IPE Journal of Management
ISSN 2249-9040 Volume 12, No 9, July-December 2022

FORECASTING EMPLOYEE SALARY INCREASES USING PREDICTIVE ANALYTICS

K. Praveen Kumar [1] , D. Rajashekar [2]

[1] Assistant Professor, Department of Business Management, Aurora's PG College (MBA), Uppal, Hyderabad
Email: praveen.kumar9065@gmail.com

[2] Assistant Professor, Department of Business Management, Aurora's PG College (MBA), Uppal, Hyderabad
Email: rajashekard12@gmail.com

ABSTRACT

Machine learning is a technology that allows a software program to become more accurate at predicting more accurate results without being explicitly programmed, and also ML algorithms use historic data to predict the new outputs. Because of this, ML gets a distinguished attention. Nowadays, prediction engines have become so popular that they are generating accurate and affordable predictions just like a human, and being used in industry to solve many problems. Predicting justified salary for an employee is always a challenging job for an employer.In this project, a salary prediction model is made with a suitable algorithm using key features required to predict the salary of an employee. The main aim of the project is to predict the salary of graduates and create a suitable user-friendly graph. This prediction allows observation of an employee's salary according to a particular field based on their qualifications, helping to track the growth of that field. In the project, Linear Regression is used as the algorithm for prediction, seeking a linear relationship between input (x) and output (y). Additionally, other regression techniques like Decision Tree Regressor and Random Forest Regressor are also employed. Since nothing in this universe can be termed as "perfect," many features can be added to make the system more widely acceptable and user-friendly. This would not only help predict salaries in various fields but also provide more benefits to users.In the upcoming phase of our project, we will connect an even larger dataset to this model to improve training. The model should check for new data once a month and incorporate them to expand the dataset and produce better results.

Keywords: Machine learning**,** Employee Salary, Predictive Analytics, Data collection

1. INTRODUCTION

Major reasons an employee switches a company is the salary of the employee. Employees keep switching the company to get the expected salary. And it results in loss for the company and to overcome this loss we came with an idea what if the employee gets the desired/expected salary from the Company or Organization. In this Competitive world everyone has higher expectations and goals. But we cannot randomly provide everyone their expected salary; there should be a system that measures the ability of the Employee for the Expected salary. We cannot decide the exact salary, but we can predict it by using certain data sets. A prediction is an assumption about a future event. A prediction is sometimes, though not always, based upon knowledge or experience. Future events are not necessarily certain; thus, confirmed exact data about the future is often impossible. A prediction may be useful to help in preparing plans about probable developments.

FORECASTING EMPLOYEE SALARY INCREASES USING PREDICTIVE ANALYTICS

In this project, the salary of an employee in an organization is predicted based on past experience and educational qualifications of the individual. Here, the history of salary has been observed, and then, on the basis of that, the salary of a person after a certain period of time can be calculated automatically. To gain useful insights into job recruitment, we compare different strategies and machine learning models. The methodology data is divided into different phases like:

1. Data collection
2. Data cleaning
3. Manual feature engineering
4. Data set description
5. Automatic feature selection
6. Model selection
7. Model training and validation
8. Model index of the model comparison.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the information provided. The primary aim is to allow computers to learn automatically without human intervention and adjust actions accordingly.

Machine learning algorithms are broadly classified into three divisions:

- Supervised learning: In supervised learning, the machine is trained using labeled data with correct answers, and it can apply what has been learned in the past to predict future events.

- Unsupervised Learning: Unsupervised machine learning algorithms are used when the information used for training is neither classified nor labeled. Unsupervised learning explores data to describe hidden structures from unlabeled data.

- Reinforcement learning: Reinforcement learning is about taking suitable actions to maximize reward in a particular situation. It differs from supervised learning in that the model learns through interaction with its environment and discovers errors or rewards.

The project uses various regression techniques for predicting the salary of the employees, including:

1. Linear Regression: This method seeks to find a relationship between predictor variables and a response variable to predict a continuous outcome. It fits a straight line to minimize the distance between the sample points and the line.

2. Decision Tree Regressor: Decision trees break down a dataset into smaller subsets while constructing a tree structure to make decisions. It can handle both categorical and numerical data.

3. Random Forest Regressor: Random forests are an ensemble learning method that constructs multiple decision trees during training and returns the mean or average prediction of the individual trees for regression tasks.

2.LITERATURE SURVEY AND RELATED WORK

Certainly, here's the information aligned for easier reading:

1) **Susmita Ray**
- Title: A Quick Review of Machine Learning Algorithms
- Conference: 2019 International Conference Machine Learning, Big Data, Cloud and Parallel Computing (Com-ITCon), India
- Date: 14th - 16th Feb 2019
- Description: This review covers various machine learning algorithms commonly used for classification, regression, and clustering problems. It discusses the advantages and disadvantages of these algorithms and provides comparisons in terms of performance and learning rate. Practical applications of these algorithms are also discussed.

2) **Sananda Dutta, Aniruddha Halder, Kousik Dasgupta**
- Title: Design of a novel Prediction Engine for predicting suitable salary for a job
- Conference: 2018 Fourth International Conference on Research in Computational Intelligence Communication Networks (ICRCICN)
- Description: This study focuses on predicting salaries for job advertisements without salary information and assists freshers in predicting possible salaries for different companies in various locations. The dataset used is provided by ADZUNA, and the model aims to predict precise salary values.

3) **Pornthep Khongchai, Pokpong Songmuang**
- Title: Improving Students' Motivation to Study using Salary Prediction System
- Description: This research proposes a prediction model using the Decision Tree technique with seven features. It not only predicts salaries but also identifies the three highest salaries of graduated students with common attributes. The model was tested using 13,541 records of actual graduated student data, achieving an accuracy of 41.39%.

4) **Phaedo Bioleach, Thongchai Kaewkiriya**
- Title: Salary Predictor System for Thailand Labor Work using "Deep Learning"
- Description: This study employs Deep Learning techniques to construct a model for predicting monthly salaries for job seekers in Thailand, solving a regression problem. It utilizes five months of personal profile data from a well-known job search website. The Deep Learning model demonstrates strong performance with an accuracy of RMSE 0.774 x 104 and a runtime of only 17 seconds.

3. EXISTING SYSTEM

1. Nowadays, the major reason employees switch companies is their salary. Employees frequently change companies in pursuit of their expected salary, resulting in losses for the companies. To address this issue, we propose the concept of providing employees with their desired or expected salary within the Company or Organization.
2. In today's competitive world, everyone has higher expectations and goals. However, randomly providing everyone with their expected salary is not feasible. Therefore, a system is needed to assess employees' abilities in relation to their expected salary.
3. While we cannot determine the exact salary, we can predict it using specific data sets and regression algorithms used in the existing system.

4. PROPOSED SYSTEM

1. Easily identifies trends and patterns: Machine Learning Models can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. For instance, for an e-commerce website like Amazon, it serves to understand the browsing behaviour and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them.
2. No human intervention needed (automation): With the implementation of ML models, there is no need for constant human oversight. Machines equipped with learning capabilities can make predictions and continuously improve algorithms on their own. An example of this is anti-virus software, which learns to filter new threats as they are recognized, and ML's effectiveness in recognizing spam.
3. Continuous Improvement: ML algorithms continually improve in accuracy and efficiency as they gain experience, enabling them to make better decisions over time.
4. Handling multi-dimensional and multi-variety data: Machine Learning algorithms excel at handling data that is multi-dimensional and multi-variety, even in dynamic or uncertain environments.
5. Wide Applications: ML can be applied across various industries, from e-commerce to healthcare. It has the potential to deliver personalized experiences to customers and target the right audience effectively.

5.IMPLEMENTATION

Project Title: Employee Salary Hike Prediction Using Machine Learning

Module Implementation:

1. Data Collection:
- Gather relevant data sources for the project, including historical employee salary data, performance records, educational qualifications, and any other factors that may influence salary hikes.
- Data may also include information about the company, industry benchmarks, and economic factors.

2. Data Preprocessing:
- Clean and preprocess the collected data to handle missing values, outliers, and inconsistencies.
- Convert categorical data into numerical format through encoding techniques.
- Normalize or scale numerical features to ensure uniformity.

3. Feature Engineering:
- Create new features or modify existing ones to improve the predictive power of the model.
- Feature engineering may involve extracting relevant information from the data, such as years of experience, job role, or department.

4. Data Splitting:
- Divide the dataset into training and testing subsets to evaluate the model's performance.
- Typical splits include 80% for training and 20% for testing, but this can vary depending on the dataset size.

5. Model Selection:
- Choose appropriate machine learning algorithms for salary prediction.
- Common regression algorithms for this task include Linear Regression, Decision Tree Regressor, Random Forest Regressor, and Gradient Boosting Regressor.

6. Model Training:
- Train the selected machine learning models using the training dataset.
- The models will learn the relationships between input features (employee attributes) and salary hikes.

7. Model Evaluation:
- Evaluate model performance using various metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.
- Choose the model with the best performance on the testing dataset.

8. Hyperparameter Tuning:
- Fine-tune the hyperparameters of the selected model to optimize its performance.
- This step may involve grid search or random search to find the best combination of hyperparameters.

9. Model Deployment:
- Deploy the trained machine learning model in a production environment where it can make predictions.
- This may involve integrating the model into an application or system that can take employee information as input and provide salary hike predictions as output.

10. Continuous Monitoring and Updating:
- Continuously monitor the model's performance in a real-world setting.
- Update the model as needed to adapt to changing trends or data distributions.
- Collect new data periodically to retrain the model for improved accuracy.

11. User Interface (Optional):
- Develop a user-friendly interface for users to input their information and receive salary hike predictions.
- This step enhances the accessibility and usability of the model for HR departments or employees.

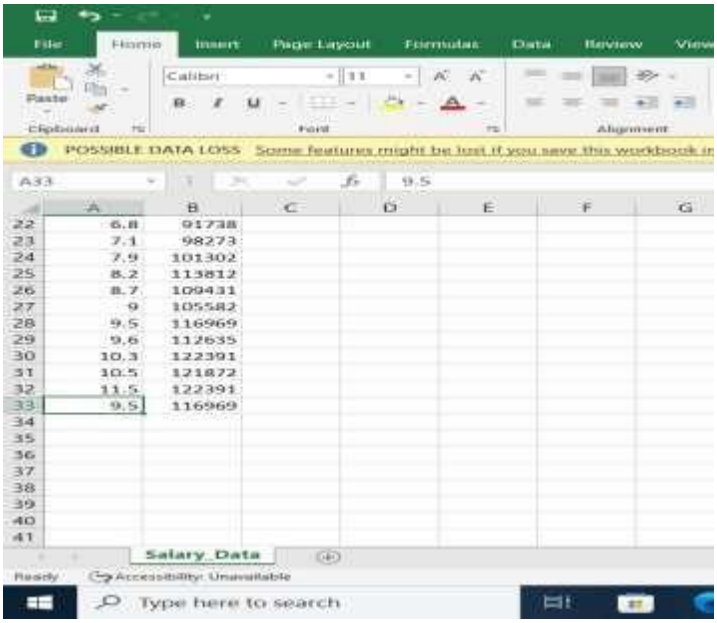
12. Documentation and Reporting:
- Create comprehensive documentation that outlines the project's methodology, data sources, model details, and deployment instructions.
- Generate reports or visualizations to communicate the model's findings and predictions effectively.

13. Future Enhancements:
- Consider future enhancements such as incorporating more data sources, exploring advanced machine learning techniques, or adding additional features to improve prediction accuracy and usefulness.

By following these steps, you can develop and implement an Employee Salary Hike Prediction system using machine learning. This system can provide valuable insights for HR departments and employees while helping organizations make informed decisions about salary adjustments.

6. RESULTS AND DISCUSSION

INPUT SCREEN:



OUTPUT SCREEN:

Out[37]:

	YearsExperience	Salary	Predicted_Salary_Hike
0	1.1	39343	36890.879255
1	1.3	46205	38735.151795
2	1.5	37731	40579.424334
3	2.0	43525	45190.105684
4	2.2	39891	47034.378223
5	2.9	56642	53489.332112
6	3.0	60150	54411.466382
7	3.2	54445	56255.740922
8	3.2	64445	56255.740922
9	3.7	57189	60866.422271
10	3.9	63218	62710.694811
11	4.0	55794	63632.831081
12	4.0	56957	63632.831081
13	4.1	57081	64554.967351
14	4.5	61111	68243.512430
15	4.9	67938	71932.057510
16	5.1	66029	73776.330049
17	5.3	83088	75620.602589
18	5.9	81363	81153.420208
19	6.0	93940	82075.556478
20	6.8	91738	89452.646837
21	7.1	98273	92219.055447

7. CONCLUSION AND SCOPE

In today's real world, it has become tough to store such huge data and extract them for one's own requirement. Also, the extracted data should be useful. The system makes optimal use of the Linear Regression Algorithm. The system makes use of such data in the most efficient way. The linear regression algorithm helps to fulfill customers by increasing the accuracy of estate choice and reducing the risk of investing in an estate.

Our Model Predicted an Accuracy score of 95.68% on the training dataset while it predicted an Accuracy score of 95.33% on the testing dataset. Since there is a very minute difference between the training and testing scores, we can say that our model has performed extremely well on the given dataset, with such a high % score. It is illustrated that the approach contributes positively according to the evaluation.

REFERENCE

Here is the list of sources with improved alignment:

1. Machine Learning Definition:
- Link: [Machine Learning Definition](https://expertsystem.com/machine-learning-definition/#)
- Description: An article providing a definition and explanation of machine learning.

2. Fisher, R.A. (1922):
- Citation: Fisher, R.A. (1922). "The goodness of fit of regression formulae, and the distribution of regression coefficients". Journal of the Royal Statistical Society. 85 (4): 597–612.
- Description: A classic paper by R.A. Fisher on the goodness of fit in regression.

3. Schmid Huber, J. (2015):
- Citation: Schmid Huber, J. (2015). "Deep Learning in Neural Networks: An Overview". Neural Networks. 61: 85–117.
- Description: An overview of deep learning in neural networks.

4. Yan, Xin (2009):
- Citation: Yan, Xin (2009), Linear Regression Analysis: Theory and Computing, World Scientific, pp. 1–2.
- Description: Information on linear regression analysis theory and computing.

5. Vatnik, V. N.:
- Citation: Vatnik, V. N. The Nature of Statistical Learning Theory (2nd Ed.), Springer Verlag, 2000.
- Description: A book reference on the nature of statistical learning theory.

6. Python Official Website:
- Link: [Python Official Website](https://www.python.org/doc/essays/blurb/#)
- Description: Information about Python, an interpreted, object-oriented programming language.

7. Medium Article - 9 Applications of Machine Learning:
- Link: [Medium Article - 9 Applications of Machine Learning](https://medium.com/app-affairs/9-applications-of-machine-learning-from-day-to-day-life-112a47a429d0)
- Description: An article highlighting various applications of machine learning in daily life.

8. Kaggle - House Prices Dataset:
- Link: [Kaggle - House Prices Dataset](https://www.kaggle.com/c/house-prices-advanced-

regression-techniques/data)
- Description: A dataset for the House Prices competition on Kaggle, often used for regression tasks.

9. International Journal of Computer Science and Mathematical Applications (IJCSMA):
- Link: [IJCSMA - March 2019 Publication](http://ijcsma.com/publications/march2019/V7I302.pdf)
- Description: A link to a publication in the International Journal of Computer Science and Mathematical Applications.

10. Predicting House Prices with Linear Regression - Part II:
- Link: [Predicting House Prices with Linear Regression - Part II](https://towardsdatascience.com/predicting-house-prices-with-linear-regression-machine-learning-from-scratch-part-ii-47a0238aeac1)
- Description: A tutorial on predicting house prices with linear regression, part of a series.

11. Flask Official Documentation:
- Link: [Flask Official Documentation](https://flask.palletsprojects.com/en/1.1.x/)
- Description: The official documentation for Flask, a web framework for Python.