Plagiarism - Report

Originality Assessment

24%

Overall Similarity

Date: Apr 30, 2024
Matches: 1541 / 4585 words
Sources: 41

Remarks: Moderate similarity detected, you may improve the document (if needed).

Verify Report:

Smart Trading with NLP: A Journey through Scopus Research on Stock Market Analysis

Prathamesh Birajdar 1, Lavanya Vanga 2, Sakshi Dama3

1Department of CSE(AI&DS), Shree Siddheshwar College Of Engineering, Solapur, India.

birajdarprathmesh877@gmail.com

2Department of CSE(AI&DS), Shree Siddheshwar 36 College of Engineering, Solapur, India

lavanyavanga2005@gmail.com

3Department of CSE(AI&DS), Shree Siddheshwar College Of Engineering, Solapur, India

sakshidama198@gmail.com

Abstract

This research paper provides a comprehensive review of the use of Natural Language Processing (NLP) in stock market analysis, based on a systematic exploration of Scopus-indexed literature. The primary focus is on how NLP technologies have been applied to extract and analyze textual data from financial news, reports, and social media to forecast market trends and guide trading decisions. Through a meta-analysis of documented case studies and empirical research, this paper evaluates the effectiveness of various NLP techniques, including sentiment analysis, topic modelling, and event detection, in enhancing market prediction models compared to traditional quantitative approaches. Results indicate that while NLP presents a significant advantage in interpreting unstructured data, its integration into trading algorithms also poses challenges such as data inconsistency, model overfitting, and the need for adaptive learning mechanisms. 37 The paper concludes by identifying future research directions that focus on improving the accuracy, robustness, and real-time capabilities of NLP applications in financial trading, underscoring the potential of NLP to transform financial market analytics.

Keywords

NLP (Natural Language Processing), Data Mining, Tokenization, Deep Learning, Reinforcement Learning, Supervised Learning.

1. Introduction

In the rapidly evolving field of financial markets, the ability to predict stock movements with high accuracy remains a paramount goal for investors and analysts alike. Traditional 1 stock market analysis has relied heavily on quantitative data; however, the digital age has ushered in an era where unstructured textual data is equally influential. News articles, financial reports, and increasingly, social media content play a crucial role in shaping market perceptions and movements. This has led to the integration of advanced computational techniques such as Natural Language Processing (NLP) to mine and interpret this vast amount of textual data.

[29] Natural Language Processing, a subfield of artificial intelligence, offers tools to understand and manipulate human language, turning text into actionable data. Its application in financial analytics is relatively recent but rapidly gaining traction [1] due to its potential to unearth insights that are not readily available through traditional numerical data analysis. For instance, [29] sentiment analysis can capture market mood from news headlines or social media feeds, which often precedes shifts in market dynamics. Likewise, topic modelling can identify emerging trends that might influence investment decisions.

Despite the promising advantages, [1] the application of NLP in stock market analysis is fraught with challenges. The ambiguity of language, the speed at which market-relevant news impacts stock prices, and the sheer volume of data to be processed require sophisticated models and algorithms. Moreover, the integration of NLP into trading strategies must be done carefully to avoid overfitting and to ensure [11] models are robust against the noise inherent in large datasets.

This paper reviews the current scope of NLP [1] in stock market analysis through a detailed examination of Scopus-indexed research. It aims to synthesize the findings from various studies to evaluate the effectiveness of NLP techniques and identify gaps in the current literature. Furthermore, it compares [4] the performance of NLP-based models to traditional models, providing a nuanced understanding of where NLP adds value and where its limitations lie. By conducting this meta-analysis, the paper seeks [11] to contribute to the growing field of AI in finance, offering insights into future directions for research and practical implementations.

2. Literature Review

FERNANDO G. D. C. FERREIRA1, AMIR H. GANDOMI [1], they present [16] a systematic review of the literature on Artificial Intelligence applied to investments in the stock market based on a sample of 2326 papers from the Scopus website between 1995 and 2019. [25] These papers were divided into four categories: portfolio optimization, stock market prediction using AI, financial sentiment analysis, and combinations involving two or more approaches. For each category, the initial introductory research to its state-of-the-art applications is described. In [17] addition, an overview of the review leads to the conclusion that this research area is gaining continuous attention and the literature is becoming increasingly specific and thorough.

2.1 Natural Language Processing (NLP)

M. IZADI AND M. N. AHMADABADI [2], they [18] reviewed how NLP-based models for SE problems are being evaluated by researchers. The findings indicate that currently there is no consistent and widely-accepted protocol for the evaluation of these models. [8] While different aspects of the same task are being assessed in different studies, metrics are defined based on custom choices, rather than a system, and finally, answers are collected and interpreted case by case. Consequently, there is a dire need to provide a methodological way of evaluating NLP-based models to have a consistent assessment and preserve the possibility of fair and efficient comparison.

[5] PRIYANKAR BOSE, SATYAKI ROY, PREETAM GHOSH [3], in this work, they applied natural language processing (NLP) based approaches on scientific literature published on COVID-19 to infer significant keywords that have contributed to our social, economic, demographic, psychological, epidemiological, clinical, and medical understanding of this pandemic. We identify key terms appearing in COVID literature that vary in representation when compared to other virus-borne diseases such as MERS, Ebola, and Influenza. We also identify countries, topics, and research articles that demonstrate that the scientific community is still reacting to the short-term threats such as transmissibility, health risks, treatment plans, and public policies, underpinning the need for collective international efforts towards long-term immunization and drug-related challenges.

R. SONBOL, G. REBDAWI AND N. GHNEIM [4], their [3] survey shows that the research direction has changed from the use of lexical and syntactic features to the use of advanced embedding techniques, especially in the last two years. Using advanced embedding representations has proved its effectiveness in most RE tasks (such as requirement analysis, extracting requirements from reviews and forums, and semantic-level quality tasks). However, representations that are based on lexical and syntactic features are still more appropriate for other RE tasks (such as modelling and syntax-level quality tasks) since they provide the required information for the rules and regular expressions used when handling these tasks. In addition, they identify four gaps in the existing literature, why they matter, and how future research can begin to address them.

2.2 Data mining

XIANGPENG WAN, MICHAEL C. LUCIC, HAKIM GHAZZA, YEHIA MASSOUD [5], In this work, they develop an automated Natural Language Processing (NLP)-based framework to empower and complement traffic reporting solutions by text mining social media, extracting desired information, and [2] generating alerts and warning for drivers. They employ the fine-tuned Bidirectional Encoder Representations from Transformers classification model to filer and classify data. Then, they apply the Question-Answering model to extract necessary information characterizing the reported incident such as its location, occurrence time, and nature of the incidents. Afterwards, they convert the collected information into alerts to be integrated into personal navigation assistants.

R. ESPINOSA, L. GARRIGA, J. J. ZUBCOFF AND J. -N. MAZÓN [6], [1] In this work, they present an infrastructure that allows non-expert users to (I) apply user-friendly data mining techniques on big data sources, and (ii) share results as Linked Open Data (LOD). [12] The main contribution of this work is an approach for democratizing big data through reusing the knowledge gained from data mining processes after being semantically annotated as LOD, then obtaining Linked Open Knowledge. Their work is [11] based on a model-driven viewpoint in order to easily deal with the wide diversity of open data formats.

2.3 Supervised Learning

P. C. CHHIPA, R. UPADHYAY, G. G. PIHLGREN, R. SAINI, S. UCHIDA AND M. LIWICKI [7], in this work they present a novel self-supervised pre-training method to learn efficient representations without labels on histopathology medical images utilizing magnification factors. Other state-of-the-art 4 works mainly focus on fully supervised learning approaches that rely heavily on human annotations. However, the scarcity of labelled and unlabelled data is a long-standing challenge in histopathology. Currently, representation learning without labels remains unexplored in the histopathology domain. The proposed method, Magnification Prior Contrastive Similarity (MPCS), enables self-supervised learning of representations without labels on small-scale breast cancer dataset Break His by exploiting magnification factor, inductive transfer, and reducing human prior. The proposed method matches fully supervised learning state-of-the-art performance in malignancy classification when only 20% of labels are used in fine-tuning and outperform previous works in fully supervised learning settings for three public breast cancer datasets, including Break His.

YUANHONG CHEN; FENGBEI LIU; MICHAEL ELLIOTT; CHUN FUNG KWOK; CARLOS PEÑA-SOLORZANO; HELEN FRAZER [8],In this paper, they introduce a new deep-learning diagnosis framework, called Internal, that is designed to be highly accurate and interpretable. Internal 6 consists of a student-teacher framework, where the student model is an interpretable prototype-based classifier (Protopines) and the teacher is an accurate global image classifier (GlobalNet). The two classifiers are mutually optimised with a novel reciprocal learning paradigm in which the student ProtoPNet learns from optimal pseudo labels produced by the teacher GlobalNet, while GlobalNet learns from ProtoPNet's classification performance and pseudo labels.

2.4 Deep Learning

DIONYSIS GOULARAS; SANI KAMIS[9], 7 In this work they evaluate and compare ensembles and combinations of CNN and a category of RNN the long short-term memory (LSTM) networks. Additionally, we compare different word embedding systems such as the Word2Vec and the global vectors for word representation (GloVe) models. For 14 the evaluation of those methods, they used data provided by the international workshop on semantic evaluation (SemEval), which is one of the most popular international workshops on the area. This study contributes to the field 7 of sentiment analysis by analysing the performances, advantages and limitations of the above methods with an evaluation procedure under a single testing framework with the same dataset and computing environment.

G. KARATAS, O. DEMIR AND O. KORAY SAHINGOZ [10], 19 In this paper, it is aimed to survey deep learning-based intrusion detection system approach by making a comparative work of the literature and by giving the background knowledge either in deep learning algorithms or in intrusion detection systems.

3. METHODOLOGY

3.1 Working of implemented System

Figure 1: Working of implemented model

1 3.2 Data Collection :

To collect data for our stock market trading analysis project, we utilized several methods. We used web scraping techniques to extract data from websites, such as financial news sites, social media platforms, and forums. Tools like Beautiful Soup (for Python) can be helpful for this.

□ APIs: 38 Many financial data providers offer APIs that allow us to access their data programmatically. We use these APIs to retrieve real-time and historical data for analysis. Examples include the Alpha Vantage API, the Yahoo Finance API, and the Quandy API.

□ Academic Databases: We use academic databases like Scopus to search for and access research papers related to NLP 1 in stock market analysis.

□ Surveys and Questionnaires: Conducting surveys or questionnaires to gather data from traders, investors, or experts in the field of stock market analysis to gather insights or validate your findings.

□ Data Purchasing: In some cases, you may need to purchase specialized datasets that are not publicly available. This can include datasets on financial transactions, market sentiment, or other relevant information.

□ Social Media Monitoring: Monitor [2] social media platforms like Twitter, Stock Twits, and Reddit for discussions and sentiments related to stocks and financial markets. Tools like Tweepy (for Twitter) can be used for this purpose.

□ [41] Historical Data Archives: Access historical data archives, such as those provided by financial data providers or academic institutions, to retrieve past stock prices, company financials, and economic indicators for backtesting and analysis.

□ Manual Collection: In some cases, we may need to manually collect data from sources [2] that do not provide an API or allow web scraping. This can be time-consuming but may be necessary for certain types of data.

3.3 Data Preprocessing :

[26] In the data preprocessing stage ,we need to clean and prepare the data for analysis. [28] Data preprocessing is a crucial step in the data analysis process that involves cleaning, transforming, and preparing raw data into a format that is suitable for analysis. The goal of data preprocessing is [2] to ensure that the data is accurate, complete, and properly formatted to improve the performance and reliability of data analysis and machine learning models. Techniques used in data preprocessing :-

□ Text Data Cleaning: To remove special characters, numbers, and punctuation. [1] And to convert all text to lowercase to ensure consistency. Remove stopwords (common words that do not carry much meaning) using a predefined list or NLP library.

□ Tokenization: Tokenize the text into words or phrases to prepare it for further analysis. Use NLP libraries like NLTK or spaCy for tokenization.

□ Data Normalization: Scaling the data to a standard range to ensure that different features contribute equally to the analysis. Common normalization techniques include Min-Max scaling and z-score normalization.

□ Data Deduplication: Identify and remove duplicate data points from the dataset. Ensure that each data point is unique and does not skew the analysis.

□ Data Sampling: Random Sampling: Select a random subset of the data for analysis. Stratified Sampling: Ensure that the sample is representative of the population by maintaining the same class distribution as the original data.

□ Data Augmentation: Generate new data points by applying transformations such as rotation, flipping, or scaling to existing data points. Commonly used in image processing 26 and natural language processing tasks.

□ Handling Missing Data: Remove rows or columns with missing values. Fill in missing values with the mean, median, or mode of the column. Use machine learning algorithms to predict missing values based on other data.

3.4 Data Splitting:

30 Data splitting is the process of dividing a dataset into multiple subsets for different purposes, such as training a machine learning model, validating the model, and testing the model's performance. Data splitting techniques 2 are used to partition historical data into training, validation, and test sets. These sets are then used to develop, fine-tune, and evaluate trading strategies and models.

□ Fixed Windows: 1 The historical data is divided into fixed-size windows, such as days, weeks, or months. Each window is used as a separate subset for training, validation, and testing.

□ Rolling Windows: Similar to fixed windows, but the windows move forward in time by a fixed interval. This allows the model to be trained on more recent data while still using older data for validation and testing.

□ Time-Based Splitting: 2 The data is split based on a specific date or time point. Data before the split point is used for training, data at the split point is used for validation, and data after the split point is used for testing.

□ Cross-Validation: The data 10 is divided into k folds, with each fold used once as a validation while the k - 1 remaining folds form the training set. This process is repeated k times, with each fold used as the validation set exactly once.

□ Walk-Forward Validation: Similar to rolling windows, but the model is retrained on the updated training set at each time step. This allows the model to adapt to changing market conditions over time.

□ Train-Test Split: A simple approach where a fixed percentage of the [4] data is used for training and the remaining data is used for testing. This is useful for quick model evaluation but may not capture the temporal nature of stock market data.

3.5 Model Training :

Model training is the process of teaching [13] a machine learning algorithm to recognize patterns and make predictions based on input data. During training, the algorithm learns the relationships [39] between the input data (features) and the target variable (the variable to be predicted). The goal of training is [13] to minimize the difference between the predicted outputs of the model and the actual outputs in the training data.

□ Supervised Learning: This technique involves training a model [1] on historical data where the inputs are features (e.g., stock prices, trading volumes) and the output is a target variable (e.g., buy/sell signals, price movements). Common [26] supervised learning algorithms include:

-Regression: Predicting a continuous value, such as stock prices or returns.

-Classification: Predicting a discrete value, such as whether to buy, sell, or hold a stock.

□ [1] Time Series Analysis: Stock market data is often treated as a time series, where each data point is indexed by time. Time series analysis techniques, such as autoregressive integrated moving average (ARIMA) and exponential smoothing, are used to model and predict future stock prices or returns.

□ Reinforcement Learning: This technique involves training a model to make sequential decisions (e.g., when to buy or sell stocks) based on rewards or penalties received from the environment (e.g., stock market). Reinforcement learning algorithms, such as Q-learning and deep reinforcement learning, can be used to develop trading strategies.

□ Ensemble Learning: Ensemble learning involves training multiple models and combining their predictions to improve accuracy and robustness. Techniques like bagging (bootstrap aggregating) and boosting [11] are commonly used in stock market trading analysis to reduce overfitting and improve performance.

□ Neural Networks: [1] Deep learning techniques, such as artificial neural networks (ANNs) and convolutional neural networks (CNNs), can be used to analyze complex stock market data and extract meaningful patterns. These models can learn from large amounts of data and capture nonlinear relationships.

□ Feature Engineering: Feature engineering involves creating new features from existing data to improve model performance. In stock market trading analysis, features like moving averages, relative strength index (RSI), and MACD (moving average convergence divergence) are commonly used to capture trends and momentum in stock prices.

□ Backtesting: After training a [2] model, it is important to evaluate its performance using historical data. Backtesting involves simulating trades based on the model's predictions and analyzing the results to assess its profitability and risk.

□ Support Vector Machines (SVM): SVM is a supervised learning algorithm that can be used for both classification and regression tasks. It works by finding the hyperplane that best separates different classes or predicts a continuous value.

3.6 Model Selection :

Model selection is the process of choosing the best machine learning model or algorithm for a given problem based on its performance on a dataset. It involves evaluating multiple models with different configurations and selecting [10] the one that performs best according to predefined criteria. The goal of model selection is to find a model that generalizes well to new, unseen data and achieves the desired performance metrics. Here are model selection techniques used in stock market trading analysis:

□ Cross-Validation: Cross-validation [32] is a technique used to assess how well a model will generalize to an independent dataset. In stock market trading analysis, k-fold cross-validation is commonly used, where [10] the dataset is divided into k subsets. The model is trained on k-1 subsets and validated on the remaining subset, repeating the process k times. The average performance across all folds is used to evaluate the model.

□ Grid Search: Grid search is a technique used to find the optimal hyperparameters for a model.

It [15] involves defining a grid of hyperparameter values and evaluating the model's performance for each combination of values. The combination of hyperparameters that results in the best performance is selected as the final model.

□ Random Search: Random search is similar to grid search but instead of evaluating all possible combinations of hyperparameters, it randomly samples a subset of them. [27] Random search is often more efficient than grid search for high-dimensional hyperparameter spaces.

□ Bayesian Optimization: Bayesian optimization is a [40] sequential model-based optimization technique that uses probabilistic models to select the next set of hyperparameters to evaluate. It uses [27] the results of previous evaluations to update the probabilistic model and focus the search on promising regions of the hyperparameter space.

□ [13] Ensemble Methods: Ensemble methods combine multiple models to improve prediction accuracy and robustness. Common ensemble methods used in stock market trading analysis include bagging (bootstrap aggregating) and boosting. Bagging [21] involves training multiple models on different subsets of the data and averaging their predictions, while boosting involves sequentially training models to correct the errors of previous models.

□ Feature Selection Techniques: Feature selection techniques are used to identify the most relevant features for the model. This helps reduce overfitting and improve model performance. Common [35] feature selection techniques include recursive feature elimination (RFE), which recursively removes features and evaluates the model's performance, and feature importance scores from tree-based models like random forests.

□ Model Comparison: Finally, [1] models can be compared based on various metrics such as accuracy, precision, recall, F1-score, or profitability metrics specific to stock market trading. The model that performs best on these metrics is selected as [21] the final model.

4. System Implementation and Result

4.1 Implementation

- Hardware Requirement

Processor: i5

Reason: Processor i5 is used for best performance. The i5 [1] is a powerful machine suitable for developing machine learning model.

Ram: 8GB and above

Reason: With 8GB of RAM there would be enough memory to run a few lightweight programs at a time and can open a handful of browser tabs, and many other operations at a time

• Software Requirement

Operating System: Windows 8/10 or above, ubuntu

Reason: It can support to any operating system of windows but for fast execution [21] we are using ubuntu.

Library: scikit learn, TensorFlow, Pandas, NumPy, Seaborn

Reason: scikit-learn: This is [11] a machine learning library that provides simple and efficient tools for data mining and data analysis.

TensorFlow: TensorFlow is an open-source machine learning framework used for building and training [1] neural network models. It provides tools for creating deep learning models, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs).

Pandas: [33] Pandas is a powerful data manipulation and analysis library in Python. It provides data structures [34] such as Data Frame and Series, which are ideal for working with structured data like stock prices and trading volumes

NumPy: It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays.

Seaborn: [24] Seaborn is a data visualization library based on matplotlib that provides a high-level interface for drawing attractive and informative statistical graphics.

Toolkit: NLTK [3] (Natural Language Toolkit)

Reason: NLTK (Natural Language Toolkit) is a library in Python that is widely used for natural language processing (NLP) tasks. While NLTK is not typically used directly [1] in stock market trading analysis, it can be beneficial in certain aspects of financial analysis where text data is involved.

4.2: Result:

The [22] F-measure, also known as the F1-score, is a metric commonly used in machine learning and statistics to evaluate the performance of a model, particularly in binary classification tasks. [31] It is the harmonic mean of precision and recall and provides a balance between these two metrics.

The [23] traditional F measure is designed as follows:

F-Measure = ((Two * Accuracy * Memory))/ (Accuracy + Memory)

It can first define the accurate positives (TP), wrong positives (FP), and wrong negatives (FN) as follows in order to simplify the equation:

accuracy [20] = TP / (TP + FP)

memory = TP / (TP + FN)

Substituting these equations in the F-measure equation,

F-measure = 2 *((TP / (TP + FP) * TP / (TP + FN)))/ (TP / (TP+ FP) + TP / (TP + FN)) Simplifying this equation,

F-measure = 2TP / (2TP + FP + FN)

Calculated accuracy is [9] as follows:

Precision = Accurate Positives / (Accurate Positives +Wrong Positives)

Precision = 100 / (100 + 70).

Precision = 0.588.

It can analyze the memory as follows:

Recall = Accurate Positives / (Accurate Positives + Wrong Negatives)

Recall = 100 / (100 + 15).

Recall = 0.869.

This shows that the model has poor precision, but excellent recall. Finally, it can analyze the F-Measure as follows:

F-Measure = ((2 * Accuracy * Memory))/ (Accuracy+ Memory)

F-Measure = ((2 *0.588 * 0.869))/ (0.588+0.869).

F-Measure = ((2 * 0.509))/ 1.454.

F-Measure = 1.018 / 1.454.

F-Measure=0.700.

Name of journal	Precision	Recall	F-Score
Scopus	0.588	0.869	0.700

Table 1. Parameters like precision, recall and f score for corresponding journal

5. CONCLUSION

Our project showcases [3] the effectiveness of utilizing natural language processing (NLP) to analyze stock market trends and sentiments using research articles from Scopus. Through the application of NLP techniques, we have been able to extract valuable insights from a large corpus of textual data, providing a deeper understanding of market dynamics. [1] One of the key findings of our project is the ability of NLP to accurately gauge market sentiment, which has significant implications for traders and investors. By analyzing sentiment from research articles, we can identify emerging trends and sentiments that may impact market behaviour. This information can be used to inform trading strategies and make more informed investment decisions. Furthermore, our project highlights [2] the importance of data-driven decision-making in the financial markets. By leveraging NLP to analyze textual data, we can uncover patterns and insights that may not be apparent from numerical data alone. This [4] can lead to more effective risk management strategies and improved investment outcomes. Looking ahead, there are several opportunities for future research in this area. One avenue for exploration [2] is the use of more advanced NLP techniques, such as deep learning, to further improve the accuracy and efficiency of sentiment analysis. Additionally, integrating data from other sources, such as social media and financial news, could provide a more comprehensive view of market sentiment. Overall, our project demonstrates the potential of NLP to enhance 1 stock market analysis and decision-making processes. By leveraging NLP techniques, we can gain valuable insights into market trends and sentiments, ultimately leading to more informed and effective trading strategies.

References

[1] F. G. D. C. Ferreira, A. H. Gandomi and R. T. N. Cardoso, "Artificial Intelligence Applied to Stock Market Trading: A Review," in IEEE Access, vol. 9, pp. 30898-30917, 2021, doi: 10.1109/ACCESS.2021.3058133.

[2] M. Izadi and M. N. Ahmadabadi, "On the Evaluation of NLP-based Models for Software Engineering," 2022 IEEE/ACM 1st International Workshop on Natural Language-Based Software Engineering (NLBSE), Pittsburgh, PA, USA, 2022, pp. 48-50, doi: 10.1145/3528588.3528665.

[3] P. Bose, S. Roy and P. Ghosh, "A Comparative NLP-Based Study on the Current Trends and Future Directions in COVID-19 Research," in IEEE Access, vol. 9, pp. 78341-78355, 2021, doi: 10.1109/ACCESS.2021.3082108.

[4] R. Sonbol, G. Rebdawi and N. Ghneim, "The Use of NLP-Based Text Representation Techniques to Support Requirement Engineering Tasks: A Systematic Mapping Review," in IEEE Access, vol. 10, pp. 62811-62830, 2022, doi: 10.1109/ACCESS.2022.3182372.

[5] X. Wan, M. C. Lucic, H. Ghazzai and Y. Massoud, "Empowering Real-Time Traffic Reporting Systems With NLP-Processed Social Media Data," in IEEE Open Journal of Intelligent Transportation Systems, vol. 1, pp. 159-175, 2020, doi: 10.1109/OJITS.2020.3024245.

[6] R. Espinosa, L. Garriga, J. J. Zubcoff and J. -N. Mazón, "Linked Open Data mining for democratization of big data," 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 2014, pp. 17-19, doi: 10.1109/BigData.2014.7004479

[7] P. C. Chhipa, R. Upadhyay, G. G. Pihlgren, R. Saini, S. Uchida and M. Liwicki, "Magnification Prior: A Self-Supervised Method for Learning Representations on Breast Cancer Histopathological Images," 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2023, pp. 2716-2726, doi: 10.1109/WACV56688.2023.00274.

[8] C. Wang et al., "An Interpretable and Accurate Deep-Learning Diagnosis Framework Modeled with Fully and Semi-Supervised Reciprocal Learning," in IEEE Transactions on Medical Imaging, vol. 43, no. 1, pp. 392-404, Jan. 2024, doi: 10.1109/TMI.2023.3306781.

[9] D. Goularas and S. Kamis, "Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data," 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), Istanbul, Turkey, 2019, pp. 12-17, doi: 10.1109/Deep-ML.2019.00011.

[10] G. Karatas, O. Demir and O. Koray Sahingoz, "Deep Learning in Intrusion Detection Systems," 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), Ankara, Turkey, 2018, pp. 113-116, doi: 10.1109/IBIGDELFT.2018.8625278.

Sources

1	https://www.researchgate.net/publication/375414050_Machine_learning-based_approaches_for_financial_market_prediction_A_comprehensive_review INTERNET 3%
2	https://www.researchgate.net/publication/347021661_Empowering_Real-Time_Traffic_Reporting_Systems_With_NLP-Processed_Social_Media_Data INTERNET 2%
3	https://www.researchgate.net/publication/344152757_Towards_a_Semantic_Representation_for_Functional_Software_Requirements INTERNET 2%
4	https://www.researchgate.net/publication/359254577_Magnification_Prior_A_Self-Supervised_Method_for_Learning_Representations_on_Breast_Cancer_Histopathological_Images INTERNET 2%
5	https://doaj.org/article/f9cbe2466e594bf3b0b9731f2359a116 INTERNET 1%
6	https://ieeexplore.ieee.org/document/10225391/ INTERNET 1%
7	https://www.semanticscholar.org/paper/Evaluation-of-Deep-Learning-Techniques-in-Sentiment-Goularas-Kamis/e3663bf5eb16fdcbcab74e1b45de03e24c843fc5 INTERNET 1%
8	https://ieeexplore.ieee.org/document/9808680 INTERNET 1%
9	https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/ INTERNET 1%
10	https://www.geeksforgeeks.org/cross-validation-machine-learning/ INTERNET 1%
11	https://link.springer.com/article/10.1007/s43546-023-00618-x INTERNET 1%
12	https://ieeexplore.ieee.org/document/7004479/ INTERNET 1%
13	https://ai-jobs.net/insights/model-training-explained/ INTERNET 1%
14	https://ieeexplore.ieee.org/document/8876896 INTERNET 1%

15	https://www.datacamp.com/tutorial/parameter-optimization-machine-learning-models INTERNET 1%
16	https://www.researchgate.net/publication/349189858_Artificial_Intelligence_Applied_to_Stock_Market_Trading_A_Review INTERNET 1%
17	https://ieeexplore.ieee.org/abstract/document/9350582/ INTERNET 1%
18	https://arxiv.org/pdf/2203.17166.pdf INTERNET 1%
19	https://ieeexplore.ieee.org/document/8625278 INTERNET 1%
20	https://stackoverflow.com/questions/35365007/tensorflow-precision-recall-f1-score-and-confusion-matrix INTERNET 1%
21	https://vivek-murali.medium.com/ensemble-learning-combining-models-for-better-predictions-53a4c80cb18a INTERNET <1%
22	https://klu.ai/glossary/accuracy-precision-recall-f1 INTERNET <1%
23	https://www.jetir.org/papers/JETIR2308230.pdf INTERNET <1%
24	https://medium.com/geekculture/python-seaborn-statistical-data-visualization-in-plot-graph-f149f7a27c6e INTERNET <1%
25	https://ieeexplore.ieee.org/document/9350582/ INTERNET <1%
26	https://www.tandfonline.com/doi/full/10.1080/15427560.2021.1974443 INTERNET <1%
27	https://www.linkedin.com/pulse/model-optimization-machine-learning-random-vs-neves-junior-phd-x459f#:~:text=Efficiency in High-Dimensional Spaces: Random Search is often,it can find good solutions with fewer iterations. INTERNET <1%
28	https://medium.com/@abuzarzulfikar/a-comprehensive-guide-to-data-preprocessing-219634f3247b INTERNET <1%

29	https://www.ibm.com/topics/natural-language-processing INTERNET <1%
30	https://medium.com/data-and-beyond/how-to-split-data-in-machine-learning-5-simple-strategies-and-python-examples-a500c3f2f750 INTERNET <1%
31	https://towardsdatascience.com/performance-metrics-confusion-matrix-precision-recall-and-f1-score-a8fe076a2262 INTERNET <1%
32	https://medium.com/@redeaddiscolll/advanced-machine-learning-and-deep-learning-techniques-for-stock-market-analysis-06e5b8c6b62c INTERNET <1%
33	https://learntodatascience.com/python-pandas-for-data-manipulation-and-analysis/ INTERNET <1%
34	https://www.geeksforgeeks.org/dataframe-vs-series-in-pandas/ INTERNET <1%
35	https://www.analyticsvidhya.com/blog/2021/06/feature-selection-techniques-in-machine-learning-2/#:~:text=Feature selection techniques in machine learning involve selecting,best set of features for a given dataset. INTERNET <1%
36	http://www.sswcoe.edu.in/ INTERNET <1%
37	https://www.sciencedirect.com/science/article/pii/S0963869523002451 INTERNET <1%
38	https://christianmartinezfinancialfox.medium.com/how-to-access-financial-data-using-python-and-apis-cc01af9b4cc7 INTERNET <1%
39	https://databasetown.com/supervised-learning-algorithms/ INTERNET <1%
40	https://arxiv.org/pdf/2101.02289.pdf INTERNET <1%
41	https://www.sciencedirect.com/science/article/pii/B9780128158746000125 INTERNET <1%

EXCLUDE CUSTOM MATCHES OFF

EXCLUDE QUOTES ON

EXCLUDE BIBLIOGRAPHY ON