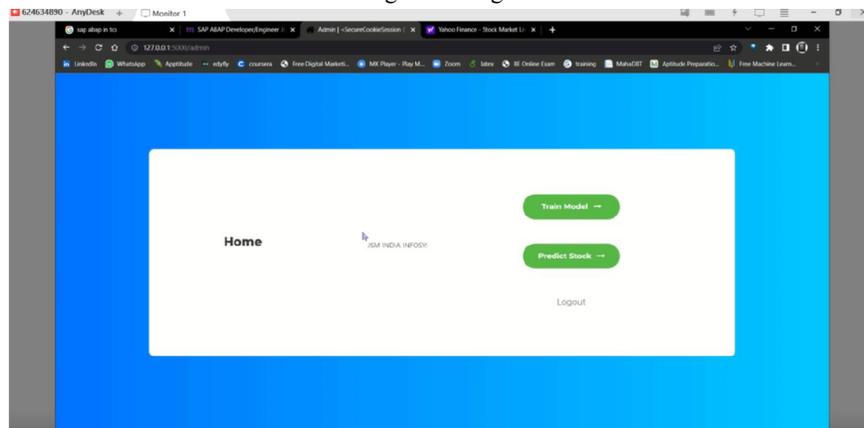


Fig: User Home

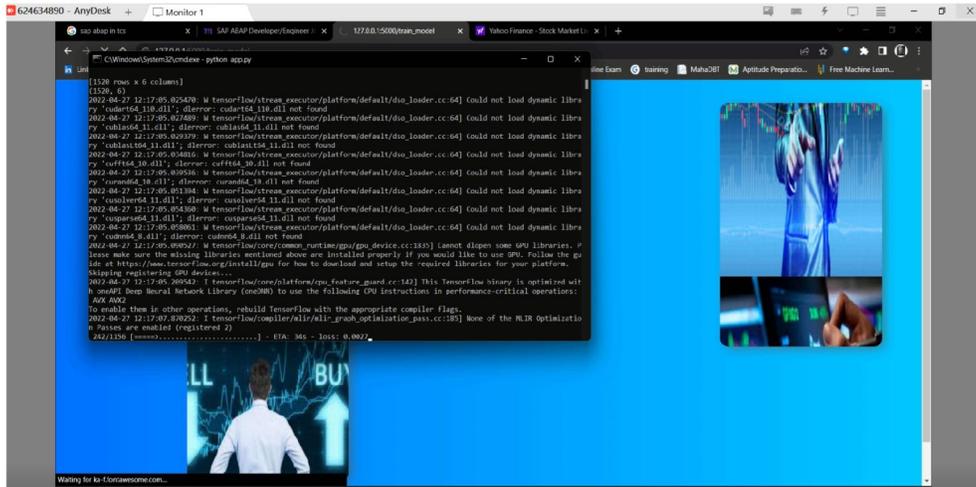


Fig. Train Model

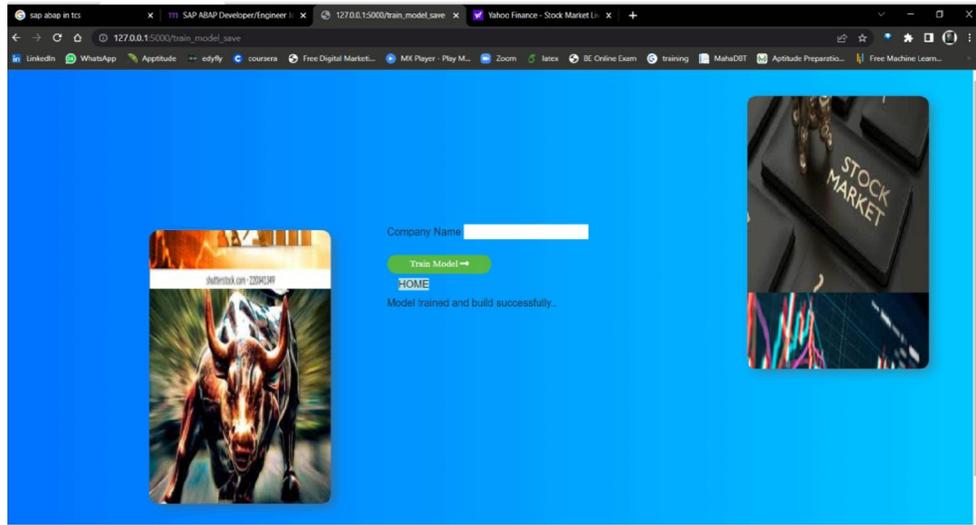


Fig. Model Trained

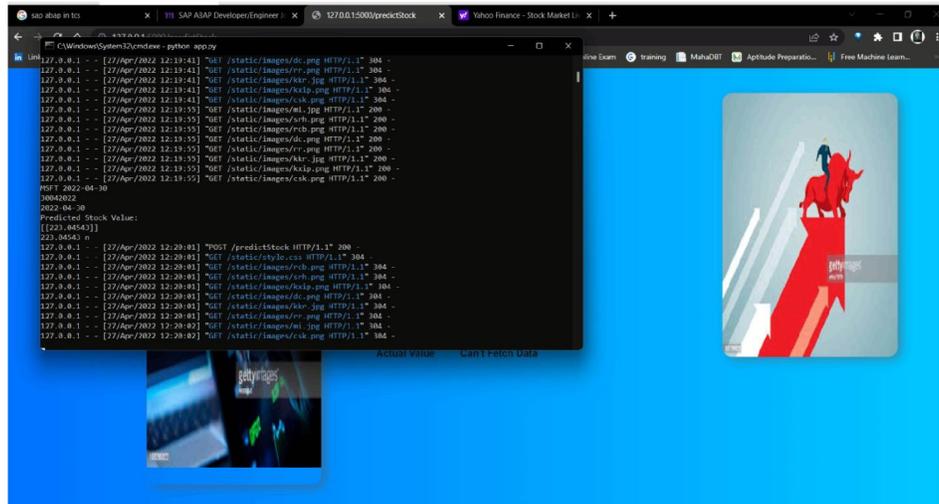


Fig. Prediction

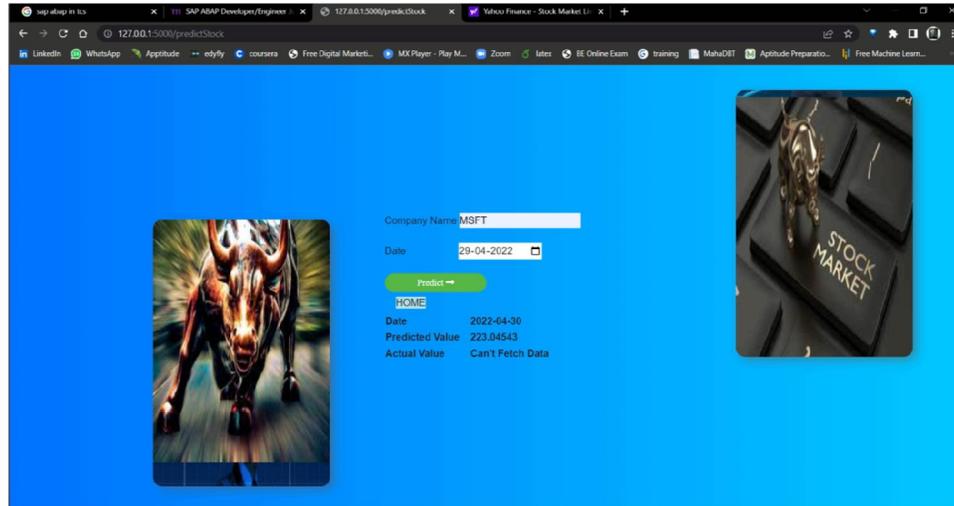


Fig. Prediction Result

VIII. APPLICATIONS

Applications:-

- Educational sector
- Government Organizations
- Banking Sector.
- Personal Use

IX. CONCLUSION & FUTURE WORK

Thus we have implemented a system for stock market prediction using ML. Programming language used will be python. We have used online data-set from yahoo finance for building model.

Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. The more the parameters are taken into account more will be the accuracy. The algorithms can also be applied for analyzing the contents of public comments and thus determine patterns/relationships between the customer and the corporate employee. The use of traditional algorithms and data mining techniques can also help predict the corporations performance structure as a whole.

ACKNOWLEDGEMENTS

The completion of our project brings with it a sense of satisfaction, but it is never complete without those people who made it possible and whose constant support has crowned our efforts with success. One cannot even imagine our completion of the project without guidance and neither can we succeed without acknowledging it. It is a great pleasure that we acknowledge the enormous assistance and excellent co-operation to us by the respected personalities.

REFERENCES

- [1]. Masoud, Najeb MH. () "The impact of stock market performance upon economic growth." International Journal of Economics and Financial Issues 3 (4) : 788–798.
- [2]. Murkute, Amod, and Tanuja Sarode. () "Forecasting market price of stock using artificial neural network." International Journal of Computer Applications 124 (12) : 11-15.
- [3]. Hur, Jung, Manoj Raj, and Yohanes E. Riyanto. () "Finance and trade: A cross-country empirical analysis on the impact of financial development and asset tangibility on international trade." World Development 34 (10) : 1728-1741.



- [4]. Li, Lei, Yabin Wu, Yihang Ou, Qi Li, Yanquan Zhou, and Daoxin Chen. (2017) "Research on machine learning algorithms and feature extraction for time series." IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC): 1- 5
- [5]. Seber, George AF and Lee, Alan J. () "Linear regression analysis." John Wiley & Sons 329
- [6]. Reichek, Nathaniel, and Richard B. Devereux. () "Reliable estimation of peak left ventricular systolic pressure by M-mode echo graphic determined end diastolic relative wall thickness: identification of severe valvular aortic stenosis in adult patients." American heart journal 103 (2) : 202-209.
- [7]. Chong, Terence Tai-Leung, and Wing-Kam Ng. () "Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30." Applied Economics Letters 15 (14) : 1111-1114.
- [8]. Zhang, G. Peter. (2003) "Time series forecasting using a hybrid ARIMA and neural network mode." Neuro computing 50 : 159-175.
- [9]. Suykens, Johan AK, and Joos Vandewalle. () "Least squares support vector machine classifiers." Neural processing letters 9 (3) : 293-300.
- [10]. Liaw, Andy, and Matthew Wiener. () "Classification and regression by Ran- dom Forest." R news 2 (3) : 18-22.