A Project Report

on

Comprehensive Report on "Predicting New Employee Salary Based on Current Employee Attributes Using Multiple Regression Model

College of Computing and Mathematic

جامعة الملك فهد للبترول والمعادن

King Fahd University of Petroleum & Minerals

Submitted By

Abd-ur-Rahman

Abdibasid

Ahmad

Submitted To

Dr. Nassir Abbas

College of Computing and Mathematic

Department of Mathematics, King Fahd University of Petrolium & Minerals, Dhahran

DECEMBER – 2023

Abstract

In an era where employee turnover due to salary dissatisfaction is increasingly prevalent, the need for accurate and fair salary prediction models has become paramount for organizations. This project report presents a comprehensive study on predicting new employee salaries based on current employee attributes using a Multiple Linear Regression (MLR) model. The primary objective of this research is to develop a model that not only aligns with organizational goals but also aids in attracting and retaining top talent while ensuring financial sustainability.

The study employs a Multiple Linear Regression framework, analyzing variables such as years of experience and geographical distance from the workplace, while considering the age of employees. The methodology involves a rigorous statistical analysis to validate the model's assumptions, including linearity, independence, normality, and homoscedasticity, ensuring the reliability of the predictions.

The results indicate that years of experience and distance are significant predictors of salary, while age does not show a significant impact. The final model demonstrates a high predictive accuracy with a predicted R-square of 95.66%, indicating its effectiveness in salary forecasting. The model's adequacy is confirmed through various statistical tests, ensuring its applicability in real-world scenarios.

This study contributes to the field of human resource management by providing a robust and interpretable framework for salary prediction. It enables organizations to make informed, data-driven compensation decisions, thereby addressing the challenges of employee retention and financial management. The findings of this research offer a solid foundation for future studies and practical applications in the realm of compensation management.

Introduction

In the contemporary business landscape, the strategic management of employee compensation has emerged as a critical factor influencing organizational success and employee satisfaction. The ability to accurately predict salaries plays a pivotal role in this context, balancing the need to attract and retain talent with the imperative of financial sustainability. This project report delves into the development of a sophisticated salary prediction model using Multiple Linear Regression (MLR), aimed at addressing the challenges faced by organizations in compensation management.

Background Information on Salary Prediction

Salary prediction is an essential aspect of human resource management, directly impacting employee motivation, job satisfaction, and retention. In an era marked by rapid job mobility and a competitive talent market, organizations are increasingly recognizing the importance of offering fair and competitive compensation packages. Salary prediction models are instrumental in achieving this goal, as they provide a data-driven approach to determining compensation based on various employee attributes. These models not only help in maintaining internal equity and external competitiveness but also ensure adherence to budgetary constraints and financial planning.

Statement of the Problem

One of the primary challenges in human resource management is the high turnover rate attributed to salary dissatisfaction. Employees often seek opportunities elsewhere when they perceive their compensation as inadequate or unfair compared to market standards or their contributions. Traditional methods of salary determination, which may rely on subjective assessments or outdated market data, often fail to meet the dynamic needs of the modern workforce. There is a pressing need for a more analytical and objective approach to salary prediction, one that can adapt to changing market conditions and individual employee characteristics.

Purpose of the Study

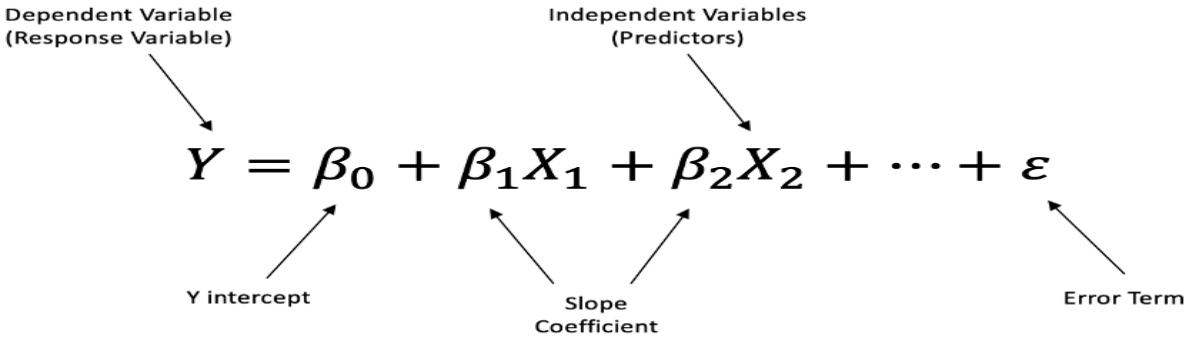
The purpose of this study is to develop and validate a Multiple Linear Regression model for predicting employee salaries. By incorporating various employee attributes such as years of experience, geographical distance, and age, the model aims to provide a comprehensive and accurate tool for salary determination. This study seeks to bridge the gap between employee expectations and organizational capabilities, thereby reducing turnover rates and enhancing employee satisfaction. Additionally, the model aims to assist organizations in strategic decision-making regarding compensation policies, ensuring they are both equitable and financially viable.

Scope of the Study

The scope of this study encompasses the development and analysis of a Multiple Linear Regression model tailored to predict employee salaries in a corporate setting. The research focuses on identifying and quantifying the impact of specific employee attributes on salary levels. While the study primarily considers years of experience and geographical distance as predictors, it also explores the relevance of age as a potential factor. The scope includes a thorough examination of the model's assumptions, its statistical validation, and its applicability in real-world scenarios. The study aims to provide valuable insights for human resource professionals, compensation analysts, and organizational leaders in designing effective and equitable salary structures.

In the contemporary employment landscape, dissatisfaction with salary remains a principal catalyst for employee turnover. The frequent occurrence of employees changing companies in pursuit of their desired salary not only results in talent loss but also poses financial challenges for organizations. In response to this prevalent issue, the project seeks to proactively predict employee salaries based on qualifications and merit, utilizing the potent framework of a Multiple Linear Regression model.

Multiple Linear Regression:



The Multiple Linear Regression model stands out as an invaluable tool for predicting salaries, taking into account an array of factors or predictors. This model allows for a nuanced examination of how multiple independent variables—such as years of experience, distance, and age—collectively influence an individual's salary.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

It is imperative to underscore that, during the meticulous construction of the model, various assumptions must be considered, including linearity, independence of errors, normality, homoscedasticity, and the absence of multicollinearity among predictors.

A meticulous analysis of the multiple regression model not only provides invaluable insights into the factors contributing to salary differentials but also empowers decision-makers in the realms of compensation, human resources, and strategic workforce planning.

Data and Variables: The fundamental variables utilized in the analysis include:

1. **Experience Years:** This variable encapsulates the number of years of professional experience.
2. **Distance:** A crucial factor representing the geographical distance between an employee's location and a reference point.
3. **Age:** Considering the age of the employee, this variable holds relevance in demographic understanding and workforce planning.

Analysis and Results

The Multiple Linear Regression model for our data is.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

```
> vif(model)
      x1       x2       x3
1.286386 1.195048 1.155984
```

```
lm(formula = y ~ x1 + x2 + x3, data = data)

Residuals:
   Min     1Q Median     3Q    Max
 -7162  -3962   -925   3193   8818

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  39120.8    12048.3   3.247  0.00253 **
x1            9696.8      355.8  27.252  < 2e-16 ***
x2             250.1      277.4   0.902  0.37327
x3            -271.4      109.4  -2.480  0.01793 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5219 on 36 degrees of freedom
Multiple R-squared:  0.9627,    Adjusted R-squared:  0.9595
F-statistic: 309.4 on 3 and 36 DF,  p-value: < 2.2e-16
```

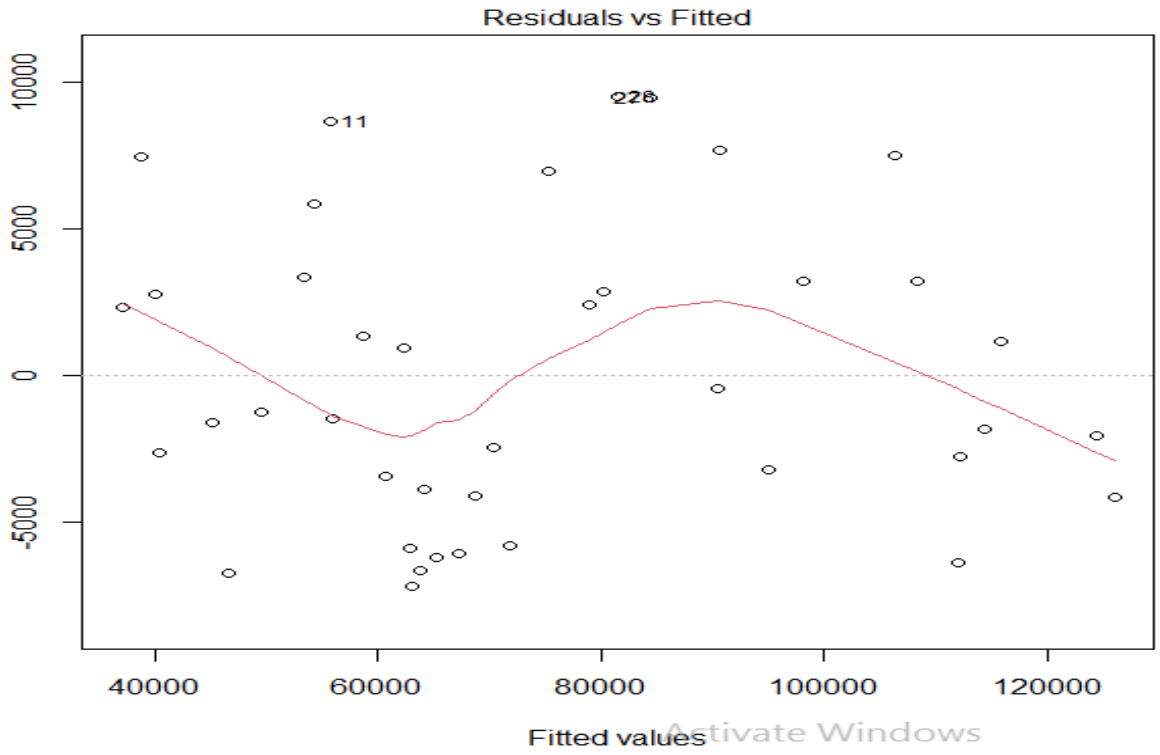
The rigorous analysis of the dataset reveals that years of experience and distance emerge as statistically significant predictors of salary, while age does not exert a significant influence. Consequently, the age variable is judiciously excluded from the model.

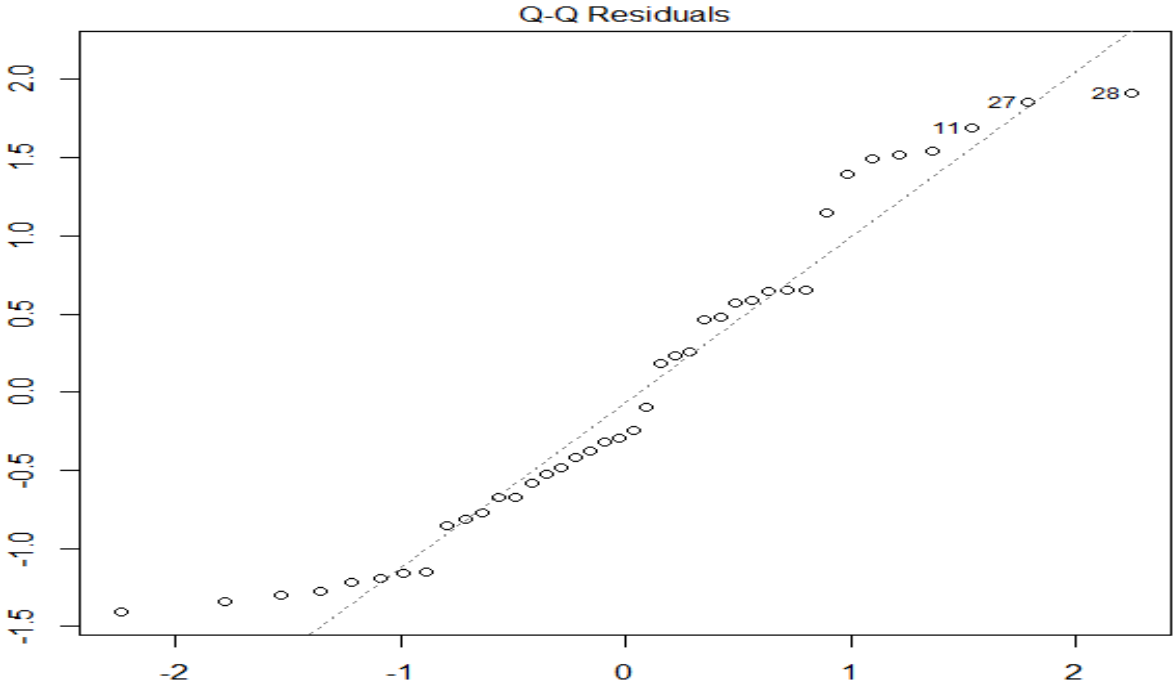
Final Model:

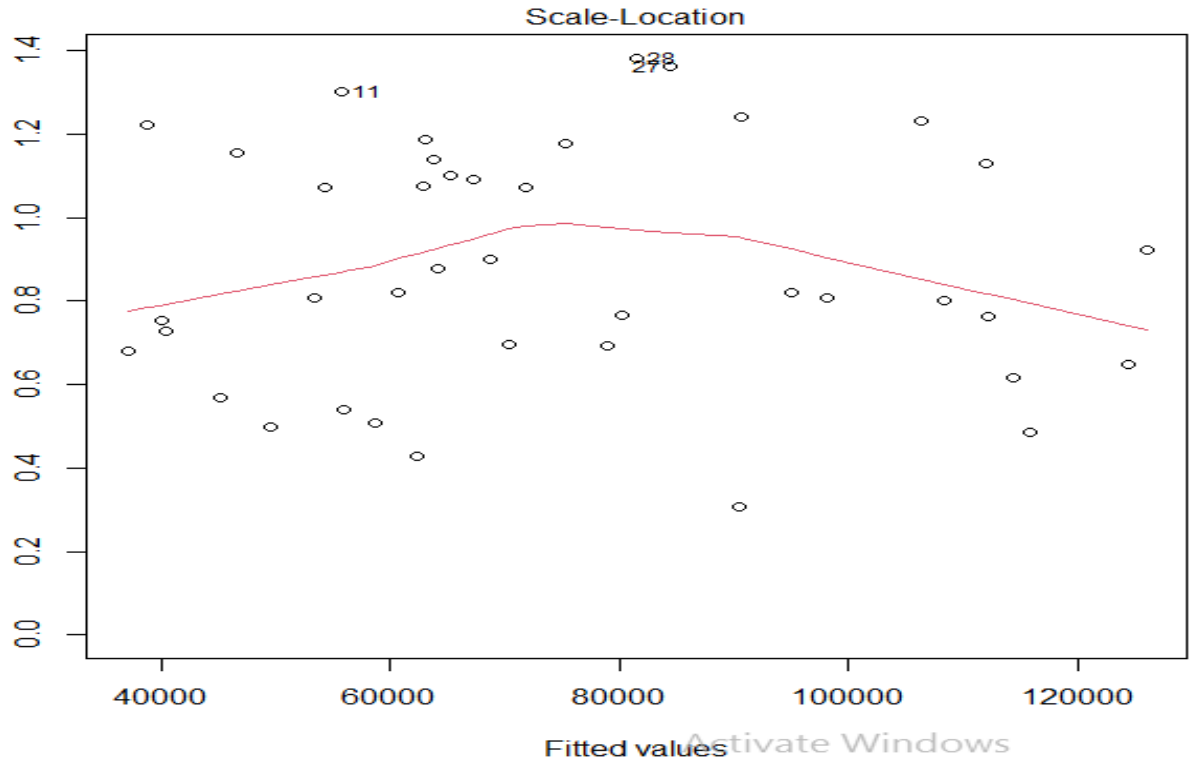
```
lm(formula = y ~ x1 + x3, data = data)

Residuals:
   Min     1Q Median     3Q    Max
 -7209  -3955  -1379   3211   9481

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  46463.0     8856.8   5.246 6.58e-06 ***
x1            9804.8      334.2  29.341  < 2e-16 ***
x3            -259.9      108.4  -2.398   0.0216 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5205 on 37 degrees of freedom
Multiple R-squared:  0.9618,    Adjusted R-squared:  0.9598
F-statistic:   466 on 2 and 37 DF,  p-value: < 2.2e-16
```

The resultant final model boasts a predicted R-square of 95.66%, underscoring its robustness in forecasting future employee salaries.

Model Adequacy:







Ensuring the model's adequacy, it is discerned that all foundational assumptions—Linearity, Normality, Constant variance, and Independence—are satisfactorily met. This robust confirmation cements the reliability of the model for precise salary predictions.

Prediction for New Data Set:

```
        fit       lwr      upr
1 55082.38 44293.86 65870.9
```

For an illustrative scenario involving an employee with 3 years of experience located 80 km away, the estimated salary stands at $55,082.38. Bolstering confidence in this prediction, a 95% confidence interval indicates that the actual salary is likely to fall within the range of $44,923.86 to $65,870.9.

Conclusion:

In summation, this project introduces a highly sophisticated salary prediction system deploying the Multiple Linear Regression algorithm. The discernment of years of experience and distance as the most impactful predictors underscores the system's efficacy. By offering an interpretable framework for salary forecasting, the system facilitates organizations in making not only fair but also data-driven compensation decisions. As organizations navigate the multifaceted landscape of compensation management, this research stands as a steadfast foundation for

informed decision-making, nurturing the cultivation of equitable and data-driven salary practices.

Reference

- D.M Lothe,Prakash Tiwari, Nikhil Patil, Sanjana Patil, Vishwajeet Patil, “SALARY PREDICTION USING MACHINE LEARNING” 2021 6 International Journal Of Advanced Scientific Research and Engineering Trends (IJASRET)
- R. Voleti and B. Jana, "Predictive Analysis of HR Salary using Machine Learning Techniques", *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 1, pp. 34-37, 2022.
- K. Gopal, A. Singh, H. Kumar and S. Sagar, "Salary Prediction Using Machine Learning", *International Journal of Innovative Research in Technology (IJIRT)*, vol. 8, no. 1, pp. 380-383, 2021.
- Pornthep Khongchai, Pokpong Songmuang, “Improving Students’ Motivation to Study using Salary Prediction System” 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)
- Mangui Wu, Shunmin Shu,” Top Management Salary, Stock Ratio and Firm Performance: A Comparative Study of State-owned and Private Listed Companies in China”