ISSN No:-2456-2165

Salary Prediction Model using Principal Component Analysis and Deep Neural Network Algorithm

Habibu Aminu[1]
Department of Mathematical Sciences
Abubakar Tafawa Balewa University
Bauchi, Nigeria

Badamasi Imam Yau[2]
Department of Mathematical Sciences
Abubakar Tafawa Balewa University
Bauchi, Nigeria

Fatima Umar Zambuk[3]
Department of Mathematical Sciences
Abubakar Tafawa Balewa University
Bauchi, Nigeria

Emanuel Ramsom Nanin[4]
Department of Computer Science
Abubakar Tafawa Balewa University
Bauchi, Nigeria

Abdulmutallib Abdullahi[5]
Department of Computer Sciences
Federal University Calabar
Cross River, Nigeria

Ismail Zahraddeen Yakubu[6]
Department of Computing Technologies
SRM Institute of Science and Technology
Chennai, India

Abstract:- Machine learning implementations are growing in every organization for forecasting their employee's working nature and competence by calculating the time taken by the employee to complete the task. Numerous recent research have been published to evaluate the effectiveness of various clssification algorithms for predicting the employee salary classes. From the machine learning perception, Salary prediction is a difficult task due to the small sample size, relatively high dimensionality, and presence of noise. To address this, to find more useful features, deeper architectures are required. Additionally, more data analysis and data processing can be pragmatic to make the prediction model go beyond the correlation and precision standards by feature extraction techniques. Hence, this study proposes an enhanced method for salary prediction that selects a subset of characteristics from all available data using a PCA system and a deep neural network (DNN) model for the classification process. Upon assessment with other classical machine learning methods such as DT and RF. Better classification accuracy, precision, recall, and F-score are achieved by the proposed DNN model. Furthermore, the proposed DNN model achieves the highest MAE of 94.9% as compared to DT and RF, which attain an MAE score of 89.6% and 76.4% respectively. This result suggests that the proposed model has a prediction error of 5.1% which is fewer when compared to DT and RF which has prediction error of as much as 10.4% and 23.6%, thereby, signifying the dominance of deep learning algorithm over conventional machine learning algorithms in salary classification and prediction task.

Keywords:- *Principal Component Analysis, Random Forest, Deep Neural Network, Machine Learning, Deep Learning, and Decision Tree.*

I. INTRODUCTION

A prediction is an assumption about something that might happen. Though not usually, a prediction is founded on knowledge or past events. Future events are not always guaranteed, hence it is frequently impossible to establish precise data regarding the future, in order to prepare plans regarding likely developments, prediction may be helpful [1]. A prediction engine is a technology that uses a collection of historical observations to forecast a future result. Due to its ability to produce forecasts that are nearly as accurate and affordable as those made by people, prediction engines have grown in popularity in recent years. Another benefit of utilizing a prediction engine is that it makes predictions rather than making decisions on its own, leaving decision-making to humans or users. Today, industry and scholars use these forecast engines to foresee an abundance of issues [2].

Today's research indicates that every company operating for a year or more most likely has "a ton of data" in its file that might help them make better decisions [2]. But a human cannot reasonably make a choice based on this volume of data, which is why we require a model that can accomplish this for us in the most pertinent way imaginable [cite]. In addition, these models must estimate the financial and economic implications of the choice or phenomenon, which are inherently opaque to some degree [cite]. Machine learning techniques can be utilized to develop a model or prediction engine that have the intelligence to accomplish a task. By evaluating all the variables based on historical data, this type of prediction engine determines the most advantageous result after receiving a few critical features as input [2].

The work economy has experienced a notable impact due to the growth of the Internet [cite]. Determining the most desirable and coveted aspects of job offers is crucial for both recruiters and prospects. Since hiring decisions have a significant impact on a company's competitiveness and productivity, hiring decisions are a crucial component of any business strategy. [cite]. The objective of selecting the best experts for any company's requirements has prompted the development of recruitment processes into intricate tasks that now involve thorough candidate evaluations and interviews. The introduction of the web and the Internet has made e-Recruitment a crucial component of all hiring practices [3].

Machine learning implementations are growing in every organization for predicting their personnel working nature and the ability of the employee by calculating the time taken by the employee to ample the task [4]. A lot of studies [3, 5-8] have been published recently to evaluate how well different classification algorithms fare in forecasting employee salary classifications. Among these algorithms are Bayesian belief networks, naïve Bayes, support vector machines, decision trees, and neural networks [9]. Every one of the previously mentioned algorithms has demonstrated encouraging outcomes, as evidenced by the research in [1, 4, 5, 10].

However, the prediction determines such pay amounts using a variety of data regarding a potential or present employee, is a challenge that businesses encounter on a regular basis [cite]. While HR managers usually communicate with relevant department-level managers when addressing such wage forecast and negotiating issues, they would greatly benefit from any automated system that have such a capability [cite]. In light of an employee's characteristics (current or prospective), which provides details about her qualifications and performance level in addition to her demographic profile etc., the salary class can be predicted using a number of well-known classification algorithms. [cite].

From a machine learning standpoint, predicting salary is a difficult task due to the limited number of samples, relatively high dimensionality, and noise [cite]. To address this, to locate features that are more informative, deeper architectures are required. Furthermore, by using features extraction techniques, further data analysis and processing can be used to extend the prediction model's capabilities beyond correlation and accuracy values. Thus, this study presented an improved technique for salary prediction that makes use of a PCA system to choose a subset of characteristics from all required features and a deep neural network (DNN) model for classification. The following summarizes the remaining part of the paper: The related work is presented in Section 2, the methodology is explained in Section 3, the results are presented in Section 4, and the study is concluded in Section 5.

II. RELATED WORK

For growth and success in knowledge-based enterprises, compensation planning is an essential strategic area. An ideal wage offer is essential for keeping top performers on staff. Companies frequently have to determine such wage estimates based on a plethora of information about a previous or prospective employee. While HR managers usually interact with relevant department-level executives to address wage forecast and negotiating difficulties, they would greatly benefit from any computerized system with this capability. Considering the attributes of a worker (potential or present), encompassing her demographic profile and additional details like performance level and criteria etc., It is possible to forecast the salary class using a variety of well-known classification algorithms. However, it is incorrect that these employee data specifics for any establishment are typically not available to the public for the purpose of assessing the effectiveness of classification algorithms [9].

However, this constraint is partially overcome, for example, [9] An analysis comparing the effectiveness of several classification systems in forecasting employee wage groups. These algorithms include; Neural networks, decision trees, support vector machines, naïve Bayes, and Bayesian belief networks. Decision trees and Bayesian belief networks outperform the other three algorithms—naïve Bayes, support vector machines, and neural networks—according to experimental results. Nevertheless, the study is unable to add performance attributes to the UCI census dataset (using random generation or some other heuristics) thus, diminishing the study's achieved classification accuracy.

In [5] a data mining technique was used to create a wage prediction model for graduate students based on people with similar training characteristics. In order to determine which data mining technique is best for salary prediction, an experiment was conducted to evaluate Decision Trees ID3, C4.5, and Random Forest. Key parameters were adjusted to improve the accuracy of the results. At 90.50 percent, Random Forest provided the highest accuracy, whereas Decision Trees ID3 and C4.5 produced results with lower accuracy, at 61.37 percent and 73.96 percent, respectively. However, Random Forest models lack interpretability and are akin to opaque boxes. For really large data sets, the size of the trees can take up a lot of memory. The tendency is to overfit.

Because employee turnover has a negative effect on workplace productivity and long-term growth goals, it has been recognized as a major concern for firms. To solve this problem of overfitting [10] Examine how the Extreme Gradient Boosting (XGBoost) technique, whose regularization formulation makes it more resilient, can be applied. XGBoost's much improved accuracy for staff turnover prediction is demonstrated by comparing it to six supervised classifiers from the past using data from a large retailer's HRIS. A well-thought-out network with an adequate number of hidden layers may increase accuracy;

ISSN No:-2456-2165

yet, scalability and real-world application issues must also be considered.

An important consideration for both the employer and the employee is paying a fair and reasonable wage for any given employment, which has always been a challenging issue. Salary may not be the most important consideration for someone looking for work, but it is necessary to satisfy one's fundamental human needs when combined with other considerations. Therefore, [2] suggested machine learning techniques to automate and create a proposed wage prediction model. Based on a few key parameters, the proposed prediction model can forecast the salary. The proposed method involves fitting a raw dataset into ensembles and decision trees. The results attained are encouraging and with high accuracies.

Recently, in India, [7] a quantitative approach to forecasting the factors influencing an individual's pay should be developed. Aspiring Minds' Employability Outcomes (AMEO-2015) dataset, which includes job explorer personal and employment details of Indian undergraduates, as well as the Aspiring Minds Computer Adaptive Test (AMCAT) score, has been measured for the study. It has been noted that the best indicators of compensation are rational and quant scores. The developed model's relative squared error root is 82.3056%.

Furthermore, [8] proposed Salary Prediction using the Bidirectional GRU-CNN Model and compare the outcomes with the most advanced CNN, RCNN, Bid-LSTM, and Res Net models. Experience has shown that the proposed model performs better than the other models currently in use, including Text CNN, RCNN, Bid-LSTM, and Res Net. Nevertheless, the study does not investigate additional options for text regression using the combination model.

Regression analysis was used to forecast salaries and evaluate the outcome using polynomial and linear regression. in [1]. It was experimental that, It was suggested to use regression analysis to predict salaries and evaluate the outcome using polynomial and linear regression by [1].

Additionally, [4] examined and contrasted multiple machine learning methods, including Naïve Bayes, Random Tree, Random Forest, and REP Tree, for pay status predictions. When compared to Random Forest, Naïve Bayes, and Random Tree classifiers, it is found that the precision of the REP Tree approach is the most accurate for income dataset prediction. Though, the study is susceptible to disparities in the datasets.

More recently 1, (Lothe et al., 2021) proposed machine learning-based salary prediction, which is then tested against a linear regression technique employing second-order polynomial transformation. The proposed model achieved uppermost accuracy of 76%. Nonetheless, the primary shortcoming of the research is that it solely examined algorithms for the fundamental model with a pair of features. This could result in the proposed model's accuracy being lower than intended.

III. METHODOLOGY

Utilizing a PCA approach to choose a subset of features from all specified features, the proposed method uses a DNN model for classification. The stages outlined in the proposed strategy are shown in Figure 1 and are explained below.

- Preprocessing is done on the salary dataset to eliminate noise from the instances.
- Apply PCA to determine the dataset's top four features.
- The best feature was ranked, retrieved, and other superfluous features were removed using PCA.
- Classify the dataset with a deep neural network framework.

Due to the restraint of training sample with various features, this study uses a demographic information dataset to predict the salaries. The demographic datasets empower us to have a broader angle and more diversity in the inputs which will be mirrored to have an improved assessment and more accurate results. Moreover, the relatively large dimensionality and the existence of noise in the datasets are addressed in this work by utilizing PCA. The context of the proposed method is presented in Figure 1 This investigate proposes the framework to improve salary prediction using demographic data collected for Kaggle repository portal. To expand prediction accuracy, the proposed is collected of 4 major steps, which includes data analysis stage, feature extraction stage, training, learning and prediction.

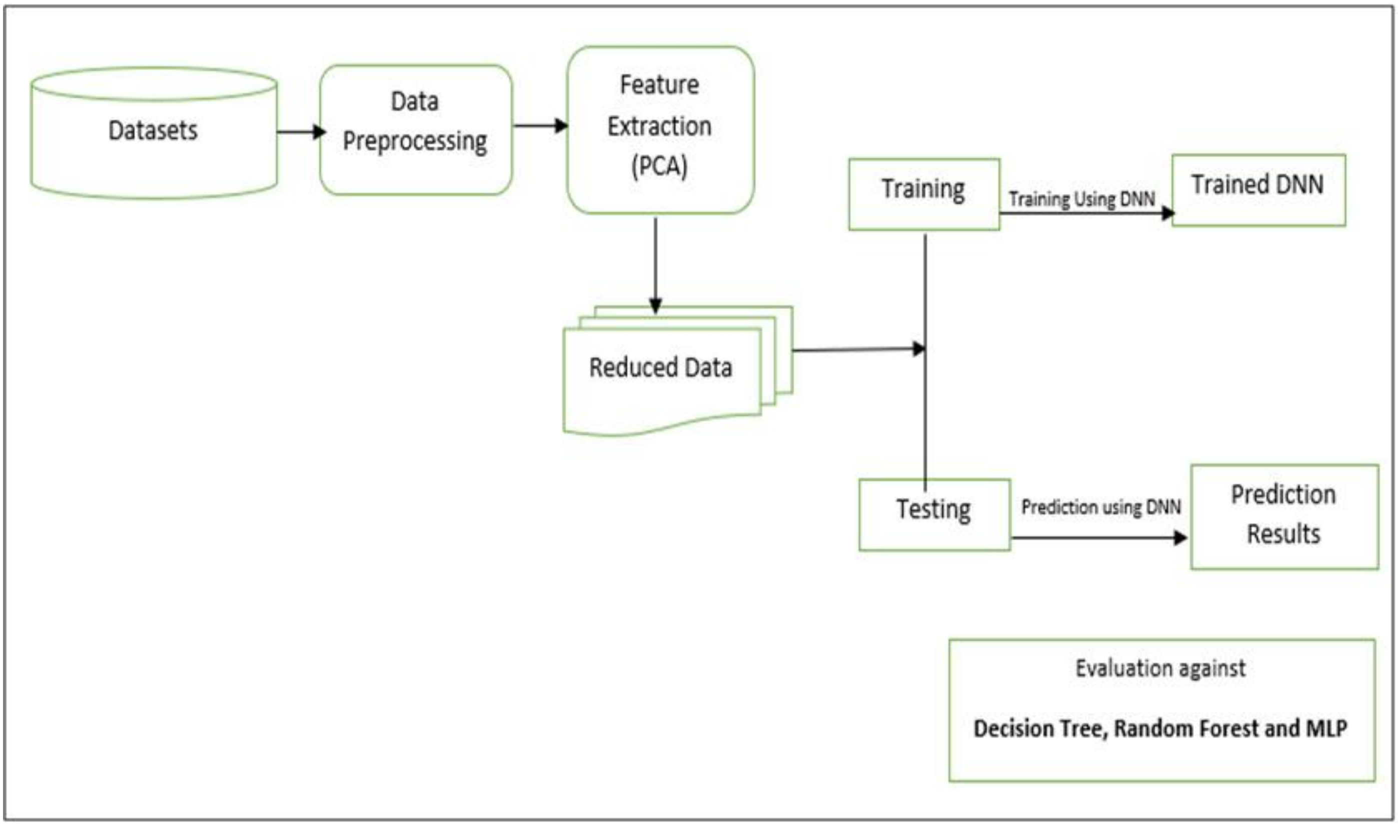

Fig 1 Architecture of the Proposed PCA-DNN Model

A. Data Pre-Processing and Analysis Stage

Both machine learning and deep learning models greatly increase their performance through the crucial phase of data pre-processing. First, the datasets were normalized using MinMax normalization technique. The motive for the choice of MinMax normalization technique is that Min-max normalization conserves the relationships among the original data values [cite]. The cost of having this bounded range is that we will end up with reduced standard deviations, which can overwhelm the effect of outliers.

Additionally, in machine learning, data splitting is typically utilized to divide data into train, test, and validation sets. Every algorithm separated the data into two subgroups: validation and training. The validation set will be used for evaluation, while the training set will be used to fit the model. Hence, For experimental purposes, the system splits the dataset into 80–20, 70–30, and 60–40 percent train-test splits. The dataset is divided at random, without regard to any particular order.

B. Feature Engineering

In this study, to help extract more information from the data that already exists, we use PCA. In terms of new features, fresh information is extracted. The variance in the training data may be more accurately described by these properties. improved model correctness as a result.

C. Principal Component Analysis

The main reason for using PCA is to reduce the number of model features to fewer, rarer components in order to help visualize data patterns and speed up the model's execution. By eliminating features with strong correlation, PCA also lessens the possibility of the model being overfit [cite]. Furthermore, When dealing with high dimensionality and highly correlated variables, PCA can enhance the prediction model's accuracy.

D. Deep Nueral Network

In this model, the unfairness value, which is often set to 1 in neural networks to prevent network outcomes that are nullified, will be allocated as 1. At achieve different results from the model, the default learning rate will be set to 0.15 and thereafter altered at random through trial and error. The network can determine the initial weight of the nodes at random, modify it during back propagation by calculating the error rate, and update it on a regular basis following each epoch. The number of inputs and data size determine the total number of hidden layers and nodes in each hidden layer. The number of epochs reached or achieving the desired outcome from the learning model is referred to as the network's finish condition. It will require more time and resources to train the model if a network has more layers and nodes.

E. Model Evaluation and Validation Stage.

To assess each algorithm's prediction results. Using the same datasets and various algorithms—DT, RF, MLP, and DNN—we will first conduct a number of experiments. We will next use the current study to assess the developed model's performance as part of our primary goal in [2].

F. Datasets Description

The Kaggle repository portal provided the open-source datasets used in this study. Unlike the previous datasets utilized in this framework, the salary datasets encompass diverse features with several classes which makes it an improvement on the existing datasets thereby appropriate for

ISSN No:-2456-2165

the implementation of the proposed method. The dataset comprises demographic data that can be utilized to forecast an individual's annual income as either over $50,000 or under $50,000. Each adult's demographic information elements, including age, work class, fnlwgt, education status, marital status, occupation, relationship, race, sex, hours per week, native country, and pay, are listed in each row. The individual's salary is displayed in the final column, indicating if their annual compensation is $50,000 or more. These exquisite qualities are appropriate for using the proposed technique.

G. Evaluation Metric

Four distinct metrics; Accuracy, Precision, Recall, F1-score, MAE, and MSE will be used to evaluate the classifiers' performance. These assessment parameters were used in other to assess the proportion of correct classifications among all classifications thus, allowing the researcher select the finest performing model. The metrics can be stated as:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$MAE = \frac{|(y_i - y_p)|}{n} \quad (4)$$

$$\text{F1} = 2* \frac{\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \quad (5)$$

$$MSE = \sum_{i=1}^{n} (Y_i - Y_i)^2 \quad (6)$$

In this case, n represents the total number of observations in the data, yi denotes the observation's true value, and yi and yp represent the observation's predicted value.

IV. EXPERIMENTAL RESULTS

The proposed method is implemented using WEKA 3.9 that support the wekaDeeplearning4j package. The Weka wrappers for Deeplearning4j gives users access to filters and classifiers built with the Deeplearning4j library. Numerous layers, activations, loss functions, and other components can be used to construct deep neural networks. The choice was made based on its flexibility in execution artificial intelligence algorithms with the neural network toolbox. The software environment required includes windows operating system and WEKA version 4.0. For all the machine learning, WEKA was used to implement. The Minimum hardware requirements are 1.2 MHz, 2 Gig RAM, and 120 GB HDD. Using stratified sampling, the datasets are split 70% for training and 30% for testing to maintain the class distribution as much as feasible. Additionally, To account for the impact of features with varying scales, min–max normalization is used to normalize all datasets.

The datasets have the following features: age, work class, fnlwgt, education, education-num, marital status, race, sex, hours per week, hours gained, hours lost, and salary. PCA test was used to confirm if all the features are contributive equally to the overall variance or eliminate those variables that are unrelated to the salary prediction model. From the preprocessing stage, it was observed that some features such as the education, work class and hours of work have higher influence that the others feature such as the relationship, race and marital status. Figure 1 depicts the distribution of the datasets in relation to the salary class of the workers.

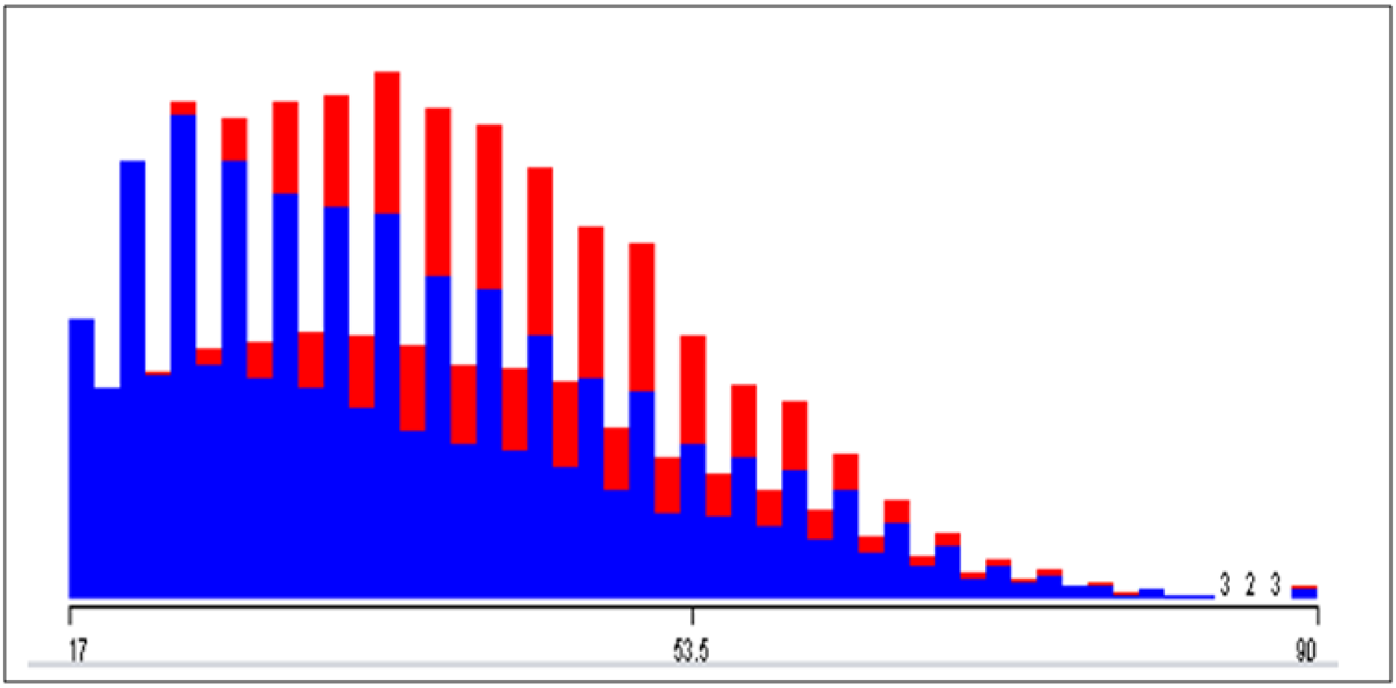

Fig 2 Workers Age in Relation to the Salary

From Figure 1 it can be noticed that the workers age has huge impact in relation to their salaries. It is noticed that the workers within the age of 17 years to 53.5 years have higher salaries than those within the age of 53.5years to 90 years irrespective of the occupation. Similarly, Figure 2 depict the marital status class distribution.

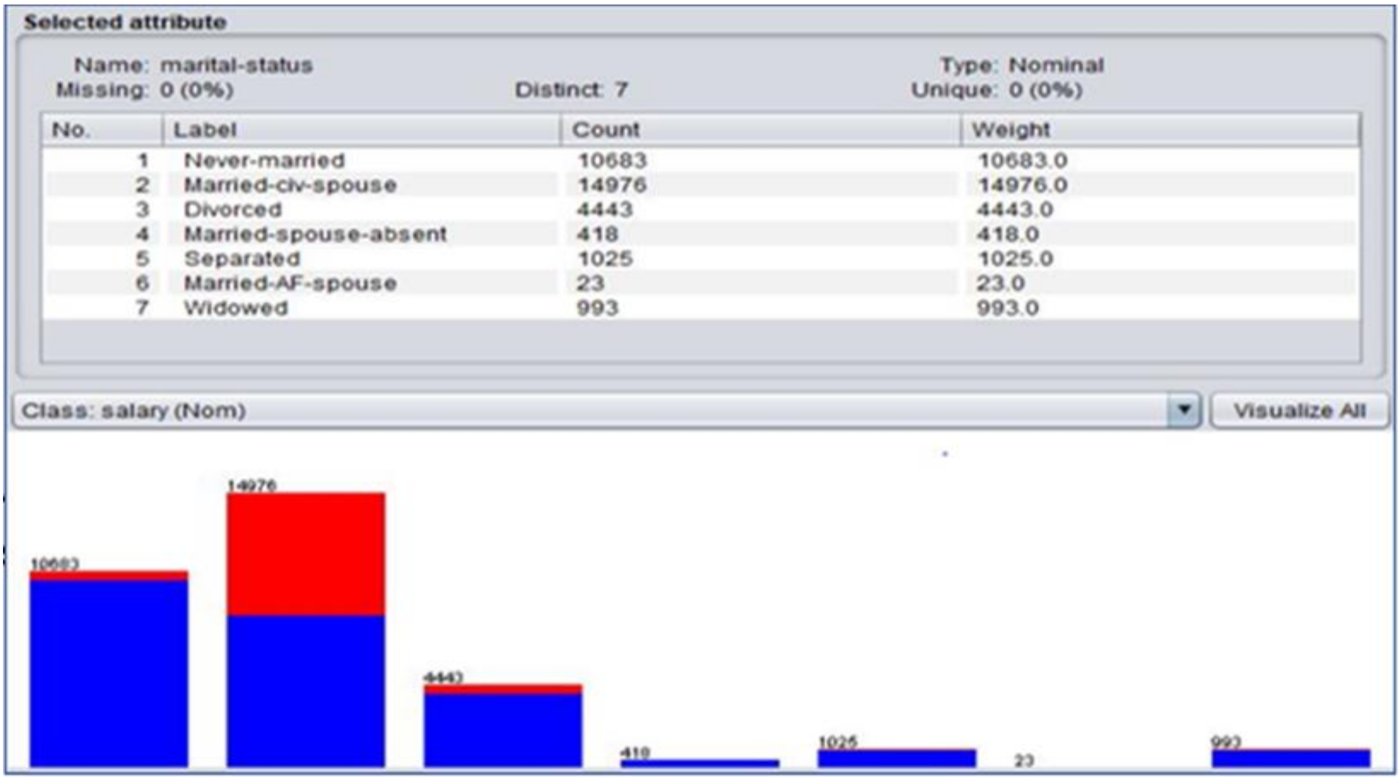

Fig 3 Marital Status Class Distribution

- Similarly, Figure 3 depict the different class of occupation and the respective salary class distribution.
- Similarly, Figure 4 depict the work class distribution.

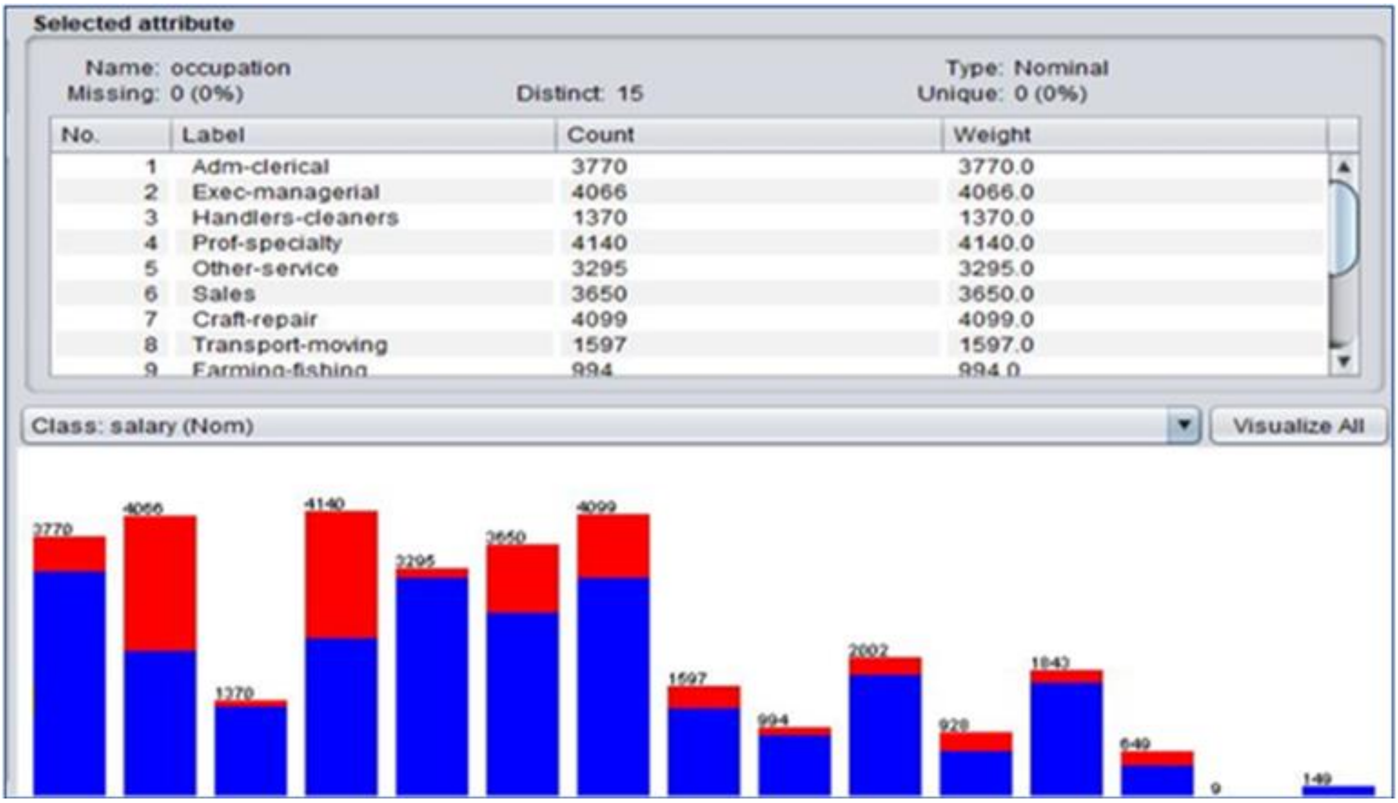

Fig 4 Different Class of Occupation

ISSN No:-2456-2165

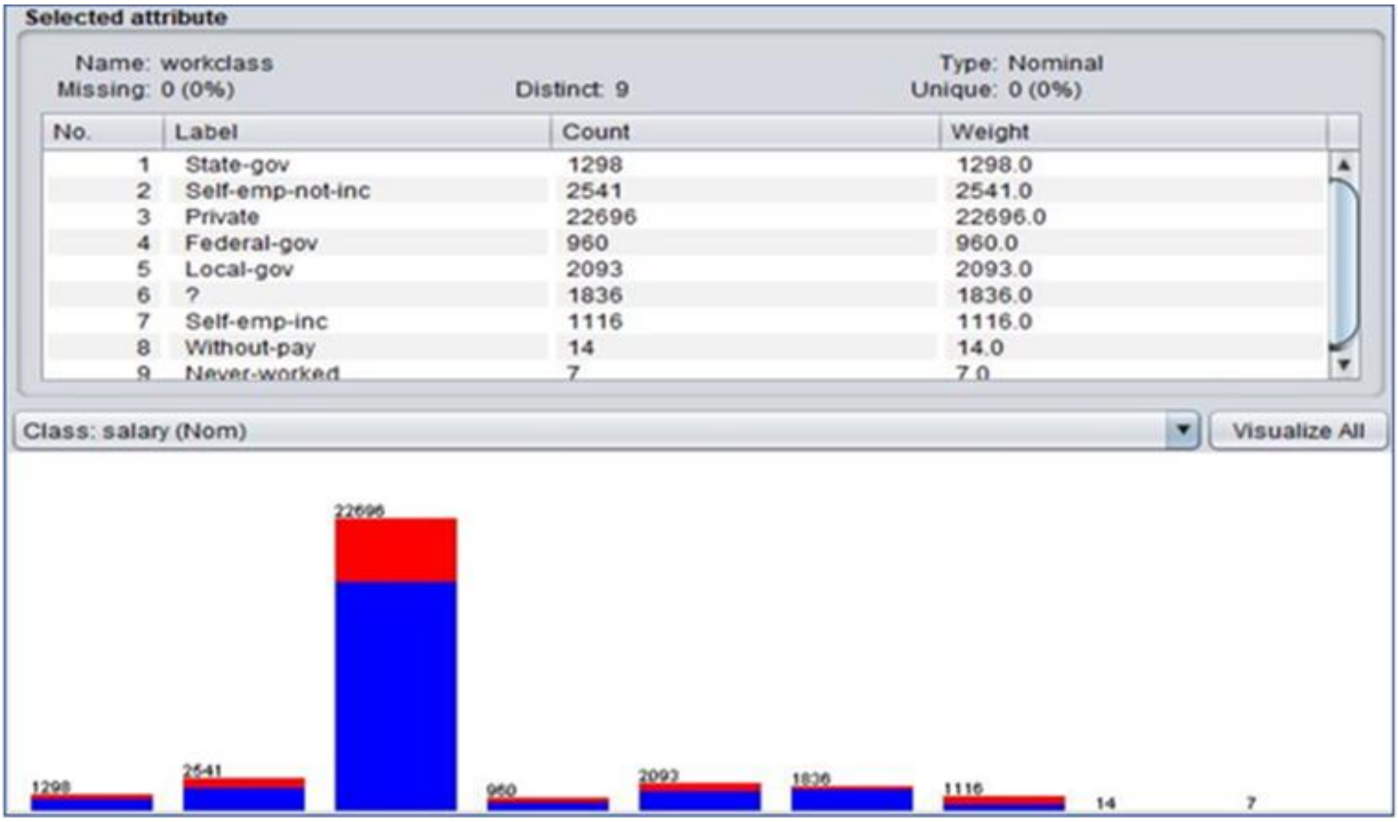

Fig 5 Work Class Distribution

- Similarly, Figure 5 depict the education class distribution

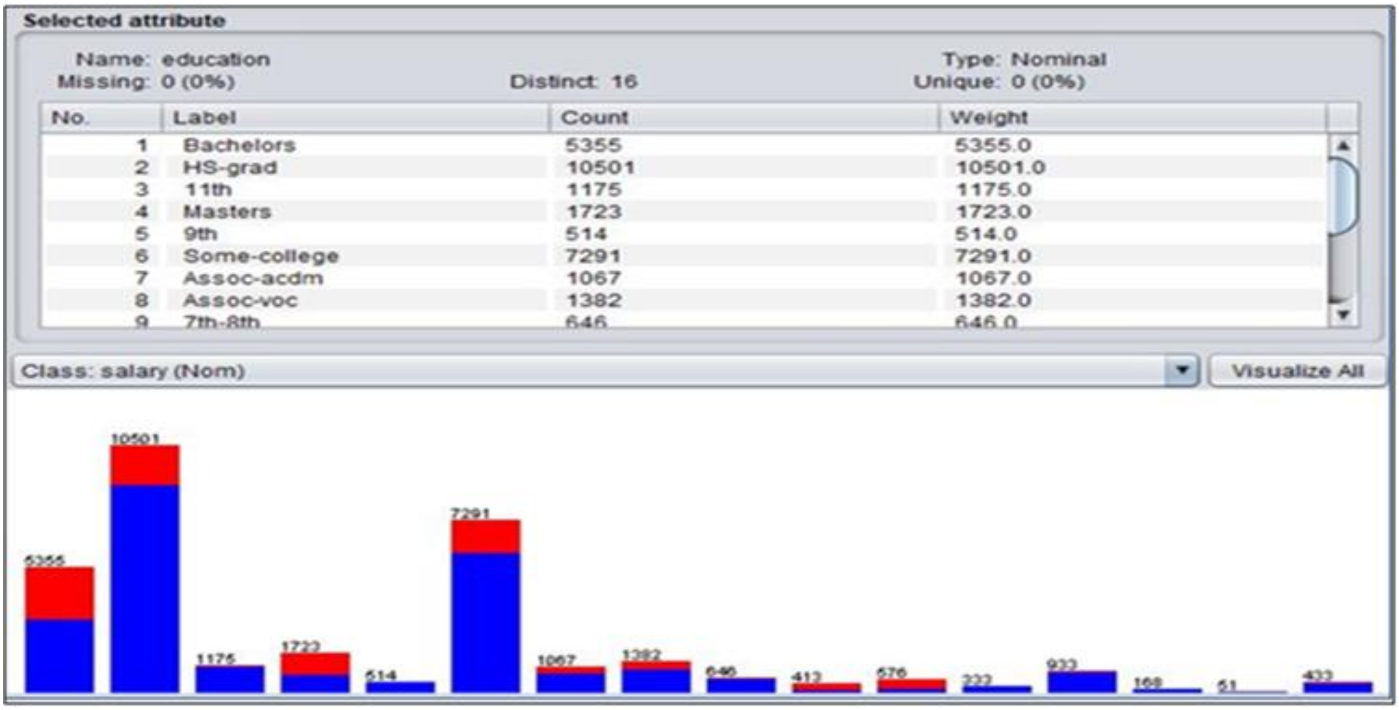

Fig 6 Education Class Distribution

Finally, Figure 6 depict the class distribution of the employees base on the features and determining factors. The distribution of the datasets was such that 24720 workers were classes under wage of less than 50k and 7841 were classified under wages of above 50k.

Figure 7 depict the WEKA implementation environment for the proposed DNN model. We used cross validation during the training phase and set the fold to 10. The parameter settings are depicted in Table 1.

ISSN No:-2456-2165

Table 1 Parameter Settings

Parameters	**Settings**
Bias Initialization	0.0
Learning Rate	0.001
Learning Rate Schedule	Constant Schedule
Minibatch	True
Optimization Algorithm	Stochastic Gradient Descent
Gradient Normalization Threshold	1.0
Training Workspace Mode	ENABLED
Weight Noise	Disabled

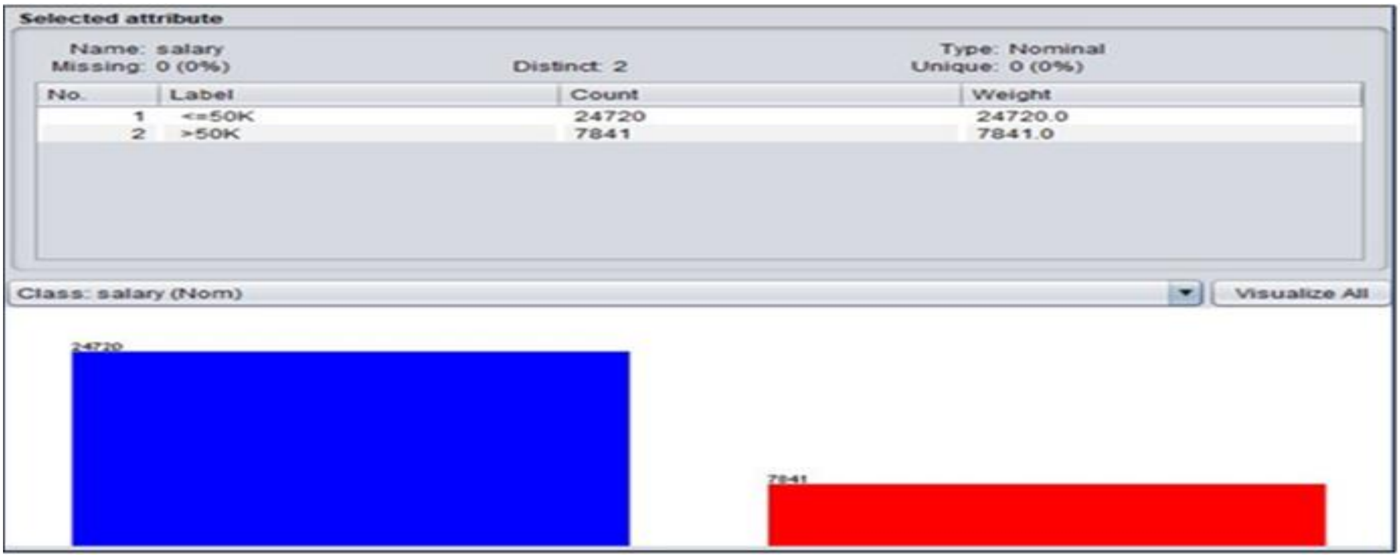

Fig 7 Class of Wages Among Worker

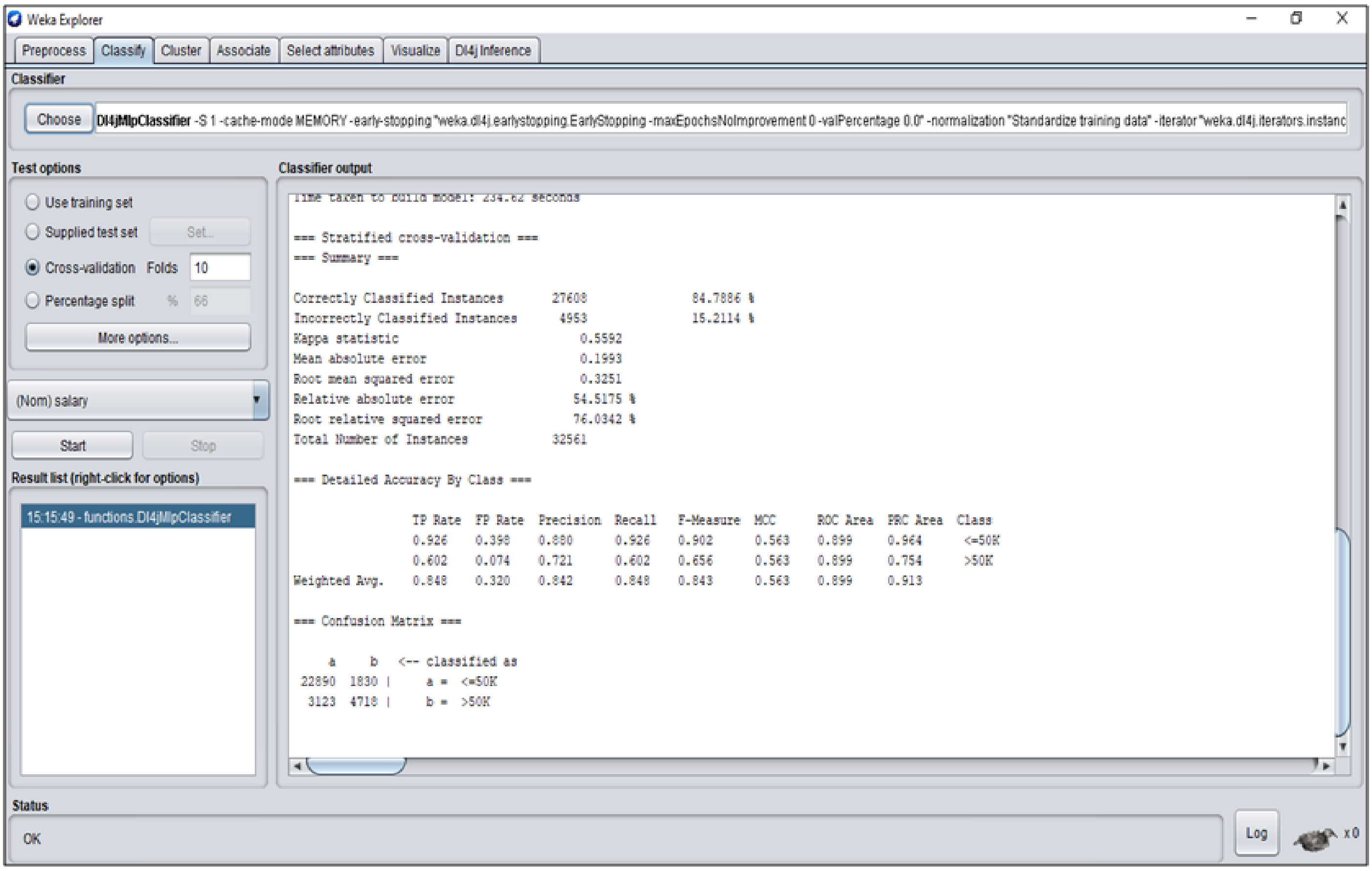

Fig 8 Depict the Implementation Environment

ISSN No:-2456-2165

From Figure 7, the result achieved by the each of the algorithms is depicted using the same datasets. The results are presented and analysis in the next subsection.

A. Results Presentation

This phase's objective is to ascertain whether the data that is been analyzed will be acceptable to flow through the network based on the results that is been gotten from the classification of the numerous categories. After implantation of the each of the model on the same salary dataset, we found the following results as shown in the Table 2.

Table 2 Overall Results for Binary Classifications of Wages

Algorithm	Accuracy	Precision	Recall	F-1	MAE	MSE
Proposed DNN	92.5	91.7	89.6	90.2	87.7	94.9
Decision Tree	79.9	75.9	75.9	86.3	80.2	89.6
Random Forest	84.8	84.2	84.8	84.4	53.9	76.4

Table 2 below shows the outcomes attain by the proposed system when compared with the benchmark paper. The table further shows the value obtained for both precision and recall from the simulation results gained after several iterations.

B. Classification Accuracy

This performance metric deals with the precise prediction made by the model. Figure 8 depict the performance reached by the proposed model against state-of-the-art methods.

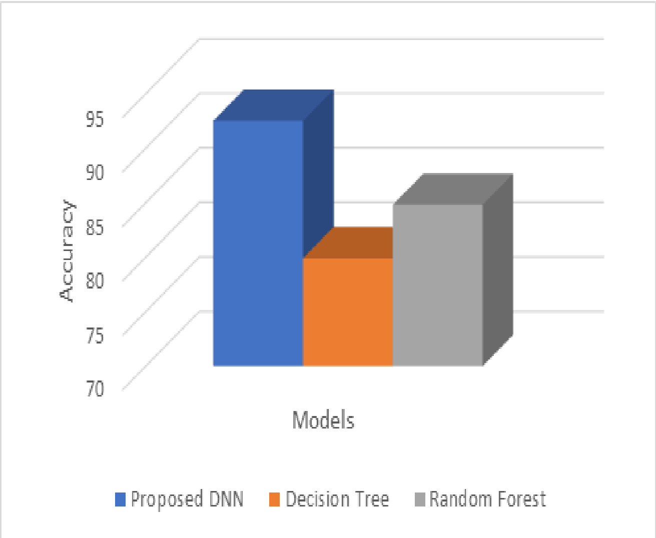

Fig 9 Classification Accuracy for all Method

From Figure 8 above, the proposed system achieved a improved classification accuracy of 92.5% as against the existing DT and RF algorithms, which attained 79.9% and 84.8% respectively. This suggest that the proposed model predicted the correct instances with higher accuracy compared to the state of the art.

C. Precision

Figure 9 depicts the performance achieved by the proposed method against the existing method for the precision.

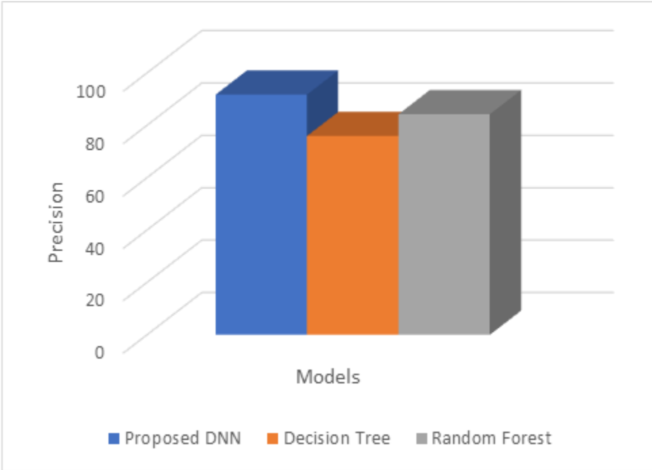

Fig 10 Precision for all Method

From Figure 9 above, the proposed system achieved improved precision of 91.7% as against the existing DT and RF algorithms, which attained 75.9% and 84.2% respectively. This illustrates how the proposed approach is better than the state of the art.

D. Recall

As stated earlier, the precision and recall help us additional understand how strong the accuracy shown holds true for a particular problem. In comparison to the current recall strategy, Figure 10 shows the performance attained by the proposed approach.

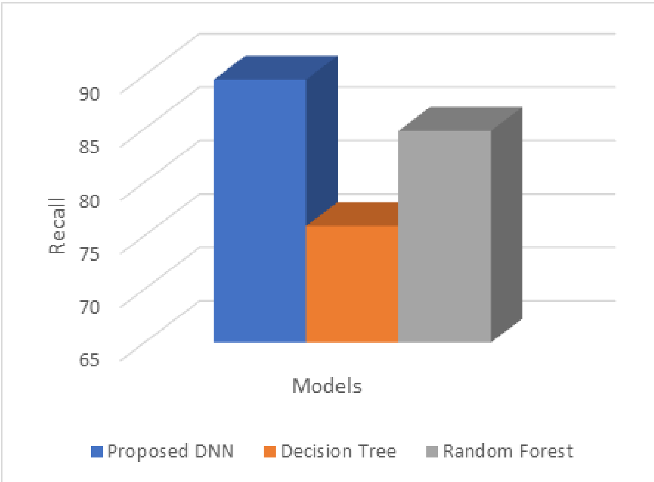

Fig 11 Recall for all Method

ISSN No:-2456-2165

From Figure 10 above, the proposed system achieved improved precision of 89.6% as against the existing algorithm, which are DT and FR with 75.9% and 84.8% respectively. Thus, the proposed model has further demonstrated its overall superiority in categorizing the instances of predicting salary as compared to the state of the art by reaching higher values in both precision and recall.

E. F-Measure

From Figure 11 above, the proposed system achieved improved precision of 90.2 % as against the existing algorithms, which are DT and RF with 86.3% and 84.4% respectively. Therefore, the proposed model has further demonstrated its overall superiority in classifying the workers' wage classes as opposed to the state of the art by reaching higher values in terms of precision, recall, and F-score.

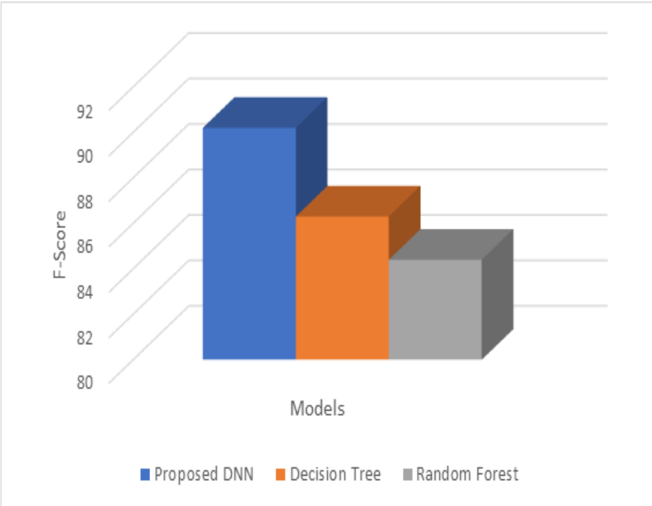

Fig 12 Performance Comparison with Existing Method for F-Measure

F. Results on MAE and MSE Performance

In statistics, Regression model performance is assessed using a statistic called mean absolute error (MAE). It can be defined as the mean absolute difference between the model's predicted values and the actual values of the data. Figure 12 depict the MAE achieved by each of the algorithms.

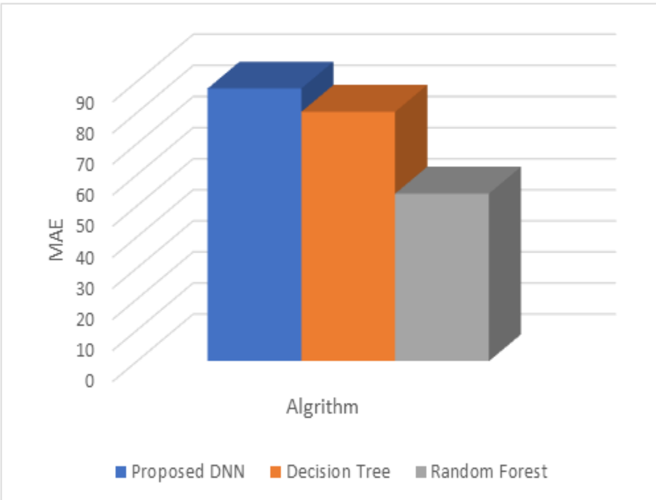

Fig 13 MAE for all Method

For the MAE and MSE, a predictor's performance improves with a decreasing value. Hence, from Figure 18, the proposed DNN model attains the highest prediction performance of 87.7% as compare with the DT and RF which achieves an MAE score of 80.2% and 53.9% respectively. Technically, this result means that the proposed model has prediction error of 12.3% which is less when compared to DT and RF which has prediction error of as much as 19.8 and 46.1% respectively.

Figure 13 depicts the prediction performance achieved by each of the algorithms used in the study.

Fig 14 MSE for all Method

As shown in Figure 14, the proposed DNN model attains a higher prediction performance of 94.9% compared with the DT and RF, which attain an MAE score of 89.6% and 76.4% respectively. Technically, this result means that the proposed model has prediction error of 5.1% which is less than that of DT and RF with prediction errors of 10.4 an 23.6% respectively.

Therefore, from the results presented and analyzed above, it may be noticed that the proposed deep learning attains the first place in all the cases followed by RF. Furthermore, it is also clear that the proposed deep learning attains the higher level of performance by obtaining the better scores in terms of decision support metrics. The result has further established the superiority of deep learning algorithm over conventional machine learning algorithms in salary classification and prediction task, which is dependable with the literature.

V. COCLUSION

This research employed PCA to aids preprocess and extract relevant information from the demographic data with high impact on prediction model. Thus, for salary prediction, the research uses a robust deep neural network approach (PCA-DNN) to increase prediction performance over existing machine learning classifiers like DT and RF. A better classification accuracy was attained using the proposed PCA-DNN algorithms in terms of accuracy, precision, recall and F-score. Additionally, the proposed DNN model attains higher prediction performance of 94.9% as compared to DT and RF which reaches an MAE score of 89.6% and 76.4% respectively. The outcome of the result implies that the proposed model has prediction error of 5.1%, which is less compared to DT and RF with prediction error of 10.4 an 23.6% respectively. This establishes the superiority of deep learning algorithm over conventional machine learning algorithms in salary classification and prediction task.

However, the limitation of this study is that the study uses only one dataset to evaluate the algorithms. This was due to the difficulty in accessing the wages datasets in Nigeria that is not freely accessible from online repository portals. In the future, we recommend the implementation of the proposed system on larger datasets. In addition, the proposed models can be used to additional classification tasks to guarantee the model's generalizability.

ACKNOWLEDGEMENT

We are grateful to our supervisors, Drs. B.I. Yau and F.U. Zambuk, for their guidance and expert support. We also wish to express our gratitude to other researchers in this field who have given us access to pertinent materials that have made our research successful.

REFERENCES

[1]. Das, S., R. Barik, and A. Mukherjee, *Salary prediction using regression techniques.* Proceedings of Industry Interactive Innovations in Science, Engineering & Technology (I3SET2K19), 2020.

[2]. Dutta, S., A. Halder, and K. Dasgupta. *Design of a novel Prediction Engine for predicting suitable salary for a job*. in *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*. 2018. IEEE.

[3]. Martín, I., et al., *Salary Prediction in the IT Job Market with Few High-Dimensional Samples: A Spanish Case Study.* International Journal of Computational Intelligence Systems, 2018. **11**(1): p. 1192-1209.

[4]. Srivastava, S., D. Sharma, and P. Sharma, *Comparing various Machine Learning Techniques for Predicting the Salary Status*. 2020, EasyChair.

[5]. Khongchai, P. and P. Songmuang. *Random forest for salary prediction system to improve students' motivation*. in *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. 2016. IEEE.

[6]. Khongchai, P. and P. Songmuang. *Implement of salary prediction system to improve student motivation using data mining technique*. in *2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*. 2016. IEEE.

[7]. Pawha, A. and D. Kamthania, *Quantitative analysis of historical data for prediction of job salary in India-A case study.* Journal of Statistics and Management Systems, 2019. **22**(2): p. 187-198.

[8]. Wang, Z., S. Sugaya, and D.P. Nguyen, *Salary Prediction using Bidirectional-GRU-CNN Model.* Assoc. Nat. Lang. Process, 2019.

[9]. Chakraborti, S., *A Comparative Study of Performances of Various Classification Algorithms for Predicting Salary Classes of Employees.* vol, 2014. **5**: p. 1964-1972.

[10]. Ajit, P., *Prediction of employee turnover in organizations using machine learning algorithms.* algorithms, 2016. **4**(5): p. C5.