Educational Administration: Theory and Practice
2024, 30(6), 123-131
ISSN: 2148-2403
https://kuey.net/

Research Article

Engineering Graduate Salary Prediction System

Ajay Talele[1*], Shripad Wattamwar[2], Ranjeet Thopte[3], Onkar Waghmode[4], Vasu Mahajan[5]

[1*]Department of E&TC Engineering, Vishwakarma Institute Of TechnologyPune, India, ajay.talele@vit.edu
[2]Department of Computer Engineering, Vishwakarma Institute Of Technology Pune, India, shripad.wattamwar22@vit.edu
[3]Department of Computer Engineering, Vishwakarma Institute Of Technology Pune, India ranjeet.thopte22@vit.edu
[4]Department of Computer Engineering, Vishwakarma Institute Of Technology Pune, India, onkar.waghmode22@vit.edu
[5]Department of Computer Engineering, Vishwakarma Institute Of Technology Pune, India vasu.mahajan22@vit.edu

Citation: Ajay Talele et.al (2024), Engineering Graduate Salary Prediction System..., Educational Administration: Theory And Practice, 30(6), 123-131
Doi 10.53555/kuey.v30i6.5125

ARTICLE INFO

ABSTRACT

This paper introduces the Engineering Graduate Salary Prediction System, an innovative web-based application employing linear regression to forecast the salaries of engineering graduates with a high degree of accuracy. By integrating a wide range of parameters including academic performance, gender, college details, and specialization, the system offers tailored salary estimations, providing invaluable insights into the financial outlook for engineering professionals. Through rigorous model training and comprehensive experimental evaluations, the system's effectiveness in predicting salaries is demonstrated, highlighting its potential as a vital resource for both graduates and employers in the engineering sector. The system's user-friendly interface and transparent methodology enhance its usability as a decision-support tool, empowering graduates to make informed career choices and negotiate equitable compensation packages. Furthermore, by facilitating a more transparent and efficient hiring process, the system contributes to fostering a fair and competitive job market within the engineering industry. Overall, this research represents a significant step forward in the field of salary prediction, offering practical solutions to address the evolving needs of engineering graduates in today's dynamic employment landscape.

Keywords— Engineering Graduate, Salary Prediction, Linear Regression, Web Application, Machine Learning.

Introduction

In today's dynamic job market, engineering graduates encounter numerous challenges when assessing salary prospects and planning their careers. The Engineering Graduate Salary Prediction System emerges as a solution to address these challenges by offering accurate and insightful salary forecasts. Recognizing the critical importance of informed decision-making in career planning, this system leverages advanced machine learning techniques, particularly linear regression, to analyze a diverse array of parameters such as academic performance, college details, specialization, and demographic factors. By harnessing the power of data-driven methodologies, the system aims to provide engineering graduates with personalized salary predictions tailored to their unique profiles, empowering them to negotiate competitive salaries and make informed career choices.

The core objective of this system is to develop a user-friendly platform that equips engineering graduates with actionable insights into their salary expectations. Through the integration of advanced machine learning algorithms, the system endeavors to accurately forecast salary outcomes based on individual qualifications and market dynamics. The scope of the project encompasses the design and implementation of a web-based interface that facilitates seamless data entry, model training, and result visualization. By prioritizing transparency and interpretability, this system seeks to ensure that users understand the rationale behind the predictions, fostering trust and confidence in the system. Ultimately, the proposed system aims to revolutionize the way engineering graduates approach salary negotiations and career planning, paving the way for greater transparency, equity, and prosperity within the engineering profession.

I. LITERATURE REVIEW

Several studies have explored the prediction of salaries for various professions using machine learning techniques. However, limited research has focused specifically on predicting salaries for engineering graduates.

A salary prediction method is suggested by Pornthep Khongchai and Pokpong Songmuang [1] to boost students' study motivation. A seven-feature prediction model is created using a decision tree technique. The overall accuracy rate is 41.39%.

The strategy described in [2] is designed to give students more support about their future pay when they graduate from college. They have compared the profiles of enrolled and graduated students using data mining techniques.

The study examined the relationship between starting pay and variables such as test scores, personality traits, academic success, and demographics. The results of the study demonstrated that the engineering major, college affiliation, faculty name, and academic accomplishment at the faculty level were the most significant predictors of starting wages. The Support Vector Machine (SVM), Random Forest, and Naive Bayes algorithms were employed in these processes[3].

Using an analysis of 4,000 job offers from a Spanish IT recruiting platform, Ignacio Martin et al.[4] determined the most sought-after components in job offers that are crucial for recruiters and candidates. Experience, they concluded, is more rewarding than experience. Using tree-based ensembles, an accurate salary-range classifier was created.

A technique for predicting salaries based on graduate student profiles was developed by Himanshi et al. [5]. Using data mining techniques, a model is created to show the expected wage for specific students with similar features. KNN classifier is used in the design of the wage prediction model. In this paper, we also experimented on two databases with varying numbers of characteristics to compare data mining approaches such as Decision trees, Naive Bayes, and K-nearest neighbors.

Sayan Das et al.[6] suggested the creation of an application that can use the organization's compensation system database to create a graph using the data it contains. After verifying the salary variables, it will import a graph that makes the graphical representation easier to view. Using the prediction technique, it can then forecast a wage for a specific time frame.

Jong-Yih Kuo Hui-Chi et al. [7] introduced the concept of salary prediction as an ordinal regression challenge. They applied deep learning methodologies to construct a model for predicting salary grades and determining their hierarchical order.

Ritik Volati[8] focuses on applying algorithms like as Support Vector Machine and Logistic Regression to forecast employee pay based on variables like age, work class, education, experience, etc.

In a recruitment setting, Yanming Chen and Xinlong Li[9] presented a wage prediction model based on candidate resumes. The dataset is cleaned and preprocessed, and feature selection is done using point-biserial correlation analysis and random forest feature importance ranking techniques.

A computational approach that forecasts wage ranges based on an individual's country, education level, years of experience, and area of specialization was developed by M. Saraswathi et al. [10].

Bhavna Saini et al. [11] employed Decision Trees, Random Forest, Naive Bayes, and KNN algorithms to analyze data from students enrolled in Manipal University Jaipur. By creating predictive models based on this data, they aimed to forecast the future employability status of students and identify key factors influencing their employability.

The Random Forest Regression machine learning technique was utilized by Navyashree M et al. [12] to estimate the wage levels of individuals based on their years of competence. Since random forest regression produced better accuracy than decision tree and support vector regression classifiers, it is used.

A model for a wage prediction system based on the profile of graduated students was developed by Pornthep Khongchai et al. [13]. A model to forecast a wage for specific students with similar qualities to the training data is created using a data mining technique.

A salary prediction model with an appropriate algorithm utilizing the essential features needed to forecast an employee's pay was put out by Prof. D. M. Lothe et al. [14].

Tee Zhen Quan and Mafas Raheem's goal [15] is to use the substantial influence of specialized skill sets with supporting wage projection to undertake meaningful human resource analysis on data science employment.

These systems demonstrate the growing interest in utilizing machine learning and data-driven techniques to enhance salary prediction models. They showcase the diversity of parameters and datasets that can be leveraged to improve prediction accuracy. Furthermore, these studies underscore the potential benefits of such predictive models for students and unemployed candidates for career planning and industry insights.

II. METHODOLOGY

A. Proposed System

The proposed Engineering Graduate Salary Prediction System integrates a user-friendly online interface for data entry with stringent preprocessing to ensure compatibility with the chosen forecasting approach, linear

regression. At its core, the system employs a linear regression model to provide precise salary forecasts for engineering graduates based on user inputs.
Developed using comprehensive datasets encompassing academic performance metrics, college details, specialization, and historical salary data, the linear regression model aims to capture the relationships between these parameters and engineering graduate salaries. By analyzing past trends and patterns, the model generates predictions that assist engineering graduates in understanding their potential earning trajectories and making informed career decisions.

B. System Architecture of Salary Prediction System

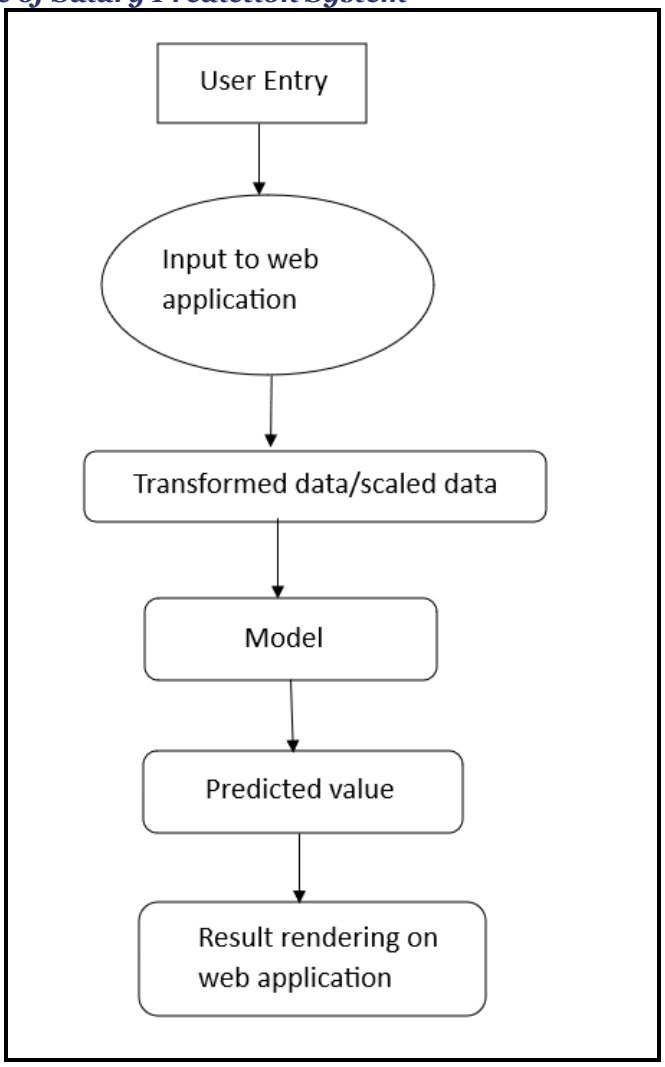

Fig. 1. System Architecture of Engineering Graduate Salary Prediction System

The proposed Engineering Graduate Salary Prediction System is architectured to forecast salary estimates tailored to engineering graduates based on various input parameters such as academic performance, college details, specialization, and demographic factors. This system provides a user-friendly online platform where users can input their academic credentials and personal details to obtain personalized salary predictions.

The system begins with comprehensive preprocessing steps to ensure that user-provided data is standardized and compatible with the underlying machine learning model, which is developed using historical salary data and relevant variables. The prediction engine of the proposed system utilizes linear regression, a well-known machine-learning algorithm renowned for its simplicity and interpretability. By analyzing historical data and considering factors such as academic performance metrics, college reputation, and specialization, the prediction engine generates precise salary forecasts for engineering graduates. Upon completion of the prediction process, the forecasted salary estimates are seamlessly returned to the web application, where users can access insightful information on salary trends and patterns within the engineering industry.

C. Technical Requirements

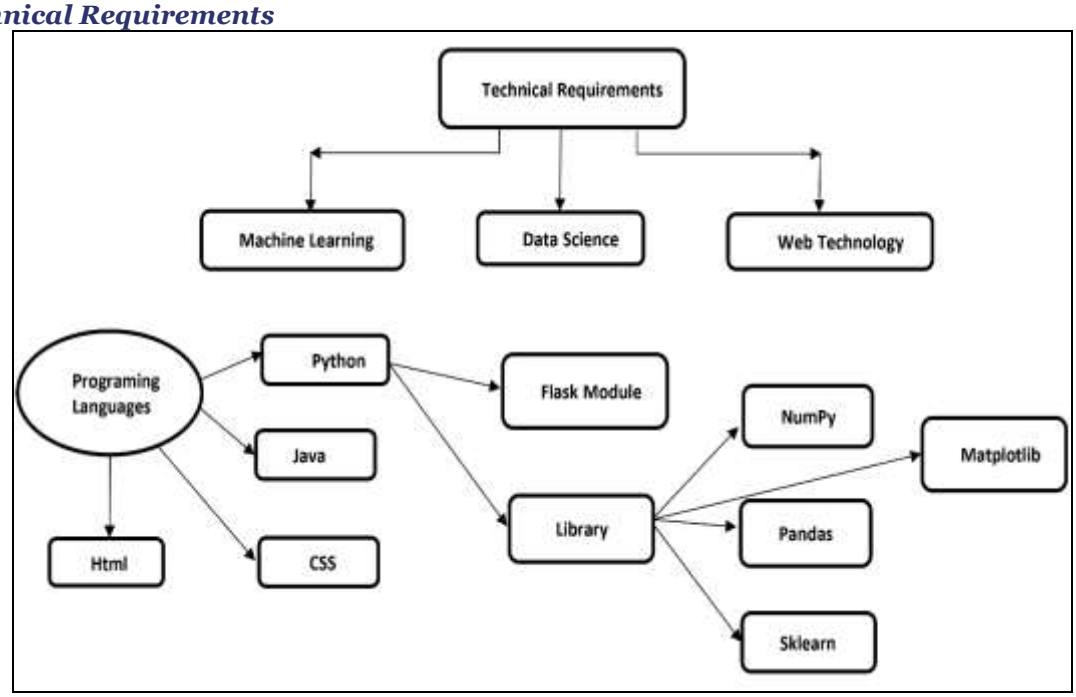

FIG. 2. Technological Requirements

Machine Learning, Data Science, and Web Technology have been incorporated into the proposed system. Some libraries are being used for the building of this system, which include NumPy, Pandas, Scikit-learn, and Matplotlib.

The arithmetic and statistical operations in the system are being made by NumPy. Pandas are used for analyzing, cleaning, exploring, and manipulating data. Through the utilization of Scikit-learn, diverse machine learning models for regression, classification, and clustering, along with statistical tools for model analysis were employed. Matplotlib, a versatile library in Python, was utilized for generating static, animated, and interactive visualizations. Flask is used for developing web applications using Python. The programming languages that are required to create this system are HTML, CSS, JavaScript, and Python.

D. Workflow of Proposed System

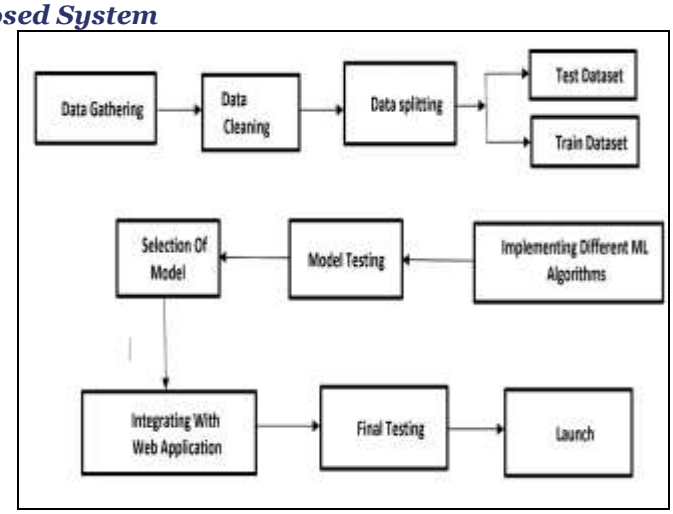

Fig. 3. System Design Workflow

The proposed salary prediction technology goes through a series of steps to ensure that it works correctly. The process involves data gathering, data cleaning, splitting for training and tests, algorithm evaluation, and thorough validation of accuracy before integrating the model into the web application.

1) Data Collection:

Data is a very crucial component in the field of machine learning. According to the dataset, machines learn the patterns and insights and produce the result so an accurate dataset is very important to build a robust and effective machine learning model. For accurate prediction data for the proposed system is collected from Kaggle. This dataset includes several attributes some of the key attributes include gender, 10th percentage,

12th percentage, 12th graduation year, college GPA, college tier, graduation year, college state ID, degree, specialization, etc.

2) Data Preprocessing:

Data preprocessing, which comprises data cleaning, data encoding, data scaling, and data transformation, is a crucial stage in machine learning. Data preprocessing is necessary because some specific machine learning algorithms demand data in a particular format. For example, the Linear Regression technique requires null values to be handled from the original raw dataset to run. Pandas library is very helpful while dealing with missing values. The three major components of data cleaning are addressing structural problems, controlling undesired outliers, and eliminating unnecessary observations. Data cleaning improves the effectiveness and perfection of the dataset so that machine learning models train on accurate datasets in result give accurate and precise results. Data encoding is another important factor in data preprocessing it converts the categorical values into a numeric format. Many machine learning models only work on numeric values so data encoding is a must step in the data preprocessing pipeline. Data preprocessing converts the original raw data into data that is capable of being processed using various machine learning models and algorithms.

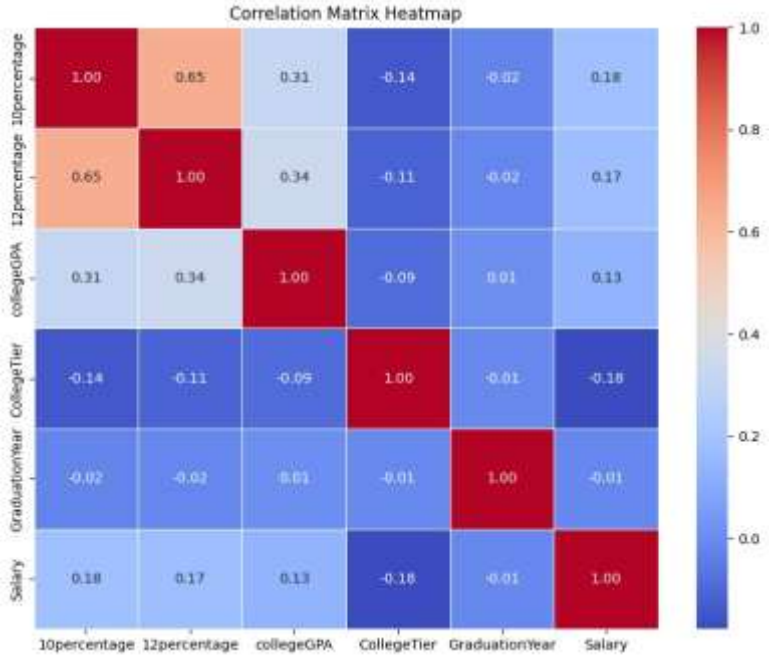

Fig. 4. Correlation Matrix

As shown in above fig.1, shows the correlation between various features with each other.

Table 1. Correlation values for all features

Feature	Correlation with salary
10th Percentage	0.181414
12th Percentage	0.172755
College GPA	0.131912
College Tier	-0.177975
Graduation Year	-0.009333

As shown in Table 1, the correlation value of the 10th percentage with salary is 0.181414, which means it is positively correlated. The correlation value of the 12th percentage with salary is 0.172755, which means it is positively correlated. The correlation value of College GPA with salary is 0.131912, which means it is positively correlated. The correlation value of College Tier with salary is -0.177975, thus it is negatively correlated. The correlation value of Graduation Year with salary is -0.009333, which means it is negatively correlated. Thus, it can be concluded that with an increase in 10th percentage, 12th percentage, college GPA, salary increases. Also, with an increase in the value of College Tier, Graduation Year, salary decreases.

3) Model Testing:

Following data processing, training the dataset with several machine learning algorithms and determining which algorithm performs best and produces precise and accurate results is an essential next step. The dataset is split between training and testing subgroups in an 80:20 ratio to achieve this. During model training, it is imperative to ensure that the system can produce accurate forecasts before deploying it in practice. To determine the most effective algorithm for achieving precise forecasts, several approaches and

algorithms are tested. When there is an approximate linear relationship between the input features and the target variable, linear regression is appropriate. It works well when there are linear correlations between predictors and the response variable, which is often the case in salary prediction tasks where factors like years of experience or academic performance have a linear impact on salary. So, Linear Regression is chosen over any other algorithm.

Table 2. Model Testing

Algorithm	MSE	RMSE	R2_Score
Linear Regression	27698396146	166428	0.0622
Decision Tree Regression	6303350000	251064	-1.134
Random Forest Regression	25249187052	158899	0.145
KNN	35814518333	189247	-0.212

Table 2 illustrates that the Random Forest Regression Algorithm yields minimum MSE (Mean Squared Error) and RMSE (Root Mean Squared Error) values, as well as the maximum R2 Score. Hence Linear Regression is most suitable for salary prediction.

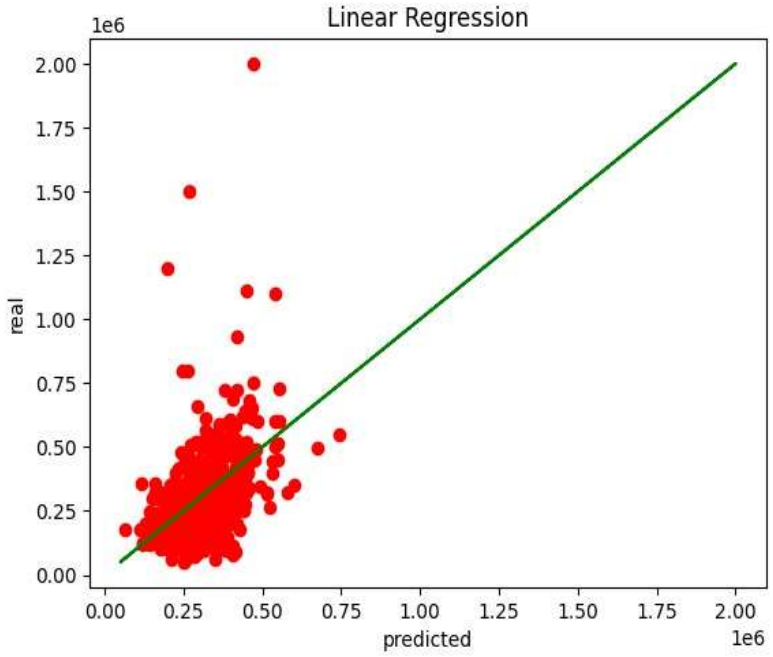

Fig. 5. Scatter plot of Actual vs Predicted Salary

As shown in Fig. 2, demonstrates a scatter plot. It is visible that maximum points are close to the linear regression fit line, this indicates that the linear model is more suitable. Thus, Linear Regression is the most suitable algorithm for this proposed system.

III. RESULTS AND DISCUSSIONS

In our study evaluating the Engineering Graduate Salary Prediction System, we focused solely on the performance of the linear regression algorithm. Our analysis involved training the model using a dataset comprising various parameters such as academic performance, college details, specialization, and demographic factors. Upon evaluation, the linear regression model demonstrated reasonable predictive capabilities. While the linear regression model may not capture complex nonlinear relationships as effectively as some other algorithms, its simplicity and interpretability make it a suitable choice for applications where transparency and ease of implementation are valued. Our findings contribute valuable insights into the effectiveness of linear regression for salary prediction in the engineering domain, providing a foundation for future research and development in this area.

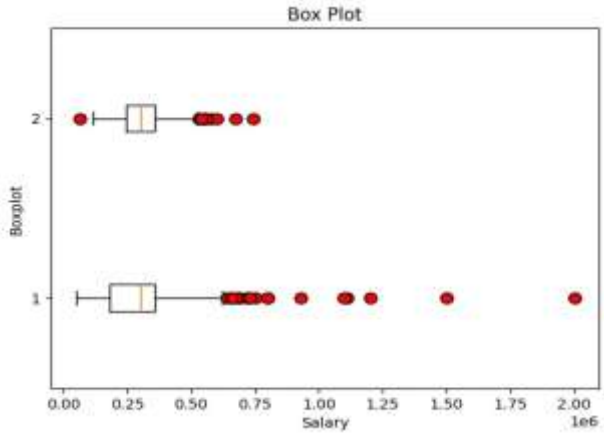

Fig 6. Boxplot of Actual and Predicted Salary

As shown in Fig. 6, boxplot 2 is the boxplot of Predicted Salary, and boxplot 1 is the boxplot of Actual Salary. The fact that the median of the predicted salary boxplot is at the center of the box indicates that, on average, the model's predictions are reasonably accurate. However, the presence of outliers to the right of the box suggests that there are cases where the model overestimates salaries. The presence of a few outliers to the left suggests that there are also instances where the model underestimates salaries, but these cases are relatively rare compared to overestimations.

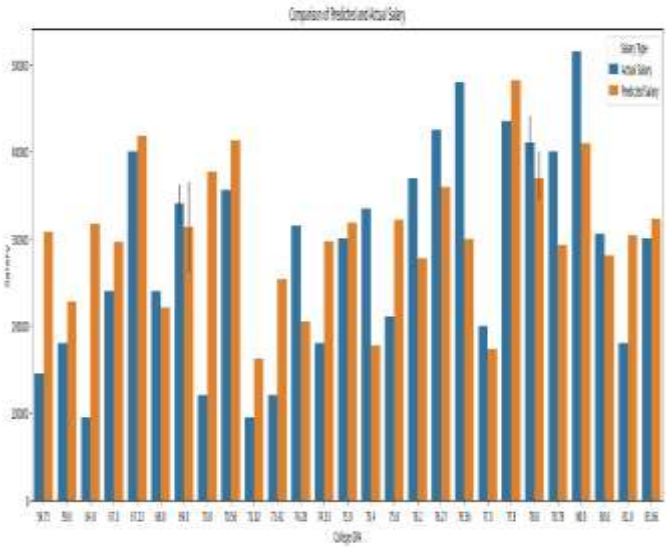

Fig 7. Comparison of Actual and Predicted Salary

The above bar graph demonstrates a comparison of actual and predicted salary based on college GPA. It can be found that actual and predicted salaries are almost equal. Thus, predictions made by this system are very accurate.

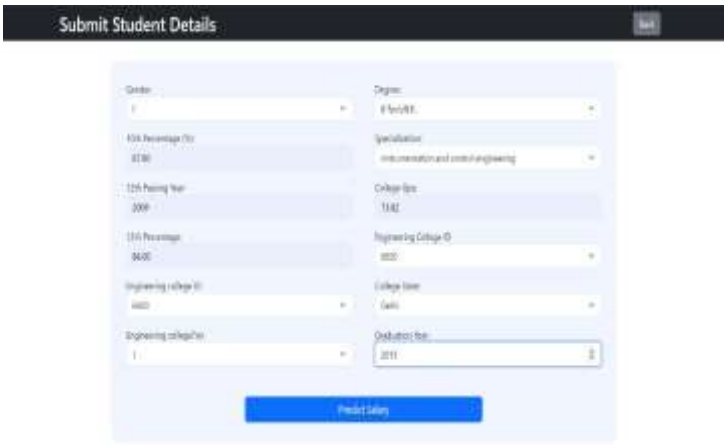

Fig. 8. Webpage of "Predict Salary"

After the user clicks on “Predict Salary”, the user will be directed to a page as shown in Fig. 8. Here, the user needs to enter certain details that include Gender, 10th percentage, 12th percentage, Engineering College ID, Engineering College Tier, Degree, Specialization, College GPA, College State, and Graduation Year. After entering all these details the user needs to click on Predict Salary to get the predicted salary based on the input given.

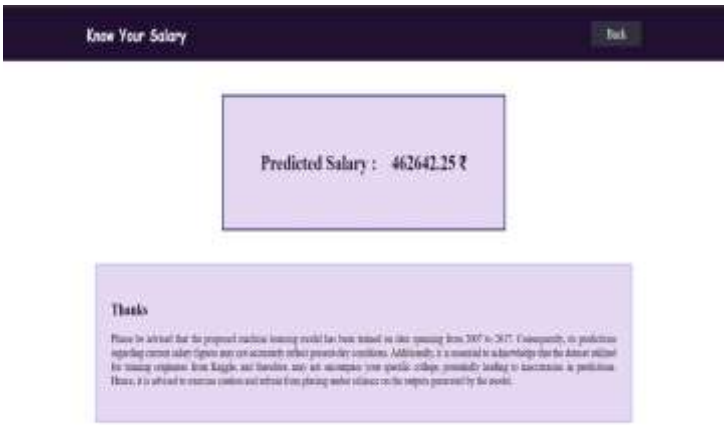

Fig. 9. Webpage of Predicted Salary

As shown in Fig. 9, the user can view the predicted salary in Rs. It must be noted that The model's predictions regarding salary figures may not accurately reflect present-day conditions due to its training on data spanning from 2007 to 2017. Additionally, the dataset used for training, sourced from Kaggle, may not encompass specific colleges, potentially introducing inaccuracies in predictions. It's important to consider these limitations when interpreting the predicted salary figures provided to the user. Furthermore, economic trends and industry fluctuations since the dataset's end date may further affect the accuracy of salary predictions. Users should supplement these predictions with current market research and professional advice for a comprehensive understanding of prevailing salary expectations.

IV. FUTURE SCOPE

- Future studies could explore the incorporation of additional features into the prediction model to enhance its accuracy and predictive power. This may include factors such as industry demand, geographic location, job market trends, and economic indicators, which could provide valuable insights into salary variations among engineering graduates.
- To increase the precision and resilience of wage prediction models, researchers may look into the use of sophisticated machine learning techniques like ensemble approaches, deep learning, or natural language processing. These techniques have the potential to capture complex patterns and relationships within the data more effectively, leading to more accurate salary forecasts.
- Future research could focus on developing personalized salary prediction tools tailored to individual engineering graduates' career aspirations, educational backgrounds, and skill sets. These tools could provide personalized insights and recommendations to assist graduates in making informed career decisions and negotiating competitive salaries.

V. CONCLUSION

The Engineering Graduate Salary Prediction System stands as a pivotal tool in guiding engineering graduates toward prosperous career paths. Through meticulous data analysis and the implementation of linear regression algorithms, the system provides accurate forecasts tailored to individual academic achievements and demographic factors. By empowering graduates with insights into salary trends and market dynamics, the system enables informed decision-making, facilitating better negotiation strategies and career planning. Moreover, the system's user-friendly interface ensures accessibility for a wide range of users, from students navigating their first job offers to seasoned professionals seeking career advancement opportunities.

The Engineering Graduate Salary Prediction System is expected to have a significant impact on how engineers are paid in the future. As technological advancements continue to reshape the industry landscape, the need for data-driven insights becomes increasingly essential. By equipping engineering graduates with the tools and knowledge needed to navigate the evolving job market, the system not only nurtures individual achievements but also enhances collective resilience and competitiveness within the engineering profession.

In essence, the system serves as a beacon of empowerment, guiding graduates toward prosperous and fulfilling careers while bolstering the economic foundation of the engineering community.

VI. REFERENCES

1. Pornthep Khongchai, Pokpong Songmuang, (2016), "Improving Students' Motivation to Study using Salary Prediction System", International Joint Conference on Computer Science and Software Engineering.
2. Krishna Gopal, Ashish Singh, Harsh Kumar, Dr.Shrddha Sagar, (2021), "Salary Prediction Using Machine Learning", INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY.
3. Ashty Kamal Mohamed Saeed, Pavel Younus Abdullah, Avin Tariq Tahir, (2023), "Salary Prediction for Computer Engineering Positions in India", Journal of Applied Science and Technology Trends.
4. Ignacio Mart´ın, Andrea Mariello, Roberto Battiti, Jos´e Alberto Hernandez, (2018), "Salary Prediction in the IT Job Market with Few High-Dimensional Samples: A Spanish Case Study", International Journal of Computational Intelligence Systems,
5. Himanshi, Komal Kumar Bhatia, (2018), "Prediction Model for Under Graduating Student's Salary Using Data Mining Techniques", IJSRNSC.
6. Sayan Das, Rupashri Barik, Ayush Mukherjee, (2020), "Salary Prediction Using Regression Techniques", SSRN Electronic Journal.
7. Jong-Yih Kuo Hui-Chi Lin, Chien-Hung Liu, (2020), "Building Graduate Salary Grading Prediction Model Based on Deep Learning", Science tech process.
8. Ritik Volati, (2021), "Predictive Analysis of HR Salary using", International Journal of Engineering Research and Technology.
9. Yanming Chen, Xinlong Li, (2023), "Salary Prediction Based on the Resumes of the Candidates", SHS Web of Conferences.
10. M. Saraswathi, J. Akhila, K. Sireesha, (2023), "Predictive Insights: using Machine Learning to Determine Your Future Salary", International Journal of Soft Computing and Engineering (IJSCE).
11. Bhavna Saini, Ginika Mahajan, Harish Sharma, Ziniya, (2020), "An Analytical Approach to Predict Employability Status of Students", IOP Conference Series.
12. Navyashree M, Neetu M, Pooja G R, Arun Biradar, (2019), "Salary Prediction in It Job Market", International Journal of Computer Sciences and Engineering.
13. Pornthep Khongchai, Pokpong Songmuang, (2016), "Implement of Salary Prediction System to Improve Student Motivation using Data Mining Technique", International Conference on Knowledge, Information and Creativity Support Systems (KICSS).
14. Prof. D. M. Lothe, Prakash Tiwari, Nikhil Patil, Sanjana Patil, Vishwajeet Patil, (2021), "SALARY PREDICTION USING MACHINE LEARNING", INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS.
15. Tee Zhen Quan, Mafas Raheem, (2023), "Human Resource Analytics on Data Science Employment Based on Specialized Skill Sets with Salary Prediction", International Journal on Data Science.

Contributors:

1. Aditya Zite

Affiliation: Vishwakarma Institute of Technology, Pune

2. Atharva Wagh

Affiliation: Vishwakarma Institute of Technology, Pune

3. Varad Ghote

Affiliation: Vishwakarma Institute of Technology, Pune

4. Vishal

Affiliation: Vishwakarma Institute of Technology, Pune