

Project Review

Notes

[Slide 1] INTRODUCTION

Hello everyone, today we're going to present our project and its details. Our project is about **Stock Market Prediction using Machine Learning** and we're doing this project under the guidance of our HOD, Dr. John Peter.

[Slide 2]

Stock Market is one of the many ways in which companies, businessmen and even common citizens earn money by investing in the share market. Stock prices are predicted at the end of the day to determine the future value of companies, and this price is called the closing stock price. The share market is extremely volatile and changes and is affected by many factors such as political events, general economic conditions, and traders' expectations. Because the stock market is extremely prone to change it becomes necessary to be able to predict the prices so that investors can have maximum profit in return.

[Slide 3]

To predict stock prices, we used statistical methods such as Time Series Forecasting and other systematic analysis that involved linear models. Due to the growing emergence of Artificial Intelligence, and due to its various applications, Machine Learning and its methods can be applied in order to predict the stock prices.

Before we move on to see how this can be achieved, let's briefly understand what Machine Learning is.

[Slide 4] Machine Learning

Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.

Let us consider the example of playing the game of checkers. If we were to create a computer game to play Checkers, a traditional approach would be to explicitly program the rules of the game and compute every single possibility that allows the computer to win the game against a human.

But using Machine Learning and AI, we can easily "teach" the system how to play and allow it to practice and learn without computing all combinations.

Formally, ML is formally defined as *[read from the ppt only the second f]*

[Slide 5] ML in Stock Prediction

Now that we know what ML is, we can now use its concepts to predict stock price.

And one of the most widely used methods in ML, is **The Support Vector Machine**

SVMs can be used to perform Linear Regression on previous stock data to predict the future closing price any number of days ahead.

SVMs can be used to perform fast computation and it has good generalization capacity and hence can be used for new data.

[Slide 6] EXISTING SYSTEM

In existing systems, the stock prices were predicted by performing time series forecasting, and this was done manually by plotting the prices on a graph on a timeline and making predictions based on the company's performance.

In the proposed system, the prediction is done using a **Sliding-window** method where the dataset is time-sliced and prediction is done for a subset of the data and the window is moved to newer set of data.

In Machine Learning, it has been proved that Support Vector Machines are best used for stock price prediction since it can solve nonlinear regression and are computationally very fast.

A variation of the SVM is the LSSVR, also called the Least Squares Support Vector Regression was proposed, which reduces computational complexity and was more efficient than the standard SVM.

The Firefly algorithm which is an optimization algorithm is used to optimize the hyperparameters of the LSSVR. We will talk about these algorithms in the coming slides.

[Slide 7] DISADVANTAGES

The main disadvantages of the existing systems are that the proposed method predicts the stock price for a single country, namely Taiwan and its companies.

The method does not generalise for markets for all countries and other markets such as US, UK and India.

The system does not allow the raw importing of data and uses old records of small dataset.

Lastly, the system does not have a graphical user interface, and without this the proposed system cannot be used by users for general purposes.

[Slide 8] PROPOSED SYSTEM

Since the existing system uses the method to predict only stock markets in Taiwan, our proposed system generalizes the usage and applies to all major markets over the world.

The proposed system also includes the development of a user interface such as a web app that allows the application to be widely used by businessmen who are unfamiliar with the underlying complexity of the program

The proposed system also includes real-time importing of stock to perform predictions on live data. This makes the proposed system easily adaptable to online learning.

[Slide 9] ALGORITHMS

As stated earlier, the two main algorithms used in the proposed system are the LSSVR and the Firefly Algorithm.

The LSSVR is computationally faster because it solves for linear equations. The disadvantage of the LSSVR is that it requires extreme tuning of the hyperparameters such as the regularization parameters and the kernel function. A traditional approach would require an experienced Machine Learning engineer to fit the parameters either by previous experience or by trial and error.

To avoid this, the existing system suggests an optimization algorithm called as the Firefly Algorithm. It is a nature-inspired method inspired by the behaviour of fireflies.

A firefly attracts other fireflies by its natural flash light. The brighter the firefly, the more fireflies that it can attract and this is also proportional to the distance.

Using this concept, the firefly algorithm performs the optimization of its hyperparameters without doing it manually.

[Slide 10] Architecture.

This is the architecture of the entire Stock Market System.

1. Initial hyperparameters are set to be given as input to the Firefly algorithm
2. A test set is selected or imported from the data. This is the training set and is a subset of the whole data.
3. If there are huge numbers or differences between the numbers, they can be difficult to plot on the graph. Feature scaling and normalization is performed to scale down the data and its size so it is computationally easier.
4. The data must be trained initially on a smaller set and this data is used to train the model.
5. The hyperparameters are optimized using the firefly algorithm.
6. The optimized parameters are used by the LSSVR to prepare a training model.
7. The model can now be saved for forecasting.

8. The trained model is used to predict the stock prices of any new data/
9. The model can be tested using the Cross validation set and the Test Set.

[Slide 11] EXPECTED RESULTS

Let us consider a small dataset from the Coca Cola Company whose stock prices are distributed over 3 months.

The initial part of the graph is the training data that is used to train the LSSVR and prepare a model.

The model is used on the second part of the data that is the test data to perform predictions.

The actual value and the predicted stock price determines the Accuracy of the system.

The orange bar below represents the predicted closing stock price for the given dataset.

The error increases with the number of days in advance for which the prediction is made due to the volatile nature of the stock market.

[Slide 12] Disadvantages

Disadvantage of the Proposed system

The obvious disadvantages of the proposed system are as follows

1. LSSVR is computationally slower when used along with sliding-window method.
2. Each time the window is moved across, the LSSVR must learn for the new dataset
3. The LSSVR must be run each time the window slides and thus the window size affects the performance of the system
4. LSSVR uses many parameters that must be handled and taken care of.
5. The system is best suited for short term prediction and error rate increases with number of days of prediction

[Slide 13] THANKYOU

And thus, by using a sliding-window method, we can use Machine Learning to predict the stock prices days in advance

Thankyou