Fake News Detection using ML and DL Approaches

G.Srinivas
Department of IT
Aditya Engineering College
Surampalem, India
srinivas.g@aec.edu.in

A.Lakshmanarao
Department of IT
Aditya Engineering College
Surampalem, India
laxman1216@gmail.com

S.Sushma
Department of IT
Aditya Engineering College
Surampalem, India
sushma.sunkara@aec.edu.in

Dr.M.Vamsi Krishna
Department of IT,
Aditya Engineering College
Surampalem, India
vamsikrishnam@aec.edu.in

S.Neelima
Department of IT
Aditya Engineering College
Surampalem, India
neelima.sadineni@aec.edu.in

2023 International Conference on Circuit Power and Computing Technologies (ICCPCT) | 979-8-3503-3324-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCPCT58313.2023.10245398

Abstract—**Advances in industry and technology have created numerous new work opportunities for job seekers and the unemployed. The number of job openings advertised on the internet has also increased. It takes time and effort to apply for a job online. It may also cause concern because they have access to our personal, professional, and academic data. As a result, it is critical to determine whether the job we are looking for is legitimate or not. This model is being developed to detect bogus job postings. We employed Deep Neural Network in this project. For comparison, Machine learning algorithms like K Nearest Neighbour, NB classifier, decision tree, SVM, and RF classifiers were also used in this work.**

Keywords— Fake job, Kaggle, Machine Learning, Deep Learning.

I. INTRODUCTION

Advances in industry and technology have created numerous new work opportunities for job seekers and the unemployed. The number of job openings advertised on the internet has also increased. It takes time and effort to apply for a job online. It may also cause concern because they have access to our personal, professional, and academic data. As a result, it is critical to determine whether the job we are looking for is legitimate or not. This model is being developed to detect bogus job postings.Fake online job postings are more than prepared to coastal particular and licensed report from professional candidate. They may attempt to illegally collect money on occasion. According to a recent survey conducted by Action Fraud in the United Kingdom, more than 67 percent people who hunt for work online but are unaware of bogus job postings are putting themselves in grave danger. With the increasing ability to disseminate job advertisements, every ratio of fraudulent task registration has increased, causing harassment to job seekers. As a result, public are less interested in fresh job listings in order to maintain the bond along with flexibility about its special,intellectual as well as competent data. As a result, gaining people's trust and confidence in legitimate job advertisements through social and electronic media is incredibly difficult. Many works have been done through forecast whenever the position panel is bogus or else real. They have used content clarification applying Deep training Miniature, feature selection applying Support Vector Machine and other works used various machine learning models.

To increase the accuracy we are using Deep Neural Network. For classification we have made a comparative study using NB technique, Logistic Regression , DTR, RF Classifier.

Deep Neural Network(DNN)

DNNs are upgraded report of the typical ANN with numerous layers. The DNN images are newly becoming famous The architecture of DNN is shown in figure 1.

Input Layer– Initial is the input coping. Here coping obtain every testimony also cut into vacation about every system.

Hidden Layers– Further category belonging to hidden layer. Hidden layers do separate or else extra count as one neural structure. Hidden layers persist that truly answerable whereas every accomplished work along with complication about neural networks. They execute various activities appearing in equal era parallel like evidence renewel, mechanical element formation .

Output layer– End layer is the output layer. The output layer influence every product or the output of the dilemma. Rough apperance grab acknowledged to the input layer and pickup output in the output layer.

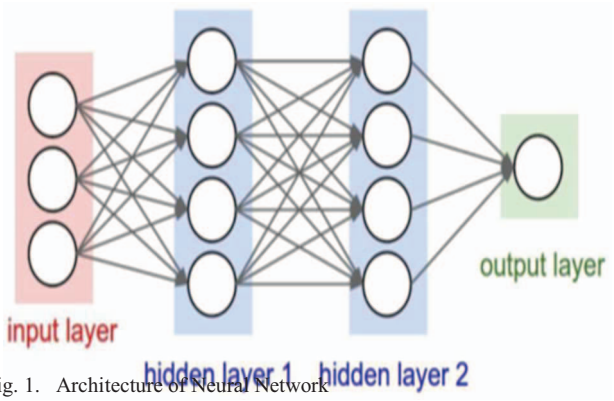

Fig. 1. Architecture of Neural Network

Deep neural networks (DNN) is a classy automobile reasearch innovation is comparable through artificial neural network including plan through imitator info clarification belonging to every intellect. DNN brush further than single unknown sheet positioned among every aid as well as output sheets.Each layer consist of a obsessed sum of entity (neurons)

979-8-3503-3324-4/23/$31.00 ©2023 IEEE

that handle a positive practical revolution toward information. The particular category about image can relative every nature about objective.

Random forest

Random Forest is a attractive appliance reasearch method that apply to the managed information procedure. They can continue recycled as the pair of Classification and Regression obstacles in ML. It is positioned about every theory as regards ensemble training, that is a movement based on connecting collective classifiers toward clarify every complicated issue moreover to enhance the achievement related to every copy. The larger figure of trees in the forest edge to above accuracy and avoid the trouble of overfitting.

II. LITERATURE SURVEY

Vidros et al. [1] comprehensible post hypocrites just as forged online trade announcer. They begin data around multiplied absolute moreover esteemed community including company which formed forged spot publicity either void net among ill-cause. They attempted about adopting a few analysis methods like NB, RF, Zero R, One R etc. RF exhibit every finest achievement against the info arranged with 89.5% analysis efficiency. They launch Logistic Regression operating perfect strapped close to evidence agreed. One R classifier accomplished strong while we equivalent the info together with adventured against this. In [2], they suggested a miniature to disclose scam disclosure in an Online enrollment Scheme. They created testimony in ternary stairs- info pre-clarification, aspect collection and deceit exposure applying classifier. In the clarification step, they ejected blast along with html strip from the info, In order that the familiar content arrangement survive dehydrated. They enforced feature selection approach into diminish every figure about description adequate and afferent. Support Vector Machine continue as feature selection and Ensemble Classifier adopting Random Forest recycled toward notice bogus position against evaluation info. Random Forest assume one timber ordered classifier whatever treated in the process of ensemble classifier including bulk polling technique . The present classifier exhibit 90.4% designation skill through expose forged carrer bar. In [3], authors scheduled into distinct Deep Neural Network images related Bi-GRU-LSTM CNN and Bi-GRU CNN whatever pre-qualified including content info. We fashioned about distributing business statistics. We experienced trade info about TextCNN copy subsist away from convolution sheet, pooling sheet and entirely linked sheet. The present copy skilled info over convolution and pooling sheets. Later every competent burden were depressed along with approved toward fully united sheet. This standard recycled soft max action as grade approach. They again worn ensemble classifier adopting superiority polling procedure toward increment analysis skill. They construct 66% grade skill using TextCNN along with 70% skill as Bi-GRU-LSTM CNN independently. The present analysis effort accomplished finest along ensemble classifier having an efficiency 72.4%. Zhang et al.[4] designed mechanical forged distinguish model to recognize among accurate and fabricated report accepting content clarification. They recycled characteristic info about broadcast material displayed through PolitiFact network . The present info worn toward teach every recommended GDU diffusive entity copy. Acquiring opinion against collective causes concurrently, the indicated competent miniature accomplished strong in the process of mechanized fraudulent pointer copy. F. Alharby et al.[5] proposed an imaginative copy for Online trade Fraud Detection. The online service organization exploit strong positioned machinery related social internet or intranet toward volunteer team. Trade import countless assests as rigid achievement aside obtaining finest competition in moment, presenting efficient demands, estimate nominees responsive freshly elected researchers. Additionally, the present mechanism generate every reserving action new economical along with energetic after expensing high payment. Every analytical contents about online service encompass hunting every condition of contestants, trade entry, along with civil structure. Y. Kim et al. [6] proposed Convolutional neural networks as judgement analysis. This paper recommended unique convolutional neural network with debate structure to raise the completion of sentence classification and manifest the planned model approximately exceed the traditional CNN model and attain ambitious work with the ones that adventure affluent acceptable features. H.Sharma et al.[7]proposed inspection about Decision Tree method about analysis in Data Mining. This paper studied various decision tree algorithms. The ability of disparate decision tree result could considered placed about their skill along with era arrested toward extract every stock. In [8], [9] and [10] the authors applied various ML classifiers and achieved good results.

III. METHODOLOGY

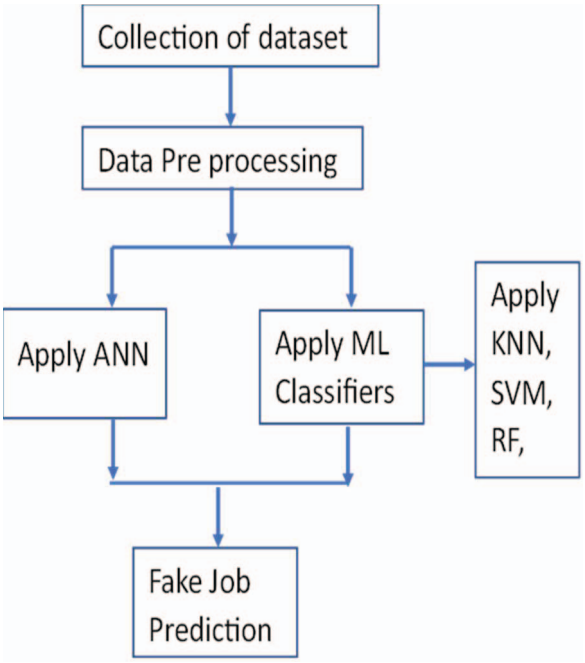

Fig. 2. Denoising image (Guassian denoising)

The proposed method is shown in figure 2. First, a dataset from kaggle is downloaded. Later, preprocessing steps applied on dataset. After that, a deep learning ANN proposed for fake job detection. For comrining the proposed model, Several ML classifiers also applied.

IV. EXPERIMENTS AND RESULTS

A. Dataset

A fake news dataset from Kaggle is collected. It consists of both fake and true jobs. The number of fake jobs is 1000 and true jobs is 866. It is shown in figure 3.

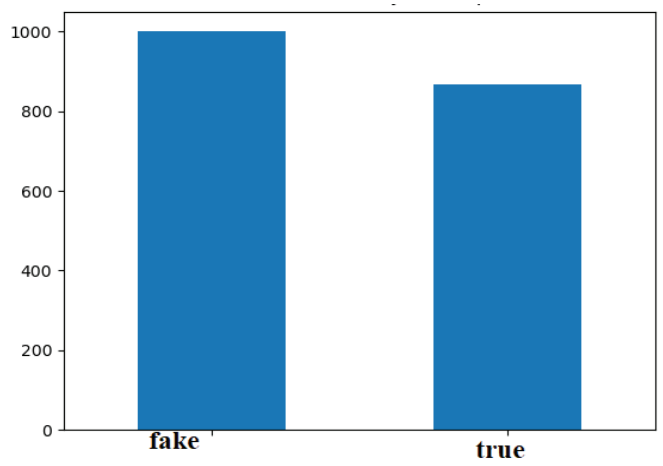

Fig. 3. Fake vs True jobs

B. *Data preprcessing*

Next, data preprocessing done with several techniques like stop word removal, stemming and lemmatizer.

Stop words are the words those are not useful in text analysis. Example stop words are “is” “was” etc. Steemming is a process where word prefixes and suffixes are removed. For example, a word “playing” is changed as “play”. After applying all these preprocessing steps, the given input data is in clean state(figure -3) . At this stage, there is a need to vectorize the input data. For this purpose, a TF-IDF vectorized used,

Tf-Idf vectorization creates floats for each word that down-weight common words and emphasize unusual words. This will devalue ordinary terms like "the," "a," , "and," etc.

Fig. 4. *Data after Preprocessing*

C. *Applying ML classifiers*

Later, Several Machine Learning classifiers applied for fake job detection. The ML algorithms applied are KNN, RF, SVM, Decision Tree and Naïve Bayes classifier. Before applying ML classifiers, the dataset is divided in to training and testing sets with 80% and 20% split. The number of samples in train set is 1492 and number of samples in test set is 374.

The accuracy, precision, recall, FM measure values after applying the ML classifiers is shown in the table-1 and figure-5.

TABLE I. ALGORITHM COMPARISION

Alg	Accuracy(%)	Precision(%)	Recall(%)	FM(%)
KNN	75.4	81.8	77	74.5
Naïve Bayes	88.7	89.7	89.3	88.7
DTC	89.5	90.3	90.6	89.5
RF	88.2	89	89.2	88.2
SVM	90.1	90.8	90.3	90

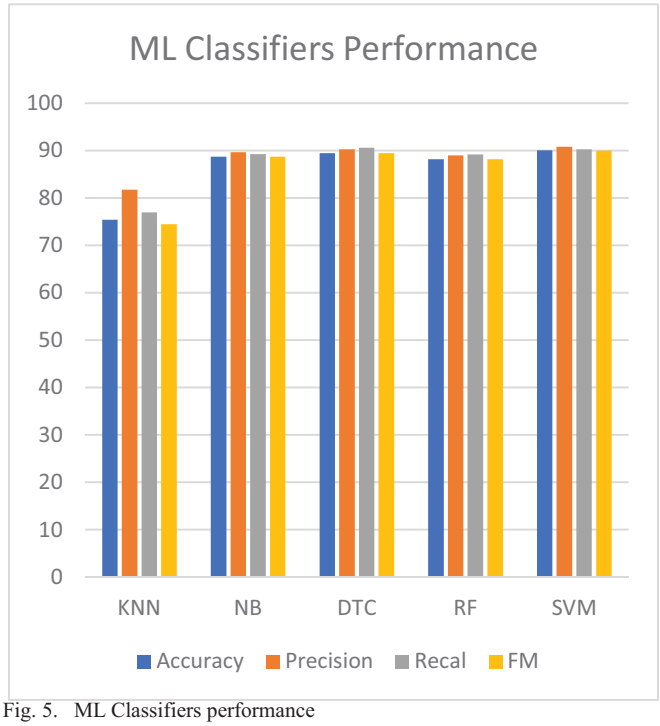

Fig. 5. ML Classifiers performance

D. *Applying Deep Neural Networks*

The maximum accuracy achieved with ML classifiers is 90.1% with support vector machine. To further increase accuracy, a deep neural network (CNN) applied. The proposed CNN given accuracy of 94.35, precision of 94.9%, Recall of 94.4% and FMeasure of 94.3%. So, proposed deep learning model achieved more accuracy than conventional ML classifiers. The accuracy comparison of ML classifiers and DNN was shown in Figure 6.

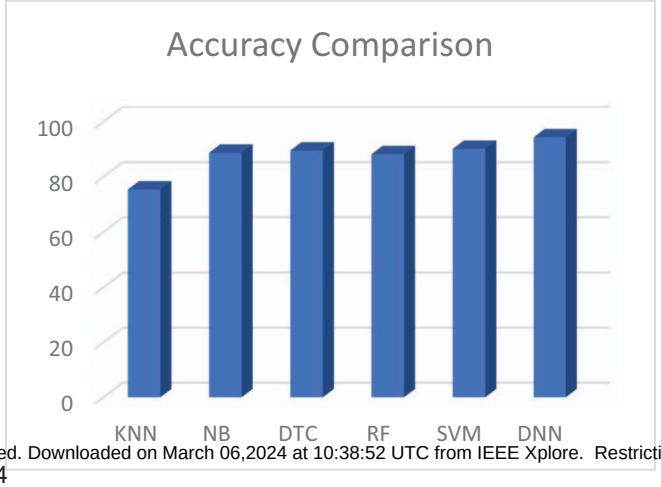

Fig. 6. Accuracy Comparison

E. Creating Web App

The proposed work is created as a webapp using FLASK.The figures 7,8 shows the working of Fake job detection web app.

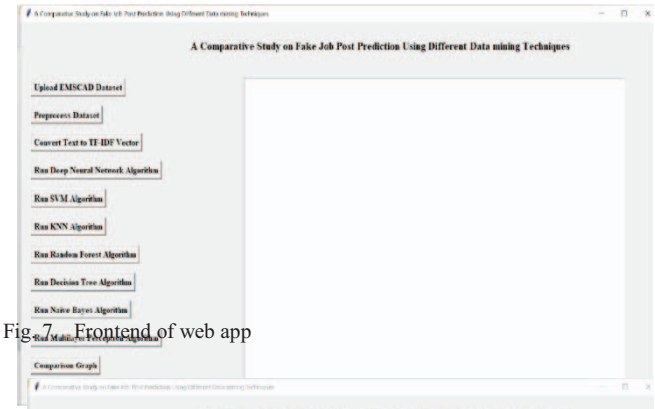

Fig. 7. Frontend of web app

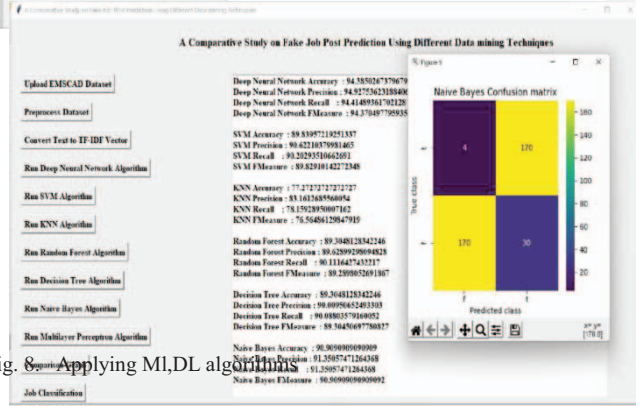

Fig. 8. Applying Ml,DL algorithms

V. CONCLUSION

In this work, Using DNN and pre-trained models, we presented a method for determining if a job posting is fraudulent or legitimate. Previously, many algorithms were tried to forecast job postings. However, accuracy is less in those initiatives. As a result, we conducted a comparison study of several algorithms and discovered that DNN is the most effective. Finally, they trust a particular model given in previously mentioned examination has hidden do used for prediction.

REFERENCES

[1] S. Vidros et al., "Automatic Detection of Online Recruitment Frauds: Characteristics, Methods, and a Public Dataset," Future Internet, vol. 9, no. 1, p. 6, Mar. 2017, doi: 10.3390/fi9010006.

[2] Alghamdi.B et al., "An Intelligent Model for Online Recruitment Fraud Detection. Journal of Information Security," 10, 2019.

[3] R.Rafique, "Deep fake detection and classification using error-level analysis and deep learning", Sci Rep 13, 7422 (2023).

[4] Rish, Irina, "An Empirical Study of the Naïve Bayes Classifier" ,IJCAI 2001 Work Empir Methods Artif Intell.

[5] F. Alharby,"An Intelligent Model for Online Recruitment Fraud Detection," J. Inf. Secur., vol. 10, no. 03, pp. 155–176, 2019.

[6] Y. Kim, "Convolutional neural networks for sentence classification," arXivPrepr. arXiv1408.5882, 2014.

[7] H. Sharma et al.,"A Survey on Decision Tree Algorithms of Classification in Data Mining, " International Journal. Sci. Res., vol. 5, no. 4, pp. 2094–2097, 2016.

[8] A.Yasin et al., "An Intelligent Classification Model for Phishing Email Detection", International Journal of Network,Security& Its Applications,2016.

[9] Jiawei Zhang et al., "Fakedetector: Effective Fake News Detection with Deep Diffusive Neural Network", International Conference on Data Engineering , 2020.

[10] A.Lakshmanarao et al., "An Effecient Fake News Detection System Using Machine Learning", International Journal of Innovative Technology and Exploring Engineering, Volume-8 Issue-10, August 2019.