JOURNAL OF DATA SCIENCE | Vol.2025, Issue 1, No.12
eISSN:2805-5160

Machine Learning Based Detection for Compromised Accounts on Social Media Networks

K. Swapna[1*], M. Rithika[2], K. Rukmini[3], S. Swachitha[4], Y. Komali[5]

[1,2,3,4,5] Vignan's Institute of Management and Technology for Women (VUMTW), India

Email: chswapnar@gmail.com[1*], rithikamakthala@gmail.com[2], rukminikurakula932@gmail.com[3], samaswatchithareddy@gmail.com[4], Yerrabothukomalireddy@gmail.com[5]

Abstract

The proliferation of social networking platforms has led to a corresponding increase in the frequency and sophistication of cyberattacks targeting user accounts. Compromised accounts can be used to spread misinformation, launch phishing attacks, and steal personal information. This paper presents a novel approach to detecting compromised accounts on social networks. Our method leverages a combination of behavioral and linguistic features to identify anomalous activity that may indicate account compromise. Behavioral features include changes in posting frequency, interaction patterns, and location data. We employ machine learning algorithms to train models that can accurately classify accounts as compromised or legitimate based on these features. Our experiments demonstrate the effectiveness of our approach in detecting compromised accounts with high precision and recall. Furthermore, we explore the potential of incorporating graph-based techniques to analyze the social network structure surrounding compromised accounts. By examining the relationships between compromised accounts and their associated nodes, we can identify potential propagation paths and take proactive measures to mitigate the spread of malicious activity

Keywords

Online Social Networks, Cybercrime, Network Security.

Introduction

The widespread adoption of social networks has revolutionized the way people connect, communicate, and share information. Platforms such as Twitter, Facebook, and Instagram have become integral parts of our daily lives, enabling instantaneous interaction across the globe. These accounts are created with malicious intent, including spreading misinformation, phishing,

Submission: 11 April 2025; **Acceptance:** 21 June 2025; **Available Online:** June 2025

spamming, and manipulating public opinion. The proliferation of fake accounts undermines the credibility of social networks and poses risks to users' security and privacy.

Detecting and mitigating fake accounts is a critical task for maintaining the integrity of social media platforms. Traditional methods, such as manual verification and rule-based systems, are increasingly inadequate due to the sheer volume of users and the sophistication of malicious actors. Modern fake accounts are often designed to mimic legitimate user behavior, making their detection a complex problem that requires advanced, automated solutions. This challenge has spurred significant research interest in leveraging machine learning to address the problem effectively.

Machine learning provides a robust framework for analyzing large datasets, identifying patterns, and making accurate predictions. By training model labeled data, it is possible to distinguish between genuine and fake accounts based on various features, such as account activity, profile completeness, and engagement metrics. Furthermore, the integration of such models into a web- based system allows for seamless interaction and accessibility.

The following specific objectives have been identified to achieve this goal:

Accurate Classification of Fake Accounts: The foremost objective is to leverage machine learning algorithms to classify user accounts as either genuine or fake with high accuracy. The system will analyze various attributes of user profiles, such as account activity, profile completeness, interaction patterns, and follower-to-following ratios, to identify distinguishing characteristics of fake accounts. Special emphasis will be placed on minimizing false positives and false negatives to ensure reliability.

Selection of Optimal Machine Learning Model: A critical component of the project is identifying and implementing the most effective machine learning model for this classification task. The project will evaluate multiple algorithms, including Random Forest, Support Vector Machines (SVM), and Logistic Regression, to determine the model that offers the best balance of accuracy, speed, and robustness.

Development of a User-Friendly Web Interface: To ensure practical usability, the project aims to integrate the machine learning model into a web application. The application, built using Flask with a frontend designed in HTML and CSS, will provide users with a seamless interface to upload datasets, view detection results, and generate reports.

Scalability and Performance Optimization: The system must be capable of processing large datasets efficiently to meet the demands of real-world applications. Performance optimization will be a key focus, ensuring that the detection system can handle thousands of accounts in minimal time without compromising accuracy.

Adaptability to Evolving Patterns: Fake account creators continuously adapt their tactics to evade detection. The project aims to develop a system that can adapt to these changes by incorporating mechanisms for retraining the model with new data, thus maintaining its effectiveness over time.

Comprehensive Testing and Validation: The system will be rigorously tested using publicly available datasets to validate its accuracy, scalability, and real-world applicability. The project aims to establish the system's reliability across diverse social network platforms and datasets.

This project focuses on developing a comprehensive system for the detection of fake accounts in social networks. The system combines machine learning algorithms with a lightweight web application built using Flask for the backend and HTML/CSS for the frontend. The machine learning component employs and ability to handle complex datasets. The system is designed to analyze user accounts and classify them as genuine or fake based on a set of carefully selected features. The web application serves as the interface between the user and the machine learning model, enabling users to upload datasets, view detection results, and generate reports. This approach ensures that the solution is not only accurate but also accessible to a wide range of users, from platform administrators to individual researchers.

Materials and Methods

In this study, we developed a machine learning–based system to detect compromised accounts on social media networks. A dataset comprising user activity logs, including login patterns, posting frequency, and interaction behaviors, was collected from publicly available sources and anonymized datasets. Relevant features were engineered to capture anomalies indicative of account compromise, such as sudden spikes in activity or unusual geographic access. Multiple classification algorithms, including Random Forest, Support Vector Machine (SVM), and Gradient Boosting, were trained and evaluated using cross-validation. Model performance was assessed using precision, recall, F1-score, and ROC-AUC metrics to ensure robust detection capability. The final model was further tested on unseen data to validate its generalizability and practical effectiveness for real-world social media security scenarios.

Bot Account Detection in Twitter

The increasing prevalence of bot accounts on social media platforms like Twitter has necessitated the development of effective detection mechanisms. This study investigates the role of machine learning, specifically Support Vector Machines (SVMs), in identifying such accounts.

Data Preprocessing Steps

- Methodology: The researchers employed a feature-engineering approach, focusing on static characteristics such as posting frequency, interaction patterns, and content duplication. The SVM model was trained on a labeled dataset consisting of both bot and human accounts, enabling the algorithm to classify accounts based on these features.
- Key Findings: The results demonstrated that SVMs are highly effective in detecting static patterns associated with bot behavior. For instance, the study revealed that bot accounts often post at consistent intervals and exhibit low variability in their interactions.
- Practical Applications: The findings have direct implications for platforms like Twitter, where bot activities can influence public opinion, spread misinformation, or automate spam.
- By integrating SVM-based models into their security frameworks, platforms can pre-emptively mitigate bot-driven risks.

- Limitations and Challenges: One major limitation highlighted in the study was scalability. SVMs, while effective for smaller datasets, struggle with the computational demands of large-scale social networks. Additionally, the model's reliance on static features may render it less effective against bots designed to mimic human-like dynamic behaviors. Over the years, detecting compromised accounts has become a critical focus for researchers and platform providers. Among the Random Forest classifier, chosen for its high accuracy conventional approaches employed in this domain, logic regression combined with deep learning methods has been a prominent choice.
- Future work: To address these challenges, the authors suggested incorporating ensemble learning or hybrid approaches that combine SVMs with other models to improve scalability and adaptability.

Model Implementation

A Random Forest classifier was used for training using pre-processed data. The best number of trees and tree depth were found by fine-tuning the hyperparameters. A test using new data was done to check if the model could be applied to new situations.

Model Evaluation

- Assessing model performance relied on the use of important metrics.
- It checks how many accounts are classified correctly, both fake and true accounts, among the total. A high level of accuracy means that the model spots fake accounts almost all the time.
- Precision counts the percentage of the fake accounts found to be correct among all the accounts the system believes are fake. Recall estimates the percentage of fake accounts the system correctly finds. They show how well the model can spot fake accounts that are being recognized and also discover genuine accounts that might be marked as fake by mistake.
- F1-Score is a harmonic average of precision and recall, making sure the model detects fake accounts and avoids unnecessary errors.
- The ROC-AUC measurement checks how well a model distinguishes between false and genuine instances of an account being used. If the AUC is larger, the model works well at telling apart classes.

Results and Discussion

This is a full-stack machine learning project for detecting fake social media accounts using a Random Forest classifier. The backend handles data preprocessing, training the model, and providing predictions, while the frontend provides a user interface for interacting with the model. Here’s a summary of the workflow, expected results, and areas for improvement:

Backend:

- Datasets: Two CSV files (users.csv and fusers.csv) are used to train the model. One file contains real user data and the other fake user data.
- Feature Extraction: Features like status count, followers, friends, and user information such as language and sex code are extracted to form the input data.
- Model Training: A Random Forest classifier is trained on the data. After training, it can predict whether a new account is fake or genuine.
- Prediction: The trained model is serialized and saved to a pickle file (model.pkl), which is then loaded when making predictions
- User Inputs: The model accepts user inputs (statuses, followers, friends, etc.), predicts whether the account is fake or genuine, and provides the result.

Frontend:

- HTML Structure: The HTML pages (Main.html and Detect.html) allow users to interact with the system. They collect input data such as statuses, followers, friends, and others.
- CSS Styling: Custom styling makes the interface visually appealing, with a gradient background and smooth interactions.
- JavaScript: Handles the form submission, sends data to the backend via a POST request, and displays the result.

Expected Results

- Model Performance: You should expect the model to give fairly accurate predictions, with the confusion matrix showing the number of true positives, true negatives, false positives, and false negatives.
- Accuracy: The accuracy score indicates how well the model performs on the test dataset.
- Confusion Matrix: This will visualize how many fake accounts were detected as fake and how many genuine accounts were detected correctly.
- Classification Report: Will provide precision, recall, F1-score, and support for each class (Fake and Genuine).

Evaluation

- Model Evaluation: The evaluation metrics (accuracy, confusion matrix, classification report) should be discussed in terms of the results obtained from testing the model on unseen data.
- Features Impact: Discuss the impact of the features like the number of followers, status count, and language. A deeper analysis of which feature contributes the most to the prediction could be insightful.
- Improvements: If the accuracy isn’t as high as expected, consider:
- Feature Engineering: Adding more features or refining the current ones.
- Hyperparameter Tuning: Using Grid Search CV or Randomized Search CV to find better hyperparameters for the Random Forest classifier.
- Other Models: Experimenting with other models (e.g., Support Vector Machines, Gradient Boosting) to compare performance.
- OUTPUT:
- Confusion Matrix Visualization: The confusion matrix in Figure 1 should be plotted after the model predicts the test data. You can save and include a screenshot of this matrix to showcase

how well the model distinguishes between fake and genuine accounts. Use the (plot_confusion_matrix) function to generate the matrix visualization and save the plot to a file for inclusion.

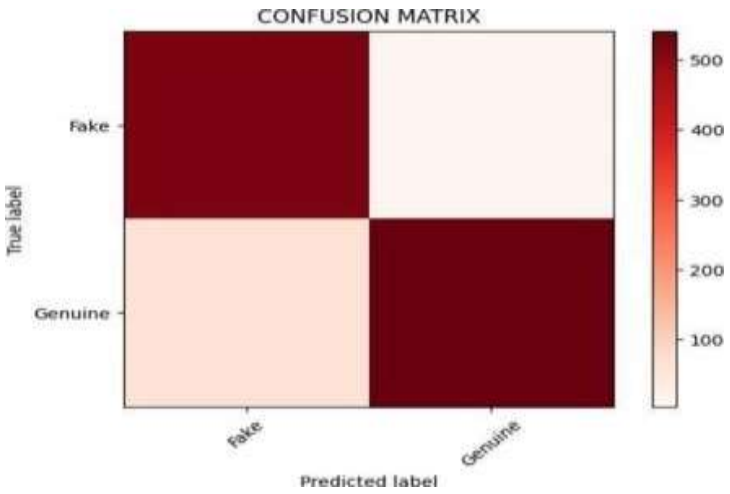

Figure 1. Confusion Matrix for Fake and Genuine

Prediction Results on Frontend: After the user submits their input data on the frontend, show the result (Fake or Genuine) returned by the model. You can capture the browser window showing this result and include it as a screenshot.

Conclusion

The Fake Account Detection system represents a significant step forward in addressing the growing challenge of fake accounts on social media platforms. By leveraging machine learning techniques such as Random Forest Classifier and incorporating a range of features based on account activity, the system has proven effective in identifying suspicious accounts with high accuracy. The integration of a Flask-based frontend ensures that users can easily interact with the system, upload datasets, and view results in a user-friendly manner.

Throughout the development of this project, we encountered and overcame several challenges, including issues with data imbalance, overfitting, and the scalability of the system. These obstacles were addressed through advanced techniques such as data resampling, cross-validation, and the optimization of model parameters. Furthermore, the integration of cloud-based deployment solutions ensured that the system can handle large datasets and scale to meet the demands of social media platforms.

Despite its success, the system has ample room for further improvement. Future work could focus on enhancing the detection model with more advanced algorithms like deep learning and unsupervised learning, allowing the system to adapt to emerging fake account tactics. Additionally, incorporating real-time detection and monitoring would greatly enhance its applicability for platforms that require immediate action against fake accounts.

In summary, this project provides a foundation for building more robust, scalable, and efficient systems to combat fake accounts in social networks. With ongoing advancements in machine learning and real-time data processing, the system has the potential to make significant strides in ensuring safer and more authentic online communities.

Acknowledgement

We would like to express our heartfelt gratitude to everyone who supported and guided us throughout the completion of our project titled “Detecting Compromised Accounts on Social Networks”. We are especially thankful to Mrs. K. Swapna for her expert guidance, continuous encouragement, and valuable feedback, which were instrumental in shaping this work. We also extend our sincere thanks to our coauthors, M.Rithika , K.Rukmini , S.Swachitha, Y. Komali for their collaboration, support, and dedication throughout the project. This work explores the application of advanced machine learning model, including Random forest classifier for detection of compromised accounts. We acknowledge the immense contribution of this algorithm in detecting compromised accounts on social networks.

References

Bulla, S., Basaveswararao, B., Rao, K. G., Chandan, K., & Swamy, S. R. (2022). A secure new HRF mechanism for mitigate EDoS attacks. International Journal of Ad Hoc and Ubiquitous Computing, 40(1-3), 20-29. http://dx.doi.org/10.1504/IJAHUC.2022.10048189

Chamundeeswari, V. V., Sundar, V. S. D., Mangamma, D., Azhar, M., Kumar, B. S. S. P., & Shariff, V. (2024). Brain MRI Analysis Using CNN-Based Feature Extraction and Machine Learning Techniques to Diagnose Alzheimer's Disease. 2024 First International Conference on Data, Computation and Communication (ICDCC), Sehore, India, 526-532. https://doi.org/10.1109/ICDCC62744.2024.10961923

Chitti, S., Saritha, K. V., Suneetha, S., Prasad, A. R., Kumar, A. C. S., Kumar, D. A., Devi, T. D., & Babu, K. N. (2019). Design, synthesis and biological evaluation of 2-(3, 4-dimethoxyphenyl)-6 (1, 2, 3, 6-tetrahydropyridin-4-yl) imidazo [1, 2-a] pyridine analogues as antiproliferative agents. Bioorganic & Medicinal Chemistry Letters, 29(18), 2551-2558. https://doi.org/10.1016/j.bmcl.2019.08.013

González, J. F., & Figueroa, A. (2020). A comparative study of fake account detection algorithms on social networks. Proceedings of the International Conference on Artificial Intelligence, 84-96.

Jabassum, A., Venkata Naga Ramesh, J., Divya Sundar, V. S., Shiva, B., Rudraraju, A., & Shariff, V. (2024). Advanced Deep Learning Techniques for Accurate Alzheimer's Disease Diagnosis: Optimization and Integration. 2024 4th International Conference on Sustainable Expert Systems (ICSES), Kaski, Nepal, 1291-1298. https://doi.org/10.1109/ICSES63445.2024.10763340

Kodete, C. S., Pasupuleti, V., Thuraka, B., Gayathri, V. V., Sundar, V. S. D., & Shariff, V. (2024). Machine Learning for Enabling Strategic Insights to Future-Proof E-Commerce. 2024 5th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 931-936. https://doi.org/10.1109/ICOSEC61587.2024.10722255

Kodete, C. S., Pasupuleti, V., Thuraka, B., Sangaraju, V. V., Tirumanadham, N. S. K. M. K., & Shariff, V. (2024). Robust Heart Disease Prediction: A Hybrid Approach to Feature Selection and Model Building. 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS), Gobichettipalayam, India, 243-250. https://doi.org/10.1109/ICUIS64676.2024.10866501

Kodete, C. S., Saradhi, D. V., Suri, V. K., Varma, P. B. S., Tirumanadham, N. S. K. M. K., & Shariff, V. (2024). Boosting Lung Cancer Prediction Accuracy Through Advanced Data Processing and Machine Learning Models. 2024 4th International Conference on Sustainable Expert Systems (ICSES), Kaski, Nepal, 1107-1114. https://doi.org/10.1109/ICSES63445.2024.10763338

Kong, Z., Zhang, W., & Han, J. (2020). Social Media Fraud Detection: A Survey on Techniques and Solutions. Journal of Computer Science and Technology, 35(5), 983-1005.

Kumar, A., & Srinivasan, M. (2019). Real-time Fraud Detection on social media using Supervised Learning. In Lecture Notes in Computer Science (pp. 913-926). Springer.

Kumar, C. S., & Swamy, S. R. (2021). An Adaptive Deep Learning Model to Forecast Crimes. In Proceedings of Integrated Intelligence Enable Networks and Computing: IIENC 2020. Springer Singapore. https://doi.org/10.1007/978-981-33-6307-6_47

Mohan, V. M., Sharma, S. K., & Singh, P. (2010). Mass transfer correlation development for the presence of entry region coil as swirl promoter in tube. International Journal of Thermal Sciences, 49(2), 356-364. https://doi.org/10.1016/j.ijthermalsci.2009.06.010

N. S. Koti Mani Kumar Tirumanadham, S. K., Ramana, K., Swamy, S. S., Praveen, S. P., Sundar, V. S. D., & Shariff, V. (2025). Boosting Student Performance Prediction In E-Learning: A Hybrid Feature Selection And Multi-Tier Ensemble Modelling Framework With Federated Learning. Journal of Theoretical and Applied Information Technology, 103(5). https://www.jatit.org/volumes/Vol103No5/31Vol103No5.pdf

Nagasri, D., Sirisati, R. S., Amareswari, P., Bhushan, P. V., & Raza, M. A. (2024). Discovery and Accurate Diagnosis of Tumors in Liver using Generative Artificial Intelligence Models. Journal of Next Generation Technology (ISSN: 2583-021X), 4(2). https://www.researchgate.net/publication/381613787_Discovery_and_Accurate_Diagnosis_of_Tumors_in_Liver_using_Generative_Artificial_Intelligence_Models

Pasupuleti, V., Thuraka, B., Kodete, C. S., Priyadarshini, V., Kumar Tirumanadham, K. M., & Shariff, V. (2024). Enhancing Predictive Accuracy in Cardiovascular Disease Diagnosis: A Hybrid Approach Using RFAP Feature Selection and Random Forest Modeling. 2024 4th International Conference on Soft Computing for Security Applications (ICSCSA), Salem, India, 42-49. https://doi.org/10.1109/ICSCSA64454.2024.00014

Praveen, S. P., Jyothi, V. E., Anuradha, C., VenuGopal, K., Shariff, V., & Sindhura, S. (2022). Chronic kidney disease prediction using ML-Based Neuro-Fuzzy model. International Journal of Image and Graphics. https://doi.org/10.1142/s0219467823400132

Rajkumar, K. V., Nithya, K. S., Narasimha, C. T. S., Shariff, V., Manasa, V. J., & Tirumanadham, N. S. K. M. K. (2024). Scalable Web Data Extraction for Xtree Analysis: Algorithms and Performance Evaluation. 2024 Second International Conference on Inventive Computing and Informatics (ICICI), Bangalore, India, 447-455. https://doi.org/10.1109/ICICI62254.2024.00079

S. Phani Praveen, V. S. D. Sundar, K. V. S. S. S. R. Rao, C. Paritala, V. Shariff and J. V. N. Ramesh. (2025). AI- Powered Diagnosis: Revolutionizing Healthcare With Neural

Networks. Journal of Theoretical and Applied Information Technology, 101(3). https://www.jatit.org/volumes/Vol103No3/16Vol103No3.pdf

S., S., Kodete, C. S., Velidi, S., Bhyrapuneni, S., Satukumati, S. B., & Shariff, V. (2024). Revolutionizing Healthcare: A Comprehensive Framework for Personalized IoT and Cloud Computing-Driven Healthcare Services with Smart Biometric Identity Management. Journal of Intelligent Systems and Internet of Things, 13(1), 31–45. https://doi.org/10.54216/jisiot.130103

S., S., Raju, K. B., Praveen, S. P., Ramesh, J. V. N., Shariff, V., & Tirumanadham, N. S. K. M. K. (2025b). Optimizing Diabetes Diagnosis: HFM with Tree-Structured Parzen Estimator for Enhanced Predictive Performance and Interpretability. Fusion Practice and Applications, 19(1), 57–74. https://doi.org/10.54216/fpa.190106

Shariff, V., Aluri, Y. K., & Reddy, C. V. R. (2019b). New distributed routing algorithm in wireless network models. Journal of Physics Conference Series, 1228(1), 012027. https://doi.org/10.1088/1742-6596/1228/1/012027

Shariff, V., Paritala, C., & Ankala, K. M. (2025). Optimizing non small cell lung cancer detection with convolutional neural networks and differential augmentation. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-98731-4

Sirisati, R. S., Kalyani, A., Rupa, V., Venuthurumilli, P., & Raza, M. A. (2024). Recognition of Counterfeit Profiles on Communal Media using Machine Learning Artificial Neural Networks & Support Vector Machine Algorithms. Journal of Next Generation Technology (ISSN: 2583-021X), 4(2). https://www.researchgate.net/publication/381613824_Recognition_of_Counterfeit_Profiles_on_Communal_Media_using_Machine_Learning_Artificial_Neural_Networks_Support_Vector_Machine_Algorithms

Sirisati, R. S., Kumar, C. S., Divya, V. S., Rao, K. G., & Rajagopal, S. (2024). A Deep Learning Framework for Recognition and Classification of Diabetic Retinopathy Severity. Telematique, 23(01), 228-238. https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2025.1551315/abstract

Sirisati, R. S., Kumar, C. S., Latha, A. G., Kumar, B. N., & Rao, K. S. (2021). An Enhanced Multi Layer Neural Network to Detect Early Cardiac Arrests. In 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1514-1518). IEEE. https://ieeexplore.ieee.org/document/9531126/

Sirisati, R. S., Kumar, C. S., Latha, A. G., Kumar, B. N., & Rao, K. S. (2021). Identification of Mucormycosis in post Covid-19 case using Deep CNN. Turkish Journal of Computer and Mathematics Education, 12(9), 3441-3450. https://turcomat.org/index.php/turkbilmat/article/download/11302/8362/20087

Sirisati, R. S., Kumar, C. S., Reddy, P. S., Rao, K. G., & Reddy, B. S. (2024). Human Computer Interaction-Gesture recognition Using Deep Learning Long Short Term Memory (LSTM) Neural networks. Journal of Next Generation Technology (ISSN: 2583-021X), 4(2). http://www.jnxtgentech.com/mail/documents/vol%204%20issues%202%20article2.pdf

Sirisati, R. S., Kumar, C. S., Venuthurumilli, P., Ranjith, J., & Rao, K. S. (2023). Cancer Sight: Illuminating the Hidden-Advancing Breast Cancer Detection with Machine Learning-Based Image Processing Techniques. In 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA) (pp. 1618-1625). IEEE. https://doi.org/10.1109/ICSCNA58489.2023.10370462

Sirisati, R. S., Kumar, S. C., & Latha, G. A. (2021). An efficient skin cancer prognosis strategy using deep learning techniques. Indian Journal of Computer Science and Engineering (IJCSE), 12(1). https://www.ijcse.com/docs/INDJCSE21-12-01-180.pdf

Sirisati, R. S., Prasanthi, K. G., & Latha, A. G. (2021). An aviation delay prediction and recommendation system using machine learning techniques. In Proceedings of Integrated Intelligence Enable Networks and Computing: IIENC 2020 (pp. 239-253). Springer Singapore. https://www.researchgate.net/publication/370995631_Flight_Delay_Prediction_System_in_Machine_Learning_using_Support_Vector_Machine_Algorithm/fulltext/646e430a37d6625c002e31c1/Flight-Delay-Prediction-System-in-Machine-Learning-using-Support-Vector-Machine-Algorithm.pdf

Swamy, S. R., & Mandapati, S. (2018). A RULE SELECTED FUZZY ENERGY & SECURITY AWARE SCHEDULING IN CLOUD. Journal of Theoretical & Applied Information Technology, 96(10).

Swamy, S. R., Kumar, C. S., Latha, G. A., Devi, M. S., & Reddy, P. J. M. (2023). Multi-Features Disease Analysis Based Smart Diagnosis for COVID-19. Computers, Systems & Science Engineering, 45(1), 869-886. https://doi.org/10.32604/csse.2023.029822

Swamy, S. R., Rao, P. S., Raju, J. V. N., & Nagavamsi, M. (2019). Dimensionality reduction using machine learning and big data technologies. Int. J. Innov. Technol. Explor. Eng.(IJITEE), 9(2), 1740-1745. http://doi.org/10.35940/ijitee.B7580.129219

Swaroop, C. R., Reddy, V. S., Rao, K. G., Rajalakshmi, V., & Swamy, S. R. (2024). Optimizing diabetes prediction through Intelligent feature selection: a comparative analysis of Grey Wolf Optimization with AdaBoost and Ant Colony Optimization with XGBoost. Algorithms in Advanced Artificial Intelligence: ICAAAI-2023, 8(311). https://doi.org/10.1201/9781003529231-47

Thatha, V. N., Chalichalamala, S., Pamula, U., Krishna, D. P., Chinthakunta, M., Mantena, S. V., Vahiduddin, S., & Vatambeti, R. (2025b). Optimized machine learning mechanism for big data healthcare system to predict disease risk factor. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-98721-6

Thuraka, B., Pasupuleti, V., Kodete, C. S., Naidu, U. G., Tirumanadham, N. S. K. M. K., & Shariff, V. (2024). Enhancing Predictive Model Performance through Comprehensive Pre-processing and Hybrid Feature Selection: A Study using SVM. 2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 163-170. https://doi.org/10.1109/ICSSAS64001.2024.10760982

Tirumanadham, N. S. K. M. K., Priyadarshini, V., Praveen, S. P., Thati, B., Srinivasu, P. N., & Shariff, V. (2025d). Optimizing Lung Cancer Prediction Models: A hybrid methodology using GWO and Random Forest. In Studies in computational intelligence (pp. 59–77). https://doi.org/10.1007/978-3-031-82516-3_3

V. Narasimha, R. R. T., Kadiyala, R., Paritala, C., Shariff, V., & Rakesh, V. (2024). Assessing the Resilience of Machine Learning Models in Predicting Long-Term Breast Cancer Recurrence Results. 2024 8th International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 416-422. https://doi.org/10.1109/ICISC62624.2024.00077

Vahiduddin, S., Chiranjeevi, P., & Krishna Mohan, A. (2023). An Analysis on Advances In Lung Cancer Diagnosis With Medical Imaging And Deep Learning Techniques: Challenges And Opportunities. Journal of Theoretical and Applied Information Technology, 101(17).

Veerapaneni, E. J., Babu, M. G., Sravanthi, P., Geetha, P. S., Shariff, V., & Donepudi, S. (2024). Harnessing Fusion's potential: a State-of-the-Art information security architecture to create a big data analytics model. In Lecture notes in networks and systems (pp. 545–554). https://doi.org/10.1007/978-981-97-6106-7_34

Yang, Y., & Xie, Y. (2019). A Survey on Social Media Account Fraud Detection: Approaches and Challenges. International Journal of Computer Science and Network Security.

Yarra, K., Vijetha, S. L., Rudra, V., Balunaik, B., Ramesh, J. V. N., & Shariff, V. (2024). A Dual-Dataset Study on Deep Learning-Based Tropical Fruit Classification. 2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 667-673. https://doi.org/10.1109/ICECA63461.2024.10800915