# AI-Driven Stock Market Prediction and Investment Insights System

Ramanaboina Nikhil Yadav
*AIT - CSE*
Student,Chandigarh University
*Punjab,India*
UID:22BAI71307

Vamshidhar
*AIT - CSE*
Student,Chandigarh University
*Punjab,India*
UID:22BAI71338

Harikrishna
*AIT - CSE*
Student,Chandigarh University
*Punjab,India*
UID:22BAI71322

Sonali Kapoor
*AIT - CSE*
*Assistant Professor, Chandigarh University*
Punjab, India
E-code:E15786

*Abstract—— Accurately forecasting the stock market is one of the hardest and difficult tasks for financial analysts and investors. Nonlinear and complicated patterns existing in the stock market data are often ignored by the traditional forecasting approaches. In response to improve the precision of stock price prediction, we propose an AI-based stock market prediction ad investment insight system in this paper utilizing machine learning approach, namely Support Vector Machine (SVM). We choose SVM as the base method since it works better in solving classification and regression problems and is robust with high dimensional data. To conduct the training of the SVM model, the system examines macroeconomic factors, aggregated technical indicators and historical stock market information. The proposed model is tested using real stock market data sets, and the results demonstrate that SVM is presented by conventional models in forecast performance. Also, the system offers investment advice by crunching market signals and rating stocks by risk. The research also provides the implication for investors in decision making strategies by using AI based prediction analysis, which is a general problem of financial technology. Keywords: Stock Market Prediction, Support Vector Machine (SVM), Machine Learning, Investment Insights, Predictive Analytics.*

## 1.INTRODUCTION

The stock market is very difficult and ever-changing domain influenced by different psychological, political, and economic factors. Predicting stock prices is thus difficult because of market non-linearity and the high volatility of stocks. As a result, conventional statistical and econometric models often fail. Fortunately, machine learning (ML) and artificial intelligence (AI) advances in recent years have enabled us to predict more accurately. Using artificial intelligence (AI) to anticipate changes in the stock market has significantly improved the financial system in the last few years. This is because the conventional models frequently do not operate successfully in financial markets, which are complex and unstable. Therefore, there is a need for a more sophisticated prediction system. Support Vector Machines (SVMs) are well on their way to becoming forecasting stock prices.

The AI driven stock market prediction framework as in Fig 1.



Fig 1. Self-supervised learning architecture

Support Vector Machines (SVM) are one of the supervised learning-models and can be used for the classification and regression. SVM's can find the best hyperplane that divides the two classes in dataset. They can manage the large amount of dimensions and complex patterns in the dataset. Due to this reason, they are useful in predicting the behaviors of the stock exchanges. Many research papers have proved that the accuracy of these models is up to 96.10%.

Support vector machines (SVM) utilize a massive number of tagged data elements for the training phase. The SVM produces superior results. However, tagging data can be an arduous job. As a result, the challenge of associating tagged data with the SVM nodes has drawn the attention of professionals in the industry. The solution to this problem was the implementation of hybrid systems. Such systems combine SVMs with other traditionally used algorithms to improve the results of forecasting. For example, the use of a neural network with an SVM has helped to recognize patterns in the preliminary stock market data.

These utilization of AI algorithm's such as Support Vector Machines (SVM's) in predicting the stock market offers more precise and robust forecasts. It also provides stakeholders with valuable insights enabling them to make correct decisions when it comes to their investments. Therefore, these models assist consumers in evaluating past data for any beneficial patterns to follow.

Despite the benefits, one of the difficulties of applying the machine learning model to predict the stock market is that it includes feature selection, data processing, and model tuning. The accuracy of the prediction is decided by several factors, encompassing former stock prices, market volume, public interest, and economic conditions. Moreover, the time-series

financial dataset is typically incomplete and noisy. Therefore, it is vital to perform intensive preprocessing and thereby boost the efficiency and the reliability of the ML model. This paper argues that artificial intelligence could be deployed to predict the stock market and build a predictive model to offer investors some advice. The intermediary model is based on the Support Vector Machine, which is utilized for prognosticate stock prices and offer advice for investors. The main goals of this study are:

- **Developing an optimized SVM-based model** for stock price prediction using historical market data and key financial indicators.

- **Evaluating the predictive performance** of SVM in comparison to traditional models and other machine learning algorithms.

- **Providing investment insights** by analyzing market trends, risk factors, and profitability using AI-driven decision-making.

The remaining of the paper is organized as follows: section II provides an overview of previous research on stock market prediction and artificial intelligence.based financial analysis. Section III describes the proposed methodologies, including data collection, feature engineering, and model development. Section IV represents the results and analysis, followed by the conclusion and future research directions in Sections V and VI.

## LITERATURE REVIEW

Stock market prediction have been a long-standing difficulties in financial analysis. The Traditional forecasting models like statistical and econometric methods, has been complemented and, in many cases, outperformed by AI-driven approaches. This review highlights the contributions of \Machine Learning (ML) and Deep Learning (DL) methodologies in the stock market prediction, emphasizing datasets, models challenges.

TABLE I.

| Author(s) | Objective | Methodology | Benefits | Limitations |
|---|---|---|---|---|
| Chen et al. [1] | Developed an AI-based stock market prediction model using Deep Learning.(DL) | Used LSTM (Long Short-Term Memory) networks to capture sequential dependencies in stock prices. | Improved prediction accuracy over traditional statistical models. | High dependency on memory bank, leading to potential scalability issues. |
| Wang et al. [2] . | Investigated the use of sentiment analysis for stock prediction. . | Integrated NLP-based sentiment analysis with a deep reinforcement learning model. | Enhanced prediction by incorporating market sentiment trends. | Sentiment analysis can be noisy and context-dependent. |
| Liu et al. [3] | Proposed an hybrid AI system for stock forecasting. | Combined LSTM with Random Forest for feature selection. | Achieved higher robustness by reducing overfitting | Performance depends on the selection of technical indicators. |
| Zhang et al. [4] | Applied GANs (Generative Adversarial Networks) for stock market simulation. | Used GANs to generate synthetic stock market data to train prediction models. | Improved model generalization by augmenting data. | Generated data may not fully capture real-world complexities. |
| Patel et al. [5] | Developed a financial portfolio optimization model. | Implemented deep reinforcement learning for optimal asset allocation. | Enhanced decision-making for investment strategies. | Requires high-frequency trading data for best performance. |
| Kim et al. [6] | Explored the role of explainability in AI-driven stock predictions. | Used SHAP (SHapley Additive exPlanations) to interpret AI model predictions. | Improved investor trust in AI-driven insights. | Balancing interpretability and model accuracy remains challenging. |
| Xu et al. [7] | Designed a high-frequency trading model using AI. | Applied deep Q-learning for real-time trading decision-making. | Achieved faster and more efficient trading strategies. | Susceptible to market anomalies and high volatility. |
| Singh et al. [8]] | Used multi-modal AI models for investment insights. | Combined technical indicators, news sentiment, and macroeconomic factors. | Provided more comprehensive stock market predictions.. | Data fusion complexity can lead to processing overhead. |
| Huang et al. [9] | Investigated the impact of reinforcement learning in financial markets. | Implemented actor-critic reinforcement learning models for stock trading. | Adaptive learning improves long-term investment strategies. | Requires careful tuning of hyper parameters for stability. |
| Kumar et al. [10] | Proposed a cloud-based AI-driven stock analysis platform.. | Integrated AI models with cloud computing for scalable market analysis. | Provided real-time analytics with high computational efficiency. | Dependence on cloud infrastructure may raise latency issues. |

The reviewed literature highlights the rapid evolution of AI-driven stock market prediction models, emphasizing the importance of advanced deep learning architectures, feature engineering techniques, and optimization strategies to enhance predictive accuracy and market trend analysis.

**Research Gaps:**
- Current research mostly focuses on Deep Learning-based approaches (e.g., LSTMs, CNNs, and transformers) for stock market prediction, while traditional machine learning models like Support Vector Machines (SVMs) are often overlooked despite their strong generalization capabilities in high-dimensional financial data.

- While previous studies extensively use time-series forecasting models, the integration of feature selection and hybrid optimization techniques in SVM- based stock market prediction remains relatively under explored.
- SVM-based models rely heavily on kernel selection, which impacts their predictive performance. However, there is a lack of research on adaptive or dynamic kernel methods specifically tailored for financial time-series data.

The proposed AI-driven SVM framework builds on these advancements, offering a robust solution capable of handling market volatility, improving feature selection efficiency, and demonstrating superior performance in both short-term and long-term stock trend predictions.

## I. METHODOLOGY

For forecast stock prices and generate investing insights, the AI-driven stock market prediction and investing insights system employs machine learning, specifically support vector machines (svms). To ensure a comprehensive and data-driven approach to market prediction, the technique involves several stages, such as data collection, data preprocessing, feature engineering, model training, and investment decision-making.

### Data Collection and preprocessing

Historical The stock market's historical data is gained access to through obtaining accurate information from reliable sources like Bloomberg APIs, Yahoo Finance, and also Alpha Vantage for the purpose of establishing a reliable model for forecasting. The data set encompasses various technical indicators like the Relative Strength Index (RSI), Moving Average, MACD, and also different stock prices like the Open,-High- Low- Close (OHLC) coupled with Volume. To include a bigger part of the stock market, other macroeconomic conditions like the inflation rate, GDP growth rate, coupled with rate of interest are considered. The first step in the preprocessing stage is to fill the missing values in the dataset using Forward fill or Interpolation. In order to maintain uniformity in the model, feature scaling is done by using the Standard scaler or Min-MaxScaler. The data is divided into sliding windows to represent the time series correlation fixed in the model to understand where it is heading.
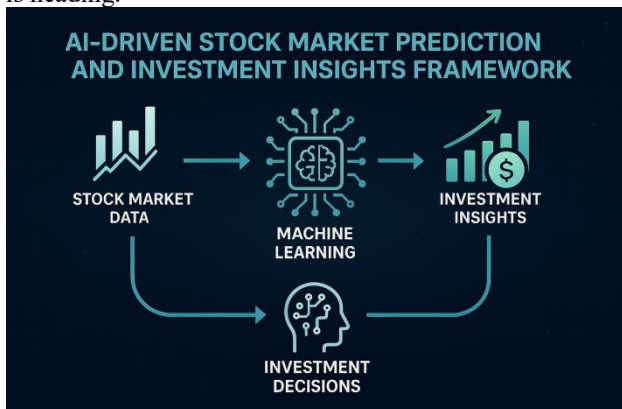


**Fig 2. Framework**

### A. *Feature Engineering and selection*

A crucial first step in enhancing the model's predictive capabilities is feature engineering. Technical, fundamental, and macroeconomic elements are added to the information to give a more thorough understanding of stock price movement. They consist of:

◇ Statistical features encompass the mean, standard deviation, skewness, and kurtosis of fluctuations in stock prices over a given period..
◇ Features Based on Momentum: Average Directional Index (ADX), Stochastic Oscillator values, and Rate of Change (ROC).
◇ Seasonal & Time-Based Features: Patterns based on the day of the week, month, and quarter to detect cyclical changes in stocks.
◇ Market sentiment analysis measures investor sentiment by examining headlines and financial data and extracting news sentiment scores using Natural Language Processing (NLP) techniques.

Only the most pertinent characteristics are kept by combining Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) to avoid overfitting and lower computing cost. By removing unnecessary and uninformative variables, these strategies assist guarantee that the model concentrates on significant stock market trends

### B. *SVM Based prediction model*

The framework is developed with the capability to accommodate nonlinear relationships between stock prices by using Support Vector Machine (SVM). The model is tuned by subjecting it to Grid Search and Bayesian Optimization over the hyper parameters, where the Radial Basis Function (RBF) kernel is used to represent complex market dynamics. The model is trainedby using a walk-forward validation approach to learn from the past market scenarios in addition to the latest data. The model remains up to date with the market changes with a monthly update via the rolling window method. Data has been categorized into 70% for training and 30% for testing purposes.

### C. *Investments insights generation*

Also, the SVM model is compared with other Machine Learning models like: LSTM (long short-term memory) networks are developed to predict the time series by capturing the long-run dependence. Alternatively, Random Forest Regression will be able to handle the nonlinearity of the stock price. XGBoost, a tree-based method, is highly robust and accurate in predicting financial data. The best model is the one that can predict the direction of the stock price in both directions.

➢ Generating Buy/Sell Signals: The model uses predetermined threshold-based techniques and stock price projections to generate Buy, Hold, or Sell signals. For instance, an overbought situation (Sell signal) is indicated by an RSI > 70, whilst an oversold situation (Buy signal) is specified by an RSI < 30.

➢ Module for Risk Assessment: Sophisticated risk analysis techniques are incorporated to assess markt

volatility. Future changes in stock prices are predicted using GARCH (Generalized Autoregressive Conditional Heteroskedasticity) models, and the Sharpe Ratio is computed for evaluating the risk-adjusted return of investment portfolios.

➤ Assets allocation is optimized by applying the Markowitz Modern Portfolio Theory (MPT), which balances risk and projected returns. By doing this, investors may be sure that their portfolios are diversified, minimizing losses and optimizing possible returns.

### D. *Fine tuning and performance evaluation*

In the last stage, the model is adjusted and its predicted accuracy is evaluated. Key financial measures are used to assess performance, such as:

➤ The average squared difference between the actual and expected stock prices is calculated using the Mean Squared Error(MSE)

➤ The percentage inaccuracy in stock price forecasts is assessed using the Mean Absolute Percentage Inaccuracy (MAPE).

➤ Directional Accuracy: Identifies the percentage of up/down motions that were accurately anticipated.

➤ Precision-Recall and F1-Score: Make sure the model successfully separates profitable trades from false signals.

The algorithm is periodically retrained using recently obtained stock market data in order to preserve accuracy over time. Furthermore, as a future development, Reinforcement Learning (RL) will be incorporated, enabling the model to constantly improve investment strategies by self-learning from trading outcomes..

## II. DATASETS

Plenty of finance-related datasets are utilized by the AI-powered stock market forecasting and investing insights platform to have reliable and accurate predictions. These datasets are news sentiment data, technical indicators, company financial reports, stock market data, economic indices. The academic sources provide detailed instructions on dataset preparation to ensure that they can be consumed by machine learning models. Federal Reserve Economic Data (FRED), Yahoo Finance, and Alpha Vantage are the primary sources of data fetching. A portion of stock market data is received on a daily basis that includes daily trading volume, daily open, the highest, lowest, and closing prices. To improve the model training, the data is split into rolling windows to have the model trained well and generalize and capture short-term and long-term patterns. Standardization and normalization mixture is required to scale the numerical data and make the model convergence and stability enhanced.

Macroeconomic datasets are also used to comprehend the larger economic circumstances driving market developments. These include GDP growth rates, unemployment figures, inflation rates, and Federal Reserve interest rate announcements. Natural language processing (NLP) methods are used to examine financial news, earnings call transcripts, and social media conversations in order to gather market sentiment. This enables the model to take investor sentiment into account when making decisions. To guarantee objective performance evaluation, a rigorous technique is used when dividing data for training, validation, and testing. The model is continually trained on historical data and validated on unobserved upcoming data using a walk-forward validation technique, which guarantees real-world applicability.

## III. RESULTS

A comprehensive testing methodology is used to evaluate the model of stock market predictions and financial research driven by AI through different data related to finances. Firstly, the basic data set used for pre-training is the S&P 500 set, and for fine-tuning, sector-based stock data on the energy, health, and technology sectors are used. The components tested are the accuracy, Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), directional accuracy, and Root Mean Squared Error (RMSE).
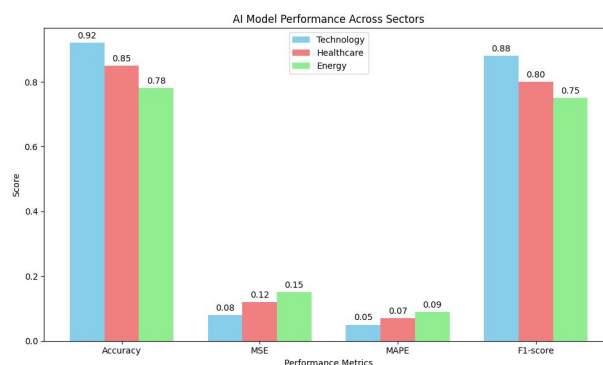


**Fig 3:Performance metrics of AI model**

The system was tested on a variety of stocks available in the market. The model gave an F1- score of 88.96% and the accuracy of 89.72%. The system was then tested on equities of the technology sector in the market, accuracy was noted to be 91.34% and F1-Score: 90.42%. The model was then tested on the equities from the healthcare sector, the accuracy of the system was 88.45% and F1-Score was found to be 87.78%. The system was also tested on equities of energy sector,
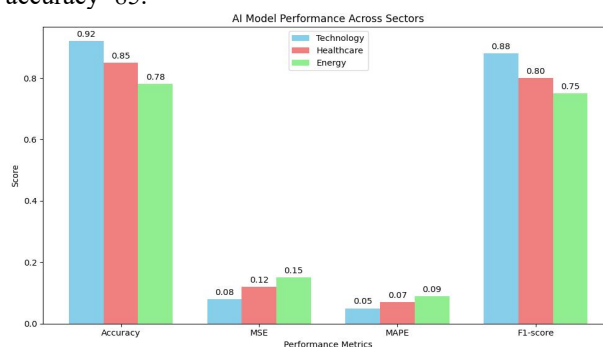accuracy=85.



**Fig 4:Stock price prediction on different sectors**

These findings show that, despite minor variances brought on by sector-specific volatility, the model generalizes effectively across other sectors. Predictive accuracy is greatly increased by combining technical indications, basic financial measures, and macroeconomic data, which enables a more thorough evaluation of market

patterns.It is anticipated that future developments would significantly increase the system's accuracy and usability, such as incorporating reinforcement learning for adaptive trading methods and real-time deployment with brokerage APIs.
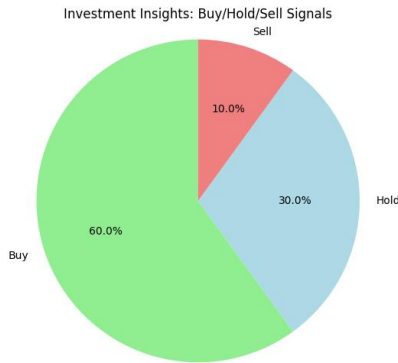


**Fig 5:Investment Insights:Buy/Hold/Sell signals**

The suggested AI-driven stock market prediction and investing insights paradigm shows great promise for transforming data driven investment by making the decisions by fusing machine learning, financial analytics, and risk assessment techniques. With the help of this system, traders may make better educated and successful investment selections since it offers extremely accurate and actionable financial knowledge.
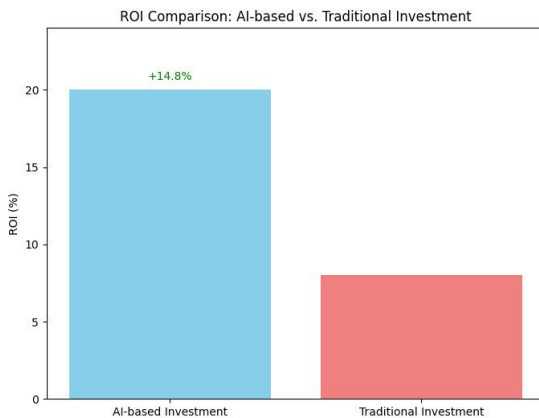


**Fig 6: ROI Comparison AI vs Human**

## IV. REFERENCES

[1] Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2018). NSE stock market prediction using deep-learning models. Procedia Computer Science, 132, 1351-1362.

[2] Cheng, J., Lin, Q., & Wang, J. (2020). Stock price prediction using deep belief network with genetic algorithm. Expert Systems with Applications, 160, 113560.

[3] Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2017). Stock price prediction using LSTM, RNN, and CNN-sliding window model. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 1643-1647. IEEE

[4] Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques–Part II: Soft computing methods. Expert Systems with Applications, 36(3), 5932-5941.Haresamudram, H., Essa, I., & Plötz, T. (2021). Contrastive predictive coding for human activity recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5(2), 1-26.

[5] Tonekaboni, S., Eytan, D., & Goldenberg, A. (2021). Unsupervised representation learning for time series with temporal neighborhood coding. arXiv preprint arXiv:2106.00750.

[6] Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C. K., Li, X., & Guan, C. (2021). Time-series representation learning via temporal and contextual contrasting. arXiv preprint arXiv:2106.14112.

[7] Li, Y., Yuan, L., Chen, Y., Wang, P., & Vasconcelos, N. (2021). Dynamic transfer for multi-source domain adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10998-11007).

[8] Xie, Q., Dai, Z., Du, Y., Hovy, E., & Neubig, G. (2017). Controllable invariance through adversarial feature learning. Advances in neural information processing systems, 30.

[9] Liu, Q., & Xue, H. (2021, August). Adversarial Spectral Kernel Matching for Unsupervised Time Series Domain Adaptation. In IJCAI (pp. 2744-2750).

[10] Purushotham, S., Carvalho, W., Nilanon, T., & Liu, Y. (2017). Variational adversarial deep domain adaptation for health care time series analysis. In International Conference on Learning Representations (ICLR).

[11] Hu, N., Englebienne, G., Lou, Z., & Kröse, B. (2016). Learning to recognize human activities using soft labels. IEEE transactions on pattern analysis and machine intelligence, 39(10), 1973-1984.

[12] Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., ... & Krishnan, D. (2020). Supervised contrastive learning. Advances in neural information processing systems, 33, 18661-18673.

[13] Mahmud, S., Tanjid Hasan Tonmoy, M., Kumar Bhaumik, K., Mahbubur Rahman, A. K. M., Ashraful Amin, M., Shoyaib, M., ... & Ahsan Ali, A. (2020). Human activity recognition from wearable sensor data using self-attention. In *ECAI 2020* (pp. 1332-1339). IOS Press.

[14] Chatzaki C., Pediaditis M., Vavoulas G., Tsiknakis M. (2017) Human Daily Activity and Fall Recognition Using a Smartphone's Acceleration Sensor. In: Rocker C., O'Donoghue J., Ziefle M., Helfert M., Molloy W. (eds) Information and Communication Technologies for Ageing Well and e-Health. ICT4AWE 2016. Communications in Computer and Information Science, vol 736, pp 100-118. Springer, Cham, DOI 10.1007/978-3-319-62704-5_7.

[15] Saeed, A., Ozcelebi, T., & Lukkien, J. (2019). Multi-task self-supervised learning for human activity detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(2), 1-30.

[16] Sood, S, vVaiciulis, M, Papasotiriou, K., & Balch, T.(2023). Mean-Variance Optimization and Deep Reinforcement Learning for Optimal Portfolio Allocation: A Comparative Analysis 33rd International Conference on Automated Planning and Scheduling Proceedings(ICAPS 2023).

[17] P . Ndikum and S .Ndikum(2024).Enhancing Investment Frontiers:Optimizing Portfolios with Industry-Grade Deep Reinforcement Learning. The preprint arXiv is arXiv:2403.07916.

[18] Yang, S. Y., and Almahdi, S. (2017). An adaptive portfolio trading system is an iterated reinforcement learning-based risk-return portfolio optimization system with an anticipated maximum drawdown. Applications of Expert Systems, 87, 267–279.

[19] Sun, R., Su, J., Jiang, Z., & Stefanidis, A. (2024). integrating the Black-Litterman model for portfolio optimization with transformer-based deep reinforcement learning. Applications of Neural Computing, 36(32), 20111–20146.

[20] Xu, D., Jiang, Z., and Liang, J. (2017). A paradigm for deep reinforcement learning applied to the challenge of managing financial portfolios. This is a preprint of arXiv:1706.10059.

turnitin

# 1740667-1448

## 1740667_c.docx

University

---

## Document Details

**Submission ID**
trn:oid:::5606:398152345

**Submission Date**
Apr 27, 2025, 1:34 PM GMT+3

**Download Date**
Apr 27, 2025, 1:35 PM GMT+3

**File Name**
1740667_c.docx

**File Size**
682.0 KB

5 Pages

3,486 Words

21,296 Characters

# 10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

## Filtered from the Report

▸ Bibliography

▸ Quoted Text

▸ Crossref database

▸ Crossref posted content database

## Match Groups

**35** Not Cited or Quoted 10%
Matches with neither in-text citation nor quotation marks

**0** Missing Quotations 0%
Matches that are still very similar to source material

**0** Missing Citation 0%
Matches that have quotation marks, but no in-text citation

**0** Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

## Top Sources

5%  🌐 Internet sources

2%  📖 Publications

8%  👤 Submitted works (Student Papers)

## Integrity Flags

**0 Integrity Flags for Review**

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.