Received: 24 February 2023 | Revised: 10 May 2023 | Accepted: 26 May 2023 | Published online: 13 June 2023

RESEARCH ARTICLE

Artificial Intelligence and Applications
2023, Vol. 00(00) 1–8
DOI: 10.47852/bonviewAIA3202795

BON VIEW PUBLISHING

Salary Prediction Model for Non-Academic Staff Using Polynomial Regression Technique

Samuel Iorhemen Ayua[1,*], Yusuf Musa Malgwi[2] and James Afrifa[3]

[1]*Mathematical Sciences, Taraba State University, Nigeria*

[2]*Computer Science, Modibbo Adama University, Nigeria*

[3]*Computer Science, Adamawa State Polytechnic, Nigeria*

Abstract: The idea of regression has increased rapidly and significantly in the machine learning domain. This paper builds a salary prediction model to predict a justifiable salary of an employee commensurate to the increase or decrease in exchange rate (XR) using polynomial regression (PR) techniques of degree 2 in Jupyter Notebook on Anaconda Navigator tool. Predicting a feasible salary for an employee by the employer is a challenging task since every employee has a high goal and hope as the standard of leaving increases without a corresponding increase in salary. This model uses a salary dataset from Taraba State University, Jalingo, Nigeria in building and training the model and XR dataset for the prediction of employee salary. The result of the research shows that since the distribution of the dataset was nonlinear and the major feature significant in determining employee's salary from the in-salary dataset was grade level and XR, this fully confirmed the use of PR algorithm. The research has immensely contributed to the knowledge and understanding of regression techniques. The researcher recommended other machine learning algorithms explored with various salary datasets and the potential applicability of machine learning fully incorporated in the financial department on the large dataset for better performance. The model performance was evaluated using R^2 scores accuracy and the value of 97.2% realized, indicating how well the data points fit the line of regression and unseen dataset in the developed model.

Keywords: polynomial regression, exchange rate, machine learning, compensation, linear regression, supervised learning, employee salary

1. Introduction

Machine learning is aimed at designing computers to learn by themselves without the explicit intervention of humans. It is a process, in which machines learn from knowledge gained and understanding of concepts or skills by studying the instruction or from experience (Archana et al., 2013). The learning process starts with the observation of data to recognize patterns in data for improved decision making (Deisenroth et al., 2020). They continue that the significant point in machine learning is data entry and training of the algorithms to find patterns that are capable to make future predictions based on new data (Deisenroth et al., 2020).

The inability of organizations to accurately predict and pay employees' salaries and compensate accordingly become a major challenge in Nigeria and has contributed to the continuous switch of employees from one organization to another. This problem leads employees to switch from one organization to another to get the expected salary, as the standard of living rises with a high exchange rate (XR) without a corresponding salary increase. These issues prompted researchers to propose a salary prediction system using a linear regression algorithm with a second-order polynomial transformation, five most relevant features, and a value of 76% R^2 score accuracy was obtained from the study (Lothe et al., 2021). In the same vein, another research on salary prediction was carried out using regression techniques (linear regression and polynomial regression (PR)) and the predicted salary was chosen from an x-y graph (Das et al., 2020). Bansal et al. (2021) compared the performance of the two regression techniques (multiple linear regression (MLR) and simple linear regression (SLR)) using a very small dataset to predict the salary of employees after certain years and to predict the price of the estate house. From the studies above, the researcher observed the accuracy of 76% R^2 considered not encouraging for feasible prediction of employee's salary. Secondly, chosen salary of an employee from the x-y graph may lead to an error in obtaining values from the graph if not carefully checked by an expert and lastly using a small dataset will not give a better picture of how the model will predict employee salary if exposed to a large dataset.

However, it is worrisome that employers cannot randomly provide employees with their expected salary; there is a need to design a system, which should measure the capability of every employee for the expected salary commensurate to the XR. The exact salary cannot be decided but predicted using certain datasets commensurate to the decrease or increase in XR. This research developed a model that will help in predicting employee salaries commensurate to the decrease or increase in XR, enabling the

*Corresponding author: Samuel Iorhemen Ayua, Mathematical Sciences, Taraba State University, Nigeria. Email: ayuasamuel@gmail.com.

organization to accurately compensate employees in Taraba State University, Jalingo, Nigeria.

Swanepoel et al. (2014) in their research state that compensation is a pay provided by an employer to its employees for services rendered (skills, time, and effort). This includes both fixed and variable pay tied to performance levels. They also continued that compensation is financial and non-financial extrinsic rewards provided by an employer for the time, skills, and efforts made available by the employee in fulfilling job requirements aimed at achieving organizational objectives.

Absar et al. (2010) reported that employee compensation is one of the major functions of human resources management. Compensation is important for both employers and employees regarding attracting, retaining, and motivating employees. Ray and Ray (2011) regarded compensation as important for employees working. According to Fu and Deshpande (2014), total compensation is the total of all rewards provided to employees in return for their services. Qasim et al. (2012) stated that monetary rewards play a major role in determining job satisfaction. Pay is one of the fundamental components of job satisfaction (Fu & Deshpande, 2014).

According to Lothe et al., (2021), PR machine learning technique takes defined features to estimate appropriate salary. In addition, future events are not known to anyone, so finding accurate data about the future is also difficult. However, to improve the accuracy of the salary prediction, a feasible R^2 score must be obtained when PR technique is used to determine the employee required salary. Given that it involves forecasting continuous numerical values, the prediction problem is a regression problem. According to Brownlee (2019), a learned program can propose new outputs given new inputs.

PR is also a statistical technique that predicts a continuous variable (response variable) taking into account the higher power of the predictor variable when the relationship between the predictor and the response is nonlinear. It is a special case of MLR because the higher power of predictors is treated as a new predictor and this makes it MLR (i.e. the higher power of x say x^2 is treated as another predictor and must have a different regression coefficient) (Das et al., 2020).

In this research, the technique used was the PR of supervised machine learning for salary prediction. This model helps in predicting employee salaries commensurate to the decrease or increase in XR and enabling the organization to accurately compensate non-academic staff of Taraba State University such that the mission and vision of the institution are achieved and in turn provide job satisfaction to employees. It increases the speed and accuracy of the proposed system since its patterns were derived from the training dataset that reduces the complexity of the traditional system to just a few features that offer better correlation to the target variable (salary). Given new data points that do not exist in the dataset, the model predicts the discrete salary range of employees based on the defined parameters. It benefits the community in terms of employment for obtaining a justifiable salary commensurate to the decrease or increase in the XR and helps the university to actualize its mission and vision since salary is one of the most important factors in job satisfaction that stimulate staff output. With the help of this machine learning model, predicted salary can easily be visualized which will help to draft company policies.

Below are the objectives of this research:

i. Carry out dataset creation and data preprocessing;
ii. Develop and train PR model for salary prediction using Python programming language and Jupyter Notebook in Anaconda Navigator cross-platform;
iii. Evaluate the model performance using k-fold cross-validation and performance accuracy metrics like R^2 score (coefficient of determinant);
iv. Validate the model performance using the test dataset;
v. Make salary predictions from the model using new data points.

2. Literature review

According to Shiqi (2023), due to COVID-19 effect on the economy, employees experience pressure in searching for a job that has an encouraging salary due to a decrease in corporate earnings, as a result of that, the researcher proposed three algorithms (MLP model, RF model, and GBDT model) to predict the salary for job recruiters. After analyzing the performance of each model, GBDT algorithm outperforms the other and was chosen such that employers can publish job information to ease job seekers for better positions.

As stated by Kablaoui and Salman (2022), salary is the primary source of job satisfaction, and finding a justifiable salary for employees will have greater benefits to both employees and employers. Hence, machine learning and data science have made salary prediction more simplified and realistic to estimate the expected salary. In this research, the researcher uses linear regression, random forest, and neural network models for the prediction of employee salary, where a 20,000-salary dataset from the USA was used to build the models. The output of the neural network model outperforms the other models with an accuracy of 83.2%.

This research uses random forest classification model for the prediction of a person's pay for both fresher and old employees. This model uses a universities placement database for model training, having a total number of 49,000 rows and 8 features (Shankar and Malik 2022).

According to Demir et al. (2022), having an insight into the salary range of job positions is paramount for employees and employers. They use feature-engineering techniques on the dataset obtained from Turkey for improving the training data and then subject the improved data to the models for the training. The researcher deduces that artificial neural networks and XGBoost achieve the highest success.

Bansal et al. (2021) examined SLR and MLR effectiveness in predicting both home prices and employee wages. According to their findings, MLR was more effective in prediction than basic linear regression in both circumstances.

Lothe et al., (2021) developed the salary prediction model utilizing a linear regression method with a polynomial second-order transformation. From their research, five features were found to be the most important. The system's output is calculated using an algorithm by contrasting it with other algorithms using common metrics such as classification accuracy, F1 score, Receiver Operating Characteristics (ROC) curve, and precision-recall curve. They kept using the fundamental model and discovered the best strategy with a maximum accuracy of 76%. They then advised the researcher to include a graphical user interface in the feature work system such that the trained model is saved and used again.

Also, according to Das et al. (2020), several regression techniques were used to forecast employee salaries, and later advise that in the future, k-nearest regression used to assess how an employee's salary has evolved over the employee's career advancement.

In Japan, Wang et al. (2019) employed web scraping techniques to collect the data and natural language processing to construct features from the scraped job advertisements to forecast the incomes of specific job adverts. Their key problems included related issues of natural language processing and training of one model to be as effective as possible rather than looking for the best model.

Dutta et al. (2018) concentrated on the issue of forecasting pay of job ads in which salary is not specified and attempted to assist fresh graduates in projecting likely salaries for various organizations in various areas. The model was able to predict an exact value using ADZUNA dataset. They employ decision models such as the decision tree and the ensemble model. The outcomes were impressive, with great precision. In addition, the researcher recommended the use of REST API for deployment of the system as a web based application for easy access and usage.

Using the deep learning technique, Viroonluecha and Kaewkiriya (2018) created a wage-predicting model to forecast the monthly salaries of Thailand employees. The data acquired via a popular online job site were over 1.7 million users. The model building makes use of the dataset from the first 5 months of the year 2018. By comparing the efficacy of deep learning with other algorithms like random forest and gradient boost trees, the approaches for selecting features applied to deep learning. The optimal outcome of integrating feature selection with transfer learning in R^2 was 0.462 with the 15.37-second fastest runtime.

Using high-dimensional samples with features of about 2,000, Martin et al. (2018) worked on estimating the salaries of certain positions in the IT labor market. They used K-means clustering to forecast the wage range of specific job posts and simple regression algorithms to forecast the continuous quantitative value of the compensation as a baseline while comparing methods like random forest and support vector machines to the baseline.

Khongchai and Songmuang (2016) suggested a forecasting model utilizing the decision tree algorithm with seven features to forecast employees' salaries and the graduated student's highest salary who share common attributes with the users. The 13,541 records were used to test the efficiency of the system, and the overall resultant accuracy was 41.39%.

Supervised learning is another method of machine learning that was developed by Hetal & Amit (2012). The goal variables or correct solutions of this method are given to the machine together with the learning data; the computer makes an effort to identify patterns in the data used to make generalized predictions that will be useful for any unknown data. Supervised learning algorithms use the input data to produce the abovementioned forecast pattern (Hetal & Amit, 2012). The most important objective of supervised learning is to locate a function that generates reliable predictions based on test data (Bekkerman et al., 2012).

The training and testing data are the two forms of data used in the processes of constructing and measuring models. The training data are utilized to create a model, fit the data, and use the model to discover patterns in the data, while the performance and accuracy of the model were verified using the testing data. The primary objective is to fit the model in a way that ensures the dependent features appropriately correspond to the independent future (observation) (James et al., 2017). They go on that classification and regression methods are the two most popular supervised learning techniques. Regression attempts to forecast infinite numerical continuous outputs, whereas classification approaches seek to categorize the predictions into already established classification output groups (Bhuvaneswari & Sarma, 2013).

Correlation and its counterpart regression are the most prevalent methods for investigating the relationship among variables. The perseverance of correlational research is to uncover and quantify correlations between variables in a population (Leedy & Ormrod, 2010). (Curtis et al., 2016). Their efforts focused on researching inheritance, which led to the circumstances for developing regression. Correlation is the measure of the linear relationship that is present between two parameters (Curtis et al., 2016). Furthermore, it is one of the widely utilized techniques for establishing the magnitude and trend of a connection between two variables. This substantiates the claim of the association between nutrient intake (N) and height increase (H) in children observed over 10 years. It established that youngsters who feed on food that are more nutritious grow faster than those who feed on less nutritious food. It highlights how an influence on one variable translates to a change in the value of another variable. They further explain that the relationship is not causal, meaning that a change in one variable does not imply a change in another. A correlation investigation cannot, in general, deduce a causation impact. Generally, a regression can be characterized as a set of approaches for estimating relationships. Linear regression uses the relationship between independent and dependent variables to fit a line of regression, which then be used to predict the dependent attributes based on the independent attributes (Han et al., 2011).

After preprocessing and cleaning, the data are ready for the modeling and commencement of the parameter-tuning phase. To generalize the predictions for a new circumstance, Alpaydin (2014) states that the model selection comprises choosing the unbiased model to forecast as-yet-unobserved observations. As a result, it is essential to consider the anticipated output type while choosing the model. Virtually every system has hyperparameters that ought to modify as claimed by Archana et al., (2013). The algorithm searches for the optimum potential parameter combinations at each iteration of the process, which is iterative. If the model is categorizing the result or forecasting an endless numeric result, the assessment of the model is different. The researcher emphasizes the assessment of regression models as they attempt to forecast an infinite numerical goal variable, known as wage (Friedman et al., 2010). From their research, both the root square error and the R^2 metric were reported in the MLR. In the root mean square error approach, the errors and residuals are squared, added, and their means computed; before taking the square root, this method is employed in the regression model for performance measurements (Friedman et al., 2010).

To overcome the problem of sampling, k-fold cross-validation provides a mechanism in which the M dataset is randomly partitioned into sub-datasets (K) (Benner, 2020). This K dataset was created from the M dataset. In addition, after the model was trained with K minus one, it was assessed using the remaining dataset from the K sub-dataset. The iterative procedure for the model training and testing continues until each new dataset served as a k-1 testing set in the built model (Kuhn & Johnson, 2013; Alpaydin, 2014).

3. Research methodology

The proposed method for salary prediction is shown below:

Step 1: The salary dataset was collected, preprocessed, cleansed, and transformed.

Step 2: The data points corresponding to the salary data of staff will be plotted on the graph to illustrate the distribution and behavior of the features and extract relevant features.

Step 3: The processed dataset will be divided into 80% training and 20% testing.

Step 4: The 80% training dataset will be used for the model building and training using the PR technique in Jupyter Notebook.

Step 5: The built model will be evaluated.

Step 6: The built model will be validated using a 20% testing dataset.

Step 7: Analyze the monthly XR dataset to compute the benchmark exchange rate (BXR).

Step 8: To predict the salary of employees commensurate to the XR using the built model, the computed BXR and the current government-approved XR will be supplied at runtime.

Step 9: After then, through the built model, we predict the salary of employees with or without consideration of the XR.
Step 10: Also, we can predict an employee's future salary with or without consideration of the XR.

3.1. Study area, data source, and size

The area of this research is Taraba State University, Jalingo, Nigeria, where records of non-academic staff of the university salary data were collected in an Excel format.

The potential source of data used to undertake this research was from Bursary Department Taraba State University, Jalingo, Nigeria with a total number of 487 data points totaling 16 features and the monthly average XR dataset obtained from (source: www.cbn.gov.ng/rates/exrate.asp) 2010 to 2021 to compute and obtain the BXR.

3.2. Design, analysis, and development tool

The research built a salary prediction model that predicts a justifiable salary of an employee commensurate to the increase or decrease in the XR. The study makes use of a PR technique of degree 2 in Jupyter Notebook on Anaconda Navigator toolkit.

The generic equation of polynomial regression of degree 2 is shown below:

$$Y_1 = M_0 + \beta_1 X + \beta_2 X^2 + \varepsilon \tag{1}$$

where

X is the independent feature (grade level) transformed that offered a better correlation to the predicted variable (salary);

X^2 is the new higher power of independent feature (grade level) for the polynomial transformation of degree two that offered a better correlation to the predicted variable (salary);

ϵ is the random error component reflecting the difference between the observed value and the predicted value;

M_0 is the intercept term of the model, which tells the expected value of the predicted value Y_1 when X is zero;

β_1 is the linear effect parameter;

β_2 is the quadratic effect parameter;

Y_1 is the response or the predicted value (salary without XR).

The equation below finally predicts the salary of employees to the fluctuation in XR

$$Y_2 = (NXR \div BXR) \times Y_1 \tag{2}$$

where

NXR is the current new exchange rate approved by the CBN from dollar to naira;

BXR is the benchmark exchange rate within an interval of 12 years from 2010 to 2021;

Y_1 is the predicted salary without considering XR;

Y_2 is the predicted salary with consideration of XR.

3.3. The proposed model

In this research, Figure 1 describes the processes that are involved in building the proposed model from the dataset collection and preprocessing to the simulation of the model.

4. Results and discussion

To make the collected data compatible with Python libraries when uploaded into the data analysis tool (Jupyter Notebook), the dataset was cleaned and properly formatted as a comma-separated value file. With the help of data preprocessing (the replacement of missing values,

Figure 1
Polynomial regression architectural model

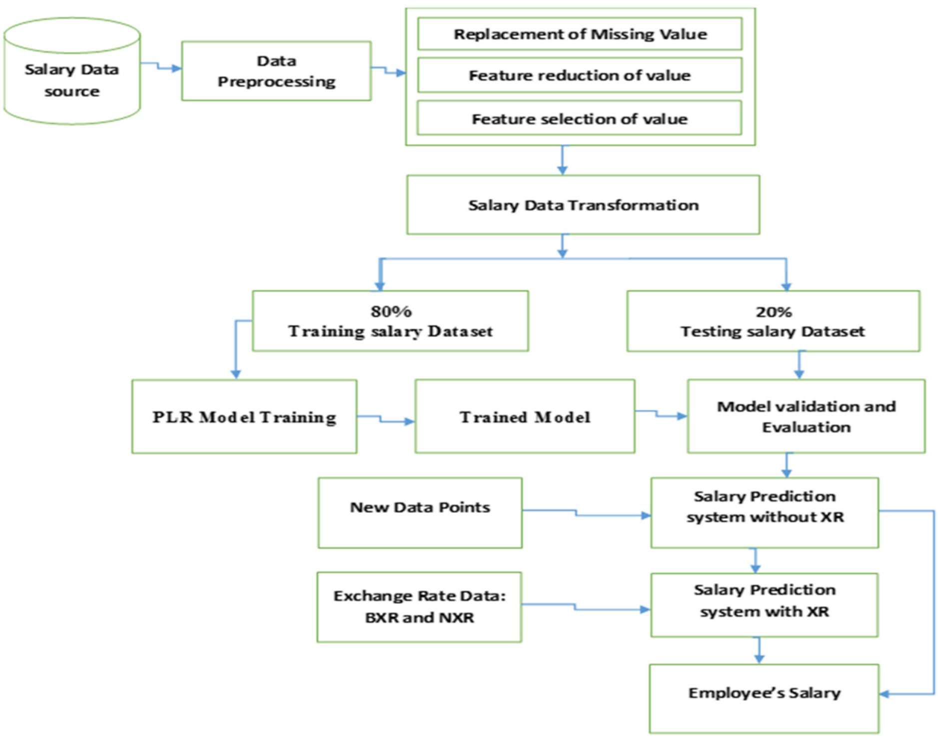

Table 1
The uploaded salary dataset

Qualification	Experience	Grade level	Step	Salary
1	6	2	10	24,759.170
3	4	4	1	29,458.330
2	12	3	6	31,091.670
4	5	7	5	101,727.400
4	29	7	10	89,434.750
1	4	1	1	24,420.830
2	5	3	4	29,036.420
4	4	7	9	114,020.000
4	9	8	6	122,072.900
4	4	7	4	98,654.250

Table 2
10-fold cross-validation

Number of predictions
0.955
0.965
0.986
0.977
0.976
0.973
0.977
0.955
0.979
0.973

feature selection, and feature reduction of values), 487 datasets totaling 16 features were reduced to four independent features, and the target variable (salary) is shown in Table 1. When the scattered plot was constructed to study the behavior of the features and its correlation to the target variable, grade level was chosen because it exhibits polynomial distribution (nonlinear), as shown in Figure 2.

The dataset was divided into 80% training and 20% testing using the train_test_split from sklearn.model_selection package in Python programming language. The researcher uses PR technique to build the salary prediction model using 80% of the training dataset on the jupyter notebook in cross-platform Anaconda Navigator toolkit, where patterns were obtained and used to predict the salary of non-academic staff.

After the training process, the performance of the trained PR model was evaluated employing k-fold cross-validation from sklearn.model_selection and r2_score from sklearn.metrics package in a Python programming language and the value of 97.2% R^2 score accuracy was obtained, which describes how well the data points fit the regression line. In this research, the number of iterations or k-fold cross-validation on the training dataset was $k = 10$ on which the built system was trained and within each fold tested as shown in Table 2, after which the test dataset was used to validate the model performance over unseen data.

R^2 Accuracy = Number of predictions/Number of k-folds

R_2 Accuracy = 97.2%

Table 3 shows the coefficients of the linear effect parameter ($\beta 1$), quadratic effect parameter ($\beta 2$), and the intercept (M0) of the PR degree two after the trained model.

Table 3
Summary of coefficients values and intercept from the trained model

Indicators	Coefficients values obtained from the model
Intercept (M_0)	32,167.443
linear effect parameter (β_1)	−4987.796
quadratic effect parameter (β_2)	1797.024

The researcher identify two approaches for the salary prediction after the model was trained. The first uses the built-in Python libraries and packages for the salary prediction after model training and/or secondary to obtain all the necessary coefficients to form the PR equation used for the salary prediction. In this research, both approaches were adapted since the second approach is the subset of the first.

The prediction system through the Python libraries can predict the salary of employees without XR and can obtain patterns or coefficients generated by the PR technique with degree 2 to form a prediction equation that predicts the salary of employees without XR. The prediction process considering XR as varying factors (NXR, and BXR) as stated in equation 2 uses the result obtained in the salary prediction process without XR as stated in equation 1 to predict the salary of employees commensurate to the increase or decrease in XR. In this process, the NXR was identified based

Figure 2
Scattered plot of salary against independent features

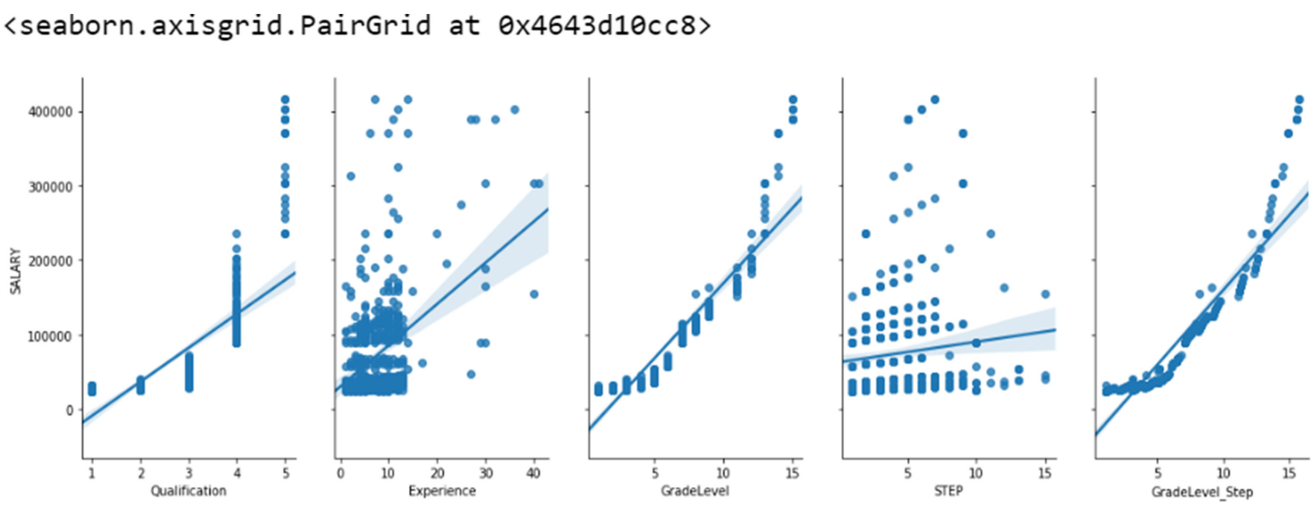

Table 4
Yearly average exchange rate of dollar to naira for 12 years

Year	Yearly exchange rate	Yearly AVG exchange rate
2010	1836.78	153.065
2011	1911.83	159.319
2012	1930.29	160.858
2013	1949.57	162.464
2014	2057.37	171.448
2015	2673.32	222.777
2016	4474.37	372.864
2017	4745.04	395.420
2018	4341.76	361.813
2019	4314.31	359.526
2020	5204.37	433.698
2021	5327.26	443.938

Table 5
Current and computed benchmark exchange rate

Year	Exchange rate	Value
2010–2021	Benchmark exchange rate	283.099
January, 2023	Current exchange rate	455.060

on the current government-approved market price of dollar to naira (445.06). And the BXR was computed using the past records of the exchange rate of dollar to naira from (2010 to 2021) as shown in Table 4

BXR considering 12-year interval = Sum of total yearly AVG exchange rate/Number of years interval

BXR = 283.0990972 naira

Table 5 shows the value of the current XR and BXR, which was used in the built model for salary prediction commensurate to the increase or decrease in XR.

The PR model predicts the employees' salary (Y1) without XR or (Y2) with XR of non-academic staff for decision-making by the management, bursary, and finance department of the university. Figure 3 shows the transformed equation of PR of degree 2 after the respective coefficient values were obtained from the trained model, which also forms the pattern that serves as a model for the salary prediction.

The 20% test dataset was subjected to the trained model to validate how well the model learn from the training dataset and the pattern generated, and the result of the predicted salaries using the test dataset is shown in Table 6.

The system predicts the salary of employees without consideration on the exchange rate as shown in Table 4 and/or predict the salaries of employees commensurate to the increase or decrease in exchange rate as shown in Table 6. Furthermore,

Figure 3
Transform PR model equation with patterns

The Model Equation/Pattern without Exchange Rate

Y_1 = (32167.44285423 - (4987.79575934 *float (GradeLevel_Step)) + (1797.02432065 * float (GradeLevel_Step) **2))

The Model Equation/Pattern with Exchange Rate

Y_2 = (float (NXR) / float (BXR)) * Y_1

Table 6
The actual and predicted salary of employees considering XR

Grade Level_ Step	Actual salary	Predicted salary	Salary with XR	Salary difference
2.2	25,614.667	29,891.890	48,050.455	−22,435.788
7.1	89,434.750	87,342.089	140,400.194	−50,965.444
1.3	25,812.917	28,720.279	46,167.121	−20,354.204
3.1	25,953.500	33,974.680	54,613.437	−28,659.937
7.7	107,873.667	100,306.987	161,240.940	−53,367.273
7.2	92,507.917	89,413.054	143,729.218	−51,221.301
6.2	60,356.833	70,320.724	113,038.782	−52,681.949
7.6	104,800.500	98,056.320	157,623.049	−52,822.549
2.4	27,325.667	30,547.593	49,104.482	−21,778.815
5.7	44,734.083	62,122.327	99,860.066	−55,125.983

Figure 4
Plot variation between the actual and the predicted salary values for the test dataset

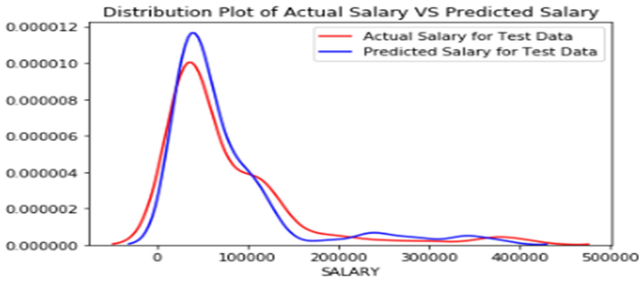

Figure 5
Model result after polynomial transformation and prediction

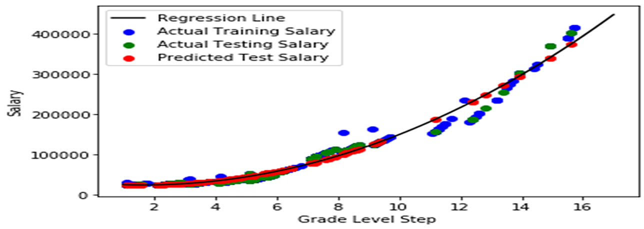

Figure 5 shows the significant predicted value and the distribution plot of both the actual training, actual testing, the predicted salary data points, and the regression line that fits the model using polynomial of degree 2 on the test dataset.

To predict employee salary from the model considering new data points, the values of GradeLevel_Step, NXR, and BXR are needed at runtime, after which the model returns the result as shown in Table 7.

Table 7
Salary prediction using new data points considering XR

GradeLevel_Step	NXR	BXR	Salary with-out XR	Salary with XR
14.5	455.060	283.099	337,668.768	542,793.986

5. Conclusion

Predicting the salary of employees using the PR technique had been done in the past, but such an outcome only predicted the salary of employees without considering the decrease or increase in the XR. Also, 76% R^2 score accuracy obtained from the study of Lothe et al. (2021) was considered not feasible for salary prediction, lastly chosen salary of an employee from the x-y graph according to (Das et al., 2020) may lead to error obtaining values from the graph if not read by an expert.

Predicting a feasible salary for the employee by the employer is a challenging task since every employee has a high goal and hope as the standard of leaving increases without a corresponding salary increase. This research developed a model to predict the justifiable salary range of employees with consideration of the increase or decrease in XR using PR techniques of degree 2 in Jupyter Notebook on Anaconda Navigator tool. The system is light and easy to use by even non-expert users. Following the objective of the study, the dataset feature was identified, selected, and preprocessed, and the model was built, trained, and tested, and a reasonable performance R^2 score accuracy of 97.2% higher compared to the 76% obtained, which tells how well the data point fit the regression line and how the model responded to unseen data points. Lastly, the predicted salary of employees was displayed for easy visualization both on browsers and even exported as an Excel file for easy identification.

Recommendations

This study recommends the application of the built model in all finance or bursary departments in predicting a justifiable salary of non-academic staff, in addition, to the use of a large dataset for the model building and training.

Acknowledgements

To begin with, the authors are grateful to the God for leading them toward this project path and for his ever-present help in times of their need. In addition, the authors thank all who worked on actualizing this paper's success.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

References

Absar, M. M. N., Azim, M.T., Balasundaram, N., & Akhter, S. (2010). Impact of human resources practices on job satisfaction. *Evidence from Manufacturing Firms in Bangladesh*, *62*, 31–42.

Alpaydin, E. (2014). *Introduction to Machine Learning. Adaptive Computation and Machine Learning*. The MIT Press.

Archana, C., Raj, K. & Savita, K. (2013). Machine learning classification techniques: A comparative study. *International Journal on Advanced Computer Theory and Engineering (IJACTE)*, *2*(4), 2319–2526. Retrieved from: https://krishi.icar.gov.in/jspui/bitstream/123456789/10464/1/ArchanaMalwaPaperinJournal.pdf.

Bansal, U., Narang, A., Sachdeva, A., Kashyap, I., & Panda, S. P. (2021). Empirical analysis of regression techniques by house price and salary prediction. *IOP Conference Series: Materials Science and Engineering*, *1022*, 012110.

Bekkerman, R., Bilenko, M., & Langford, J. (2012). *Scaling up Machine Learning*. Cambridge University Press.

Benner, J. (2020). Cross-Validation and hyperparameter tuning: How to Optimise your machine learning model. Retrieved from: https://towardsdatascience.com/cross-validationand-hyperparameter-tuning-how-to-optimise-your-machine-learning-model-13f005af9d7d

Bhuvaneswari, E., & Sarma, V. R. (2013). The study and analysis of classification algorithm for animal kingdom dataset. *Information Engineering*, *2*(1), 6–13.

Brownlee, J. (2019). A tour of machine learning algorithms. Retrieved from https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms.

Curtis, E. A., Comiskey, C., & Dempsey, O. (2016). Importance and use of correlational research. *Nurse Researcher*, *23*(6), 20–25.

Das, S., Barik, R., & Mukherjee A. (2020). Salary prediction using regression techniques. *SSRN Electronic Journal,* pp. 1–5.

Deisenroth, M. P., Aldo, F. A., & Soon, O. C. (2020). Mathematics for machine learning: Cambridge university press. *Association for Information Systems*, *19*(1), 6.

Demir, S. I.O., Ulke, U., Serhan, F. D., & Keziban, G. O. (2022). Salary prediction via sectoral features in Turkey. In *International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*. pp. 1–6, https://doi.org/10.1109/INISTA55318.2022.9894130.

Dutta, S., Halder, A., & Dasgupta, K. (2018). Design of a novel Prediction Engine for predicting suitable salary for a job. *Fourth International Conference on Research in Computational Intelligence and Communication Networks* (ICRCICN), pp. 275–279. https://doi.org/10.1109/ICRCICN.2018.8718711

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1.

Fu, W., & Deshpande, S. P. (2014). The impact of caring climate, job satisfaction, and organizational commitment on job performance of employees in a China's insurance company. *Journal of Business Ethics*, *124*(2), 339–349.

Han, J., Kamber, M. & Pei, J. (2011). Data Mining: concepts and techniques. *3rd Edition, Morgan Kaufmann Publishers, Burlington.*

Hetal, B., & Amit, G. (2012). A comparative study of training algorithms for supervised machine learning. *International Journal of Soft Computing and Engineering (IJSCE)*, *2*(4), 74–81.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *Introduction to Statistical Learning: with Applications in R*, 8th edn. New York, NY, USA.

Kablaoui, R., & Salman, A. (2022). Machine learning models for salary prediction dataset using python. In *International Conference on Electrical and Computing Technologies and Applications (ICECTA)*. pp. 143–147. https://doi.org/10.1109/ICECTA57148.2022.9990316.

Khongchai, P. & Songmuang, P. (2016). Improving students' motivation to study using salary prediction system. *International Joint Conference on Computer Science and Software Engineering* (JCSSE) pp. 1–6. https://doi.org/10.1109/JCSSE.2016.7748896

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*, *26*. New York: Springer.

Leedy, P.D. & Ormrod, J.E. (2010). Practical research: planning and design. *9th Edition, Pearson Education, Inc., Upper Saddle River,* 67.

Lothe, D. M., Prakash, T., Nikhil, P., Sanjana, P., & Vishwajeet, P. (2021). Salary prediction using machine learning, *6*(5), 199–202

Martin, I., Mariello, A., Battiti, R., & Hernández, J., A. (2018). Salary prediction in the IT job market with few high-dimensional samples: A Spanish case study. *International Journal of Computational Intelligence Systems*, *11*(2018), 1192–1209.

Qasim, S., Cheema, F. E. A., & Syed, N.A. (2012). Exploring factors affecting employee's job satisfaction at work. *International Journal of Social Sciences*, *8*, 31–39.

Ray, S. & Ray, I. A. (2011). Human resource management practices and its effect on employee's job satisfaction: A study on selected small and medium sized iron and steel firms in India. *Public Policy Admin*, *1*, 22–31.

Shankar, A., & Malik, M. (2022). Predicting person's pay using machine learning. In *4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 205–208, https://doi.org/10.1109/ICAC3N56670.2022.10074407.

Shiqi, Y. (2023). Automated employee salary prediction algorithm based on machine learning. *International Conference on Computer Vision, Application, and Algorithm (CVAA), 12613,* 126130Y. https://doi.org/10.1117/12.2673738.

Swanepoel, B. J., Erasmus, B. J., Schenk, H. W., & Tshilongamulenzhe, T. (2014). *South African Human Resource Management*. 4th edn. Juta, Cape Town: Theory and Practice.

Viroonluecha, P., & Kaewkiriya, T. (2018). Salary predictor system for thailand labour workforce using deep learning. *The 18th International Symposium on Communications and Information Technologies,* pp. 473–478.

Wang, Z., Sugaya, S., & Nguyen, P.T. (2019). Salary prediction using bidirectional-GRU-CNN model. Retrieved from: https://anlp.jp/proceedings/annual_meeting/2021/pdf_dir/F3-1.pdf.

How to Cite: Ayua, S. I., Malgwi, Y. M., & Afrifa, J. (2023). Salary Prediction Model for Non-Academic Staff Using Polynomial Regression Technique. *Artificial Intelligence and Applications* https://doi.org/10.47852/bonviewAIA3202795