ONLINE RECRUITMENT FRAUD DETECTION USING DL APPROACH

MS.M.ANITHA[1], Mr.Y.NAGA MALLESWARAO [2], PUPPALA.AJAY[3]

[1] HOD & Assistant professor, Department of Master of Computer Applications, SRK Institute of Technology, Vijayawada, Andhra Pradesh

[2] Assistant professor, Department of Master of Computer Applications, SRK Institute of Technology, Vijayawada, Andhra Pradesh

[3] MCA Student, Department ofMaster of Computer Applications, SRK Institute of Technology, Vijayawada, Andhra Pradesh

ABSTRACT

Online recruiting platforms are becoming vital resources for both businesses and job seekers in today's technologically advanced work environment. But as these platforms have grown in popularity, there has also been a notable increase in online recruitment fraud, which involves using phone employment offers to take advantage of weaker people. A more secure and sophisticated detection system is desperately needed, as shown by the thousands of cybercrime instances involving bogus jobs that are reported each year in India alone. Conventional strategies for spotting fraudulent job advertisements mostly depend on user reporting systems, static rule-based filtering, and human verification—all of which are laborious, prone to mistakes, and unable to keep up with increasingly complex scam tactics. This project's primary goal is to use Deep Learning (DL) methods to identify and categories fraudulent job posts automatically with high accuracy and little assistance from humans. This study suggests a method for analyzing job descriptions, metadata, and other relevant fields from job advertisements using Convolutional Neural Networks (CNNs). Traditional machine learning models, including Decision Tree classifiers, were first taken into consideration; however, the need for deep learning was prompted by their limited capacity for generalization and their inability to handle high-dimensional data. By automatically learning hierarchical feature representations, the deep learning method improves prediction accuracy significantly. Through a web-based platform that uses Django combined with HTML and CSS for user interface design, the system encompasses data preparation, feature engineering, model training, assessment, and deployment. By offering a dependable, scalable, and intelligent solution that benefits companies, job seekers, and platform administrators equally, the suggested method seeks to lower the number of recruitment fraud incidents. This experiment highlights how using AI-based models might improve cybersecurity in the hiring industry in the future

KEYWORDS: Convolutional Neural Network, Deep Learning, HTML, CSS.

1. INTRODUCTION

1.1 Overview

The recruiting process has been revolutionized all across the globe, especially in India, since it has shifted from using conventional newspaper ads to internet portals. The rise of online recruiting has increased the speed, accessibility, and diversity of the applicant pool. The door has been pried open,

nonetheless, for deceitful pursuits. Approximately 40% of Indian job searchers fell victim to phoney employment offers between 2018 and 2023. Con artists use deceptive job postings, phoney employment websites, and illegal consultancies to trick unwary applicants. The online recruiting area is susceptible to fraud due to the increasing dependence on digital platforms that do not have rigorous verification processes. Using Deep Learning (DL) methods to identify fraudulent actions during online recruiting procedures is the main emphasis of this title's comprehensive analysis. Enhancing the safety and reliability of the recruiting environment, DL models have the ability to automatically learn patterns from big datasets and detect suspicious actions.

Greetings and Practical Uses:By automatically detecting scams, online recruiting fraud detection utilising deep learning guarantees secure hiring practices. This safeguard prevents both businesses and job seekers from becoming victims of frauds. Online portals and recruiting firms get more reputation as a result. It creates a digital workplace that is safer and more reliable for employees.

1.2 Problem Statement

Detecting recruiting fraud relied on human judgement and manual verification prior to the adoption of deep learning and machine learning. Background checks, phone interviews, and manual resume reviews were laborious, inaccurate, and time-consuming. False applicants were often chosen due to human mistake in the verification procedure. Fraud detection was ineffective due to the datasets' incapacity to manage enormous amounts of data. With the exponential growth of online job applications, scalability became a significant concern.

1.3 What Motivates Research?

The need to automate fraud detection and make it quick, accurate, and scalable is critical. Due to the increasing complexity and volume of online recruiting fraud, traditional manual solutions were unable to handle the problem. Without human interaction, deep learning algorithms may uncover hidden patterns and outliers. Company recruiting procedures may be better protected if fraud detection is made faster and more accurate. Constructing a trustworthy system to verify the legitimacy of online job offers is a compelling reason to do research.

1.4 Mission

Using deep learning methods, we want to create a reliable system that can identify online recruiting fraud automatically. Our goal is to make things easier for humans, identify fraud in real-time, and cut down on manual labour. Data points pertaining to recruiting, such as job descriptions, recruiter information, and application trends, will be analysed by the system. Based on patterns it has learnt, it will categorise and report any suspect activity. In the end, we want to make sure that both businesses and job seekers can feel comfortable while recruiting online.

1.5 Practical Uses

Recruiters may automate the detection of fake ads by integrating this technology into online job platforms.When used throughout the screening process, it may help HR departments distinguish between legitimate and fraudulent applications.This technique may be used by government employment portals to ensure that job adverts are genuine.Using this approach, staffing firms may compile an extensive list of qualified applicants and recruiters.Professional networking sites like Naukri and

LinkedIn may benefit from the model's ability to increase trust.During campus placements, educational institutions might utilise it to check the organisations that are providing positions.New grads and seasoned workers alike might be shielded from job scams with its help.By incorporating fraud detection into their employment procedures, financial organisations may proactively combat job-related financial

2. LITERATURE SURVEY

The digital revolution seems to have been adequate, in part, because of a number of key enabling technologies, including as AI, the internet of things (IoT), and big data. Since the widespread use of the internet, banks and other financial institutions began using mobile portals and other technology to increase their customer base and revenue [1]. Still, the meteoric surge in its use due to smartphones has undoubtedly amplified online theft involving debit cards.

From fraudulent use of debit cards to data recombination and card fraud, the increasing prevalence of credit and debit card cybercrime poses a serious threat on a global scale [two]. With more than 2.700 million cards in circulation globally, con artists now have the opportunity to capitalise on the aforementioned trend. It is estimated that the worldwide total failure due to complete fraudulent credit card would reach $43 billion by 2023–24. Fraudulent credit card transactions account for 46% of all illegal operations globally. According to the department of statistics, the country saw a significant increase in fraudulent activity in 2021–2022, with a forecast deficit of almost $2.2 trillion. These statistics highlight the critical necessity of enhancing fraud detection methods to mitigate the growing threat of online credit or debit card theft, especially in rural areas where these crimes are becoming more commonplace [3]. There was a significant increase in the incidence of yeah payment online fraud among numerous Australians in 2019 and 2020. The amount of money that must have been destroyed as a result of full certificate cybercrime is an estimated sort of two trillion dollars, according to a paper from the department of statistics (abs). From 2020–2021, the total proportion of affected areas in Australia experiencing payment online fraud increased by 6.6%, and from 2022–2023, it increased by 8.7%. This chart shows a change in the way that individuals and businesses in Australia are affected by internet fraud after 2019 (from 2018 to 2019). [4].

It is crucial to identify online fraud using credit or debit cards, since it requires ongoing development to deal with evolving deceitful tactics [5]. conventional methods have their limitations, but new approaches like deep convolutional neural network (dl) and machine learning (ml) offer the potential to improve accuracy, adaptability, and overall efficacy [6, 7]. Nonetheless, both deep learning and millilitres still encounter challenges such as scalability issues, imbalanced lecture distribution, and generalisation issues [8], as well as data-related challenges such as dynamic and changing thieving patterns [9]. Instead than focussing on developing a comprehensive framework for fraud prevention, the aforementioned paper examines a single combination of ml and dl.

The goal, which involves combining millilitres and deep learning techniques into a single prototype, arises from the shortcomings of traditional brands and the need to combine their strengths. This will enable the detection of fraudulent activities. Some examples of mg/l designs include: tree-based (dt), confusion-matrix (rf), support vector machine (SVM), intense backpropagation algorithm (xgboost), classification-boosting (catboost), and regression analysis (lr) that takes into account skewed data points and complex fraud techniques, despite my inability to identify incremental but rather temporal relationships throughout transactions. The following linkages are confronted by downlink kinds ranging from completely convolutional (cnn) to input and output hard short attention span infrastructure (bilstm): they uncover hidden patterns and relationships, which are then adjusted to complete numerous distinct thieving techniques [10]. Its model's concentration on particular features is going to boost communication facilities and accuracy with an inclusion of yes attn methods [11]. By integrating the

best features of both ML and DL, this universal solution provides a robust, deployable, and relatively accurate plan to combat the growing problem of bank card fraud.

3. PROPOSED SYSTEM

Stage 1: Preparing for Theft Data

Choosing and preparing a fraud prevention data collection is the first and most important step in this research. A more comprehensive dataset, such new trustworthy computer vision kinds, seems to be crucial. The data frame utilized in this study includes several aspects such as profile pages, narratives about payments, behavioral qualities, and potential file or resume contributors, depending on the particular purpose of the investigation into the effects of fraud on children. Each document in the data collection seems to have a classification that indicates whether it is legitimate or not. it's crucial to know this similar data -set delivery early since forgeries sets of data are often incredibly skewed, with significantly fewer instances of fraud compared to honest ones. The previous step guarantees that subsequent kinds will be able to acquire its complex forms, which distinguish between unlawful and lawful actions.

2. Data frame pre-processing (term encoding, null-worth removal)

After successfully acquiring this data set, the following step is to preprocess the images with a focus on detail in order to improve the statistical standard and value. At first glance, the dataframe seems to have been checked for missing or default values, which might have an adverse effect on the model quality if left unchecked. depending on the degree of functionality (mean, mean average, phase, and forecasted imputation), zeros are either deleted or distributed using an effective technique. Sticker encrypting has been used to move data from several current sets to new statistical genres ever after incomplete data was washed. These sets include users, payment subgroups, and Facebook postings. Attempting to tailor humans to ML algorithms, the aforementioned gene encodes morphs or pas categories based on computer values gathered. Data pre-processing is a crucial part of the evolution process because it guarantees that the data collection is organised, comprehensive, and ready for tasks like accurate model construction.

stage 3: building the framework itself integrated into this data analysis (logistic stagnation, light, and gradient descent classifier)

This research not only reviews prior computer vision kinds to establish a single effectiveness baseline, but it also continues to develop after the information time to prepare has passed. regression analysis, light (light xgboost machine), and stochastic gradient (stochastic slope descent) classification are the three models that have been considered. It seems that the clarity and communication facilities along binary class work activities led to the selection of regression analysis, specifically one vintage polynomial regression. LightGBT is a powerful augmentation technique that has been hand-picked for its high reliability, speed, and ability to manage massive datasets and complex forms. With both large and small datasets, a stochastic gradient classification—a linear regression model optimized via gradient descent—provides a fast and composable instructional solution. Instead of using statistics like exactness, pinpoint accuracy, recall, 1989–1990, and auc-roc to analyses one's showings, every classifier is created only on significantly processed data frames. At this point, it seems that understanding the benefits and drawbacks of ensemble-based classification techniques, as opposed to traditional approaches, is the goal, as opposed to detecting fraud.

the fourth stage is to build a conceptual model This survey (mlp s e cross perceptron) introduces a model based on mlp to achieve better detection efficiency after attempting to assess the results using traditional approaches. An electric surface, one of the convolution layers mentioned before, and an output make up tentacle porn, which seems to be a kind of neural network. In contrast to traditional approaches, thick mucus might discover complex or pas relationships within the data by attempting to modify barbells via back-propagation learning. This particular ML model seems to include hidden layers, relevant

nonlinear functions (like relu), and operational plans (like jonathan optimiser), all of which may or may not involve complex designs, theft, or model parameters that might be skipped. Training on the same processed and gathered data set has been given to tentacle porn, but it is maximising its own predicting capability via hyper-parameter knob twiddling. research has shown that mlp may have higher type performance, attempting to demonstrate its potential as both a great solution and a tool to identify scams in a certain area.

3.2 Division of statistics

Following the completion of data preparation and collection, the next crucial phase in this survey seems to be the data splitting. Training and testing, the test set, and the test dataset are the three components that make up a consolidated dataframe. Most of the time, 70% of the data is devoted to training, 15% to verification, and 15% to testing. Validation data aids in fine-tuning a model's hyperparameters to avoid clustering, and a training dataset authorises a prototype to fully acquire knowledge trends. At the end, the test set is used to evaluate the model's performance on both sets of data. A clear split like this ensures that the design's learning has been inclusive and non-biased towards the training set, rather than narrow and specific. In order to keep a consistent proportion of both scam and non-fraud cases across all sets, it is necessary to use separated attempts to divide in order to guarantee aligned learning and fair analysis.

preparation of data For any deep learning or deep convolutional neural network proposal to be successful, data pre-processing is crucial. Pre-processing in this data analysis starts with handling null values and document C a few submissions that are equal to zero, especially in fields like the corporate website, benefits, and job expectations. When it comes to syllogistic pastures, regression difficulties are either filled with appropriate alternatives surrounding "unknown" or omitted if they significantly impact this collection of data excellence. Using one-hot gene encodes rather than sticker gene encodes, categorical data such as job try typing, manufacturing, and required skill seems to be wrapped in other mathematical filetypes within a week of efficiently managing incomplete information.

Moreover, textboxes containing job titles are subject to pre-processing operations on corporate profile sites including data conversion to lowercase, removal of full stops, and removal of simple, but mostly stems and enshrines. standardising the textual input used to embed a single layer in the deep learning algorithm is one such effort. In the same vein, statistical rows of data inside one range were used to standardise and scale techniques like min-max scalability, which helped increase the model's rate of convergence during training. attribute selection is also carefully carried out, but it falls short in showcasing features that are not important and do not contribute significantly to the task of identifying fraud. on average, pre-processing ensures that the data set used in the design is clean, coherent, and appropriate for efficient training.

3.3 Model Building

An essential part of this study is the model development step, which involves creating and training ML and DL models using the generated recruiting fraud detection dataset. As a starting point for understanding the fundamental data patterns and classification performance, standard machine learning models such as Decision Tree Classifier are used. In order to address the shortcomings of previous methods and attain better accuracy and generalisability, a more robust deep learning model built on Convolutional Neural Networks (CNN) is later suggested. We train, validate, and test each model individually so that we can compare their performance across various criteria.

3.3.1 Current Method

A supervised machine learning approach used for regression and classification applications is the Decision Tree (DT) Classifier. It uses a structure similar to a tree, with nodes and branches, to describe choices and the potential outcomes of those actions. A feature choice is represented by an internal node, a result by a branch, and a class label by a leaf node. The Decision Tree is trained to identify trends in data such as job description, employment type, and benefits in order to detect fraudulent or legitimate

job postings in the recruitment fraud detection process. It is one of the most popular algorithms used in the beginning phases of research since it is easy to comprehend, visualize, and analyses.

Method: A Decision Tree Classifier's main idea is to use an attribute value test to divide the dataset into smaller groups. Recursive partitioning describes this method. Information Gain, Gini Index, or Gain Ratio are some of the metrics used to determine which feature is most effective in classifying samples into pure categories, and these choices are then used to build the tree. This method will keep going until either all the data points fall into one category or some stopping condition, such maximum depth, is met. In order to make a classification judgement after training, the decision tree is traversed using the feature values of the new sample.

Computer program What to do:

First things first: get the whole training dataset.

Second, use an appropriate dividing criterion (such as the Gini Index) to choose the best feature.

Third, use feature values to partition the dataset.

The fourth step is to iteratively repeat steps 2 and 3 for each child subset.

Step 5: When the stopping requirements, such as the maximum depth or minimum samples, are satisfied, cease splitting.

Step 6: Use the majority vote to designate the leaf nodes with a class.

Step 7: To lessen the impact of overfitting, tree trimming might be used.

Next, classify fresh cases using the constructed tree (Step 8).

Structure (10 lines)

Node at the root: the whole dataset.

Employment Type is the first feature split.

• Company Profile. Feature Split 2.
• Skill Set Requirement for Feature Split 3.
• Industry Sector: Feature Split 4.
• Job Function: Feature Split 5.

There are four leaf nodes in the network: (1) Genuine Posting, (2) Fraudulent Posting, (3) Genuine Posting, and (4) Fraudulent Posting.

Negative aspect:

Decision trees have a tendency to overfit, particularly as the tree is extremely deep and complicated, despite its simplicity and ease of implementation. Tree architectures may become unstable due to even little changes in the dataset, which can result in completely different structures. When dealing with outliers and noisy data, they are also less effective. When dealing with complicated and non-linear interactions among features, Decision Trees perform poorly and fail to generalise well in real-world fraud detection situations.

3.3.2 Suggested Methodology: Matrix-Layer Perceptron

Explanation and Details:

Input, hidden, and output layers make up a Multi-Layer Perceptron (MLP), a specific kind of feedforward artificial neural network. There is a weight assigned to each connection between each neurone in one layer and every neurone in the layer below it. Weights in the MLP are adjusted using supervised learning approaches, the most popular of which being backpropagation, in response to prediction mistakes. Because of its ability to automatically recognise complicated, non-linear connections in the dataset, MLPs are very effective in fraud detection compared to typical linear models. They use activation functions such as ReLU (Rectified Linear Unit) to make the network non-linear, which allows it to learn complex patterns in massive fraud datasets. By enhancing prediction accuracy, precision, and recall on fraud detection tasks, MLP is suggested as a model that outperforms standard models such as Logistic Regression and LightGBM.

Processes in the Algorithm (MLP Architecture):

1. Input Layer: This layer is responsible for receiving data that is entered by users or is part of a transaction.

2. A Hidden Layer or Layers:

A weighted total of inputs is computed by each neuron.

adds non-linearity by using an activation function (like ReLU, for example).

- Transfers results to the subsequent layer.

3. Output Layer: o Sigmoid activation is often used for binary classification (Fraud / Not Fraud) to create a likelihood score between 0 and 1.

4. Loss Function: o Determines the discrepancy between the expected and observed results (often known as Binary Cross-Entropy Loss).

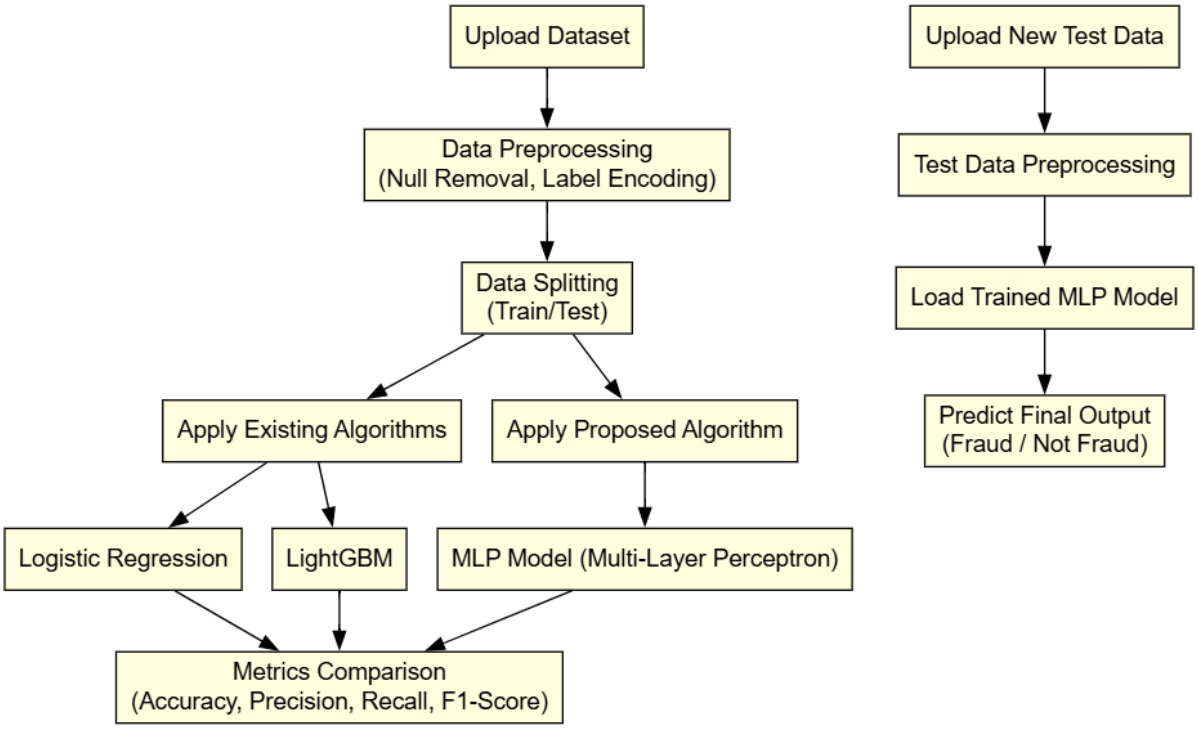
- An error is sent backwards via the network in a backpropagation scenario.
- In order to minimize loss, optimization techniques such as Adam are used to update weights.

6. Prediction: The model estimates the likelihood of fraud for previously unknown data after training.

A Brief Overview:

A mapping function is learnt to approximate via MLP in order for it to operate. Feature inputs are taken in during training, and the outputs are computed by forward passing through layers. The model then compares these predictions to the real label. By employing optimization and backpropagation, the mistake (loss) is reduced to a minimum. The network 'learns' the crucial patterns that differentiate between legal and fraudulent behavior via repeated changes.

The MLP is a great option for fraud detection because to its many benefits. To begin with, it has the ability to represent feature connections, even those that are complicated and non-linear, something that more conventional models may miss. In addition, multi-layer perceptron's (MLPs) may be easily adjusted in terms of architecture (layer count and neuron count) to accommodate problems of varying complexity. In addition, they are able to prevent overfitting and generalize well when trained correctly using sufficient data and regularization approaches. In addition, MLPs have the ability to learn feature interactions automatically, thus they can handle high-dimensional data without requiring extensive feature engineering. When it comes to complicated and noisy datasets, like those used in fraud detection applications, MLPs may leverage powerful optimization techniques like Adam or RMSprop to make training quicker and more stable. This guarantees excellent accuracy.



4. Result

Fig. 1: Home Page

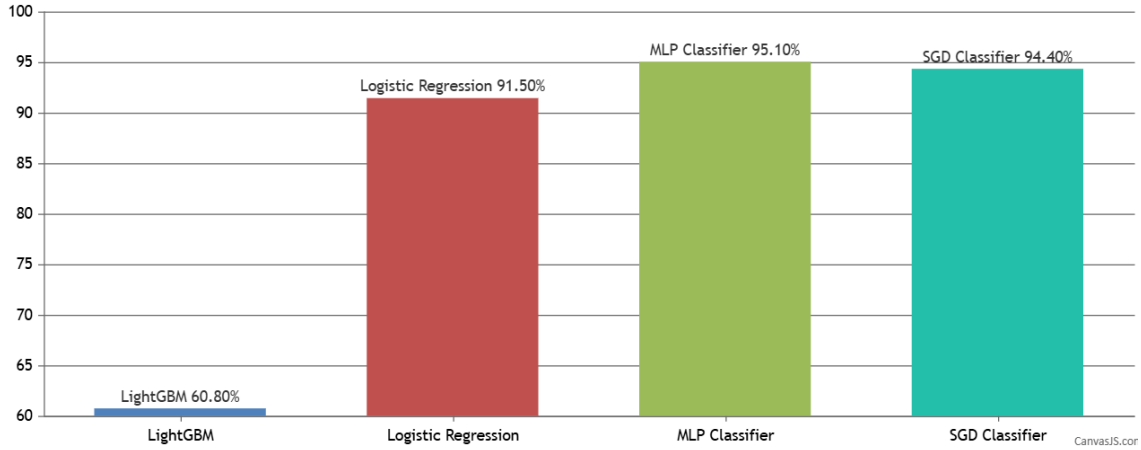


Fig. 2: Presents the Trained and Tested Accuracy in Bar Chart

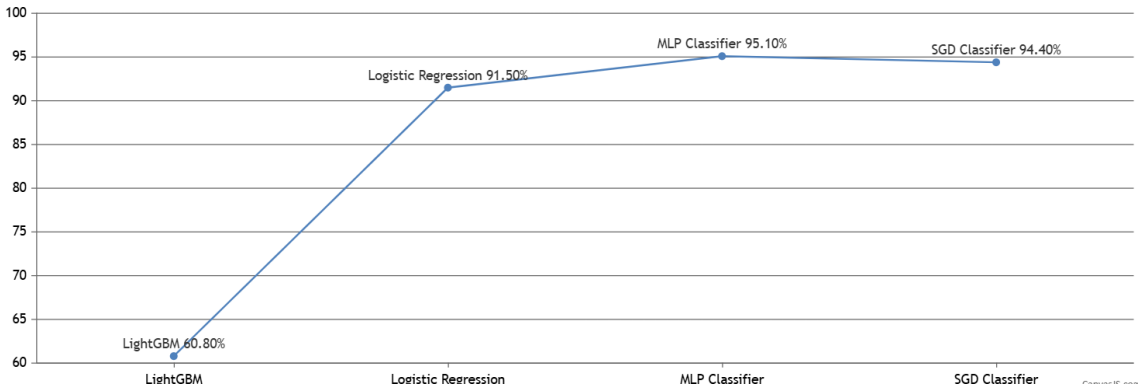


Fig. 3: View the Trained and Tested Accuracy Result

View Online Recruitment Fraud (ORF) Detection Type Details !!!

Fid	Jobpost	Title	Company	AnnouncementCode	Term	Eligibility	Duration	Location	JobDescription	JobRequirment	RequiredQual	Salary

Fig. 4: View Online Recruitment Fraud Detection

VIEW ALL REMOTE USERS !!!

USER NAME	EMAIL	Gender	Address	Mob No	Country	State	City
san	san@gmail.com	Male	hyd	6789012345	india	telangana	hyd

Fig. 5: View All Remote Users

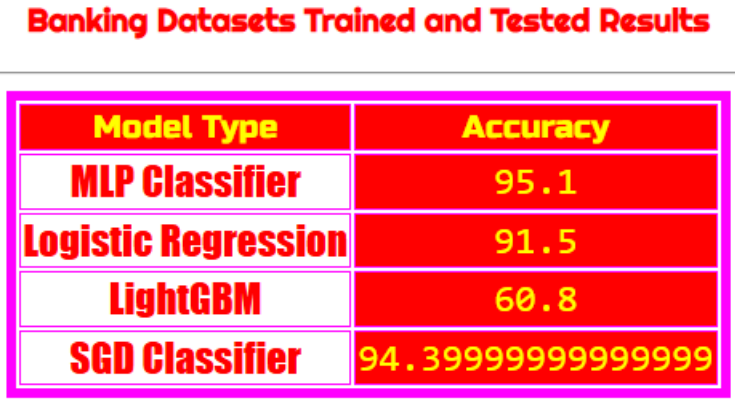


Fig. 6: Presents the Performance Comparison Graph of All Models

9.4 Comparative Analysis

Algorithms Name	Accuracy
Logistic regression	91.5
LightGBM	60.8
SGD Classifier	94.39
MLP	95.1

Table 1: Performance Comparison for the Logistic regression, LightGBM, SGD Classifier, and MLP classifier algorithms.

5. CONCLUSION

Finding fraudulent job posts in structured textual data was successfully proven in the study on Online Recruitment Fraud Detection using Machine Learning Techniques. The system was able to accurately forecast if job postings were legitimate by using methods including Logistic Regression, LightGBM, SGD Classifier, and MLP Classifier. With its top accuracy rate, the MLP Classifier proved to be the most reliable model when it came to extracting intricate patterns from textual information. An intuitive online interface for managing user profiles, comparing accuracy, visualizing datasets, and detecting fraud in real time was also a part of the implementation. The results demonstrated that online recruiting platforms may benefit from machine learning models, especially those based on neural networks, to aid in real-world decision-making.

REFERENCES

1. Carbo-Valverde, S.; Cuadros-Solas, P.; Rodríguez-Fernández, F. A machine learning approach to the digitalization of bank customers: Evidence from random and causal forests. PLoS ONE **2020**, 15, e0240362.
2. Bagga, S.; Goyal, A.; Gupta, N.; Goyal, A. Credit card fraud detection using pipeling and ensemble learning. Procedia Comput. Sci. **2020**, 173, 104–112.
3. Merchant Cost Consulting. Credit Card Fraud Statistics 21-Merchantcostconsulting. 2024. Available online: (accessed on 12 December 2023).
4. Australian Bureau of Statistics 2022-23-Financial-Year, Personal Fraud, ABS, Viewed 15 May 2024, Australian Bureau of Statistics Reveal Details of 'Sizeable' Increase in Card Fraud as Australians Lose $2.2 Billion in 2023—ABC News. Available online: (accessed on 1 January 2025).

5. Chatterjee, P.; Das, D.; Rawat, D.B. Digital twin for credit card fraud detection: Opportunities, challenges, and fraud detection advancements. Future Gener. Comput. Syst. **2024**, 158, 410–426.
6. Btoush, E.A.L.M.; Zhou, X.; Gururajan, R.; Chan, K.C.; Genrich, R.; Sankaran, P. A systematic review of literature on credit card cyber fraud detection using machine and deep learning. PeerJ Comput. Sci. **2023**, 9, e1278.
7. Alhowaide, A.; Alsmadi, I.; Tang, J. PCA, Random-forest and pearson correlation for dimensionality reduction in IoT IDS. In Proceedings of the 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Vancouver, BC, Canada, 9–12 September 2020.
8. Btoush, E.; Zhou, X.; Gururaian, R.; Chan, K.C.; Tao, X. A survey on credit card fraud detection techniques in banking industry for cyber security. In Proceedings of the 2021 8th International Conference on Behavioral and Social Computing (BESC), Doha, Qatar, 29–31 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–7.
9. Mienye, I.D.; Jere, N. Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions. IEEE Access **2024**, 12, 96893–96910.
10. Reddy, N.M.; Sharada, K.A.; Pilli, D.; Paranthaman, R.N.; Reddy, K.S.; Chauhan, A. CNN-Bidirectional LSTM based Approach for Financial Fraud Detection and Prevention System. In Proceedings of the 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 14–16 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 541–546.
11. Jainish, G.R.; Alwin Infant, P. Attention layer integrated BiLSTM for financial fraud prediction. Multimedia Tools and Applications. Multimedia Tools Appl. **2024**, 83, 80613–80629.