Investment Recommendation System using Machine learning and Deep learning

Dr. Bhargavi Peddireddy [1], G. Shirisha[2], P. Shishira[3]
[1]*Associate Professor - Computer Science and Engineering Department, Vasavi College of Engineering, Hyderabad, India*
[2,3]*Computer Science and Engineering Department, Vasavi College of Engineering, Hyderabad, India*

Abstract- Creating a personalized investment plan is crucial for anyone who wants to achieve their financial goals. Whenever a personalized plan is created, it is tailored to the people's needs and based on risk tolerance and investment goals. Without a customized plan, it is too easy to make haphazard investment decisions that do not align with overall financial objectives, leading to unnecessary risks and suboptimal returns. For a common man, investment options like Stock Market, Mutual Funds, etc. seem a very complicated and confusing task. A common man is very cautious of the money that he/she is investing on anything. The paper introduces a simplified Investment Recommendation System designed to provide basic investment guidance to users based on their risk preference, age, income, and occupation, with the help of Artificial Intelligence and Machine Learning techniques to enhance the decision- making process for investors. The primary aim is to assist users in making investment decisions aligned with their financial circumstances and risk tolerance. It gives stockinvestment recommendations, mutual funds investment recommendations, gold investment recommendations, and real estate investment recommendations.

Index Terms--LSTM Network, Random Forest Regressor, Gradient boosting regressor

I. INTRODUCTION

In today's economy, making informed investment decisions is crucial for both individual and business investors. Traditional investment strategies often rely on historical data and dry analysis. However, with the advent of advanced technologies, there is a growing interest in using machine learning techniques to enhance the accuracy and efficiency of investment recommendations. Different investment strategies offer varying levels of risk. By defining a strategy, you can choose investments that align with your risk tolerance and financial situation. This helps manage risk effectively and prevents you from taking on too much risk or investing too conservatively.

Various machine learning algorithms can be used for gold investment and real estate investment recommendations. Some are Random forest Regressor and Gradient boosting regressor. Based on the use cases, any one of the regressor should be chosen to predict the future values. But for stock investment recommendation, a variant of Artificial neural network can be used. This variant eliminates the drawback of Recurrent neural network and makes it suitable for gradient descent problems. One such powerful tool is the Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN) designed to capture and learn patterns in sequential data. LSTMs are well-suited for time-series analysis, making them an ideal candidate for modelling and predicting financial market trends. The LSTM model is trained using huge datasets. Once the model is trained, it is used for predicting future values. LSTM can analyse historical market data, identify hidden patterns, and provide valuable insights to guide investment decisions. LSTM networks are a variant of Recurrent neural networks. This variant is used to eliminate the gradient vanishing problem that occurs in RNN.

II. RELATED WORK

1) "Short-term stock market price trend prediction using a comprehensive deep learning system" by Jingyi Shen & Omair Shafiq. The authors' solution is extensive, encompassing several key components. Firstly, they undertake pre-processing of the stock market dataset to ensure data quality and consistency. Next, they employ various feature engineering techniques to extract meaningful information fromthe dataset. These techniques are tailored to enhance the predictive capabilities of their model. Lastly, they develop a custom deep learning- based system

specifically designed for stock market price trend prediction. This integrated approach underscores the thotoughness and sophistication of their proposed solution.

2) In their paper titled "Stock Trend Prediction Algorithm Based on Deep RecurrentNeural Network", Ruochen Lu and Muchao Lu introduce an innovative approach. They integrate an attention mechanism to delve deeper into predictive- line patterns. This mechanism captures valuable judgement experiences akin to those of human researchers, while elucidating the prediction logic of the hybrid neural network. Through this method, the authors aim to enhance the understanding and effectiveness of stock trend prediction algorithms.

3) "Implementation of Fund Recommendation System Using Machine Learning" Chae-eun Park and Dong-seok Lee have devised a Fund Recommendation System utilizing Machine Learning techniques. Their system suggests funds based on the user's investment inclination and matches them with appropriate risk ratings. To predict future fund prices, the system employs the Prophet model, a machine learning approach designed for time series data analysis. By understanding parameters associated with trend variations, the Prophet model forecasts forthcoming fund prices efficiently. This method simplifies predictions and enhances speed by eliminating temporal dependencies in time-series data. Furthermore, web pages have been developed to facilitate fund recommendation and future fund price prediction.

4) In his work, Krist Papadopoulos (2019) explores several cooperative filtering strategies using latent variable models to forecast advisorredemptions of big and sparse mutual funds. It evaluates the suitability of several adviser attributes inside the collaborative filtering framework. In order to prepare for more productive client salesinteractions with advisers, the paper offers financial institutions a method for anticipating significant redemptions from mutual funds. It draws the conclusion that WMF gave the best test performance across all measures by using models such as WLR- MF (Weighted Logistic Regression Matrix Factorization), WLS-MF (Weighted Least Squares Matrix Factorization), and WMF (Weighted Matrix Factorization) for prediction.

5) In their research, Pendaraki, Beligiannis, and Lappa (2016) address methods for predicting the performance and net asset value of mutual funds using genetic programmingand artificial neural networks. The ANN approachoutperforms the GP approach in predicting the net asset value of mutual funds, while the GP approach outperforms the ANN approach in predicting the returns of mutual funds, according to theircomparison of the forecasting results of the two approaches used to predict mutual fund performance. We can utilize this forecast in conjunction with our mutual fund recommendation algorithm. The study describes the ANN strategy for applying the multilayer perceptron methodology, the GP approach for prediction, and the sampling and grouping of input data.

III. DATASET

There are various .csv files containing dataset from the yahoo finance, which is compiled based on the companies included in the index. The datasets are from 2013-01-01 to 2021-05-28. The attributes ofdata sets include Date, open, high, low, close, adj close and volume. Some of the columns in this dataset contain missing values. As a result, the dataset was gathered using Yahoo Finance, a well- known resource for financial data that includes stock price history, company details, and financial news. Yahoo Finance offers a handy API that developers may use to programmatically access financial data, even if it does not directly offer a direct dataset download option

A. Dataset Description

Data sets are taken from yahoo finance, which consists of stock values from 2013-01-01 to 2021-05-28. Apart from stock values, the dataset includes additional attributes such as open, close, low, high, adjacent close and volume. This gives us insights into the maximum stock value as well as theminimum value. The "yahoo finance datasets (2013- 2021)" dataset is a financial dataset containing daily stock market data for multiple assets such as equities, ETFs, and indexes. It spans from April 1, 2018 to March 31, 2023, and contains 1257 rows and 7 columns.

B. Dataset Preprocessing

The companies were eliminated from the initial dataset due to missing values of less than 10%. There are still some inconsistencies in the dataset, such as missing columns or values. Deliverable volume and VWAP are among the columns that are removed from

the model that are not necessary. The feature column is used to fill in some of the lacking values inother columns. Data purification: When there is missing data, the tuples are ignored or the values are manually filled in using the mean or most likely values. Clustering or the binning approach is used to handle noisy data. Data transformation: Either data normalization or data standardization is used to transform data. The process of normalization involves scaling the data within a range of 0. and 1. Standardization is a technique used to transform the data to have a mean of 0 and a standard after themean has been subtracted.

C.Data Splitting

The dataset is split into training and test dataset. The training dataset consists of values from 2013-01-01 to 2021-04-28. The remaining one month dataset is used for testing the model. Testing dataset provides information about the accuracy of the model. LSTM performs better in the case of text data and Gradient Boosting regressor performs better in the case of datasets. Data is split according to the algorithm used. Data sets are fed to Gradient Boosting Regressor and Text, time series data is fed to LSTM Networks.

D.Supervised Learning Algorithms

1) Linear Regression:

Type: Supervised learning algorithm for regression tasks.

Purpose: Aims to estimate the coefficient of linear equation that minimizes the difference between the observed values of the dependent variable and the values predicted by the model. Once the model is trained using historical data, it can be used to make predictions on new and unseen datapoints.

Key Feature: Quantifies and analyzes the relationship between a dependent variable and one or more independent variables. Through this analysis , linear regression helps in understanding how changes in the independent variables are associated with changes in the dependent variables.

2) LSTM Network (Long Short-Term Memory):

Type: Recurrent Neural Network (RNN) architecture.

Purpose: Effective for sequence modeling and time-series prediction.

Key Feature: Handles long-range dependencies and captures sequential information by using memory cells with gating mechanisms.

3) Random Forest Regressor:

Type: Ensemble learning algorithm.

Purpose: Predicts a continuous outcome bycombining predictions from multiple decision trees. Key Feature: Reduces overfitting and increasesaccuracy through bagging (bootstrap aggregating) and feature randomness.

4) CNN (Convolutional Neural Network):

Type: Deep learning architecture, primarily used for image and spatial data.

Purpose: Recognizes patterns in spatial data through convolutional layers, pooling layers, and fully connected layers. Key Feature: Local receptive fields and weight sharing to capture hierarchical patterns.

5) Gradient Boosting Regressor:

Type: Ensemble learning algorithm for classification and regression tasks.

Purpose: Separates data into classes or predicts continuous outcomes by finding an optimal hyperplane in a high- dimensional space.

Feature: Uses support vectors (data points near the decision boundary) to maximize the margin between classes.

IV. PROPOSED MODEL

A.Data Collection and Preprocessing:

Gathering a diverse set of financial data, including stock prices, trading volumes, economic indicators, and sentiment analysis from news articles or social media. Preprocessing data to handle missing values, normalize features, and shape the dataset that is suitable for injecting into LSTM model. Normalizing the dataset is a crucial step in data preprocessing. There are various predefined functions to normalize the values in the dataset. Normalizing helps in understanding the data and reduces complexities. A special data structure is created using training dataset and test dataset. A data structure is created, consisting of sixty time stepsand one output.

B. LSTM Network Architecture:

Designing an LSTM model that takes into account the sequential nature of financial data. Configuring the architecture with input layers, LSTM layers, and

output layers, ensures the network can effectively learn and remember patterns over different time horizons. Initialize the Recurrent neural network with appropriate number of layers. Then start adding deep layers to the model for better accuracy.

C. Training and Validation:

Split the dataset into training and test sets to train the LSTM model. Optimize the model for accuracy, taking into consideration, factors such as learning rate, batch size, and number of epochs. Add deep layers to improve models accuracy. Train the LSTM model using training dataset. Test the model using test dataset. Plot the graph between training dataset and test dataset, to understand the accuracy of the model.

D. Feature Importance and Interpretability:

Analyzing the LSTM model to understand which features and periods contribute most to itspredictions. Enhancing transparency andinterpretability to build trust among investors andfinancial professionals.

E. Integration with Real-Time Data:

Incorporating mechanisms to adapt the model to changing market conditions by regularly updating it with real-time data. Ensuring the system can handle streaming data efficiently to provide timely recommendations.

G. Benefits:

1) Improved accuracy in predicting market trends and identifying investment opportunities.
2) Enhanced adaptability to dynamic market conditions through the use of sequential data analysis.
3) Automation of the decision-making process, reducing human bias and emotions.
4) The proposed model is best suitable for time series data

H. Challenges:

1) Overfitting and underfitting concerns require careful model tuning.
2) Handling the non-stationary nature of financial markets.
3) Ensuring robustness and interpretability in the face of complex, high-dimensional data.

V. FIGURES

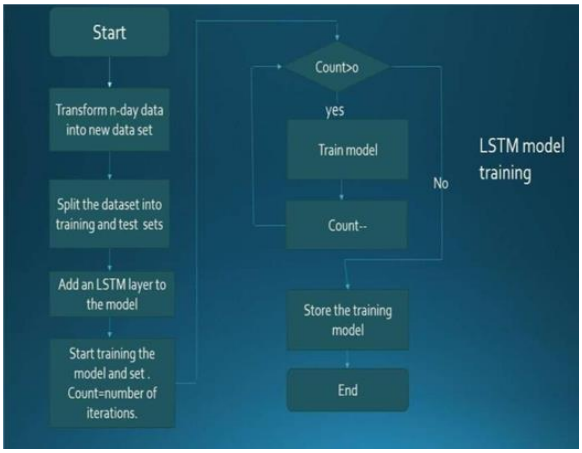

Figure 1 : LSTM training flow model

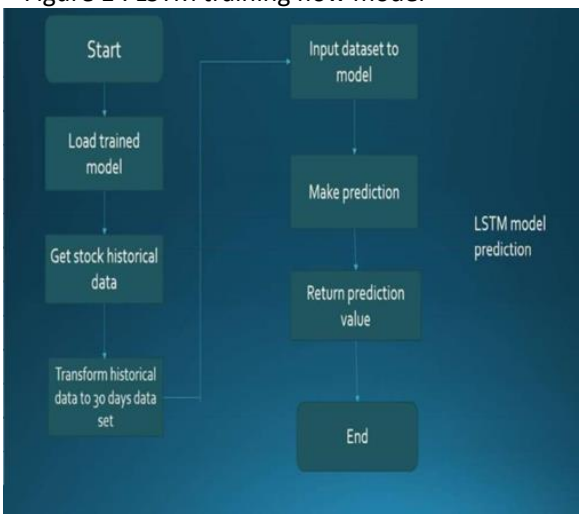

Figure 2: LSTM model prediction

VI. CONCLUSION

In conclusion, the use of Long Short-Term Memory (LSTM) networks in an investment recommendation system holds promise for enhancing decision-making in financial markets. The ability of LSTMs to capture sequential dependencies and model complex temporal dynamics makes them well-suited for analyzing historical market data and predicting future trends. The automatic feature extraction capability of LSTMs reduces the reliance on manual feature engineering, providing a valuable advantage in financial applications where relevant features may evolve. LSTM models may face challenges during extreme market events or sudden shifts in conditions, necessitating the incorporation of additional risk management strategies. The effectiveness of the recommendation system is heavily dependent on the

quality and quantity of historical data, highlighting the importance of ensuring clean, relevant, and diverse datasets. Continuous monitoring and adaptation are paramount, given the dynamic nature of financial markets. Regular updates and recalibrations are necessary to align the model with evolving market conditions and maintain its predictive accuracy. In conclusion, while an LSTM-based investment recommendation system can provide valuable insights, it should be integrated into a comprehensive investment strategy that incorporates risk management practices and human expertise, recognizing the inherent uncertainties in financial markets.

Without a personalized plan, it is far too simple to make impulsive investing choices that are at odds with overall financial goals, resulting in needless risks and less-than-ideal results. The stock market is an ever-evolving arena. It is challenging to forecast how the market will close and under what circumstances, even after thorough data analysis, because the information is always changing. A review of numerous research papers and articles about stock price prediction showed that a variety of models, including CNN, Support Vector Machines, and Random Forest Regressors, are employed in stock price analysis and prediction. A lot of researchers and developers also employ common techniques like time series data analysis toanticipate stock market trends.

After conducting research comparing LSTM networks and Gradient Boosting Regressors for stock market prediction, we This model leverages past trends observed across various companies in the Indian stock market. Notably, the suggested model demonstrates exceptional accuracy when tested against the test dataset. By enriching the original dataset with additional information, our analytic forecasts closely mirror the actual values reported in the Yahoo Finance datasets. This underscores the effectiveness of our approach in developing a reliable stock market predictionmodel.

The model's results indicate that it can serve as the foundation for a larger recommender system that can accurately forecast stock market developments. However, as previously said, the market is always shifting, so even 90% accurate forecasts cannot ensure profits in full. Moreover, expanding the dataset size and making certain speed related performance improvements can help us enhance the recommendation system and offer a more comprehensive analysis and superior recommendations. Using improved clustering techniques leads to sophisticated and scalable clustering methods. Further application of methods such as Time Series analysis can enhance the performance of the recommender system and the model.

VII. REFERENCE

[1] Research on Financial Investment Decision Based on Artificial Intelligence Algorithm, IEEE Conference Publication, IEEE Explore

[2] Stock Price Prediction using LSTM and SVR, IEEE Conference Publication, IEEEExplore.

[3] Indian Stock market prediction using Deep learning, IEEE Conference Publication, IEEE Explore

[4] An LSTM Based Model for Stock Price Prediction, IEEE Conference Publication, IEEE Explore

[5] Stock Price Prediction and Recommendation Approach Based on Machine Learning, IEEE Conference Publication , IEEE Explore.

[6] An Evaluation of Extreme Machine Learning Algorithm to Forecasting the Gold Price, IEEE Conference Publication, IEEE Explore.

[7] Real Estate Price Prediction System using Machine Learning, IEEE Conference Publication, IEEE Explore.

[8] Stock Market Investment Advice: A social network approach, IEEE Conference Publication, IEEE Explore.

[9] Stock Price Prediction using Linear Regression and Machine Learning, IEEE Conference Publication, IEEE Explore

[10] Analyzing Stock market trend prediction using Machine & Deep learning Models, IEEE Conference Publication, IEEE Explore