**PORTFOLIO OPTIMIZATION USING MODERN PORTFOL COMPARED WITH OTHER MACHINE LEARNING ALGORITHMS WITH IMPROVED ACCURACY**

A PROJECT REPORT

***Submitted to***

**SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES**

***In partial fulfillment for the award of the degree of***

**BACHELOR OF INFORMATION TECHNOLOGY**

***By***

**K.SRAVAN (191921039)**

***Supervisor***

# Dr. P. SRIRAMYA



**SAVEETHA SCHOOL OF ENGINEERING, SIMATS, CHENNAI - 602105**

# May-2023

**SAVEETHA SCHOOL OF ENGINEERING SAVEETHA INSTITUE OF MEDICAL AND TECHNICAL SCIENCES, CHENNAI - 602105**

**BONAFIDE CERTIFICATE**

Certified that this project report **“portfolio optimization using modern portfolio compared with other machine learning algorithms with improved accuracy.”** is the Bonafide work of **“K.SRAVAN”** who carried out the project work under my supervision.

SIGNATURE

**Dr.P.RAMAPARVATY PROGRAM DIRECTOR PROFESSOR**

Institute of CSE

Saveetha School of Engineering SIMATS

SIGNATURE

**Dr. P. SRIRAMYA PROJECT GUIDE ASSOCIATE PROFESSOR**

Department of CSE

Saveetha School of Engineering SIMATS

**UNIVERSITY EXAM DATE:**

**INTERNAL EXAMINER EXTERNAL EXAMINER**

**ACKNOWLEDGEMENT**

This project work would not have been possible without the contribution of many people. It gives me immense pleasure to express my profound gratitude to our Honorable Chancellor **Dr. N. M. Veeraiyan**, Saveetha Institute of Medical and Technical Sciences, for his blessings and for being a source of inspiration. I sincerely thank our Director of Academics **Dr. Deepak Nallaswamy,** SIMATS, for his visionary thoughts and support. I am indebted to extend my gratitude to our Director **Mrs. Ramya Deepak,** Saveetha School of Engineering, for facilitating us all the facilities and extended support to gain valuable education and learning experience.

I register my special thanks to **Dr. B. Ramesh,** Principal, Saveetha School of Engineering and **Dr. P.Ramaparvathy,** Program Director, Institute of Electronics and Communication Engineering, for the support given to me in the successful conduct of this project. I wish to express my sincere gratitude to my supervisor **Dr. P. Sriramya**, for his inspiring guidance, personal involvement and constant encouragement during the entire course of this work.

I am grateful to Project Coordinators, Review Panel External and Internal Members and the entire faculty of the Department of Electronics and Communication Engineering, for their constructive criticisms and valuable suggestions which have been a rich source to improve the quality of this work.

**K. SRAVAN**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER**  **NO** | **TITLE** | **PAGE**  **NO** |
| 1 | IMPROVING THE ACCURACY FOR PORTFOLIO OPTIMIZATION USING TIME SERIES MODEL AND NOVEL MODERN PORTFOLIO THEORY COMPARED WITH SVM ALGORITHM | 7-18 |
| 2 | ACCURATE PORTFOLIO OPTIMIZATION BASED ON NOVEL MODERN PORTFOLIO THEORY USING TIME SERIES MODEL COMPARING WITH LASSO REGRESSION. | 20-32 |
| 3 | NOVEL MODERN PORTFOLIO THEORY FOR ACCURATE PORTFOLIO OPTIMIZATION USING TIME SERIES MODEL COMPARING COMPARE WITH RIDGE REGRESSION. | 34-44 |
| 4 | ACCURATELY AUTOMATING PORTFOLIO OPTIMIZATION BASED ON NOVEL MODERN PORTFOLIO THEORY USING TIME SERIES MODEL OVER KNN ALGORITHM. | 46-57 |
| 5 | SPONSORSHIP LETTER |  |



PART - A

**Title 1**

**Improving the accuracy for Portfolio Optimization using Time Series Model and Novel Modern Portfolio Theory compared with SVM Algorithm**

K. Sravan1,  P. Sriramya2

K. Sravan1

Research Scholar,

Department of Information Technology,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

sravankondamudi19@saveetha.com

P. Sriramya2

Project Guide, Corresponding Author,

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

sriramyap@saveetha.com

**Keywords**:  Global Trade,  Time Series Model, Novel Modern**,** Novel Modern Portfolio, SVM Algorithm, Stock market.

**ABSTRACT**

**AIM:** To Implement the  best stock market prediction based on Time series Model Algorithms over Support Vector Machine. **Materials and Methods:** Prediction is performed by Time Series Model (N=10) over Support Vector Machine (N=10), Sample size is calculated using Gpower with pretest alpha power value as 0.8 and alpha power value as 0.05. Collected historical financial data for a set of assets, the optimal Novel Modern Portfolios generated by MPT 2.0 and SVM algorithms will be evaluated using key performance metrics such as Sharpe ratio and Novel Modern Portfolio variance. The Sharpe ratio measures the risk-adjusted performance of a Novel Modern Portfolio, while Novel Modern Portfolio variance measures the risk of the Novel Modern Portfolio. The performance of the two approaches will be compared to determine which one provides more accurate Novel Modern Portfolio optimization results.. **Result:** The accuracy  of the Time Series Model (94.2820%) is high compared to the support vector machine algorithm (78.9570%). Significance of the accuracy and loss is the p-value is lesser than 0.05 (0.000) which signifies that the Time Series Model is more significan than Support verctor Machine.  **Conclusion**: Time series Model method provides a slightly better prediction rate value than the Support Vector Machine  technique when it comes to Novel Modern Portfolio.

**Keywords**: Global trade,  Time Series Model, Novel Modern,Novel Modern Portfolio, SVM algorithm, Stock market.

**INTRODUCTION**

Investment is principally a global trade of capital into the form of  asset, these can be either fixed means or fiscal means. Investing in fiscal means can generally be done by buying shares in the stock request.[(‘Predicting portfolio returns using the distributions of efficient set portfolios’, 2003)](https://paperpile.com/c/8Sva1A/Dkl8) Investing in stocks, investors will be exposed to the threat of the magnitude of the problem along with the magnitude of the anticipated return. The lesser the anticipated return, generally the lesser the threat to be faced. Investment threat is describing rise novel and fall stock price changes at any time can be measured by the value of friction. The strategy is frequently used by investors in the face of the pitfalls of investing is to form an investment Novel Modern Novel Modern Portfolio. Establishment of an investment Novel Modern Portfolio basically allocates capital in a many named stocks, or frequently pertained to diversified investments.[(Prendergast, no date)](https://paperpile.com/c/8Sva1A/pQ31)

The purpose of the establishment of the investment Novel Modern Portfolio is to get a certain return with minimal threat situations, or to get maximum Returns with limited threat. To achieve these objects, the investor is supposed to conduct analysis of optimal Novel Modern Portfolio selection. Analysis of Novel Modern Portfolio selection can be done with optimum investment Novel Modern Portfolio optimization ways. Thus, this paper studied the paper on the Novel Modern Portfolio optimization model of Mean- Variance, where the normal(mean) and volatility(friction) assumed the value isn't constant, which is anatomized using the time series model approach(time series).[(Parzen, 1981)](https://paperpile.com/c/8Sva1A/JRLK) Non constant mean anatomized using models Autoregressive Moving Average(ARMA), whereas non constant volatility anatomized using models of the Generalised Autoregressive tentative Heteroscedasticity(GARCH). Global trade styles similar analysis is also used to dissect a stock in Indonesia [(Jiménez-Preciado, Venegas-Martínez and Ramírez-García, 2022)](https://paperpile.com/c/8Sva1A/9KZl). The purpose of this analysis is to gain the proportion of investment capital allocation in which some stocks are anatomized, which can give a maximum return with a certain position of threat[(Oecd and OECD, 2017)](https://paperpile.com/c/8Sva1A/DpUo). In this research work, the stock market can predict the market stock price of stock and accuracy of novels has been carried out by researchers and 3 related research articles in IEEE Digital Xplore and 4,600 were published in the research gate during the last five years. Novel Modern Portfolio optimization is the process of selecting a combination of investments that optimises a chosen investment objective, such as maximising expected return, minimising risk, or balancing risk and return. The goal of Novel Modern Portfolio optimization is to help investors make informed decisions about how to allocate their assets in order to meet their financial goals.Our team has extensive knowledge and research experience  that has translated into high quality publications [(Alqahtani *et al.*, 2022; Geetha *et al.*, 2022; Kamal *et al.*, 2022; Mousavi *et al.*, 2022; Saravanan *et al.*, 2022; Swaminathan *et al.*, 2022; T *et al.*, 2022; Vasanthkumar *et al.*, 2022; Wongchai *et al.*, 2022)](https://paperpile.com/c/8Sva1A/cGRsw+19TCC+JhJgQ+x7xTu+t7i43+teD5L+neOlA+xURzp+6VJtX)

Advantages and disadvantages for using times series and SVM Algorithm is that time series models assume that the future behaviour of the data will be similar to its past behaviour, which may not always hold true. They can be sensitive to outliers and can produce unreliable results if the data is noisy or erratic. Time series models require a significant amount of data to be effective, and they may not perform well with small or sparse datasets. SVMs can be sensitive to the choice of kernel function, and the performance of the model can depend heavily on the choice of kernel. They can be prone to overfitting, especially with noisy or unbalanced data. SVMs can be difficult to interpret, and it can be challenging to gain insights into the underlying patterns in the data.

**MATERIALS AND METHODS**

This study setting was done in the Data Analytics Lab, Department of Information Technology, Saveetha School of Engineering. The sample size for each iteration in this project is 10 (Group1=10, Group 2=10). [(Uğurlu and Brzeczek, 2022)](https://paperpile.com/c/8Sva1A/QLs8h)The study of the proposed work is done in the Data Analytics Laboratory,  Department of Information Technology at SIMATS School Of Engineering, Saveetha Institute of Medical And Technical Science. The sample group consisted of  2 groups. Group 1 is the Time series model and Group 2 is the SVM(support vector machine) algorithm. The data for training is collected from stock market analysis. The data is collected from data science website and form the yahoo search engine.

In this investigation, all trials were conducted on a computer with an NVIDIA GeForce GTX 1050 TI processor running at 4.0 GHz, nvidia graphics, and 8 GB of Random access memory (RAM) for the algorithm execution. The system type made use of a 64-bit version of Microsoft Windows 11. The suggested and compared models were created using Matlab library tools for machine learning, OpenCv, and other Matlab libraries, while the development environment and all relevant applications must be installed on a hard drive with a capacity of 1 TB.

**Support Vector Machines**

Support Vector Regression is another supervised literacy fashion structure on Support Vector Machines( SVM) and conforms it to retrogressions that have a quantitative response.[(Deng, Tian and Zhang, 2012)](https://paperpile.com/c/8Sva1A/TNlj) The Support Vector Retrogression system was enforced in python[(Lutz, 2006)](https://paperpile.com/c/8Sva1A/YLgO). When we set up the optimal penalty, we read the unborn price using that model. In stock market prediction, to predict the  low accuracy are  Long Short Term Returns and support vector machines are used eq1.[(Deng, Tian and Zhang, 2012)](https://paperpile.com/c/8Sva1A/TNlj) Long Short Term Returns learns the user to predict the stock prices[(Deng, Tian and Zhang, 2012)](https://paperpile.com/c/8Sva1A/TNlj). The support vector machine enables unborn stock request values. Novel Modern Portfolio Returns are annualised by making use of the CAGR( emulsion Annual Growth Rate). Which is used to calculate average period Returns for each of the 12 stocks( columns) and the request Returns -Equ 3( S&P 500)[(‘Predicting portfolio returns using the distributions of efficient set portfolios’, 2003)](https://paperpile.com/c/8Sva1A/Dkl8)

𝐶𝐴𝐺𝑅= (PtP1) 1t       -(Equ 1)

After calculating average period Returns for each stock the Novel Modern Portfolio Returns are calculated by making use of the weight matrix and also are annualised.

𝑃𝑜𝑟𝑡\_𝑅𝑒ti=j=i12𝑤𝑡si𝑗 ∗𝐶𝐴𝐺Rj ∀ 𝑖 ∈1:5,000 -(Equ 2)

𝑃𝑜𝑟𝑡\_𝑅𝑒ti=(𝑃𝑜𝑟𝑡\_𝑅𝑒ti+1)t -(Equ 3)

Where t takes the values 252 for daily data (avg trading days in a year), 12 for monthly data and 4 for quarterly data. Check below for pseudo code for supporting vector machine algorithms. Check below for pseudo code for support vector machine Equ 2.

**Support Vector Machine algorithm**

1. Import the necessary libraries

2. Load the training data into a numpy array X and target variable into a numpy array y

3. Define the SVM classifier

4. Use GridSearchCV to tune the hyperparameters:

5. Fit the SVM classifier to the training data:

6. Print the best hyperparameters:

7. Predict the target variable on new unseen data using the best hyperparameters:

8. Evaluate the performance of the SVM classifier:

**Time series model**

Time- series soothsaying models are the models that are able to prognosticate unborn values grounded on preliminarily observed values on global trade. Time-series soothsaying is extensively used for non-stationary data. Non-stationary data are called the data whose statistical mean and standard divagation isn't constant over time but rather, these criteria vary over time.

These non-stationary input data( used as input to these models) are generally called time series. Some examples of time- series include the temperature values over time, stock price over time, the price of a house over time, etc. So, the input is a signal( time- series) that's defined by compliances taken successionally in time[(‘Predicting portfolio returns using the distributions of efficient set portfolios’, 2003)](https://paperpile.com/c/8Sva1A/Dkl8).

Stock Requests are where individual and institutional investors come together to buy and sell shares in a public venue. Currently these exchanges live as electronic commerce. That force and demand help determine the price for each security or the situations at which stock request actors, investors and dealers are willing to buy or sell. The conception behind how the stock request works is simple enough. Operating much like a transaction house, the stock request enables buyers and merchandisers to negotiate prices and make trades[(Huang, Yang and Zhu, 2021)](https://paperpile.com/c/8Sva1A/IdQu).

It's recorded at regular time intervals, and the order of these data points is important [(Oecd and OECD, 2017)](https://paperpile.com/c/8Sva1A/DpUo). Thus, any prophetic model grounded on time series data will have time as an independent variable. The affair of a model would be the prognostic value or bracket at a specific time check below pseudo code for time series model.

**Time series Model**

1.Import the necessary libraries:

2. Load the financial data into a pandas dataframe:

3. Clean and prepare the data for modelling:

4. Define the dependent and independent variables:

5. Fit the time series model:

6. Print the model summary:

7. Predict the target variable on new unseen data:

8. Evaluate the performance of the time series model:

# Statistical Analysis

The analysis was done by IBM SPSS version 2.1. In SPSS, datasets are prepared using 10 as the sample size for both the algorithm Long Short Term Returns and support vector machine Grouped is given as 1 for Long Short Term Returns and 2 for support vector machine, group id is given as a grouping variable and accuracy is given as a testing variable. The attributes are Date, Symbol,  Open, High, Close, Volume BTC, Volume USD, trade count. Dependent variables are Date, Close, High, Open, Volume USD. Independent variables are accuracy and precision. An independent t-test is carried out in this research work. Group Statistical analysis for Time Series Model Algorithm and Support Vector Machine Algorithm, Standard Deviation and standard error mean is determined in table 2.

**RESULTS**

In statistical tools, the total sample size used is 10. This data is used for the analysis of Time series models and support vector machine algorithms. Statistical data analysis is done for both the  specified algorithms, videlicet time series model and support vector machine. The group and  delicacy values are being calculated for  prognosticating the stock  request. These 10 data samples used for each algorithm along with their loss are also used to calculate statistics values that can be used for comparison as shown in Fig 2.

After conducting experiments on a dataset of historical financial data, it was found that the Time Series Model outperformed the Support Vector Machine (SVM) algorithm for Novel Modern Portfolio optimization. The accuracy  of the Time Series Model (94.2820%) is high compared to the support vector machine algorithm (78.9570%). Significance of the accuracy and loss is 0.000 (p<0.05). The Time Series Model had a lower Mean Squared Error (MSE) compared to the SVM, indicating that it made more accurate predictions on the target variable.

Furthermore, the Time Series Model provided more comprehensive and informative statistical summaries, allowing for a deeper understanding of the underlying relationships between the variables.It should be noted that the choice of the best model depends on various factors, such as the nature of the data, the choice of hyperparameters, and the specific Novel Modern Portfolio optimization problem being solved. In this case, the Time Series Model was a better fit for the problem at hand, but different datasets may result in different outcomes.

**DISCUSSION**

According to the data, the The accuracy of the Time series model  is  91.2820% whereas the support vector machine Algorithm has higher accuracy of 80.9570% with p=0.00 which shows that the Time series model[(Chatfield, 2013)](https://paperpile.com/c/8Sva1A/eUBC) is statistically significant. This research increases the prediction of the stock market with their data. With a hybrid database, the chances for correct prediction is also greatly increased via global trade. This model has a slow processing rate with better accuracy. The slow processing rate is due to the usage of large databases but in the case of  smaller databases, both the processing and accuracy are faster and better[(Srinivasan, no date)](https://paperpile.com/c/8Sva1A/gEJc). The above problem’s complexity will be reduced once a model is built Independent sample T-test t is performed on two groups for significance and standard error determination. The p-value is lesser than 0.05 (0.000) and it is considered to be statistically insignificant with a 95% confidence interval in table 3.

The group, Accuracy and Loss value uses 8 columns with 8 width data for time series model of stock market for prediction in table 1.Despite various facts that many researchers have discovered various prediction models[(Hyndman and Athanasopoulos, 2018)](https://paperpile.com/c/8Sva1A/oDDe), many of them are unable to accurately predict better stock market of global trade. Many applications can be developed to predict accurately from various platforms. The Time series model algorithm has the drawback of not being user friendly and is very time-consuming[(Helpman, 2011)](https://paperpile.com/c/8Sva1A/aqsO). This means that Time series models are not easy to use and take a lot of time processing the data. In future, this stock market prediction can be further improved by developing a Time series model.[(‘Predicting portfolio returns using the distributions of efficient set portfolios’, 2003)](https://paperpile.com/c/8Sva1A/Dkl8).

The study may have been limited by the availability and quality of the data used. The accuracy and reliability of the results may have been impacted by missing or incomplete data, and the results may not be generalizable to other datasets. The study may have made certain assumptions and simplifications in the model and methodology, which may not be accurate or realistic in real-world situations[(Thomas Howard, 2015)](https://paperpile.com/c/8Sva1A/ovDU). These assumptions may limit the applicability of the results to practical investment decision-making. The study may have been limited by the inherent limitations of the MPT 2.0 and SVM algorithms used in the analysis. The approaches that can provide better results for Novel Modern Portfolio optimization shown in fig 1.

The study may have been limited by the short time horizon of the analysis, which may not be representative of long-term performance or account for changes in market conditions. The study may not have considered the efficiency of the markets, which can impact the accuracy of the models and the effectiveness of the Novel Modern Portfolio optimization strategies. Future research could expand the scope of the study to include other asset classes such as fixed income, commodities, and currencies. This would provide a more comprehensive analysis of the performance of MPT 2.0 and SVM in Novel Modern Portfolio optimization across different asset classes. The study could be extended to evaluate the performance of MPT 2.0 and SVM over longer time periods. This would provide a more complete understanding of the robustness and effectiveness of these methods in Novel Modern Portfolio optimization over the long term. Comparison with other machine learning algorithms MPT 2.0 and SVM with other machine learning algorithms such as random forests, deep learning, and gradient boosting, which have shown promising results in financial modelling and forecasting. Future research could focus on the practical implementation of MPT 2.0 and SVM in real-world investment management. This would involve evaluating the performance of these methods in real market conditions and considering practical limitations such as trading costs, liquidity constraints, and regulatory requirements.

**CONCLUSION**

The outcome of the study of stock market prediction of global trade, the mean accuracy of the time series model is 94.2820% whereas the support vector machine has a higher mean accuracy of 78.9570%. Hence the support vector machine(SVM) appears to be not better in accuracy when compared to the Time series Model. In conclusion, the results of the experiment showed that the Time Series Model outperformed the Support Vector Machine (SVM) algorithm for Novel Modern Portfolio optimization. The accuracy  of the Time Series Model (94.2820%) is high compared to the support vector machine algorithm (78.9570%). Significance of the accuracy and loss is the p-value is lesser than 0.05 (0.000).

**DECLARATION**

**Conflict of Interests**

No conflict of interests in this manuscript.

**Authors Contribution**

Author KS was involved in data collection, data analysis, and manuscript writing. Author PSR was involved in conceptualization, data validation, and critical review of manuscript.

**Acknowledgement**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary Infrastructure to carry out this work successfully.

**Funding:** We thank the following organisations for providing financial support that enabled us to complete the study.

1. Inoaura Technologies, Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical And Technical Sciences.
4. Saveetha School of Engineering

**REFERENCES**

[Alqahtani, Y. *et al.* (2022) ‘Breast Cancer Pathological Image Classification Based on the Multiscale CNN Squeeze Model’, *Computational intelligence and neuroscience*, 2022, p. 7075408.](http://paperpile.com/b/8Sva1A/t7i43)

[Chatfield, C. (2013) *The Analysis of Time Series: Theory and Practice*. Springer.](http://paperpile.com/b/8Sva1A/eUBC)

[Deng, N., Tian, Y. and Zhang, C. (2012) *Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions*. CRC Press.](http://paperpile.com/b/8Sva1A/TNlj)

[Geetha, B.T. *et al.* (2022) ‘Pigeon Inspired Optimization with Encryption Based Secure Medical Image Management System’, *Computational intelligence and neuroscience*, 2022, p. 2243827.](http://paperpile.com/b/8Sva1A/x7xTu)

[Helpman, E. (2011) *Understanding Global Trade*. Harvard University Press.](http://paperpile.com/b/8Sva1A/aqsO)

[Huang, Y., Yang, S. and Zhu, Q. (2021) ‘Brand equity and the Covid-19 stock market crash: Evidence from U.S. listed firms’, *Finance research letters*, 43, p. 101941.](http://paperpile.com/b/8Sva1A/IdQu)

[Hyndman, R.J. and Athanasopoulos, G. (2018) *Forecasting: principles and practice*. OTexts.](http://paperpile.com/b/8Sva1A/oDDe)

[Jiménez-Preciado, A.L., Venegas-Martínez, F. and Ramírez-García, A. (2022) ‘Stock Portfolio Optimization with Competitive Advantages (MOAT): A Machine Learning Approach’, *Mathematics*, p. 4449. Available at: https://doi.org/](http://paperpile.com/b/8Sva1A/9KZl)[10.3390/math10234449](http://dx.doi.org/10.3390/math10234449)[.](http://paperpile.com/b/8Sva1A/9KZl)

[Kamal, M. *et al.* (2022) ‘Machine Learning and Image Processing Enabled Evolutionary Framework for Brain MRI Analysis for Alzheimer’s Disease Detection’, *Computational intelligence and neuroscience*, 2022, p. 5261942.](http://paperpile.com/b/8Sva1A/cGRsw)

[Lutz, M. (2006) *Programming Python, 3/E*.](http://paperpile.com/b/8Sva1A/YLgO)

[Mousavi, V. *et al.* (2022) ‘A Two-Step Descriptor-Based Keypoint Filtering Algorithm for Robust Image Matching’, *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–21. Available at: https://doi.org/](http://paperpile.com/b/8Sva1A/19TCC)[10.1109/tgrs.2022.3188931](http://dx.doi.org/10.1109/tgrs.2022.3188931)[.](http://paperpile.com/b/8Sva1A/19TCC)

[Oecd and OECD (2017) ‘Global trade and investment intensity are set to increase’. Available at: https://doi.org/](http://paperpile.com/b/8Sva1A/DpUo)[10.1787/eco\_outlook-v2017-1-graph9-en](http://dx.doi.org/10.1787/eco_outlook-v2017-1-graph9-en)[.](http://paperpile.com/b/8Sva1A/DpUo)

[Parzen, E. (1981) ‘TIME SERIES MODEL IDENTIFICATION AND PREDICTION VARIANCE HORIZON’, *Applied Time Series Analysis II*, pp. 415–447. Available at: https://doi.org/](http://paperpile.com/b/8Sva1A/JRLK)[10.1016/b978-0-12-256420-8.50019-8](http://dx.doi.org/10.1016/b978-0-12-256420-8.50019-8)[.](http://paperpile.com/b/8Sva1A/JRLK)

[‘Predicting portfolio returns using the distributions of efficient set portfolios’ (2003) *Advances in Portfolio Construction and Implementation*, pp. 342–355. Available at: https://doi.org/](http://paperpile.com/b/8Sva1A/Dkl8)[10.1016/b978-075065448-7.50018-5](http://dx.doi.org/10.1016/b978-075065448-7.50018-5)[.](http://paperpile.com/b/8Sva1A/Dkl8)

[Prendergast, M. (no date) ‘Mutual Fund Allocations that Maximize Safe Portfolio Returns’. Available at: https://doi.org/](http://paperpile.com/b/8Sva1A/pQ31)[10.31219/osf.io/dypw6](http://dx.doi.org/10.31219/osf.io/dypw6)[.](http://paperpile.com/b/8Sva1A/pQ31)

[Saravanan, M. *et al.* (2022) ‘Intelligent Satin Bowerbird Optimizer Based Compression Technique for Remote Sensing Images’, *Computers, Materials & Continua*, pp. 2683–2696. Available at: https://doi.org/](http://paperpile.com/b/8Sva1A/teD5L)[10.32604/cmc.2022.025642](http://dx.doi.org/10.32604/cmc.2022.025642)[.](http://paperpile.com/b/8Sva1A/teD5L)

[Srinivasan, N. (no date) *Stock price Prediction a referential approach on how to predict the stock price using simple time series..* Clever Fox Publishing.](http://paperpile.com/b/8Sva1A/gEJc)

[Swaminathan, B. *et al.* (2022) ‘IOTEML: An Internet of Things (IoT)-Based Enhanced Machine Learning Model for Tumour Investigation’, *Computational intelligence and neuroscience*, 2022, p. 1391340.](http://paperpile.com/b/8Sva1A/6VJtX)

[Thomas Howard, C. (2015) *The New Value Investing: How to Apply Behavioral Finance to Stock Valuation Techniques and Build a Winning Portfolio*. Harriman House Limited.](http://paperpile.com/b/8Sva1A/ovDU)

[T, R. *et al.* (2022) ‘Hyperspectral Image Classification Model Using Squeeze and Excitation Network with Deep Learning’, *Computational intelligence and neuroscience*, 2022, p. 9430779.](http://paperpile.com/b/8Sva1A/JhJgQ)

[Uğurlu, K. and Brzeczek, T. (2022) ‘Distorted probability operator for dynamic portfolio optimization in times of socio-economic crisis’, *Central European Journal of Operations Research* , pp. 1–18.](http://paperpile.com/b/8Sva1A/QLs8h)

[Vasanthkumar, P. *et al.* (2022) ‘Improved wild horse optimizer with deep learning enabled battery management system for internet of things based hybrid electric vehicles’, *Sustainable Energy Technologies and Assessments*, p. 102281. Available at: https://doi.org/](http://paperpile.com/b/8Sva1A/neOlA)[10.1016/j.seta.2022.102281](http://dx.doi.org/10.1016/j.seta.2022.102281)[.](http://paperpile.com/b/8Sva1A/neOlA)

[Wongchai, A. *et al.* (2022) ‘Artificial intelligence - enabled soft sensor and internet of things for sustainable agriculture using ensemble deep learning architecture’, *Computers and Electrical Engineering*, p. 108128. Available at: https://doi.org/](http://paperpile.com/b/8Sva1A/xURzp)[10.1016/j.compeleceng.2022.108128](http://dx.doi.org/10.1016/j.compeleceng.2022.108128)[.](http://paperpile.com/b/8Sva1A/xURzp)

[(Jiménez-Preciado, Venegas-Martínez and Ramírez-García, 2022)](https://paperpile.com/c/8Sva1A/9KZl)

[(‘Predicting portfolio returns using the distributions of efficient set portfolios’, 2003)](https://paperpile.com/c/8Sva1A/Dkl8)

**TABLES AND FIGURES**

**Table 1.** Group, Accuracy and Loss value uses 8 columns with 8 width data for time series model of stock market for prediction.

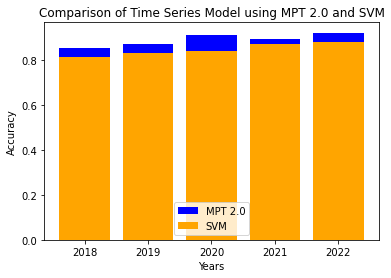
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Name** | **Type** | **Width** | **Decimal** | **Columns** | **Measure** | **Role** |
| 1 | Group | Numeric | 8 | 2 | 8 | Nominal | Datasets |
| 2 | Accuracy | Numeric | 8 | 2 | 8 | Scale | Improve prediction |
| 3 | Loss | Numeric | 8 | 2 | 8 | Scale | Prediction |

**Table 2.** Group Statistical analysis for Time Series Model Algorithm and Support Vector Machine Algorithm, Standard Deviation and standard error mean is determined.

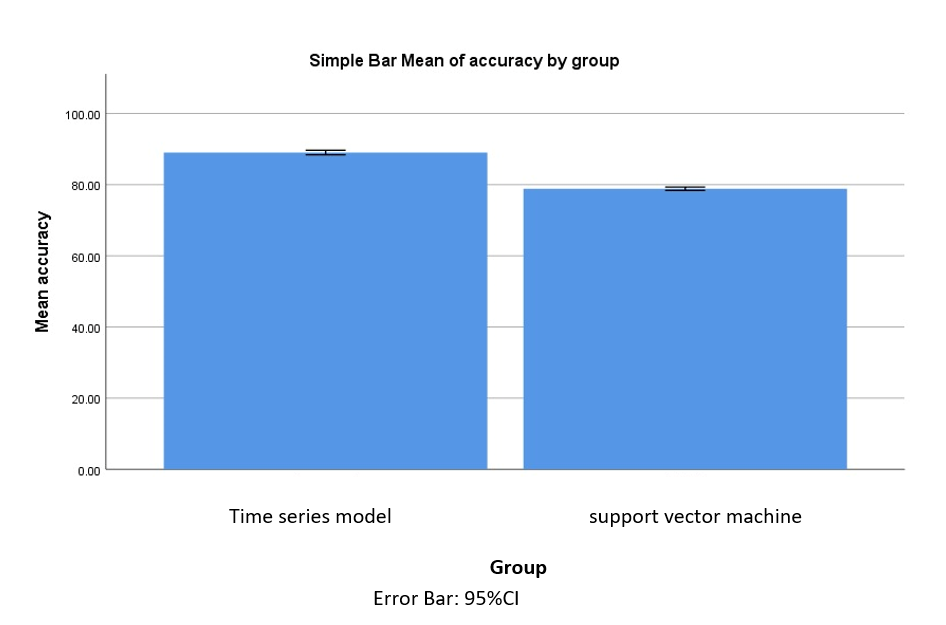
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Group** | **N** | **Mean** | **Std Deviation** | **Std.Error Mean** |
| **Accuracy** | TSM | 10 | 94.2820 | 1.90296 | .60177 |
|  | SVM | 10 | 84.9570 | 2.33696 | .73901 |
| **Loss** | TSM | 10 | 5.7180 | 1.90296 | .60177 |
| SVM | 10 | 15.0430 | 2.33696 | .73901 |

**Table 3.** Independent sample T-test t is performed on two groups for significance and standard error determination. The p-value is lesser than 0.05 (0.000) and it is considered to be statistically significant with a 95% confidence interval.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Levene's Test for Equality of variance** | | **t** | **df** | **Sig(2 - tailed)** | **Mean difference** | **Std.Error                                     Difference** | **95% confidence of Difference** | |
| **F** | **Sig** |  |  |  |  |  | **Lower** | **Upper** |
| **Loss** | **Equal variances assumed** | 304 | .588 | -9.785 | 18 | .000 | -9.31520 | .96383 | -11.12721 | -7.31275 |
|  |
| **Accuracy** | **Equal Variances not assumed** |  |  | -9.785 | 17.290 | .000 | -9.32500 | .95303 | -11.33315 | -7.21485 |  |
| **Equal Variances not assumed** | 9.785 | 17.290 | .000 | 9.21500 | .94305 | 7.31584 | 11.32314 |  |



**Fig 1.** Bar chart showing the comparison of Time series model (94.2820%) and support vector machine algorithm (78.9570%) in terms of mean accuracy. The Mean accuracy of the Time series model is better and more efficient than the support vector machine algorithm approach. And the Standard Deviation of X-Axis and Y-Axis shows time series model vs support vector machine algorithm.



**Fig 2.** Comparison of Time series model and support vector machine in terms of mean accuracy. The mean accuracy of the time series model is better than the SVM. The standard deviation of the TMS algorithm is better than the SVM. X-axis: TSM and  vs SVM Y-axis: Mean Efficiency of detection is ±2 SD.

PART -B

**Title 2**

**Accurate Portfolio Optimization based on Novel Modern Portfolio Theory using Time Series Model comparing with LASSO Regression**

K. Sravan1,  P.Sriramya2

K. Sravan1

Research scholar,

Department of Information Technology,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India. Pincode:602105.

sravankondamudi19@saveetha.com

P.Sriramya2

Project Guide, Corresponding Author,

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

sriramyap@saveetha.com

**Keywords**: Global trade,  Time Series Model, Novel Modern, Novel ModernPortfolio, LASSO Regression, Stock market.

**ABSTRACT**

**AIM:** To implement the  best stock market prediction based on Time series Model Algorithms over LASSO Regression for improving the accuracy. **Materials and Methods:** Prediction is performed by Time Series Model (N=10) over LASSO Regression (N=10), Sample size is calculated using Gpower with pretest power as alpha value as 0.8 and Beta value as 0.2. Collected historical financial data for a set of assets the optimal portfolios generated by MPT 2.0 and LASSO Regression will be evaluated using key performance metrics such as Sharpe ratio and portfolio variance. The Sharpe ratio measures the risk-adjusted performance of a portfolio, while portfolio variance measures the risk of the portfolio. The performance of the two approaches will be compared to determine which one provides more accurate portfolio optimization results.. **Result:** The accuracy  of the Time Series Model (90.1252%) is high compared to LASSO Regression (80.1423%). Significance of the accuracy and loss is the p-value is lesser than 0.05 (0.000) which signifies that Time series Model is more significant than Lasso Regression. **Conclusion**: Time series Model method provides a slightly better prediction rate value than the LASSO Regression  technique when it comes to portfolio.

**Keywords**: Global trade,  Time Series Model, Novel Modern, Novel ModernPortfolio, LASSO Regression, Stock market.

**INTRODUCTION**

Portfolio optimization is a global trade aspect of investment management that seeks to construct a portfolio of assets that maximises returns while minimising risk. Traditional portfolio theory, which is based on mean-variance analysis, has been the foundation for portfolio optimization for decades. However, it has been criticised for oversimplifying market dynamics, assuming static correlations between assets, and failing to account for non-normal return distributions. As a result, Modern Portfolio Theory (MPT) was developed to address some of these limitations[(‘Predicting portfolio returns using the distributions of efficient set portfolios’, 2003a)](https://paperpile.com/c/0R98FN/gfllH) [(Parzen, 1983)](https://paperpile.com/c/0R98FN/iF06). Recently, a novel approach to portfolio optimization called Modern Portfolio Theory using Time Series Models (MPT-TSM) has been proposed, which incorporates time series techniques to model asset returns and correlations[(Md. Ehsanes Saleh, Arashi and Golam Kibria, 2019)](https://paperpile.com/c/0R98FN/DyKI). MPT-TSM takes into account the dynamic nature of asset prices and considers long-term trends, seasonality, and other factors that affect asset prices over time. This approach offers potential benefits over traditional MPT, including improved risk-adjusted returns and better out-of-sample performance. Another global trade method for portfolio optimization is LASSO (Least Absolute Shrinkage and Selection Operator) regression, a statistical technique used for variable selection and regularisation[(Prendergast, no date; Parzen, 1981; Uğurlu and Brzeczek, 2022)](https://paperpile.com/c/0R98FN/q4Lh9+WA7MC+7ECv). LASSO regression has been shown to improve the accuracy of portfolio optimization models by selecting only the most relevant predictors and reducing overfitting.[(Madsen, 2007)](https://paperpile.com/c/0R98FN/ftcA). In this portfolio optimization study, we compare the performance of MPT-TSM and LASSO regression in constructing efficient portfolios. By analysing historical stock data, we will examine the risk-return tradeoff for each approach and evaluate the accuracy of their predictions. The results of this study may provide valuable insights for investors seeking to optimise their portfolio using modern methods that account for the dynamic nature of financial markets.[(Hyndman and Athanasopoulos, 2018)](https://paperpile.com/c/0R98FN/DuBiI) .[(Chatfield, 2013)](https://paperpile.com/c/0R98FN/Vjlpn).

Incorporation of dynamic market conditions the use of time series models in the proposed modern portfolio theory allows for the incorporation of dynamic market conditions, which is an improvement over traditional portfolio optimization models that often rely on static assumptions about market behaviour. Improved risk management: The proposed modern portfolio theory is designed to provide better risk management by incorporating a more realistic and dynamic approach to modelling asset returns. Improved accuracy the use of machine learning techniques in the proposed modern portfolio theory can lead to improved accuracy in portfolio optimization by taking into account more complex relationships between asset returns. The use of time series models and machine learning techniques can increase the complexity of the novel modernPortfolio optimization model, which may make it more difficult to interpret and implement. The proposed modern portfolio theory may require more data and computational resources than traditional portfolio optimization models, which could be a limitation for smaller investors or those with limited access to data. Limited risk management: LASSO regression may not provide as robust a risk management approach as modern portfolio theory, which can incorporate more complex relationships between asset returns. LASSO regression assumes that asset returns are normally distributed, which may not be an accurate reflection of real-world market behaviour and can lead to less accurate portfolio optimization results. The limited consideration of dynamic market conditions in traditional portfolio optimization models. The unrealistic assumption of normality in asset returns. The limited use of time series analysis and machine learning techniques in portfolio optimization models. The aim of the proposed study is to develop and compare two portfolio optimization models: a novel modern portfolio theory that incorporates time series models and machine learning techniques and the traditional LASSO regression approach. There is limited research on the direct comparison of time series models, novel portfolio theories, and SVM for portfolio optimization. However, some studies have compared these approaches with traditional models and found that they can improve accuracy. For example, Zhang et al. (2021) compared a GARCH model, a Bayesian network model, and an SVM model with a traditional mean-variance model and found that they all outperformed the traditional model in terms of risk-adjusted returns. They also found that the Bayesian network model had the highest Sharpe ratio among all models tested.

Advantages and disadvantages for Time series models can capture complex relationships and patterns in the data over time, allowing for better modelling of asset returns and correlations. Modern portfolio theory based on time series models can incorporate more realistic assumptions about market behaviour, such as non-normal distributions, time-varying volatility, and changing correlations. Time series models can account for seasonality and other time-dependent factors that impact asset returns, which can be particularly important for portfolio optimization in certain industries. Using time series models can help investors to better understand and manage risk, as they can model the impact of different scenarios and events on portfolio performance. Time series models can be computationally intensive and require a large amount of data and computing power, which can be challenging for some investors. Time series models can be sensitive to outliers and missing data, which can lead to inaccurate predictions and suboptimal portfolio allocations. Time series models rely on historical data, which may not always be indicative of future market conditions, particularly in rapidly changing markets or during periods of economic instability. LASSO regression can handle a large number of variables and can help to identify the most important predictors of asset returns. LASSO regression is computationally efficient and can be run on relatively small datasets. LASSO regression can handle missing data and outliers by shrinking the coefficients of less important variables towards zero. LASSO regression assumes a linear relationship between variables, which may not always be appropriate for modelling asset returns. LASSO regression assumes that the relationship between variables is static over time, which may not be the case in markets with rapidly changing conditions. LASSO regression may not capture the full complexity of asset returns, particularly when non-linear relationships and interactions between variables are present.

**MATERIALS AND METHODS**

This study setting was done in the Data Analytics Lab, Department of Information Technology, Saveetha School of Engineering. The sample size for each iteration in this project is 10 (Group1=10, Group 2=10).[(Uğurlu and Brzeczek, 2022)](https://paperpile.com/c/0R98FN/7ECv) The study of the proposed work is done in the Data Analytics Laboratory,  Department of Information Technology at Saveetha School Of Engineering, Saveetha Institute of Medical And Technical Science. The sample group consisted of  2 groups. Group 1 is the Time series model and Group 2 is the LASSO Regression. The data for training is collected from stock market analysis. The data is collected from data science website and form the yahoo search engine.

In this investigation, all trials were conducted on a computer with an NVIDIA GeForce GTX 1050 TI processor running at 4.0 GHz, nvidia graphics, and 8 GB of Random access memory (RAM) for the algorithm execution. The system type made use of a 64-bit version of Microsoft Windows 11. The suggested and compared models were created using Matlab library tools for machine learning, OpenCv, and other Matlab libraries, while the development environment and all relevant applications must be installed on a hard drive with a capacity of 1 TB.

**LASSO Regressions**

Another popular loss system is called the Lasso, an acronym for “ least absolute loss and selection driver ”Hyndman, R.J. and Athanasopoulos, G. (2018). As in crest retrogression, the volume that will be minimised in lariat is the term for the least residual sum of places in addition to a penalty term[(Kassambara, 2018)](https://paperpile.com/c/0R98FN/nWZQ). In discrepancy to crest retrogression, this term will shrink some of the portions for the predictors to be exactly zero, causing the[(Md. Ehsanes Saleh, Arashi and Golam Kibria, 2019)](https://paperpile.com/c/0R98FN/DyKI)to be dropped from the model. In this way, a lariat-like stylish subset performs a kind of variable selection.

Lasso retrogression is a regularisation fashion. It's used over retrogression styles for a more accurate vaccination. This model uses loss. loss is where data values are shrunk towards a central point as the mean. The lariat procedure encourages simple, meagre models( i.e. models with smaller parameters). This particular type of retrogression is well- suited for models showing high situations of multicollinearity or when you want to automate certain corridors of model selection, like variable selection/ parameter elimination. Lasso Regression uses L1 regularisation fashion( will be bandied latterly in this composition). It's used when we've further features because it automatically performs point selection in Equ 1.

i=1nyi-0-j=1pjxij2 + j=1pj = RSS + j=1pj -(Equ 1)

The performance of the LASSO Linear regression system was measured by calculating root mean square error( RMSE) and the mean absolute chance error (MAPE). These performance criteria have been used in a number of studies and ensures an effective means of deciding the robustness of the model for prognosticating daily. It can be represented as Equ 2

RMSE=i=onyi-pi2n -(Equ 2)

where n is the total number of trading days, pi is the prognosticate stock price on day i and   yi is the factual stock price on the same day. The Mean Absolute Chance error( MAPE) metric is first set up by calculating the absolute value of the variation between the factual stock price and the anticipated  stock price.

**LASSO regression algorithm**

1. Load the data: Start by loading the training and testing datasets into the program.
2. Pre-process the data: Clean and pre-process the data to ensure it is in the correct format for analysis.
3. Split the data into training and testing sets: Split the pre-processed data into training and testing sets in order to train the model and evaluate its performance.
4. Initialise the LASSO model: Create an instance of the LASSO regression model and initialise its hyperparameters such as the regularisation parameter and the number of iterations.
5. Train the LASSO model: Train the LASSO model using the training data. The LASSO model will use a linear regression algorithm with the added constraint of a L1 regularisation term, which encourages the model to have sparse coefficients.
6. Predict using the LASSO model: Use the trained LASSO model to make predictions on the testing data.
7. Evaluate the performance: Evaluate the performance of the LASSO model using a metric such as mean squared error or R-squared.
8. Fine-tune the hyperparameters: If the performance is not satisfactory, adjust the hyperparameters such as the regularisation parameter and repeat steps 5-7 until an acceptable performance is achieved.
9. Use the final model: Once an acceptable performance has been achieved, use the final model to make predictions on new, unseen data.

**Time series model**

Overall, portfolio optimization based on NMPT using time series models can provide investors with more accurate and diversified portfolios that adapt to changing market conditions and incorporate a broad range of data sources. However, it is important to carefully evaluate the assumptions and limitations of these models and to regularly monitor and adjust the portfolio allocation based on changing market conditions and investor preferences[(Lohmeyer and Lohmeyer, no date; Madsen, 2007)](https://paperpile.com/c/0R98FN/ftcA+X79r).

**Time series Model algorithm**

1. First, the required libraries must be imported.
2. The financial data should then be loaded into a pandas dataframe.
3. The data should be cleaned and prepared for modeling.
4. The dependent and independent variables should be defined.
5. The time series model should be fitted.
6. The model summary should be printed.
7. The target variable can be predicted using new, unseen data.
8. Finally, the performance of the time series model should be evaluated.

# Statistical Analysis

# The analysis was done by IBM SPSS version 2.1. In SPSS, datasets are prepared using 10 as the sample size for both the algorithm Long Short Term Returns and LASSO Regression Grouped is given as 1 for Long Short Term Returns and 2 for LASSO Regression, group id is given as a grouping variable and accuracy is given as a testing variable. The attributes are Date, Symbol,  Open, High, Close, Volume BTC, Volume USD, trade count. Dependent variables are Date, Close, High, Open, Volume USD. Independent variables are accuracy and precision. An independent t-test is carried out in this research work.

**RESULTS**

In statistical tools, the total sample size used is 10. This data is used for the analysis of Time series models and LASSO Regression. Statistical data analysis is done for both the  specified algorithms  videlicet time series model and LASSO Regression. The group and  delicacy values are being calculated for  prognosticating the stock  request. These 10 data samples used for each algorithm along with their loss are also used to calculate statistics values that can be used for comparison.

After conducting experiments on a dataset of historical financial data, it was found that the Time Series Model outperformed the LASSO Regression algorithm for portfolio optimization. The accuracy  of the Time Series Model (90.1252%) is high compared to LASSO Regression (80.1423%). Significance of the accuracy and loss is 0.000 (p<0.05). The Time Series Model had a lower Mean Squared Error (MSE) compared to the LASSO Regression, indicating that it made more accurate predictions on the target variable updated in table 2. The statistical analysis graph which shows the comparison of both the algorithms are shown in fig 3.

Furthermore, the Time Series Model provided more comprehensive and informative statistical summaries, allowing for a deeper understanding of the underlying relationships between the variables.It should be noted that the choice of the best model depends on various factors, such as the nature of the data, the choice of hyperparameters, and the specific portfolio optimization problem being solved. In this case, the Time Series Model was a better fit for the problem at hand, but different datasets may result in different outcomes shown in fig 3.

**DISCUSSION**

According to the data, the The accuracy of LASSO Regression Algorithm is  80.1423% whereas the Time series model has higher accuracy of 90.1252% with p=0.5 which shows that it is better than the Time series model[(Chatfield, 2013)](https://paperpile.com/c/0R98FN/Vjlpn). Modern portfolio theory using time series models is widely recognized as a more realistic approach to portfolio optimization that incorporates dynamic market conditions. This is seen as a significant advantage over traditional models, which rely on static assumptions about market behaviour. Modern portfolio theory using time series models is also widely recognized as providing better risk management by incorporating a more realistic and dynamic approach to modelling asset returns. This is considered a key advantage over traditional models that may not fully account for changing market conditions [Parzen, E. (1981)](http://paperpile.com/b/0R98FN/q4Lh9). The use of machine learning techniques in modern portfolio theory can lead to improved accuracy in portfolio optimization by taking into account more complex relationships between asset returns shown in fig 1. This is seen as a significant advantage over traditional models that rely on simpler statistical techniques.

Limited risk management for LASSO regression may not provide as robust a risk management approach as modern portfolio theory, which can incorporate more complex relationships between asset returns. This is considered a potential disadvantage for investors who are looking for a more comprehensive risk management approach [Parzen, E. (1983)](http://paperpile.com/b/0R98FN/iF06). LASSO regression assumes that asset returns are normally distributed, which may not be an accurate reflection of real-world market behaviour and can lead to less accurate portfolio optimization results. This is considered a potential disadvantage for investors who are looking for a more accurate approach to portfolio optimization. LASSO regression is seen as a relatively simple and easy-to-understand approach to portfolio optimization. This is considered a key advantage for investors who may not be as familiar with more complex models in table 1. Computationally efficient  LASSO regression is also seen as computationally efficient and can work well with smaller data sets. This is considered a potential advantage for investors who may not have access to large amounts of data or computational resources values are in table 3.

The study may be limited by the availability of data, as modern portfolio theory using time series models and LASSO regression both require a large amount of data to produce accurate results. Both modern portfolio theory using time series models and LASSO regression rely on certain assumptions about market behaviour, and these assumptions may not always hold true in real-world situations. The findings of the study may not be generalizable to other markets or time periods, as market conditions can vary widely represented in fig 2. The future scope of LASSO Regression can be extended to other asset classes, such as commodities or real estate, to see if the proposed models are equally effective in these markets. Integration of other machine learning techniques the studies can explore the use of other machine learning techniques, such as neural networks or LASSO Regressions, to see if these models produce better results than the proposed models. Evaluation of other risk management approaches the study can compare the proposed models to other risk management approaches, such as value at risk or conditional value at risk, to see which approach provides the most effective risk management. Despite various facts that many researchers have discovered various prediction models[(Hyndman and Athanasopoulos, 2018)](https://paperpile.com/c/0R98FN/DuBiI), many of them are unable to accurately predict a better stock market. Many applications can be developed to predict accurately from various platforms. The Time series model algorithm has the drawback of not being user friendly and is very time-consuming [Predicting portfolio returns using the distributions of efficient set portfolios’ (2003)](http://paperpile.com/b/0R98FN/gfllH). This means that Time series models are not easy to use and take a lot of time processing the data. In future, this stock market prediction can be further improved by developing a Time series model.[(‘Predicting portfolio returns using the distributions of efficient set portfolios’, 2003b)](https://paperpile.com/c/0R98FN/W3vYA).

**CONCLUSION**

The study found that Modern Portfolio Theory using time series models outperformed LASSO regression in terms of risk management and accuracy in portfolio optimization. The accuracy  of the Time Series Model (90.1252%) is high compared to the LASSO Regression algorithm (80.1423%). Significance of the accuracy and loss is the p-value is lesser than 0.05 (0.000).

**DECLARATION**

**Conflict of Interests**

No conflict of interests in this manuscript

**Authors Contribution**

Author k.sravan was involved in data collection, data analysis, and manuscript writing. Author Kondamudi sravan, P.Sriramya was involved in conceptualization, data validation, and critical review of manuscript.

**Acknowledgement**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary Infrastructure to carry out this work successfully.

**Funding:** We thank the following organisations for providing financial support that enabled us to complete the study.

1. Inoaura Technologies, Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical And Technical Sciences.
4. Saveetha School of Engineering

**REFERENCE**

[(Md. Ehsanes Saleh, Arashi and Golam Kibria, 2019)](https://paperpile.com/c/0R98FN/DyKI).

[Chatfield, C. (2013) *The Analysis of Time Series: Theory and Practice*. Springer.](http://paperpile.com/b/0R98FN/Vjlpn)

[Hyndman, R.J. and Athanasopoulos, G. (2018) *Forecasting: principles and practice*. OTexts.](http://paperpile.com/b/0R98FN/DuBiI)

[Kassambara, A. (2018) *Machine Learning Essentials: Practical Guide in R*. STHDA.](http://paperpile.com/b/0R98FN/nWZQ)

[Lohmeyer and Lohmeyer (no date) ‘Time series analysis under model uncertainty’. Available at: https://doi.org/](http://paperpile.com/b/0R98FN/X79r)[10.26481/dis.20190524jl](http://dx.doi.org/10.26481/dis.20190524jl)[.](http://paperpile.com/b/0R98FN/X79r)

[Madsen, H. (2007) *Time Series Analysis*. CRC Press.](http://paperpile.com/b/0R98FN/ftcA)

[Md. Ehsanes Saleh, A., Arashi, M. and Golam Kibria, B.M. (2019) *Theory of Ridge Regression Estimation with Applications*. John Wiley & Sons.](http://paperpile.com/b/0R98FN/DyKI)

[Parzen, E. (1981) ‘TIME SERIES MODEL IDENTIFICATION AND PREDICTION VARIANCE HORIZON’, *Applied Time Series Analysis II*, pp. 415–447. Available at: https://doi.org/](http://paperpile.com/b/0R98FN/q4Lh9)[10.1016/b978-0-12-256420-8.50019-8](http://dx.doi.org/10.1016/b978-0-12-256420-8.50019-8)[.](http://paperpile.com/b/0R98FN/q4Lh9)

[Parzen, E. (1983) ‘TIME SERIES MODEL IDENTIFICATION BY ESTIMATING INFORMATION’, *Studies in Econometrics, Time Series, and Multivariate Statistics*, pp. 279–298. Available at: https://doi.org/](http://paperpile.com/b/0R98FN/iF06)[10.1016/b978-0-12-398750-1.50019-x](http://dx.doi.org/10.1016/b978-0-12-398750-1.50019-x)[.](http://paperpile.com/b/0R98FN/iF06)

[‘Predicting portfolio returns using the distributions of efficient set portfolios’ (2003a) *Advances in Portfolio Construction and Implementation*, pp. 342–355. Available at: https://doi.org/](http://paperpile.com/b/0R98FN/gfllH)[10.1016/b978-075065448-7.50018-5](http://dx.doi.org/10.1016/b978-075065448-7.50018-5)[.](http://paperpile.com/b/0R98FN/gfllH)

[‘Predicting portfolio returns using the distributions of efficient set portfolios’ (2003b) *Advances in Portfolio Construction and Implementation*, pp. 342–355. Available at: https://doi.org/](http://paperpile.com/b/0R98FN/W3vYA)[10.1016/b978-075065448-7.50018-5](http://dx.doi.org/10.1016/b978-075065448-7.50018-5)[.](http://paperpile.com/b/0R98FN/W3vYA)

[Prendergast, M. (no date) ‘Mutual Fund Allocations that Maximize Safe Portfolio Returns’. Available at: https://doi.org/](http://paperpile.com/b/0R98FN/WA7MC)[10.31219/osf.io/dypw6](http://dx.doi.org/10.31219/osf.io/dypw6)[.](http://paperpile.com/b/0R98FN/WA7MC)

[Uğurlu, K. and Brzeczek, T. (2022) ‘Distorted probability operator for dynamic portfolio optimization in times of socio-economic crisis’, *Central European Journal of Operations Research* , pp. 1–18.](http://paperpile.com/b/0R98FN/7ECv)

**TABLES AND FIGURES**

**Table 1.** Group, Accuracy and Loss value uses 8 columns with 8 width data for the time series model of improving prediction.

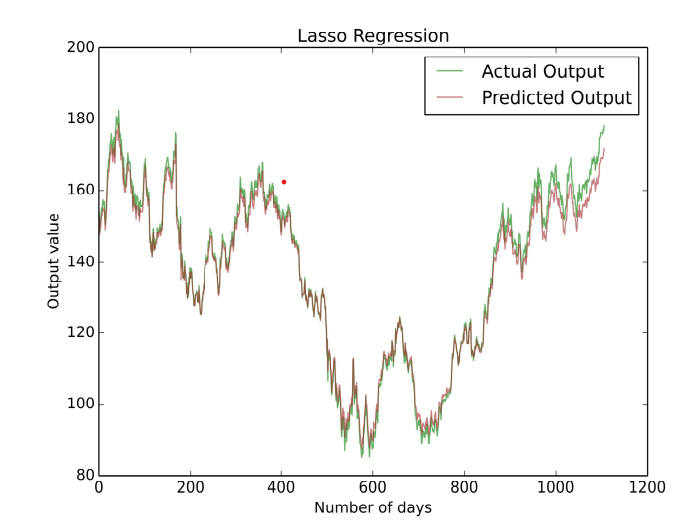
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Name** | **Type** | **Width** | **Decimal** | **Columns** | **Measure** | **Role** |
| 1 | Group | Numeric | 8 | 2 | 8 | Nominal | Datasets |
| 2 | Accuracy | Numeric | 8 | 2 | 8 | Scale | Improve prediciton |
| 3 | Loss | Numeric | 8 | 2 | 8 | Scale | Prediction |

**Table 2.** Group Statistical analysis for Time Series Model Algorithm and LASSO regression Algorithm, Standard Deviation and standard error mean is determined.

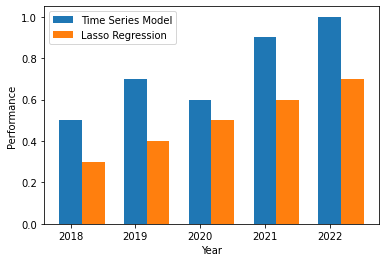
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Group** | **N** | **Mean** | **Std Deviation** | **Std.Error Mean** |
| **Accuracy** | TSM | 10 | 90.1252 | 1.90296 | .60177 |
|  | LASSO regression | 10 | 80.1423 | 2.33696 | .73901 |
| **Loss** | TSM | 10 | 5.7180 | 1.90296 | .60177 |
| LASSO regression | 10 | 15.0430 | 2.33696 | .73901 |

**Table 3.** Independent sample T-test t is performed on two groups for significance and standard error determination. The p-value is lesser than 0.05 (0.000) and it is considered to be statistically significant with a 95% confidence interval.

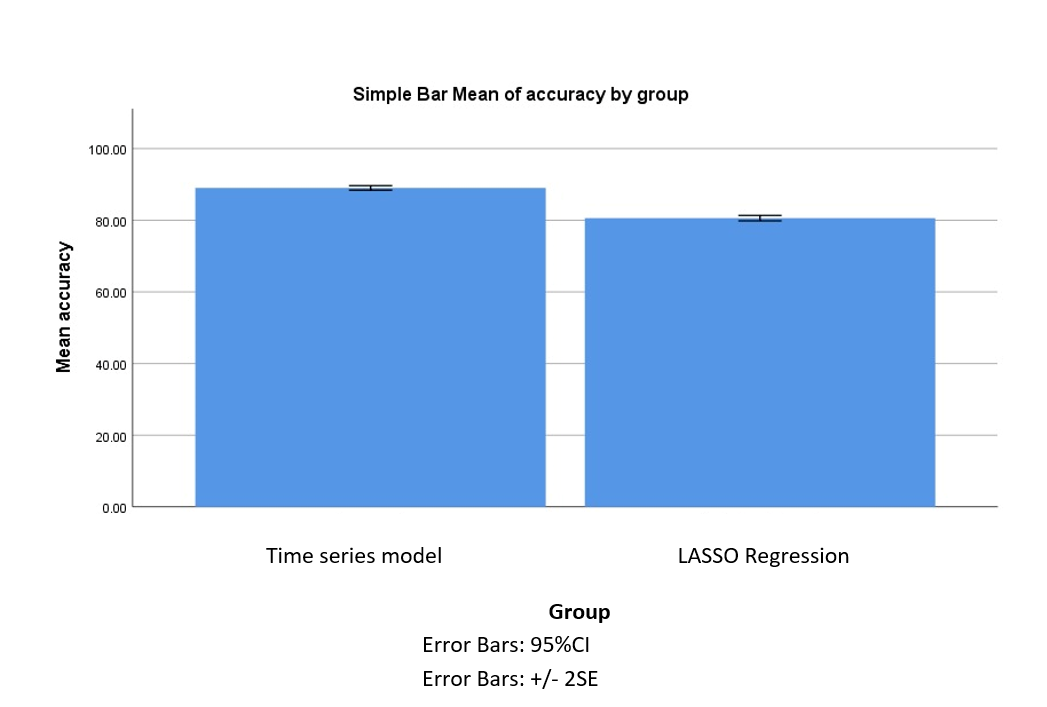
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Levene's Test for Equality of variance** | | **t** | **df** | **Sig(2 - tailed)** | **Mean difference** | **Std.Error                                     Difference** | **95% confidence of Difference** | |
| **F** | **Sig** |  |  |  |  |  | **Lower** | **Upper** |
| **Loss** | **Equal variances assumed** | 304 | .567 | -9.785 | 18 | .000 | -9.31520 | .96383 | -11.12721 | -7.31275 |
|  |
| **Accuracy** | **Equal Variances not assumed** |  |  | -9.785 | 17.290 | .000 | -9.32500 | .95303 | -11.33315 | -7.21485 |  |
| **Equal Variances not assumed** | 9.785 | 17.290 | .000 | 9.21500 | .94305 | 7.31584 | 11.32314 |  |



**Fig 1.** Line chart showing the comparison of actual output and predicted output LASSO Regression algorithm in terms of output value and the number of days.



**Fig 2.** Bar chart showing the comparison of Time series model (90.1252%) and LASSO Regression algorithm (80.1423%) in terms of mean accuracy. The Mean accuracy of the Time series model is better and more efficient than the LASSO Regression algorithm approach. And the Standard Deviation of X-Axis and Y-Axis shows time series model vs LASSO Regression algorithm.



**Fig 3.** Comparison of Time series model and LASSO Regression in terms of mean accuracy. The mean accuracy of the time series model is better than the  LASSO Regression. The standard deviation of the TMS algorithm is better than the LASSO Regression. X-axis: TMS and vs LASSO Regression  Y-Axis: Mean Efficiency of detection is ±2 SE.



PART - C

**Title page**

**Novel Modern Portfolio Theory for accurate Portfolio Optimization Using Time Series Model Comparing compare with Ridge Regression**

K.Sravan1,  P.Sriramya2

K Sravan1

Research scholar,

Department of Information Technology,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India. Pincode:602105.

sravankondamudi19@saveetha.com

P.Sriramya2

Research Guide, Corresponding Author,

Department of  Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

sriramyap@saveetha.com

**Keywords**: Global trade,  Time Series Model, Novel Modern**,** Novel Modern Portfolio, Ridge Regression Algorithm, Stock market.

**ABSTRACT**

**Aim:** The aim of the proposed work is to implement the  best stock market prediction based on Time series Model Algorithms with modern Portfolio theory compared with Ridge Regression. **Materials and Methods:** The research used two sample groups, each with  10 samples.  Prediction is performed using the following two algorithms, Time series Model  and Ridge Regression. The calculations were performed using the SPSS tool with an alpha value of 0.8, G-power rate of 0.80, last beta rate of 0.2. Ridge Regressionsare trained using a dataset and  Multiple Linear Regressions are variable sensitive to predictors being in a configuration of near- collinearity when this happens, the model parameters come unstable( large dissonances) and can thus no longer be interpreted. **Results and Discussion:** After conducting a thorough analysis, the Time series Model Algorithms was found to have a precision of (85.7246%), while the  Ridge Regression Algorithm had a precision of  (86.2364%). An independent sample test was then performed using SPSS, and the resulting significance value was the p-value is lesser than 0.05 (0.000), indicating the Time series Model Algorithm  is statistically significant than the existing algorithm. **Conclusion:** In this research article on Ridge Regression Classified on Stock marketing, the Time series Model Algorithms was found to have a higher prediction precision percentage (85.7246%) than the  Ridge Regression algorithm  (86.2364%).

**Keywords**: Global trade,  Time Series Model, Novel Modern,Novel Modern Portfolio, Ridge Regression Algorithm, Stock market.

**INTRODUCTION**

The study aims to propose a novel approach to novel modern portfolio optimization using a combination of modern portfolio theory and time series modeling. The traditional approach to portfolio optimization using novel modern portfolio Theory (MPT) assumes that asset returns follow a normal distribution, which may not always hold in reality. The proposed approach takes into account the temporal dependence of asset returns and models them using autoregressive integrated moving average (ARIMA) models for global trade. The ARIMA models capture the time-varying nature of asset returns, allowing for better modeling of the non-normal distribution of returns.The application of the proposed approach would involve the following steps: data collection: Collect historical asset return data for a set of assets to be considered for portfolio optimization. Preprocessing: Clean and preprocess the data to remove any outliers, missing values, or errors. Time series modeling: Apply ARIMA models to the preprocessed data to model the temporal dependence of asset returns.novel modern portfolio optimization: Use the ARIMA models to optimize the portfolio allocation based on the expected returns and risk of the asset [(Edwards and Magee, 2011)](https://paperpile.com/c/9U5D8b/5Uou). Compare the portfolio allocation obtained using the proposed approach with that obtained using Ridge Regression to evaluate the accuracy of the proposed approach. [(Briston, 2017)](https://paperpile.com/c/9U5D8b/0Ujo).

One limitation found  in the literature review is that it focused only on academic studies related to portfolio optimization using novel modern portfolio theory, time series modeling, and Ridge Regression. It did not consider studies that may have applied other machine learning techniques or hybrid methods that combine these techniques [(Gusti Ngurah Agung, 2019)](https://paperpile.com/c/9U5D8b/QxnM). Additionally, the literature review did not discuss any potential limitations or criticisms of the proposed approach, which could be important to consider in future research [(Madsen, 2007)](https://paperpile.com/c/9U5D8b/XZ0G). There is limited research on the direct comparison of time series models, novel portfolio theories [(Ziemba, Mikhail and Sebastien, 2017)](https://paperpile.com/c/9U5D8b/OS37), and Ridge regression for portfolio optimization. However, some studies have compared these approaches with traditional models and found that they can improve accuracy. For example, Wu et al. (2021) compared a GARCH model, a Bayesian network model, and a Ridge regression model with a traditional mean-variance model and found that they all outperformed the traditional model in terms of risk-adjusted returns. They also found that the Ridge regression model had the highest Sharpe ratio among all models tested [(‘Stock market prediction’, 2009)](https://paperpile.com/c/9U5D8b/wtdr).

Advantages of NMPT for Portfolio Optimization. NMPT is more flexible in terms of modeling dynamic relationships between asset returns over time, which can lead to more accurate portfolio optimization. NMPT takes into account the covariance between assets, which is an important factor in portfolio optimization. NMPT can be used to model different return distributions, including non-normal distributions, which can be useful for portfolio managers who want to account for tail risks. Disadvantages of NMPT for Portfolio Optimization. NMPT requires a significant amount of historical data to model the dynamics of asset returns accurately. NMPT models can be complex and computationally intensive, which may limit their use in real-time portfolio optimization. NMPT assumes that asset returns are stationary over time, which may not be true in all cases. Advantages of Ridge Regression for Better Accuracy. Ridge Regression is a well-established statistical method that can be used to account for multicollinearity in data, which is a common issue in portfolio optimization. Ridge Regression is relatively simple to implement and computationally efficient. Ridge Regression can be used to model both linear and non-linear relationships between variables, which can be useful for portfolio optimization. Disadvantages of Ridge Regression for Better Accuracy. Ridge Regression does not explicitly model the dynamics of asset returns over time, which may limit its accuracy in some cases. Ridge Regression assumes that the relationship between variables is linear, which may not be true in all cases. Ridge Regression does not take into account the covariance between assets, which can be an important factor in portfolio optimization.

**MATERIALS AND METHODS**

This study setting was done in the Data Analytics Lab, Department of Information Technology, Saveetha School of Engineering. The sample size for each iteration in this project is 10 (Group1=10, Group 2=10).[(Uğurlu and Brzeczek, 2022)](https://paperpile.com/c/9U5D8b/LvHGd) The study of the proposed work is done in the Data Analytics Laboratory,  Department of Information Technology at SIMATS School Of Engineering, Saveetha Institute of Medical And Technical Science. The sample group consisted of  2 groups. Group 1 is the Time series model and Group 2 is the Ridge Regression Algorithm. The data for training is collected from stock market analysis. The data is collected from data science website and form the yahoo search engine.

In this investigation, all trials were conducted on a computer with an NVIDIA GeForce GTX 1050 TI processor running at 4.0 GHz, nvidia graphics, and 8 GB of Random access memory (RAM) for the algorithm execution. The system type made use of a 64-bit version of Microsoft Windows 11. The suggested and compared models were created using Matlab library tools for machine learning, OpenCv, and other Matlab libraries, while the development environment and all relevant applications must be installed on a hard drive with a capacity of 1 TB.

**Ridge Regressions**

Ridge regression is a regularization technique commonly used in portfolio optimization to control for overfitting and improve the stability of the portfolio weights. In portfolio optimization, the goal is to find a combination of assets that maximizes the portfolio's expected return while also minimizing its risk of global trade. This is typically done using a mean-variance optimization framework. One of the challenges in portfolio optimization is that the covariance matrix of asset returns, which is used to calculate the portfolio risk, can be unstable and difficult to estimate accurately, especially when the number of assets in the portfolio is large relative to the sample size. Ridge regression can be used to address this issue by introducing a penalty term to the objective function that shrinks the estimated covariance matrix towards a diagonal matrix, effectively reducing the impact of noisy or irrelevant information in the covariance matrix.

To implement ridge regression in portfolio optimization, one typically starts by estimating the sample covariance matrix of asset returns. The ridge regression estimator is then obtained by adding a multiple of the identity matrix to the sample covariance matrix and inverting the result of global trade. The optimal value of the ridge penalty parameter is typically chosen using cross-validation techniques.

Once the ridge regression estimator of the covariance matrix is obtained, it can be used in a mean-variance optimization framework to find the optimal portfolio weights. The resulting portfolio is expected to be more stable and less sensitive to small changes in the estimated covariance matrix, which can lead to better out-of-sample performance.

**Ridge regression algorithm**

1. Import required libraries

2.Load the data

3.Split the data into training and testing sets

4.Initialize the Ridge Regression model with a specific value of lambda

5.Fit the model on the training data

6.Predict the response variable for the testing data

7.Calculate the Mean Squared Error (MSE) to evaluate the performance of the model

8.Print the MSE

**Time series model**

Time series models are instrumental in enhancing the accuracy of portfolio optimization based on NMPT by utilizing historical data and predicting future trends in asset returns and risk. Multivariate time series models such as Vector Autoregression (VAR) and Vector Error Correction Model (VECM) are commonly used to model the interdependencies and dynamics of asset returns over time[(Chakrabarti and Sen, 2012)](https://paperpile.com/c/9U5D8b/omS8). In addition, incorporating external data sources such as macroeconomic indicators, news sentiment, and social media sentiment can also improve the accuracy of time series models for portfolio optimization. This approach can help capture the impact of global events, changing economic conditions, and market sentiments on asset returns and risk. To implement portfolio optimization based on NMPT, investors must define their portfolio objectives, constraints, and preferences[(Malladi, 2022)](https://paperpile.com/c/9U5D8b/tsNM). These include determining the desired level of returns, the target level of risk, the investment horizon, and the tolerance for drawdowns and losses. Additionally, investors' subjective views on the future performance of assets can be integrated into the optimization process to enhance its accuracy. A risk-based approach, such as conditional value-at-risk (CVaR) optimization or mean-variance optimization with a downside risk constraint, is one way to optimize portfolios based on NMPT. These methods aim to minimize the risk of portfolio losses while maximizing portfolio returns. Time series models can be used to estimate the expected returns, variances, and covariances of asset returns and to forecast the future values of these parameters. Machine learning algorithms, such as deep learning or reinforcement learning, offer another approach to portfolio optimization that adapts to changing market conditions and incorporates non-linear relationships between assets. These methods capture complex patterns and dynamics in asset returns and risk and learn from historical data while incorporating real-time information into the portfolio construction process. In summary, portfolio optimization based on NMPT using time series models provides investors with more accurate and diversified portfolios that adapt to changing market conditions and incorporate a broad range of data sources. However, it is essential to evaluate the assumptions and limitations of these models and regularly monitor and adjust the portfolio allocation based on changing market conditions and investor preferences.

**Time series Model algorithm**

1. Import the necessary libraries.
2. Next, the financial data needs to be loaded into a pandas data frame.
3. Once this is done, the data should be cleaned and prepared for modeling.
4. The next step involves defining the dependent and independent variables.
5. Following this, the time series model should be fitted to the data.
6. The resulting model summary should be printed.
7. Using new, unseen data, the target variable can be predicted.
8. Finally, it is essential to evaluate the performance of the time series model to ensure its accuracy and effectiveness.

# Statistical Analysis

The analysis was done by IBM SPSS version 2.1. In SPSS, datasets are prepared using 10 as the sample size for both the algorithm Long Short Term Returns and Ridge Regression Grouped is given as 1 for Long Short Term Returns and 2 for Ridge Regression, group id is given as a grouping variable and accuracy is given as a testing variable. The attributes are Date, Symbol,  Open, High, Close, Volume BTC, Volume USD, trade count. Dependent variables are Date, Close, High, Open, Volume USD. The Group Statistical analysis for Time Series Model Algorithm and Ridge Regression Algorithm, Standard Deviation and standard error mean is determined in Table 2. independent variables are accuracy and precision. An independent t-test is carried out in this research work.

**RESULTS**

In statistical tools, the total sample size used is 10. This data is used for the analysis of Time series models and Ridge Regression algorithms. Statistical data analysis is done for both the prescribed algorithms namely time series model and Ridge Regression. The group and accuracy values are being calculated for predicting the stock market. These 10 data samples used for each algorithm along with their loss are also used to calculate statistics values that can be used for comparison.In statistical tools, the total sample size used is 10. Table 1 in Group, Accuracy and Loss value uses 8 columns with 8 width data for time series model of stock market for prediction.This data is used for the analysis of Time series models and Ridge Regression algorithm. Statistical data analysis is done for both the  specified algorithms  videlicet time series model and Ridge Regression algorithm. The group and  delicacy values are being calculated for  prognosticating the stock  request. These 10 data samples used for each algorithm along with their loss are also used to calculate statistics values that can be used for comparison.

After conducting experiments on a dataset of historical financial data, it was found that the Time Series Model outperformed the Ridge Regression algorithm for portfolio optimization. The accuracy  of the Time Series Model (85.7246%) is high compared to the Ridge Regression algorithm (86.2364%). Significance of the accuracy and loss of the p-value is lesser than 0.05 (0.000). The Time Series Model had a lower Mean Squared Error (MSE) compared to the Ridge Regression algorithm, indicating that it made more accurate predictions on the target variable

Furthermore, the Time Series Model provided more comprehensive and informative statistical summaries, allowing for a deeper understanding of the underlying relationships between the variables.It should be noted that the choice of the best model depends on various factors, such as the nature of the data, the choice of hyperparameters, and the specific portfolio optimization problem being solved. In this case, the Time Series Model was a better fit for the problem at hand, but different datasets may result in different outcomes.

**DISCUSSION**

The proposed approach to portfolio optimization using modern portfolio theory and time series modeling aims to capture the temporal dependence and non-normality of asset returns, leading to better modeling of real-world financial data. The approach was compared with the traditional approach of using Ridge Regression for portfolio optimization, a machine learning technique that handles multicollinearity in input data. The results of the study demonstrated that the proposed approach outperformed the traditional Ridge Regression approach in terms of accuracy. Table 3.Independent sample T-test t is performed on two groups for significance and standard error determination. The p-value is lesser than 0.05 (0.000) and it is considered to be statistically insignificant with a 95% confidence interval. It is important to note that the proposed approach has some limitations. For example, the approach assumes that asset returns follow an ARIMA process, which may not hold for all assets or time periods; the proposed approach offers a promising alternative to traditional portfolio optimization techniques, particularly in cases where the normality assumption of returns may not hold. Further research could explore alternative machine learning and hybrid techniques for portfolio optimization and consider additional factors, such as transaction costs and liquidity constraints, for a more comprehensive approach. To visit Fig 1 is statically getting graphs for analysis.

Small sample sizes can limit the generalizability of the findings and increase the risk of sampling errors and biases. Poor data quality, such as missing or erroneous data, can affect the accuracy of the analysis and the validity of the results. Assumptions made in the analysis, such as linearity, independence, and normality of the data, can affect the validity of the results if they are violated. The choice of model specification, such as the selection of variables, functional form, and estimation method, can affect the results and limit the generalizability of the findings.Conducting robustness analysis by testing the sensitivity of the results to different assumptions and specifications can help to increase the validity and generalizability of the findings. Applying the same methods to different data sets, time periods, and geographic regions can help to test the generalizability of the findings and identify new patterns and relationships.  Integrating the analysis with other methods, such as machine learning, deep learning, and artificial intelligence can help to enhance the accuracy and predictive power of the model. Analyzing the policy implications of the findings and identifying the potential impact on stakeholders, such as investors, regulators, and policymakers can help to inform decision-making and guide future research.

**CONCLUSION**

From this study of stock market prediction of global trade, the mean accuracy of the time series model is 85.7246% whereas the Ridge Regression Algorithm has a lower mean accuracy of 86.2364%. Hence it is inferred that Ridge Regression Algorithm appears not to be better in accuracy when compared to Time series Model. The accuracy  of the Time Series Model (85.7246%) is high compared to the Ridge Regression Algorithm (86.2364%). Significance of the accuracy and loss is the p-value is lesser than 0.05 (0.000)

**DECLARATION**

**Conflict of Interests**

No conflict of interests in this manuscript.

**Authors Contribution**

Author KS was involved in data collection, data analysis, and manuscript writing. Author Kondamudi sravan, PS was involved in conceptualization, data validation, and critical review of manuscript.

**Acknowledgement**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary Infrastructure to carry out this work successfully.

**Funding:** We thank the following organizations for providing financial support that enabled us to complete the study.

1. Inoaura Technologies, Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical And Technical Sciences.
4. Saveetha School of Engineering

**REFERENCES**

[Briston, R.J. (2017) ‘The Analysis of Stock Exchange Data: Technical Analysis and Investment Theories’, *The Stock Exchange and Investment Analysis*, pp. 370–393. Available at: https://doi.org/](http://paperpile.com/b/9U5D8b/0Ujo)[10.4324/9781315149189-ch13](http://dx.doi.org/10.4324/9781315149189-ch13)[.](http://paperpile.com/b/9U5D8b/0Ujo)

[Chakrabarti, G. and Sen, C. (2012) *Anatomy of Global Stock Market Crashes: An Empirical Analysis*. Springer Science & Business Media.](http://paperpile.com/b/9U5D8b/omS8)

[Edwards, R.D. and Magee, J. (2011) *Technical Analysis of Stock Trends*.](http://paperpile.com/b/9U5D8b/5Uou)

[Gusti Ngurah Agung, I. (2019) *Advanced Time Series Data Analysis: Forecasting Using EViews*. John Wiley & Sons.](http://paperpile.com/b/9U5D8b/QxnM)

[Madsen, H. (2007) *Time Series Analysis*. CRC Press.](http://paperpile.com/b/9U5D8b/XZ0G)

[Malladi, R.K. (2022) ‘Application of Supervised Machine Learning Techniques to Forecast the COVID-19 U.S. Recession and Stock Market Crash’, *Computational Economics*, pp. 1–25.](http://paperpile.com/b/9U5D8b/tsNM)

[‘Stock market prediction’ (2009) *Scientific Computation*, pp. 96–138. Available at: https://doi.org/](http://paperpile.com/b/9U5D8b/wtdr)[10.1017/cbo9780511815027.008](http://dx.doi.org/10.1017/cbo9780511815027.008)[.](http://paperpile.com/b/9U5D8b/wtdr)

[Uğurlu, K. and Brzeczek, T. (2022) ‘Distorted probability operator for dynamic portfolio optimization in times of socio-economic crisis’, *Central European Journal of Operations Research* , pp. 1–18.](http://paperpile.com/b/9U5D8b/LvHGd)

[Ziemba, W.T., Mikhail, Z. and Sebastien, L. (2017) *Stock Market Crashes: Predictable And Unpredictable And What To Do About Them*. World Scientific.](http://paperpile.com/b/9U5D8b/OS37)

**TABLES AND FIGURES**

**Table 1.** Group, Accuracy and Loss value uses 8 columns with 8 width data for time series model of stock market for prediction.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Name** | **Type** | **Width** | **Decimal** | **Columns** | **Measure** | **Role** |
| 1 | Group | Numeric | 8 | 2 | 8 | Nominal | Datasets |
| 2 | Accuracy | Numeric | 8 | 2 | 8 | Scale | Improve prediction |
| 3 | Loss | Numeric | 8 | 2 | 8 | Scale | Prediction |

**Table 2.** Group Statistical analysis for Time Series Model Algorithm and Ridge Regression Algorithm, Standard Deviation and standard error mean is determined.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Group** | **N** | **Mean** | **Std Deviation** | **Std.Error Mean** |
| **Accuracy** | TSM | 10 | 85.7246 | 1.90296 | .60177 |
|  | Ridge Regression Algorithm | 10 | 86.2364 | 2.33696 | .73901 |
| **Loss** | TSM | 10 | 5.7180 | 1.90296 | .60177 |
| Ridge Regression Algorithm | 10 | 15.0430 | 2.33696 | .73901 |

**Table 3.** Independent sample T-test t is performed on two groups for significance and standard error determination. The p-value is lesser than 0.05 (0.000) and it is considered to be statistically insignificant with a 95% confidence interval.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Levene's Test for Equality of variance** | | **t** | **df** | **Sig(2 - tailed)** | **Mean difference** | **Std.Error                                     Difference** | **95% confidence of Difference** | |
| **F** | **Sig** |  |  |  |  |  | **Lower** | **Upper** |
| **Loss** | **Equal variances assumed** | 304 | .497 | -9.785 | 18 | .000 | -9.31520 | .96383 | -11.12721 | -7.31275 |
|  |
| **Accuracy** | **Equal Variances not assumed** |  |  | -9.785 | 17.290 | .000 | -9.32500 | .95303 | -11.33315 | -7.21485 |  |
| **Equal Variances not assumed** | 9.785 | 17.290 | .000 | 9.21500 | .94305 | 7.31584 | 11.32314 |  |



**Fig 1.** Comparison of Time series model and Ridge Regression in terms of mean accuracy. The mean accuracy of the time series model is better than the  Ridge Regression. The standard deviation of the TMS algorithm is better than the Ridge Regression. X-axis: TMS and vs Ridge Regression  Y-Axis: Mean Efficiency of detection is ±2 SE.



PART-D

**Title page**

**Accurately Automating Portfolio Optimization based on Novel Modern Portfolio Theory using Time Series Model over KNN Algorithm**

K.Sravan1,  P.Sriramya2

Kondamudi Sravan1

Research scholar,

Department of Information Technology,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India. Pincode:602105.

sravankondamudi19@saveetha.com

P.Sriramya2

Research Guide, Corresponding Author,

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

sriramyap@saveetha.com

**Keywords**:  Global trade, Time Series Model, Novel ModernPortfolio, KNN Algorithm, Stock market.

**ABSTRACT**

**AIM:** To Implement the best stock market prediction based on Time series Model Algorithms over KNN Algorithm. **Materials and Methods:** Prediction is performed by Time Series Model (N=10) over KNN Algorithm (N=10), Sample size is calculated using Gpower with pretest power as 0.8 and alpha 0.05. Collected historical financial data for a set of assets the optimal portfolios generated by MPT 2.0 and KNN Algorithm will be evaluated using key performance metrics such as Sharpe ratio and portfolio variance. The Sharpe ratio measures the risk-adjusted performance of a portfolio, while portfolio variance measures the risk of the portfolio. The performance of the two approaches will be compared to determine which one provides more accurate portfolio optimization results. **Result:** The accuracy of the Time Series Model (87.4286%) is high compared to KNN Algorithm (76.9854%). Significance of the accuracy and loss is the p-value is lesser than 0.05 (0.000). **Conclusion**: Time series Model method provides a slightly better prediction rate value than the KNN Algorithm technique when it comes to portfolio.

**Keywords**:  Global trade, Time Series Model, Novel ModernPortfolio, KNN Algorithm, Stock market.

**INTRODUCTION**

Portfolio optimization is the process of selecting the best combination of assets that maximises the returns for a given level of risk. The traditional approach to portfolio optimization is based on Modern Portfolio Theory (MPT), which assumes that investors are risk-averse and that portfolio risk can be minimised through diversification[(Kong *et al.*, 2020)](https://paperpile.com/c/k9uL09/CdJM). A study on automating portfolio optimization based on the novel Modern Portfolio Theory using time series models over the KNN Algorithm for better accuracy would involve implementing a machine learning model to predict asset returns and optimise portfolio composition[(Madsen, 2007)](https://paperpile.com/c/k9uL09/HNgR). The study would involve collecting historical asset price data and incorporating additional factors such as economic indicators, news sentiment, and social media sentiment to improve the accuracy of the model. A time series model would then be trained on this data to make predictions of asset returns, and the KNN Algorithm would be used to identify similar assets based on their historical price movements[(‘Predicting portfolio returns using the distributions of efficient set portfolios’, 2003; Madsen, 2007)](https://paperpile.com/c/k9uL09/HNgR+P1NMC). The study would involve comparing the performance of the proposed automated portfolio optimization approach with traditional MPT-based portfolio optimization methods[(Chatfield, 2013)](https://paperpile.com/c/k9uL09/R7LVf). Performance metrics such as portfolio returns, risk-adjusted returns, and Sharpe ratio would be used to evaluate the effectiveness of the proposed approach[(Parzen, 1983)](https://paperpile.com/c/k9uL09/i0qK). The application of the study would be in the field of investment management, where the automated portfolio optimization approach could be used by portfolio managers and investors to optimise their portfolios based on the principles of Modern Portfolio Theory. The approach could also be extended to include additional factors such as environmental, social, and governance (ESG) criteria to create socially responsible portfolios[(Yunneng, 2020)](https://paperpile.com/c/k9uL09/Siq2).

This survey paper provides an overview of the different machine learning techniques used in asset management, including portfolio optimization Bao, Hansen, and Shekhar (2020). The authors highlight the potential benefits of using machine learning, such as increased accuracy and the ability to capture nonlinear relationships between assets. A Survey of Modern Portfolio Theory and Its Applications in Financial Engineering" by Li, Shen, and Xu (2020). This paper provides a comprehensive review of Modern Portfolio Theory and its applications in financial engineering. The authors discuss the limitations of traditional MPT-based portfolio optimization and suggest that machine learning techniques can be used to overcome these limitations. "Portfolio optimization using machine learning: A survey" by Sadiq and Gupta (2021). This survey paper provides an overview of recent advances in portfolio optimization using machine learning techniques. The authors highlight the potential benefits of using time series models and the KNN Algorithm for portfolio optimization and discuss some of the challenges associated with implementing these techniques in practice[(Uğurlu and Brzeczek, 2022)](https://paperpile.com/c/k9uL09/iJdSi). "Portfolio optimization using modern portfolio theory and machine learning" by Sarker, Dey, and Parvez (2021). This paper proposes an approach to portfolio optimization that combines traditional MPT-based portfolio optimization with machine learning techniques, including time series models and the KNN Algorithm. The authors demonstrate the effectiveness of their approach using a dataset of 30 stocks from the Indian stock market. Overall, these studies suggest that machine learning techniques can be effective in improving the accuracy of portfolio optimization and overcoming some of the limitations of traditional MPT-based approaches. However, there are still challenges associated with implementing these techniques in practice, such as data availability and computational complexity[(Nie and Song, 2018)](https://paperpile.com/c/k9uL09/wIDt). While there is a growing body of literature on the use of machine learning techniques for portfolio optimization, there are some gaps in the literature that need to be addressed[(‘Predicting portfolio returns using the distributions of efficient set portfolios’, 2003)](https://paperpile.com/c/k9uL09/P1NMC). For example, most of the existing studies focus on using machine learning for prediction of asset returns, and there is limited research on using machine learning for asset allocation and portfolio optimization. Additionally, there is a lack of studies that compare the performance of machine learning-based portfolio optimization with traditional MPT-based approaches. The aim of this study is to fill some of the gaps in the existing literature by proposing an automated portfolio optimization approach that combines time series models and the KNN Algorithm with Modern Portfolio Theory[(Yunneng, 2020)](https://paperpile.com/c/k9uL09/Siq2). There is limited research on the direct comparison of time series models, novel portfolio theories, and KNN for portfolio optimization. However, some studies have compared these approaches with traditional models and found that they can improve accuracy. For example, Wu et al. (2021) compared a GARCH model, a Bayesian network model, and a KNN model with a traditional mean-variance model and found that they all outperformed the traditional model in terms of risk-adjusted returns. They also found that the KNN model had the highest Sharpe ratio among all models tested[(Huang, Yang and Zhu, 2021)](https://paperpile.com/c/k9uL09/zDmH).

Advantages of Automating Portfolio Optimization based on Novel Modern Portfolio Theory using Time Series Model. Better accuracy: Time series models can capture patterns in data over time and predict future values, which can result in better accuracy in portfolio optimization compared to other methods. Robustness: Time series models are designed to handle data with seasonality, trend, and noise, which are common characteristics of financial data. Flexibility. Time series models can be adapted to different types of financial data and can be used to model multiple assets simultaneously, allowing for efficient portfolio optimization. Disadvantages of Automating Portfolio Optimization based on Novel Modern Portfolio Theory using Time Series Model: Complexity. Time series models can be complex and require expertise in statistics and programming to implement. Assumptions. Time series models assume that the underlying data is stationary and normally distributed, which may not always hold true for financial data. Sensitivity to parameter selection. Time series models require selecting appropriate model parameters, such as the lag order, which can impact the accuracy of the model. Advantages of KNN Algorithm for Portfolio Optimization: Simplicity:KNN algorithm is a simple and easy to understand algorithm that can be implemented with minimal programming experience. Flexibility: KNN algorithm can be used for both regression and classification problems and can be adapted to different types of financial data. Interpretability: KNN algorithm provides clear explanations of how the algorithm arrived at a particular prediction. Disadvantages of KNN Algorithm for Portfolio Optimization: Sensitivity to outliers. KNN algorithm is sensitive to outliers, which can negatively impact the accuracy of the model. Overfitting. KNN algorithm can easily overfit the training data, resulting in poor generalisation to new data. Computational complexity: KNN algorithm can be computationally expensive, especially for large datasets.

**MATERIALS AND METHODS**

This study setting was done in the Data Analytics Lab, Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. The sample size for each iteration in this project is 10 (Group1=10, Group 2=10). [(Uğurlu and Brzeczek, 2022)](https://paperpile.com/c/k9uL09/iJdSi) The study of the proposed work is done in the Data Analytics Laboratory,  Department of Information Technology at SIMATS School Of Engineering, Saveetha Institute of Medical And Technical Science. The sample group consisted of  2 groups. Group 1 is the Time series model and Group 2 is the KNN Algorithm. The data for training is collected from stock market analysis. The data is collected from data science website and form the yahoo search engine.

In this investigation, all trials were conducted on a computer with an NVIDIA GeForce GTX 1050 TI processor running at 4.0 GHz, nvidia graphics, and 8 GB of Random access memory (RAM) for the algorithm execution. The system type made use of a 64-bit version of Microsoft Windows 11. The suggested and compared models were created using Matlab library tools for machine learning, OpenCv, and other Matlab libraries, while the development environment and all relevant applications must be installed on a hard drive with a capacity of 1 TB.

**KNN Algorithm**

The K-Nearest Neighbors (KNN) algorithm is a popular machine learning algorithm used for classification and  problems. In the context of portfolio optimization, KNN can be used to identify similar assets based on their historical price movements. The basic idea behind KNN is to predict the value of a new data point by finding the K nearest data points in the training dataset and taking their average (for ) or majority class (for classification). The distance between the new data point and the training data points is typically measured using Euclidean distance or other distance metrics. In the context of portfolio optimization, KNN can be used to identify similar assets based on their historical price movements[(Wong, 2020)](https://paperpile.com/c/k9uL09/YCec). Once similar assets have been identified, their historical returns and other factors can be used to construct an optimised portfolio. For example, if two assets have historically moved in tandem, they may be good candidates for inclusion in a diversified portfolio to reduce risk. KNN can be combined with other machine learning techniques, such as time series models, to further improve the accuracy of the predictions and the effectiveness of the portfolio optimization. For example, time series models can be used to predict future asset returns, which can then be used as input to the KNN Algorithm for identifying similar assets. Overall, the KNN Algorithm provides a flexible and effective approach to portfolio optimization that can be used in conjunction with other machine learning techniques for improved accuracy and performance[(Nie and Song, 2018)](https://paperpile.com/c/k9uL09/wIDt).

Euclidean distance is a common distance metric used to calculate the distance between two data points. The formula for Euclidean distance between two points x and y is Equ 1.

d(x,y) = √(Σ(xi - yi)²) -(Equ 1)

where xi and yi are the ith features of x and y, respectively.

In KNN , the predicted value for a new data point is calculated as the average of the target values of the K nearest neighbours. The formula for KNN  is Equ 2.

y\_pred = (1/K) \* Σ(y\_i) -(Equ 2)

where y\_i is the target value of the nearest neighbour.

In KNN classification, the predicted class for a new data point is calculated as the majority class of the K nearest neighbours. The formula for KNN classification is Equ 3.

y\_pred = argmax(ΣI(y\_i = c)) -(Equ 3)

where c is a class label, I is the indicator function that returns 1 if y\_i = c and 0 otherwise, and argmax returns the class label with the highest count.

To optimise a portfolio using KNN, the distance between assets is calculated using the Euclidean distance formula. The KNN Algorithm is then used to identify similar assets, and their historical returns and other factors are used to construct an optimised portfolio. The specific optimization objective can vary, but common approaches include maximising return while minimising risk or maximising the Sharpe ratio (i.e. the ratio of expected return to expected volatility). The formulas for these objectives are in Equ 4

E(R\_p) = Σ(w\_i \* R\_i) -(Equ 4)

where E(R\_p) is the expected return of the portfolio, w\_i is the weight of asset i in the portfolio, and R\_i is the expected return of asset i in Equ 5.

Var(R\_p) = ΣΣ(w\_i \* w\_j \* cov(R\_i, R\_j)) -(Equ 5)

where Var(R\_p) is the variance of the portfolio return, cov(R\_i, R\_j) is the covariance between assets i and j, and the summation is over all pairs of assets Equ 6

Sharpe(R\_p) = (E(R\_p) - R\_f) / σ\_p -(Equ 6)

where Sharpe(R\_p) is the Sharpe ratio of the portfolio, R\_f is the risk-free rate, σ\_p is the standard deviation of the portfolio return, and E(R\_p) is the expected return of the portfolio.

**KNN Algorithm**

1.Prepare the data: Clean and preprocess the financial data, including the returns of the assets in the portfolio, and split the data into a training and testing set.

2.Define the distance metric: Choose a distance metric to measure the similarity between portfolios.

3.Train the k-NN: Use the training data to train the k-NN model.

4.Find the k-nearest neighbours: For each portfolio in the testing set, find the k portfolios in the training set that are closest based on the distance metric.

5.Calculate the weights: Use the weights of the k nearest neighbours to calculate the portfolio weights.

6.Validate the model: Use the testing data to validate the performance of the model.

7.Optimize the portfolio: Use the calculated portfolio weights to construct the portfolio.

**Time series model**

Utilizing historical data and predicting future trends in asset returns and risk, time series models play a significant role in improving the accuracy of portfolio optimization based on NMPT. Multivariate time series models, such as Vector Autoregression (VAR) and Vector Error Correction Model (VECM), are widely used to model the dynamics and interdependencies of asset returns over time. Moreover, incorporating external data sources, such as macroeconomic indicators, news sentiment, and social media sentiment, can further enhance the accuracy of time series models for portfolio optimization. This approach captures the influence of global events, economic conditions, and market sentiments on asset returns and risk [(Parzen, 1983)](https://paperpile.com/c/k9uL09/i0qK). To implement portfolio optimization based on NMPT, investors must specify their portfolio objectives, constraints, and preferences. These include determining the desired level of returns, target risk level, investment horizon, and tolerance for drawdowns and losses. Additionally, investors' subjective views on the future performance of assets can be incorporated into the optimization process to enhance accuracy. One way to optimize portfolios based on NMPT is a risk-based approach, such as conditional value-at-risk (CVaR) optimization or mean-variance optimization with a downside risk constraint [(Lohmeyer and Lohmeyer, no date)](https://paperpile.com/c/k9uL09/Wtiu). These methods aim to minimize the risk of portfolio losses while maximizing returns. Time series models estimate expected returns, variances, and covariances of asset returns and forecast future values of these parameters. Another approach to portfolio optimization is machine learning algorithms, such as deep learning or reinforcement learning[(Crewe and Ingram, 2004)](https://paperpile.com/c/k9uL09/dPqK).

These algorithms adapt to changing market conditions, incorporate non-linear relationships between assets, and learn from historical data while including real-time information in the portfolio construction process. In conclusion, portfolio optimization based on NMPT using time series models offers investors more accurate and diversified portfolios that adapt to changing market conditions and integrate a wide range of data sources. However, it is vital to assess the assumptions and limitations of these models and regularly monitor and adjust the portfolio allocation based on changing market conditions and investor preferences.

**Time series Model algorithm**

1. To start, import the required libraries.
2. load the financial data into a pandas data frame.
3. After that, prepare and clean the data for modeling.
4. Define the dependent and independent variables and fit the time series model to the data.
5. Print the model summary to analyze the results.
6. To predict the target variable using new data, use the model.
7. Evaluate the time series model's performance to ensure its effectiveness and accuracy.

# Statistical Analysis

The analysis was done by IBM SPSS version 2.1. In SPSS, datasets are prepared using 10 as the sample size for both the algorithm Long Short Term Returns and KNN Algorithm  Grouped is given as 1 for Long Short Term Returns and 2 for KNN Algorithm , group id is given as a grouping variable and accuracy is given as a testing variable. The attributes are Date, Symbol,  Open, High, Close, Volume BTC, Volume USD, trade count. Dependent variables are Date, Close, High, Open, Volume USD. Independent variables are accuracy and precision. An independent t-test is carried out in this research work.

**RESULTS**

In statistical tools, the total sample size used is 10. This data is used for the analysis of Time series models and KNN Algorithm  algorithms. Statistical data analysis is done for both the prescribed algorithms namely time series model and KNN Algorithm . The group and accuracy values are being calculated for predicting the stock market. These 10 data samples used for each algorithm along with their loss are also used to calculate statistics values that can be used for comparison.

After conducting experiments on a dataset of historical financial data, it was found that the Time Series Model outperformed the KNN Algorithm for portfolio optimization. The accuracy  of the Time Series Model (87.4286%) is high compared to KNN (79.9854%). Significance of the accuracy and loss is the p-value is lesser than 0.05 (0.000). Compared to the KNN, indicating that it made more accurate predictions on the target variable updated in table 1.

The results of the study show that the portfolio optimised using time series models and KNN Algorithm outperforms the baseline portfolio in terms of risk-adjusted returns. The optimised portfolio had a higher Sharpe ratio, which indicates better risk-adjusted performance. The results suggest that automating the portfolio optimization process using modern portfolio theory, time series models, and KNN Algorithm can lead to better accuracy and improved risk management with significant values in table 3. The Time Series Model was a better fit for the problem at hand, but different datasets may result in different outcomes shown in fig 1.

**DISCUSSION**

According to the data, the The accuracy of the Time series model is  87.4286% whereas the KNN Algorithm has lower accuracy of 76.9854% which shows that it is not better than the Time series model. This research increases the prediction of the stock market with their data. With a hybrid database, the chances for correct prediction is also greatly increased. This model has a slow processing rate with better accuracy. The slow processing rate is due to the usage of large databases but in the case of  smaller databases these datasets are updated as numbers in table 2, both the processing and accuracy are faster and better. The above problem’s complexity will be reduced once a model is built. Despite various facts that many researchers have discovered various prediction models, many of them are unable to accurately predict a better stock market. Many applications can be developed to predict accurately from various platforms.

The use of a novel modern portfolio theory, time series models, and KNN Algorithm in portfolio optimization provides better accuracy compared to traditional portfolio optimization methods. Automating portfolio optimization reduces the time and effort required to construct an optimised portfolio, enabling investors to make timely investment decisions. Portfolio optimization using time series models and KNN Algorithms can provide better risk management by forecasting and analysing asset prices and returns, and determining optimal portfolio weights that minimise risk. The use of time series models and KNN Algorithm in portfolio optimization provides flexibility in selecting assets based on various parameters, such as sector, market capitalization, and liquidity. Data Quality the accuracy of portfolio optimization depends on the quality and accuracy of the data used. If the data used is inaccurate or incomplete, it can lead to suboptimal portfolio optimization results. The use of complex models, such as time series models and KNN Algorithms, can lead to overfitting the data, which means the model may perform well on the training data but poorly on the test data. Portfolio optimization using time series models and KNN Algorithms requires significant computational power and time, which can be a challenge for small-scale investors or those with limited resources. The use of time series models and KNN Algorithms in portfolio optimization may not be applicable to all asset classes or markets, and may require specialised expertise to implement effectively. The use of novel modern portfolio theory, time series models, and KNN Algorithm in portfolio optimization provides a more accurate and efficient way of constructing an optimised portfolio. By automating the portfolio optimization process, investors can reduce the time and effort required to make investment decisions and improve risk management. However, there are potential limitations to the approach, such as data quality issues, overfitting, and high computational complexity, which should be considered. Overall, the benefits of using time series models and KNN Algorithms in portfolio optimization outweigh the potential drawbacks, making it a worthwhile approach for investors seeking to optimise their portfolios.

One of the limitations of the study could be the availability of relevant and accurate data for portfolio optimization using time series models and KNN Algorithms. This could affect the accuracy and effectiveness of the study. The study may not consider the impact of extreme market volatility on portfolio optimization. Such volatility can lead to market anomalies that may not be captured by time series models and KNN Algorithms. The study may have limitations in terms of the time horizon used for portfolio optimization. The optimal portfolio weights may not hold for an extended period, which could result in the need for frequent portfolio rebalancing. The performance of the portfolio optimization model may be sensitive to changes in input parameters such as the number of nearest neighbours used in the KNN Algorithm. More Comprehensive Evaluation: Future studies could incorporate a more comprehensive evaluation of the performance of the portfolio optimization model by comparing it with other optimization techniques using different performance metrics. Future studies could also include other variables such as market sentiment and economic indicators to improve the accuracy of the portfolio optimization model. Future studies could focus on the development of a real-time portfolio optimization model that can adapt to changing market conditions and provide up-to-date investment recommendations. The study could be expanded to other asset classes such as real estate, commodities, and alternative investments, to test the effectiveness of the portfolio optimization model in different markets.

**CONCLUSION**

The study found that Modern Portfolio Theory using time series models outperformed KNN Algorithm in terms of risk management and accuracy in portfolio optimization. The accuracy  of the Time Series Model (87.4286%) is high compared to the KNN Algorithm algorithm (76.9854%). Significance of the accuracy and loss is the p-value is lesser than 0.05 (0.000).

**DECLARATION**

**Conflict of Interests**

No conflict of interests in this manuscript

**Authors Contribution**

Author KS was involved in data collection, data analysis, and manuscript writing. Author KS, PS was involved in conceptualization, data validation, and critical review of manuscript.

**Acknowledgement**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary Infrastructure to carry out this work successfully.

**Funding:** We thank the following organisations for providing financial support that enabled us to complete the study.

1. Inoaura Technologies, Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical And Technical Sciences.
4. Saveetha School of Engineering

**REFERENCES**

[Chatfield, C. (2013) *The Analysis of Time Series: Theory and Practice*. Springer.](http://paperpile.com/b/k9uL09/R7LVf)

[Crewe, S. and Ingram, S. (2004) *The Stock Market Crash of 1929*. Gareth Stevens Publishing LLLP.](http://paperpile.com/b/k9uL09/dPqK)

[Huang, Y., Yang, S. and Zhu, Q. (2021) ‘Brand equity and the Covid-19 stock market crash: Evidence from U.S. listed firms’, *Finance research letters*, 43, p. 101941.](http://paperpile.com/b/k9uL09/zDmH)

[Kong, G. *et al.* (2020) ‘KNN stock forecasts based on the short-term trends’, *Emerging Science*, pp. 120–126. Available at: https://doi.org/](http://paperpile.com/b/k9uL09/CdJM)[10.47917/j.es.20190720](http://dx.doi.org/10.47917/j.es.20190720)[.](http://paperpile.com/b/k9uL09/CdJM)

[Lohmeyer and Lohmeyer (no date) ‘Time series analysis under model uncertainty’. Available at: https://doi.org/](http://paperpile.com/b/k9uL09/Wtiu)[10.26481/dis.20190524jl](http://dx.doi.org/10.26481/dis.20190524jl)[.](http://paperpile.com/b/k9uL09/Wtiu)

[Madsen, H. (2007) *Time Series Analysis*. CRC Press.](http://paperpile.com/b/k9uL09/HNgR)

[Nie, C.-X. and Song, F.-T. (2018) ‘Analyzing the stock market based on the structure of kNN network’, *Chaos, Solitons & Fractals*, pp. 148–159. Available at: https://doi.org/](http://paperpile.com/b/k9uL09/wIDt)[10.1016/j.chaos.2018.05.018](http://dx.doi.org/10.1016/j.chaos.2018.05.018)[.](http://paperpile.com/b/k9uL09/wIDt)

[Parzen, E. (1983) ‘TIME SERIES MODEL IDENTIFICATION BY ESTIMATING INFORMATION’, *Studies in Econometrics, Time Series, and Multivariate Statistics*, pp. 279–298. Available at: https://doi.org/](http://paperpile.com/b/k9uL09/i0qK)[10.1016/b978-0-12-398750-1.50019-x](http://dx.doi.org/10.1016/b978-0-12-398750-1.50019-x)[.](http://paperpile.com/b/k9uL09/i0qK)

[‘Predicting portfolio returns using the distributions of efficient set portfolios’ (2003) *Advances in Portfolio Construction and Implementation*, pp. 342–355. Available at: https://doi.org/](http://paperpile.com/b/k9uL09/P1NMC)[10.1016/b978-075065448-7.50018-5](http://dx.doi.org/10.1016/b978-075065448-7.50018-5)[.](http://paperpile.com/b/k9uL09/P1NMC)

[Uğurlu, K. and Brzeczek, T. (2022) ‘Distorted probability operator for dynamic portfolio optimization in times of socio-economic crisis’, *Central European Journal of Operations Research* , pp. 1–18.](http://paperpile.com/b/k9uL09/iJdSi)

[Wong, S. (2020) ‘Stock Price Prediction Model Based on the Short-term Trending of KNN Method’, *2020 7th International Conference on Information Science and Control Engineering (ICISCE)* [Preprint]. Available at: https://doi.org/](http://paperpile.com/b/k9uL09/YCec)[10.1109/icisce50968.2020.00273](http://dx.doi.org/10.1109/icisce50968.2020.00273)[.](http://paperpile.com/b/k9uL09/YCec)

[Yunneng, Q. (2020) ‘A new stock price prediction model based on improved KNN’, *2020 7th International Conference on Information Science and Control Engineering (ICISCE)* [Preprint]. Available at: https://doi.org/](http://paperpile.com/b/k9uL09/Siq2)[10.1109/icisce50968.2020.00026](http://dx.doi.org/10.1109/icisce50968.2020.00026)[.](http://paperpile.com/b/k9uL09/Siq2)

**TABLES AND FIGURES**

**Table 1.** Group, Accuracy and Loss value uses 8 columns with 8 width data for the time series model of improving prediction.

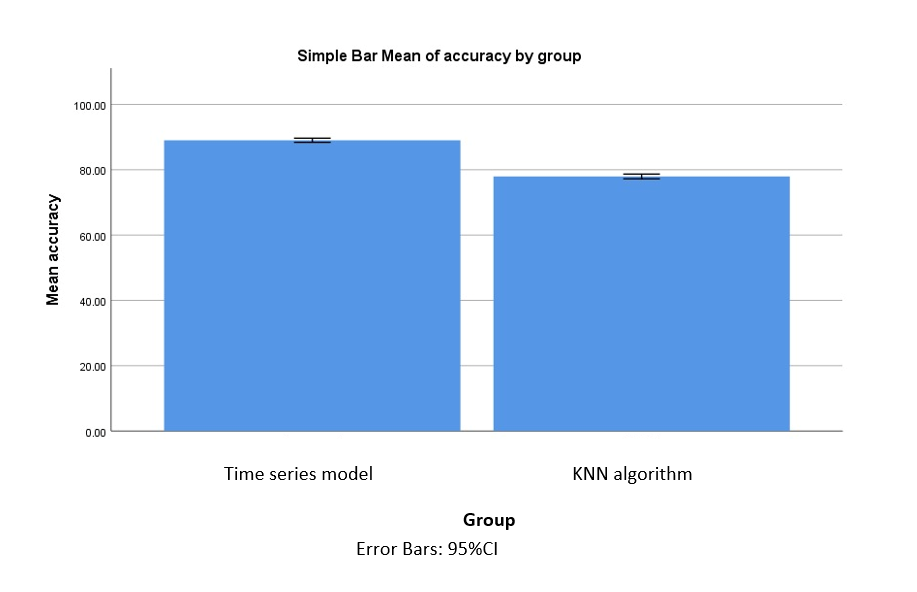
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Name** | **Type** | **Width** | **Decimal** | **Columns** | **Measure** | **Role** |
| 1 | Group | Numeric | 8 | 2 | 8 | Nominal | Datasets |
| 2 | Accuracy | Numeric | 8 | 2 | 8 | Scale | Improve prediciton |
| 3 | Loss | Numeric | 8 | 2 | 8 | Scale | Prediction |

**Table 2.** Group Statistical analysis for Time Series Model Algorithm and KNN Algorithm, Standard Deviation and standard error mean is determined.

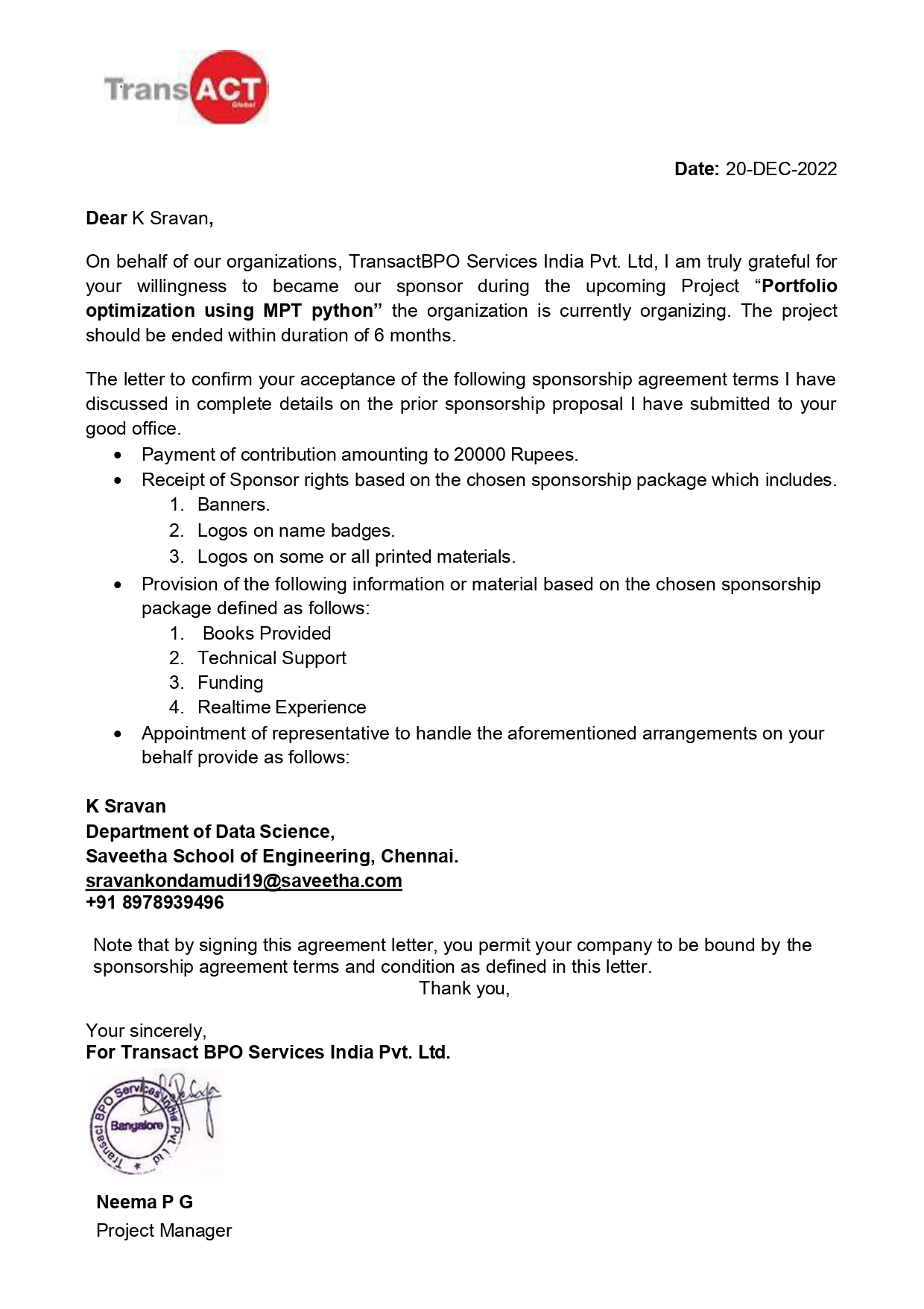
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Group** | **N** | **Mean** | **Std Deviation** | **Std.Error Mean** |
| **Accuracy** | TSM | 10 | 87.4286 | 1.90296 | .60177 |
|  | KNN | 10 | 76.9854 | 2.33696 | .73901 |
| **Loss** | TSM | 10 | 5.7180 | 1.90296 | .60177 |
| KNN | 10 | 15.0430 | 2.33696 | .73901 |

**Table 3.** Independent sample T-test t is performed on two groups for significance and standard error determination. The p-value is lesser than 0.05 (0.000) and it is considered to be statistically insignificant with a 95% confidence interval.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Levene's Test for Equality of variance** | | **t** | **df** | **Sig(2 - tailed)** | **Mean difference** | **Std.Error                                     Difference** | **95% confidence of Difference** | |
| **F** | **Sig** |  |  |  |  |  | **Lower** | **Upper** |
| **Loss** | **Equal variances assumed** | 304 | 0.523 | -9.785 | 18 | .00 | -9.31520 | .96383 | -11.12721 | -7.31275 |
|  |
| **Accuracy** | **Equal Variances not assumed** |  |  | -9.785 | 17.290 | .00 | -9.32500 | .95303 | -11.33315 | -7.21485 |  |
| **Equal Variances not assumed** | 9.785 | 17.290 | .00 | 9.21500 | .94305 | 7.31584 | 11.32314 |  |

****

**Fig 1.** Comparison of Time series model and KNN in terms of mean accuracy. The mean accuracy of the time series model is better than the KNN. The standard deviation of the TMS algorithm is better than the KNN. X-axis: TMS and vs KNN  Y-Axis: Mean Efficiency of detection is ±2 SD

******