

Enhanced K-Means Clustering Implementation for Web-Based Suspicious Profiles Detection

R. PRASANTH REDDY¹, NAGAVELLI YOGENDER NATH², GATTU RAMYA³, SYED ABDUL HAQ⁴

¹Assistant Professor, Department of Computer Science & Engineering, RSR Engineering College

²Assistant Professor, Department of Computer Science & Engineering (AI&ML), Sumathi Reddy Institute of Technology for Women, Hyderabad.

³Assistant Professor, Department of Computer Science & Engineering, Vignan Institute of Management and Technology for Women, Hyderabad.

⁴Assistant Professor, Department of Computer Science & Engineering, Malla Reddy Engineering College, Hyderabad.

Abstract- Fraud detection often requires a hybrid approach combining both human expertise and AI techniques. While AI models can process large amounts of data and detect intricate patterns, human intervention is sometimes necessary to refine rules, especially when dealing with ambiguous or doubtful cases. This is particularly important when distinguishing between fraudulent and legitimate behavior, as subtle differences can exist. Another challenge is the multilingual nature of the web. Fraud can occur across different languages and cultural contexts, making it difficult to build comprehensive fraud detection models that account for all variations. In cases where linguistic resources (such as parsers or language models) are not readily available, ML tools can be adapted to learn from the data itself. However, for such models to function effectively, it is crucial to develop a rich set of features, ensuring that the data is representative and inclusive of the varied aspects of web-based fraud. This paper compares the performance of various clustering algorithms and develops an enhanced k-means algorithm utilizing the Minkowski metric. This modified algorithm, applied to an unlabeled dataset from an online dating site, effectively clusters users into authentic and suspicious categories. Supervised machine learning techniques are subsequently employed to validate the proposed model using another labeled online dating fraud dataset from the US. This methodology underscores the potential of combining traditional clustering methods with machine learning techniques to enhance anomaly detection across various domains, offering a practical tool for administrators and security experts.

Index Terms—Deep Learning, Approaches Web-based Fraud, Fraud Detection, k-means algorithm

I. INTRODUCTION

In the last few years, social media sites like Facebook, Instagram, and X (Formally Twitter) have progressively become important channels for information dissemination. People of all ages spend most of their time on these social networking sites and a huge amount of data is being created and shared globally. As reported in [1], about 4 billion people use social media sites to stay in touch with friends, family, and co-workers. However, these expanding interests have also given rise to fake users who take advantage of genuine users by scamming them. Even though online social networks present various security risks, fake profiles are especially dangerous when it comes to dating websites. These deliberately made fake profiles have certain goals in mind; they could be used to spread gossip, steal, or obtain personal data [2]. The most prevalent ways fake accounts obtain information are through amicable online interactions or by abusing the online data made available on social media platforms. Therefore, it's crucial to recognize and remove such kinds of profiles to preserve the confidence of online dating platforms and users [3]. Traditional methods for detecting fraudulent profiles rely on supervised Machine Learning (ML) techniques that use labeled data for training purposes. Such methods detect fake and genuine profiles on online dating sites, focusing on account- and content-based features, and semantic analysis of text or a combination of these approaches [4-5]. Conversely, handling unlabeled data presents a significant

challenge in fake profile detection, particularly on online dating platforms where the lack of ground truth labels hampers supervised techniques. Labeling these accounts manually is impractical, making it imperative to explore unsupervised ML methods that can recognize fake profiles without prior labeling.

Unsupervised learning, particularly clustering algorithms such as k-means, DBSCAN, etc., offers a promising alternative by grouping similar data points based on inherent patterns within the data. However, standard k-means has limitations in handling high-dimensional data and often requires enhancements for improved performance.

This study proposed an enhanced k-means clustering approach coupled with the Minkowski distance function to address the above-mentioned limitations, enabling the effective identification of suspicious profiles on online dating platforms. It uses a publically available OkCupid-based unlabelled dataset, obtained from Kaggle, which includes a comprehensive range of user attributes. Effective preprocessing and feature engineering are performed with enhanced k-means clustering to identify suspicious users and underlying patterns indicative of fraudulent behavior. Further, various evaluation metrics are employed, including the silhouette score, Davies-Bouldin, and Calinski-Harabasz indexes for assessing the performance of obtained clusters.

Extensive experimentation revealed that the flexibility and efficiency of enhanced k-means with the Minkowski metric made it perform better than other clustering techniques. The next step in the proposed approach involves labeling the obtained clusters using a combination of statistical variance, cluster attributes, and content elements. The aforementioned procedures have resulted in the labeling of the OkCupid-based4 into two clusters (suspicious and authentic users), paving the way for the application of supervised training and testing. Figure 1 represents the proposed cluster then label approach applied in this work.

The main contribution of this study lies in its hybrid methodology, which integrates unsupervised clustering for initial grouping, followed by labeling the obtained clusters, and finally employing supervised learning for validation.

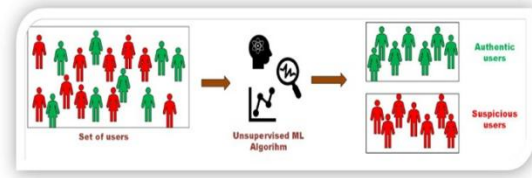


Figure 1: Clustering using unsupervised ML algorithms

Additionally, to validate the effectiveness of our proposed hybrid approach, we utilize a labeled dataset from a US-based online dating platform. This validation step is crucial in demonstrating the generalizability and reliability of our approach across different datasets and OSNs platforms.

The major highlights of this paper are as follows:

- Performance comparison of various clustering algorithms based on the Silhouette score, Calinski-Harabasz index, and Davies-Bouldin index.
- Development of an enhanced k-means algorithm that makes use of the Minkowski metric and exhibits superior performance.
- Effectively clustering users' profiles as authentic or suspicious using an unlabeled online dating profile dataset.
- Validating the proposed clustering model using supervised machine learning techniques.
- An additional validation phase is performed to authenticate clustering performance using US-based labeled online dating fraud datasets.

II. LITERATURE WORK

Detecting fake or suspicious profiles on online dating platforms is critical for ensuring user safety and trust. This section provides a comprehensive overview of the existing research, methodologies, and findings related to fake profile detection on OSNs. Initial research efforts primarily rely on supervised learning models to handle tasks such as classifying phishing emails, identifying spam accounts based on user peer acceptability, and detecting fake profiles and email spam [6]. These models, when trained on labeled datasets, have proven effective in identifying fraudulent activities. Researchers focused on feature engineering to enhance classifiers' performance. In 2015, Xiao and his team developed a supervised learning approach to detect fake accounts in OSNs by

classifying clusters of accounts as either malicious or legitimate based on user-generated text fields. Their model achieved an AUC of 98% using the RF algorithm, successfully identifying over 250,000 fake accounts on LinkedIn. [7] addressed online romance scams, a widespread form of mass-marketing frauds, by detecting fake profiles on dating platforms. Their study examined demographics, profile descriptions, and images from both genuine and scam profiles for text categorization and image analysis. Their ensemble model, employing an SVM classifier, detected scam profiles with 97% accuracy. [8] proposed a supervised ML approach to detect fraudulent Instagram accounts, using decision tree-based bagging classifiers exploiting user- and content based features. In their experimentation, the RF algorithm achieved the highest accuracy of 98%.

In another work, [9] further advanced phishing detection by proposing a model that integrates hierarchical clustering techniques with six classification algorithms: RF, SVM, NB, NN, DT, and LR. Among these, LR achieved the highest accuracy of 99.8%. Additionally, [10] developed an effective method for detecting spam accounts on Twitter by combining clustering and classification techniques. They identified similarities among spam accounts using principal component analysis and a refined K-means algorithm for clustering. Their study achieved high accuracy in identifying spam accounts using the RF classifier, utilizing both labeled and unlabeled data for training and analysis.

III. CLUSTERING METHODOLOGY

It has been observed that k-means clustering is preferred over other techniques for clustering online users' profiles into suspicious and authentic primarily due to its efficiency with large datasets and clear separation between clusters. Additionally, k-means' iterative centroid approach is well-suited for numerical data, making it ideal for the dynamic and rapidly changing environment of online dating platforms. Thus, k-means is a computationally efficient algorithm suitable for large datasets, requiring simple steps to assign points to the nearest centroid and recalculate them. Furthermore, k-means provides hard clustering, ensuring each point belongs to exactly one cluster for easy interpretation [11-13].

Various research revealed that integrating k-means with the Minkowski distance metric can significantly enhance clustering performance [14]. The Minkowski distance provides a generalized and customizable distance metric, allowing better adaptation to various data shapes and distributions. Empirical studies show that k-means with Minkowski distance can outperform standard k-means and other clustering algorithms in various scenarios [15]. This flexibility leads to improved cluster separation, robustness to noise and outliers, and overall higher accuracy.

IV. PROPOSED MODEL FOR CLUSTERING

This part proposes a clustering-based model designed to efficiently transition from a raw, unlabeled dataset to refined labeled clusters. The following subsections detail each of the six phases of this model, presenting a cohesive approach that spans from initial data clustering to the final labeling of clusters. Figure 2 illustrates the phases of the proposed model including data collection, data preprocessing, feature engineering, model building using unsupervised ML, experimentation and result evaluation, and cluster labeling [16].

Algorithm 1: Enhanced k-means

Input: Features Set (Ff) extracted from dataset D

Output: Suspicious and Authentic Clusters

Start procedure

- 1 $I_x \leftarrow \text{InputFeature}(Ff)$; // The algorithm starts by taking the Extracted Features (Ff) as input.
- 2 Define $\text{MinkowskiDistance}()$; // The Minkowski distance function is defined to measure the distance between data points and centroids.
- 3 for each data points i in Ff
- 4 Calculate the $\text{MinkowskiDistance}()$ from the data point i to each centroid.
- 5 Assign the data point i to a cluster C_{i1} based on the computed distance.
- 6 for each feature j within the data point i
- 7 Update the cluster assignment C_{i2} for the feature j .
- 8 end for
- 9 Check for convergence. // The centroids are adjusted based on convergence criteria.
- 10 $\text{DenseLayer}(Ff, \text{sigmoid_activation})$; // Once the clusters have converged, a dense layer with sigmoid activation is applied to the features Ff , resulting in whether the clusters are suspicious or authentic.

11 end for
 12 return suspicious and authentic cluster
 End procedure

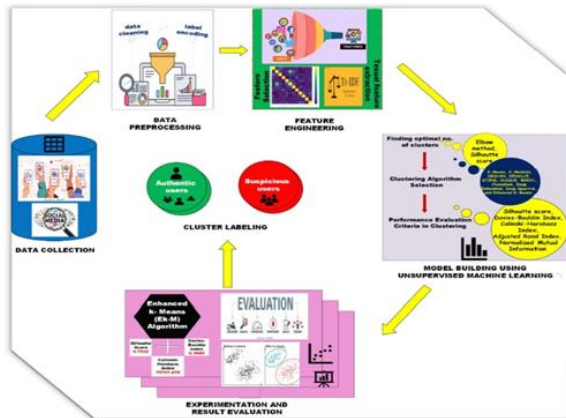


Figure 2: Phases of the proposed model for clustering [17]

Dataset Collection

The dataset OkCupid profiles employed in this research has been sourced from Kaggle and was made publicly available by [18]. This unlabeled dataset contains information on 59,946 users of the dating site OkCupid, specifically from San Francisco. It includes users' demographic details, such as age, gender, and location, as well as their responses to various questions about interests, hobbies, and personal beliefs. Additionally, the dataset provides information on users' dating preferences, such as their preferred age range and gender for potential matches. Researchers across fields, including ML and social science, have widely utilized this dataset to explore patterns and trends in online dating behavior [19]

Data Preprocessing

Data preprocessing is an essential step in unsupervised learning tasks, which involves transforming raw data into a suitable format. Data preprocessing involves cleaning the data, i.e., removing missing, duplicated, or corrupted data [20]. The next step is data transformation, which consists of transforming the categorical data into numerical data using the label encoding technique. The main objective is to ensure the data is consistent and prepared for building accurate models. Label encoding technique is applied to ensure the data is consistent and prepared for building accurate learning models.

Feature Engineering

Feature engineering is the process of creating new features or extracting meaningful information from existing ones in a dataset, and involves combining features, generating new features, scaling, and selecting the most relevant ones. This phase involves various steps:

Derived Feature Creation: This work uses the original feature age to create a new feature called agegroup, dividing the dataset into five groups: teen, young adult, mid-age, over-mid-age, and old, each labeled from 0 to 4.

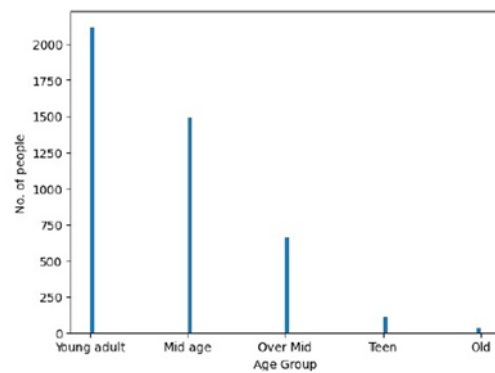


Figure 3: Deriving age group feature from the original age feature of users

Feature Selection

Feature selection is crucial in unsupervised learning, enhancing model performance, reducing training time, and increasing interpretability. It involves identifying and selecting a subset of original features in a dataset that is most relevant to the problem, reducing the dimensionality of the dataset. Filter methods such as Pearson correlation, assess features based on their statistical properties [21]. It is a common filter method used to identify the most relevant features by evaluating the linear relationship between each feature and the target variable. During the feature selection phase, certain features such as location, body_type, diet, drinks, drugs, last_online, pets, sign, and smokes were removed as they were deemed redundant and unlikely to affect clustering performance. In contrast, features such as status, sex, orientation, education, ethnicity, height, income, job, offspring, religion, speaks, and age group were retained for further experimentation. Pearson correlation ensures that selected features are not linearly dependent on each

other, allowing for a better focus on the most relevant ones, and leading to improved results, as illustrated in the heat map of Figure 4. Among the text-based features, only essay 0 was retained from essays 0 to 9. The essay 0 includes user description, which is potentially useful in determining whether a user is suspicious or authentic. The other essays were excluded due to a higher number of null values and their focus on aspects of text not relevant to this work [22].

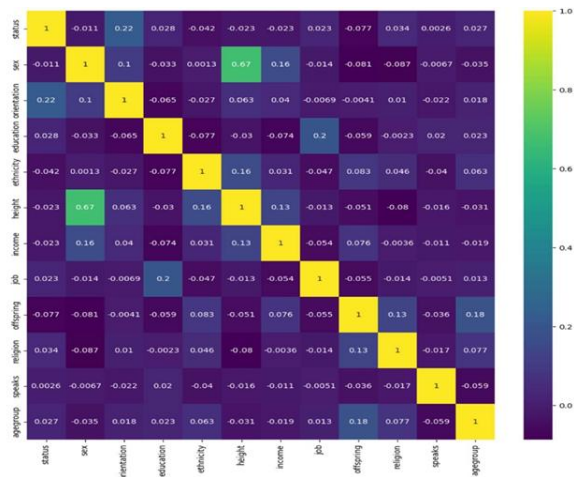


Figure 4: Heat map of the selected features

Clustering Algorithm Selection

The selection of an appropriate clustering algorithm is crucial for effectively creating meaningful clusters of user profiles on online dating platforms. This research employs a diverse set of clustering algorithms, including k-means, K-Medoids, DBSCAN, DENCLUE, STING, CLIQUE, BIRCH, Chameleon, Deep Embedded, Deep Spectral, and proposed enhanced k-means to evaluate their efficacy in forming distinct clusters that can later be analyzed to identify patterns indicative of genuine or suspicious profiles [23].

V. CLUSTERING PERFORMANCE VALIDATION USING SUPERVISED ML

As previously stated, this work employs a hybrid technique in which unlabeled users' profile data is first clustered using an enhanced k-means algorithm in two clusters and labeled them as authentic or suspicious

[24]. Second, to test the effectiveness of our clustering approach, we used supervised classification algorithms on the same dataset, appending retrieved cluster information as a label attribute in the users' profile dataset. We have a strong assumption that if the enhanced k-means clustering correctly distinguishes between legitimate and suspicious users' profiles, the supervised classifiers will perform well on this dataset in terms of classification performance metrics such as accuracy, precision, recall, and F1 score. In other words, high supervised classification performance suggests that the clustering features successfully divide the two classes, resulting in correct label validation. Thus, clustering performance validation is proceeded further by applying supervised learning using a variety of ML classifiers, including LR, SVM, XGBoost, and LightGBM, and the user profile dataset is divided into 70:30 training and testing sets. The selected classifiers are trained on the training set, and evaluated on the testing set using various performance metrics [25]. The experimental outcomes of this step have been shown using Table 1, in which supervised classifiers demonstrate excellent performance in distinguishing between genuine and suspicious profiles obtained using the proposed enhanced k-means clustering approach.

Table 1: Clustering performance validation using supervised classifiers

Classifiers	LR	SVM	XGB	LGBM
TP	1115	1113	1145	1146
TN	1618	1629	1643	1645
FP	34	38	3	2
FN	28	19	4	3
Accuracy	0.99	0.98	0.99	0.99
Precision	0.9774	0.9808	0.9974	0.9986
Recall	0.9769	0.9785	0.9976	0.9986
F1 score	0.9770	0.9797	0.9976	0.9985
CV mean accuracy	0.9786	0.9820	0.9998	0.9989
Standard Deviation	0.0044	0.0042	0.0005	0.0012

During the experiment, classifiers did not show signs of overfitting and achieved high accuracies in both training and testing sets. The cross-validation mean accuracy (with CV=10) values further support these

findings. All classifiers display minimal variance, as indicated by their low standard deviation values. Thus, consistency in performance metrics, coupled with high True Positive (TP) and True Negative (TN) rates and minimal False Positives (FP) and False Negatives (FN), underscores the reliability and effectiveness of our proposed enhanced k-means clustering approach.

Model Building using Proposed Enhanced K-means

To obtain clusters from the aforementioned labeled dataset, the label field was first masked such that the data was treated as unlabeled. Second, the Elbow and Silhouette methods were used to find the optimal number of clusters, i.e. 2.

To extract clusters from the preprocessed dataset, the enhanced k-means clustering technique was applied with masked label attribute. In line with [26-29] internal validation criteria, specifically the Silhouette score and Davies-Bouldin index, were used to assess the clustering results. A high Silhouette score of 0.7089 and a low Davies-Bouldin value of 0.3955, demonstrate the correct identification of two clusters and prove the effectiveness of the proposed enhanced k-means algorithm on the labeled dataset.

The surge in online dating platform users has increased suspicious profiles, posing serious risks to both user safety and platform trust. This research addressed the challenge of identifying suspicious profiles in large and unlabeled datasets, which are common in real-world scenarios. By leveraging an enhanced k-means clustering approach integrated with the Minkowski distance metric, and further combined with supervised machine learning techniques for validation, this work demonstrated a hybrid strategy for classifying profiles as either authentic or suspicious. Extensive data preprocessing, feature selection, and textual feature extraction methods are applied to optimize the clustering process. The final feature set contained 812 features. Further, the Elbow and Silhouette score methods were employed to determine two clusters. Internal cluster performance metrics, including the Silhouette score, Davies-Bouldin, and Calinski-Harabasz indices, were used to assess the quality of the clustering, ensuring a clear distinction between genuine and suspicious profiles. Utilizing the Minkowski distance metric for clustering, our analysis uncovered distinct behavioral

patterns associated with suspicious profiles, reinforcing the efficacy of the proposed enhanced k-means approach. Our hybrid approach proceeds further by labeling the obtained clusters and utilizing supervised ML classifiers to validate the clusters' performance. Additionally, experimental results were validated using a labeled dataset of genuine and fake profiles sourced from two US-based online dating sites. The findings from this study have important implications for the development of a reliable fraud detection system, not only in the context of online dating platforms but also in broader domains such as ecommerce, cybersecurity, etc. By combining unsupervised learning for initial grouping and supervised learning for validation, this hybrid approach offers a comprehensive and adaptable framework for anomaly detection in various fields.

VI. CONCLUSION

The research introduces an enhanced k-means clustering methodology to improve the detection of suspicious users in unlabeled datasets. The main challenge in clustering analysis is determining the number of clusters and labeling them effectively. The study integrates feature selection techniques, linguistic and functional aspects of user profiles, and statistical and performance-based metrics to optimize the clustering process. The enhanced k-means algorithm integrated with Minkowski distance, demonstrated superior performance in distinguishing between authentic and suspicious profiles. Further, the methodology incorporates supervised techniques to validate the approach. This study emphasizes the practical implications of improved user profiling in applications like social networking, ecommerce, and cybersecurity. The proposed methodology offers an efficient solution that can be adapted to different datasets and use cases. It contributes to user profiling by providing a systematic solution for handling unlabeled datasets. Future work could explore the application of the proposed approach to other fraud detection tasks, and integrate ML advances to enhance detection accuracy.

REFERENCES

- [1] Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey.

- Journal of Network and Computer Applications, 68, 90-113.
- [2] Adewole, K. S., Han, T., Wu, W., Song, H., & Sangaiah, A. K. (2020). Twitter spam account detection based on clustering and classification methods. *The Journal of Supercomputing*, 76, 4802-4837.
- [3] Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60, 19-31.
- [4] Al Sarah, N., Rifat, F. Y., Hossain, M. S., & Narman, H. S. (2021). An efficient android malware prediction using Ensemble machine learning algorithms. *Procedia Computer Science*, 191, 184-191.
- [5] Alom, Z., Carminati, B., & Ferrari, E. (2018). Detecting spam accounts on Twitter. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 1191-1198.
- [6] Ayoobi, N., Shahriar, S., & Mukherjee, A. (2023). The looming threat of fake and llm-generated LinkedIn profiles: Challenges and opportunities for detection and prevention. In *Proceedings of the 34th ACM Conference on Hypertext and Social Media*, 1-10.
- [7] Bauder, R., da Rosa, R., & Khoshgoftaar, T. (2018). Identifying medicare provider fraud with unsupervised machine learning. In 2018 IEEE international conference on Information Reuse and Integration (IRI), IEEE, 18, 285-292.
- [8] Bharne, S., & Bhaladhare, P. (2022). Investigating online dating fraud: An extensive review and analysis. In 2022 International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC), 141-147.
- [9] Billah, M. M., Bhuiyan, M. N., & Akterujjaman, M. (2021). Unsupervised method of clustering and labeling of the online product based on reviews. *International Journal of Modeling, Simulation, and Scientific Computing*, 12(02), 2150017.
- [10] Chiu, C. C., & Tsai, C. Y. (2004). A web services-based collaborative scheme for credit card fraud detection. In *IEEE International Conference on e-Technology, e-Commerce and e-Service*, 2004. IEEE'04. 2004, 177-181.
- [11] Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2015). Fame for sale: Efficient detection of fake Twitter followers. *Decision Support Systems*, 80, 56-71.
- [12] Mr DD Sarpate, "Design of Dual Band Microstrip for satellite Applications", 2nd International Conference on Recent Innovations in Engineering & Technology 2020
- [13] Debener, J., Heinke, V., & Kriebel, J. (2023). Detecting insurance fraud using supervised and unsupervised machine learning. *Journal of Risk and Insurance*, 90(3), 743-768.
- [14] Erşahin, B., Aktaş, Ö., Kılınc, D., & Akyol, C. (2017). Twitter fake account detection. In 2017 international conference on computer science and engineering (UBMK), 388-392.
- [15] Usman, M., Zubair, M., Hussein, H. S., Wajid, M., Farrag, M., Ali, S. J., ... & Habeeb, M. S. (2021). Empirical mode decomposition for analysis and filtering of speech signals. *IEEE Canadian Journal of Electrical and Computer Engineering*, 44(3), 343-349.
- [16] Singh, A., Gupta, M., Raj, A., Gupta, S. K., & Habeeb, M. S. (2020, December). TWDM-PON: The Enhanced PON for Triple Play Services. In 2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE) (pp. 1-5). IEEE.
- [17] Naveen Sai Bommina, Nandipati Sai Akash, Uppu Lokesh, Dr. Hussain Syed, Dr. Syed Umar, "A Hybrid Optimization Framework for Enhancing IoT Security via AI-based Anomaly Detection", *International Journal on Recent and Innovation Trends in Computing and Communication*, (2023) ISSN: 2321-8169 Volume: 11 Issue: 3.
- [18] Habeeb, M. S., & Babu, T. R. (2022). Network intrusion detection system: a survey on artificial intelligence-based techniques. *Expert Systems*, 39(9), e13066
- [19] Uppu Lokesh, Naveen Sai Bommina, Nandipati Sai Akash, Dr. Hussain Syed, Dr. Syed Umar. (2021). Deep Reinforcement Learning with Genetic Algorithm Tuning for Intrusion Detection in IoT Systems. *International Journal*

- of Communication Networks and Information Security (IJCNIS), 13(3), 582–595.
- [20] Uppu Lokesh, Naveen Sai Bommina, Nandipati Sai Akash, Dr. Hussain Syed, Dr. Syed Umar, "Designing Energy-Efficient and Secure IoT Architectures Using Evolutionary Optimization Algorithms", *International Journal of Applied Engineering & Technology*, Vol. 4 No.2, September, 2022.
- [21] Mr. Dikshendra Daulat Sarpate, and Dr. B.G Nagaraja, "CONVOLUTION NEURAL NETWORK-BASED SPEECH EMOTION RECOGNITION USING MFCCS", *International Journal of Communication Networks and Information Security*, 2023/12/10
- [22] Naveen Sai Bommina , Nandipati Sai Akash, Uppu Lokesh , Dr. Hussain Syed , Dr. Syed Umar, "Multi-Objective Genetic Algorithms for Secure Routing and Data Privacy in IoT Networks", *International Journal of Communication Networks and Information Security (IJCNIS)*, (2020), 12(3), 632–643.
- [23] Nandipati Sai Akash, Naveen Sai Bommina, Uppu Lokesh, Hussain Syed, Syed Umar, "Optimized Block Chain-Enabled Security Mechanism for IoT Using Ant Colony Optimization", *International Journal on Recent and Innovation Trends in Computing and Communication*, (2023), 11(10), 1226–1233.
- [24] Naveen Sai Bommina , Nandipati Sai Akash, Uppu Lokesh , Dr. Hussain Syed , Dr. Syed Umar, "Privacy-Preserving Federated Learning for IoT Devices with Secure Model Optimization", *International Journal of Communication Networks and Information Security (IJCNIS)*, (2021), 13(2), 396–405.
- [25] K Sankar, Divya Rohatgi, S Balakrishna Reddy, "COX Regressive Winsorized Correlated Convolutional Deep Belief Boltzmann Network for Covid-19 Prediction with Big Data", *Grenze International Journal of Engineering & Technology (GIJET)*, Grenze ID: 01.GIJET.9.1.547, © Grenze Scientific Society, 2023.
- [26] Naveen Sai Bommina, Uppu Lokesh, Nandipati Sai Akash, Dr. Hussain Syed, Dr. Syed Umar, "Optimized AI Models for Real-Time Cyberattack Detection in Smart Homes and Cities", *International Journal of Applied Engineering & Technology*, Vol. 4 No.1, June, 2022.
- [27] K. Kartheeban, K. Kalyani, S. K. Bommavaram, D. Rohatgi, M. N. Kathiravan, and S. Saravanan, "Intelligent Deep Residual Network based Brain Tumor Detection and Classification," in 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Dec. 2022, pp. 785–790. doi:10.1109/ICACRS55517.2022.10029146.
- [28] Umar, Syed, Bommina Naveen Sai, Nagineni Sai Lasya, Doppalapudi Asutosh, and LohithaRani. "Machine Learning based Sentiment Analysis of Product Reviews Using DeepEmbedding." *Journal of Optoelectronics Laser* 41, no. 6(2022): 108-113.
- [29] Divya Rohatgi, Dr. Tulika Pandey, "Regression Test Selection Framework for Web Services", *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 9, ISSUE 03, MARCH 2020*.