Title: Studying complex relationships between employees’ compensation and other organizational factors using Machine Learning techniques

Authors: Davon Carvalho, Sarik Dhungel, Nikhil Reddy Gatla, Saiphani Chandra Vuppala, Sai Unnathi Nandula

Team 6

**Abstract:** In this study we provide insights that are useful to aspiring job applicants and organizations through various statistical analyses of the employees’ compensation data openly available at the official governmental website of the USA [1]. It is crucial to understand complex relationships between employees’ compensation and several other workforce dynamics for effective Human Resource Management (HRM) which assists in employees’ retention and overall organizational growth. For this reason, we employed various statistical analyses along with Artificial Intelligence (AI) and Machine Learning (ML) algorithms for understanding and evaluating such complex relationships. First, we extracted temporal patterns associated with the employees’ compensation over the past decade. It was found that with the passage of time, people are getting a higher salary in general. However, still there is an unequal distribution of money within an organization resulting in very few people making a high amount of money within an organization. Furthermore, we also studied the dataset to understand which jobs are more likely to get paid better and how salaries are distributed across different organizational groups. One of these insights includes 4089 junior clerk jobs out of 4699 having salary less than 35K i.e., more than 87% of jobs in this category have pay lower than 35K. Similarly, employees associated with general city responsibilities, culture and recreation organizational groups are more likely to get lower amount of salary in comparison to other organizational groups. Insights like these are crucial to organizational policy makers for proper management of employees as well as to aspiring job applicants to make better career choices. Moreover, to understand the relation between different organizational factors and salary, regression analysis was done with 3- fold cross validations and hyperparameters tuning using linear regression, decision tree regressor, and random forest regressor ML models. The result showed that salary had a significant positive correlation with various organizational factors with RMSE value of 15808.6, 16234.6, 15808.6 and R squared value of 0.82, 0.81, 0.82 for linear regression, decision tree regressor and random forest regressor model respectively. In summary, in this study we have opened new avenues for investigation of employees’ compensation patterns across various organizations. While the observed accuracy is good, it is important to understand the forthcoming challenges while trying to acquire similar performance across diverse organizational dataset. Hence, we plan to extend this work in future by training the models with lots of optimized hyperparameters and dataset across different organizations with the major goal of enhancing model’s generalizability across various organizational data.

***Keywords: Machine Learning, Linear Regression, Decision Tree, Random Forest, Human Resource Management, Hyperparameters***

**REFERENCES**

1. Data Gov. (2023). Employee Compensation [Dataset]. data.sfgov.org. <https://catalog.data.gov/dataset/employee-compensation>