sciendo INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS

Research Article | DOI: 10.2478/ijssis-2025-0034 Issue 1 | Vol. 18 (2025)

Predicting AI job market dynamics: a data mining approach to machine learning career trends on glassdoor

Renuka Agrawal[1], Aditi Nayak[1], Preeti Hemnani[2], Barish Chetia[1], Ishaan Bhadrike[1], Jil Kapadia[1], Usha A. Jogalekar[1,*] and Safa Hamdare[3]

[1]Department of Computer Science and Engineering, Symbiosis Institute of Technology, Pune, Symbiosis International (Deemed University), Pune, India

[2]Department of Electronics and Telecommunication Engineering, SIES Graduate School of Technology, Mumbai, India

[3]Department of Computer Science, Nottingham Trent University, Nottingham NG11 8NS, UK

*E-mail: usha.jogalekar@sitpune.edu.in

Received for publication March 17, 2025.

Abstract

In today's highly competitive job market, many qualified candidates face significant difficulties in securing positions that align with their skills and career aspirations. This challenge is further compounded by the dynamic nature of the global economy, which continually reshapes the demand for specific job roles and skill sets. Simultaneously, employers encounter obstacles in efficiently identifying and recruiting applicants who possess the most relevant competencies for their organizational needs. To address these dual challenges, this study introduces a predictive system based on an ensemble learning model, which demonstrates superior performance over traditional machine learning algorithms in forecasting both job titles and salary ranges. The proposed model leverages key job-related features, including company size, ratings, income levels, and skill requirements, to provide accurate predictions. Utilizing a curated dataset comprising 956 software job postings, the study conducts a comprehensive analysis to uncover patterns and insights that inform the prediction process. Beyond prediction, the system is designed to offer personalized career guidance to job seekers by analyzing their individual profiles, technical skills, and professional preferences. By integrating data-driven analytics with user-centric recommendations, the platform aims to bridge the gap between job seekers and employers, enhance employment outcomes, and support more strategic decision-making for both parties in the hiring process.

Keywords

ML Algorithms, Global Economy, Personalized career guidance, Ensemble Model, Performance Metrics

I. Introduction

The exponential growth of technology and its establishment in various sectors have cultivated an acute need for skilled professionals in machine learning (ML) and allied domains. As industries continue to transition to data-driven solutions, it becomes necessary for ML practitioners to build algorithms that enable systems to learn and make informed decisions using data available in the public domain. Companies of all sizes, many industries from large technology firms, as well as aspirational start-ups have commenced aggressive searches for qualified ML employees who can build competitive products and improve workflow efficiency [1, 2]. On one hand, companies are looking for qualified and capable professionals to become a part of the team and contribute to the development of industry; on the other hand, prospective job seekers are compelled to visit individual company job opening portals and search for whether their skills match with the requirements.

This is not only time-consuming but also leads to missed opportunities for both candidates and companies. A customized portal to assist job seekers for

the right job and companies for suitable candidates is missing. The objective of this study is to focus on predicting the availability and demand for ML job postings from Glassdoor. A thorough analysis of data and implementation of ML frameworks for prediction of trends in ML job listings [3] is proposed in this work. This research would be invaluable to job seekers trying to sharpen their specific skill sets [4]. It offers empirically grounded insight into recruitment initiatives for organizations interested in acquiring talent, as well as for educational institutions based on market need.

Classification of jobs is an approach used by researchers working in similar domains. It enables to classify jobs posted in the public domain and categorize them based on requirements [5–7]. The authors of [7] have successfully leveraged Random Forest classifiers for detecting and categorizing genuine and fake job postings with an accuracy of 97.8%. Classification of jobs in different categories as per requirements has also been achieved with good accuracy. However, in addition to classification, job seekers are also more interested to predict the scope of job they might be able to apply for career advancements. Job seekers are looking for an interactive user interface that enables a user to input qualifications to receive job title-based tailored predictions [8, 9]. Being an interactive process [10], it enhances usability and functionality and therefore makes it a useful instrument for job seekers and professionals who target refined job search strategies. This research provides improved methodologies for career development and job search techniques [11–13]. It uses a rich dataset and a trained Random Forest regressor to finally return probably the best job positions an individual would qualify for. It further goes on to investigate the related factors that include company rating, size, year founded, revenue, and skills required. The goal would be to give the job applicant the tool, which turns out to be most useful for determining which positions one would be suited for, given certain input parameters. The purpose of this work is to connect job seekers with employers, provide contextualization of ML jobs based on statistical analysis, predictive modeling, serve an impactful approach for career advancements, workforce recruitment strategies, and future projections to assist in mastery of the rapidly evolving sector of ML that has increasingly penetrated all aspects of society.

II. Literature Review

In a competitive job market—especially in aspects of ML—applicants find it hard to find work offers that fall in the category of matching knowledge and skills. On their part, employers experience problems trying to determine the potential candidates who would fulfill the skills required by a job related to ML. This project will predict titles of jobs and salary predictions leveraging ML algorithms. To this end, analyzing important features of a job, like company ratings, size, foundation year, revenue generated, and skills, means that the project intends to create a tool for job seekers for identifying which job titles are well-matched with their qualification profiles. Furthermore, the tool will enable candidates to submit their information so that customized job predictions can be generated, hence linking the job seekers with the job opportunities in ML (Table 1).

III. Proposed Methodology

The methodology adopted for predicting jobs posted in Glassdoor or any other public domain comprises collecting dataset, preprocessing, and exploratory data analysis (EDA). The next stage is feature engineering, model building, and performance evaluation. Predictions of job or salary as required is the final stage. The methodology is outlined in Figure 1 and is covered in detail in the following.

a. Data collection

The dataset is collected from the Kaggle Glassdoor website and has 1,500 samples, with each sample consisting of 12 features. The dataset is basically a customized dataset posted in Glassdoor for computer science and Information Technology (IT) professionals. The authors ascertain that the proposed model will be equally effective for predictions of jobs postings in different domains as well. The dataset's attributes and their descriptions are provided in Table 2.

b. Data preprocessing

Before adopting ML algorithms, the gathered data is required to go through preprocessing. This comprises cleaning the data, eliminating erroneous entries or duplicates, and taking care of missing values and outliers. Certain columns of the dataset included missing or placeholder values. As a result, they were replaced with an average rating to fill the gaps. Similarly, outliers, if any, are also taken care of; they are either removed or replaced with an average value [14, 15].

As shown in Figure 2, Company Rating feature follows the Gaussian/Normal Distribution, but as there are few outliers, ranging from –1 to 5; –1 values are

Table 1: Work done by different researchers in similar domain

Ref. No.	Methodology used	Domain	Dataset used	Performance/outcome
[1]	Linear regression, Lasso, random forest	Salary prediction for Data Science Job	Kaggle—Glassdoor	MAE: For random forest—11.22, for linear regression—18.86, for ridge regression—19.67
[2]	SVM	Skill based job recommendation system	Job portals, company websites, scraping data from other online sources	Accuracy, precision, recall, and F1 score was calculated
[3]	Bidirectional, decoder-encoder, stacked, Conv LSTM	Trend analysis system to predict future job markets using historical data	Web scraping, manually collecting data, government sources	Accuracy: for bidirectional LSTM—95.71%, for decoder–encoder LSTM—91.56%, for stacked LSTM—87.24%, for Conv LSTM—83.7%
[4]	NB, KNN, NBST	Predictive analysis	Student employment in the employment market of Chongqing S colleges and universities in the past 3 years	Mean value [test time (ms)]: NB—18.607, KNN—22.224, NBST—49.026
[5]	MNB, SVM, DT, KNN, RF	Job posting classification	Kaggle, titled by "[real or fake] fake job posting prediction"	For MNB 95.6%, for SVM 97.7%, for DT 97.4%, for KNN 97.8%, for 98.2%, for RF 98.2%
[6]	LR, SVM, KNN, DT, RF, AdaBoost(DT), GB, voting classifier soft & hard, XGBoost	Campus placement analyzer: Using supervised machine learning algorithms	Training and placement department of MIT which consists of all the students of Bachelor of Engineering (B.E) from three different colleges of their campus	Accuracy: Logistic Regression 58%, support vector machine 69%, KNN 63.22%, decision tree 69%, random forest 75.25%, AdaBoost(DT) 77%, gradient boosting 77%, voting classifier soft 69.11%, voting classifier hard 68.43%, XGBoost 78%
[7]	Voting classifier	Ensemble approach for classifying job positions	Glassdoor website	For voting classifier soft—100%
[8]	NB, SGD, LR, KNN, RF classifier	Detecting and preventing fake job offers	Kaggle—real/fake job posting prediction	For random forest classifier—97.48%
[9]	NLP, KNN	Resume-based job recommendation system using NLP and deep learning	Combined from multiple sources	Improving the efficiency and success rate of the hiring process

DT, Decision Tree; GB, Gradient Boost; KNN, K Nearest Neighbor; LR, Logistic Regression; MAE, mean absolute error; NB, Naive Bayes; NBST, Naive Bayes Support Tree, NLP, Natural Language Processing; RF, Random Forest; SGD, Stochastic Gradient Descent; SVM, Support Vector Machine.

replaced with the median of the distribution. A similar procedure is done with other attributes as well.

There were columns containing textual information such as job descriptions, firm names, and locations. Cleaning these text columns involved removing and standardizing job location information. Extracted keywords were used to categorize positions based on job title, seniority level, and other relevant features (e.g., "Data Scientist," "Data Engineer"). This categorization reduced the number of title options and offered more structured information needed for modeling. This step of data preprocessing further refines

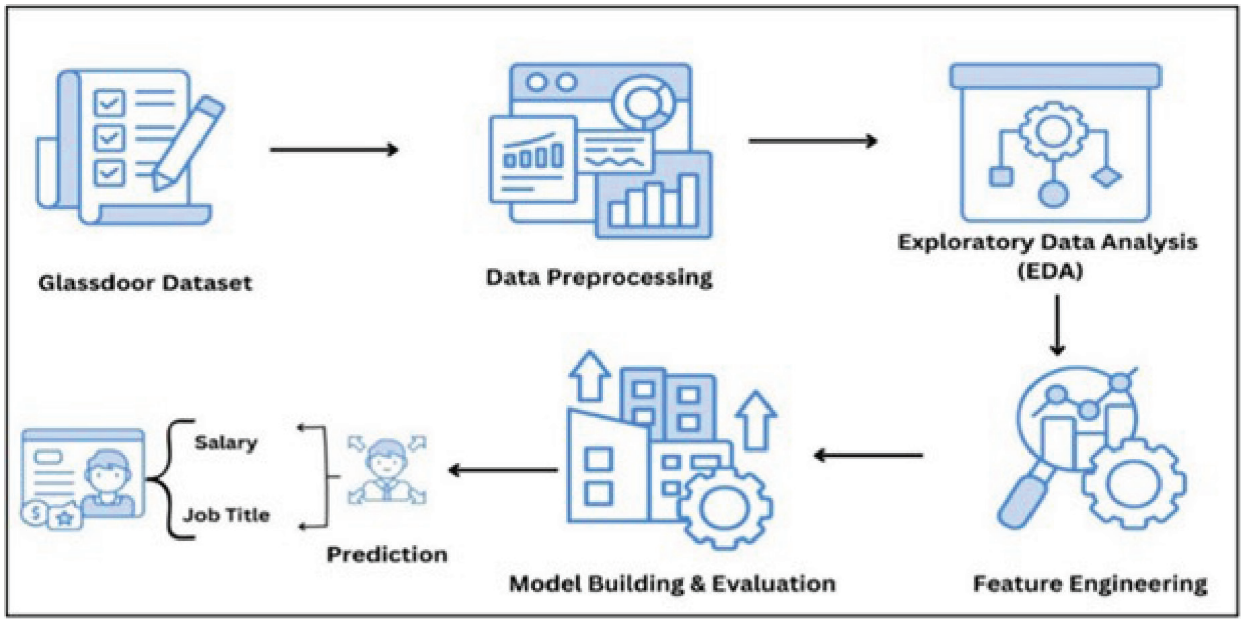

Figure 1: Methodology of proposed work. EDA, exploratory data analysis.

Table 2: Description of attributes present in dataset

S. No.	Name of attribute	Description
1.	Job title	The designation of the job being listed. E.g., data scientist, data engineer, other, manager, Director, Machine Learning Engineer
2.	Salary estimate	The estimated salary range for the job provided by Glassdoor/Employer
3.	Job description	The full description of the job, including roles, responsibilities, and qualifications
4.	Rating	Rating of the company, from employee reviews on Glassdoor. Initial reviews range from –1 to 5
5.	Company name	The name of the company offering the job. E.g., IBM, New York (United States of America), Adobe, Microsoft etc.
6.	Location	The location of the job E.g., Remote (San Jose, CA, USA), (Atlanta, GA, USA)
7.	Size	The number of employees at the company E.g., 1,001–5,000 employees, 10,000+ employees
8.	Founded	The year the company was founded
9.	Type of ownership	The ownership structure of the company E.g., private, public, government
10.	Industry	The specific industry the company operates in E.g., Telecommunications Services, Chemical Manufacturing, Computer Hardware Development
11.	Sector	The broader sector associated with the company's operations E.g., Education, Information Technology, Manufacturing
12.	Revenue	The estimated annual revenue of the company in US$

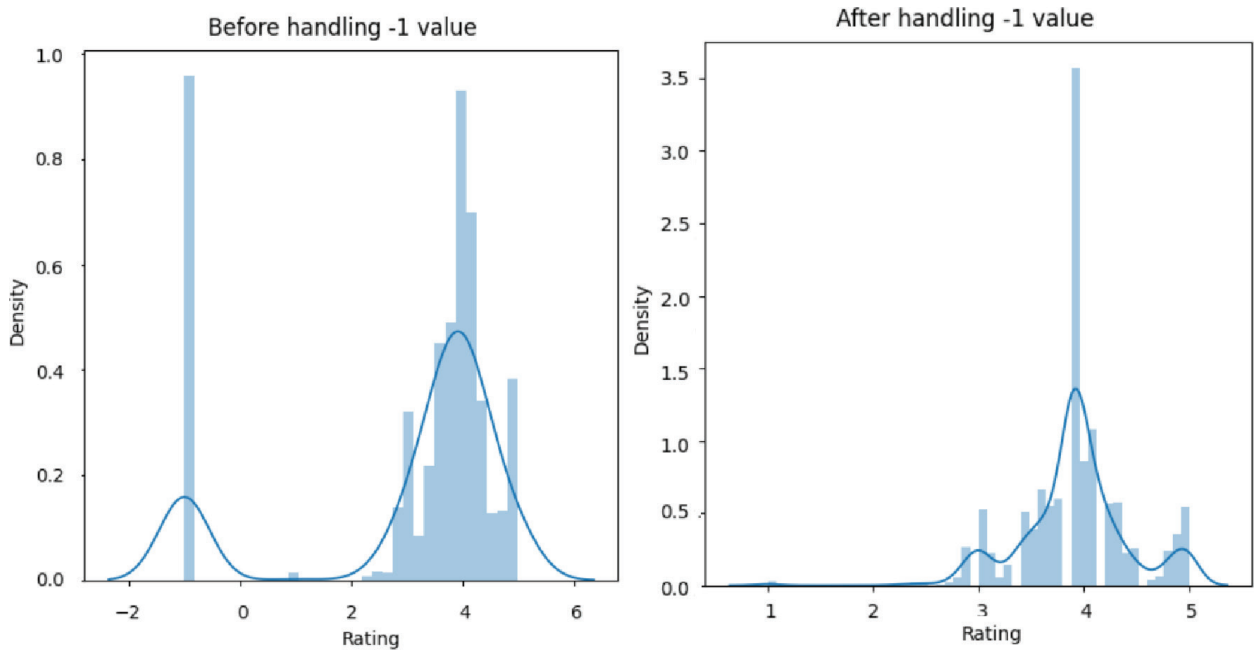

Figure 2: Representation pre and post handling of outliers in current dataset.

and clears the data before they are further processed and explored for gaining insights. The more data are organized and systematically arranged during the course of preprocessing, the better the results expected from models deployed for the data.

c. EDA

EDA involves describing the data using statistical and visual aids to highlight key elements for additional examinations.

In the process, it is critical that one has a thorough comprehension of the characteristics of the data, including its statistical features and schema and the data quality, including irregular data types and missing values and the data's capacity for prediction, as demonstrated by the feature-to-target correlation [16]. Figure 3 shows the plots of attribute "Job Title" and its various categories, whereas Figure 4 shows that of attribute "company revenue."

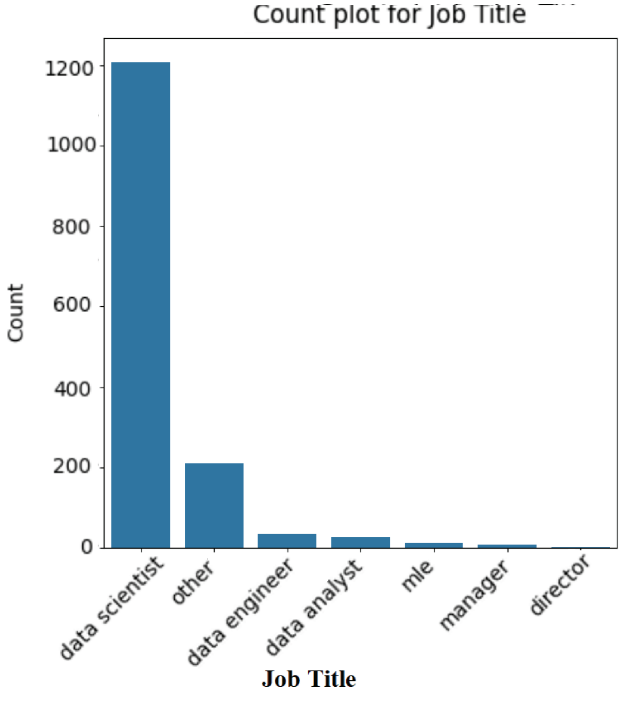

Figure 3: Count plot of attribute "Job Title."

d. Feature engineering

Feature engineering is the process of creating new features or altering existing ones to better indicate the data's underlying patterns [17]. New features were developed to show whether a job advertisement was placed at the company's headquarters. This tool can disclose whether employment role and wages differ based on proximity to the main office. Columns for critical technical skills have been added, including Python-Software Foundation, Excel-Microsoft Corporation, SQL-Salesforce, and Tableau-Salesforce. Feature engineering in the current research involves the following.

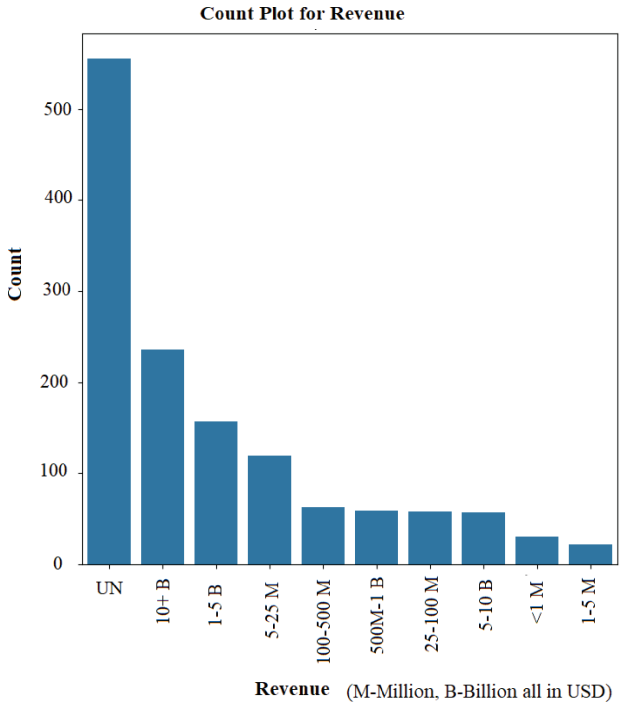

Figure 4: Count plot of attribute "Company revenue" (UN-unknown).

a.i. Creating new features from existing features

For current dataset columns such as Python, Excel, and SQL, tableau jobs are created to have a broader knowledge of features. This is shown in Figure 5. It can be seen that there is only one entry for Tableau.

a.ii. Normalizing the column salary

The salary column consists of data represented in different forms. A few examples are shown in Table 3. Glassdoor has either estimated the salaries where the employer has not mentioned it or it is stored as -1 or Not a Number (NaN). It can be seen that the salaries are represented as a range, a constant value, salary per hour, etc. All these are converted to a single number, that is, salary per annum. The NaN values are replaced by the median salary as explained earlier, the range of salaries is replaced by the average. The salary per hour is converted to salary per annum.

a.iii. Trimming columns

Features having more than 10 categories are trimmed to reduce the dimensionality. This will further reduce the complexity and response time of the final model. Figure 6 gives a representation of

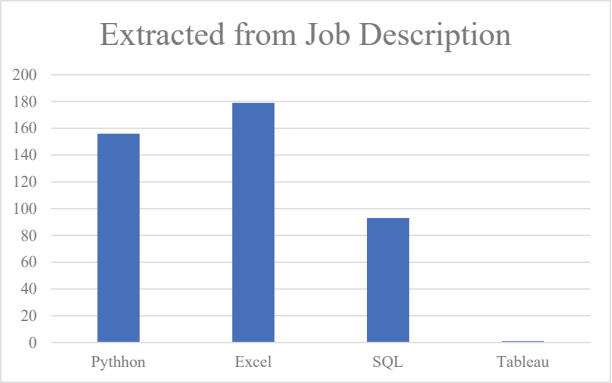

Figure 5: Number of jobs posted on different domains.

Table 3: Normalization of column—salary estimate

Original value	Value after normalization
-1	116.0 (median)
$100 K–$151 K (Glassdoor est.)	125.5
Employer provided salary: $100 K–$120 K	110
Employer provided salary:$107 K	107
Employer provided salary: $60.00 per hr	140.4
Employer provided salary: $53.62–$64.58 per hr	138.3

"Type_of_Ownership" before and after trimming. In this case, the similar categories are combined to form a single category. For example, the category Private Practice/Firm is merged with Company-Private.

a.iv. Handling ordinal and nominal categorical features

Ordinal features belong to categorical variables that have some ordering (size, rating, remark, etc.). Therefore, these features were ordered, necessitating the mapping of these categories to numerical values. By contrast, Nominal features are categorical variables that hold no numerical meaning (employment, name, etc.). Therefore, these features were ordered, with the mapping of these categories to numerical values [18]. For example, there are 257 unique job locations in the dataset, while most of them have overlapping content. San Jose (CA, USA), Los Angeles (CA, USA),

Type of ownership	count
Company - Private	618
Company - Public	488
Nonprofit Organization	70
-1	47
Hospital	43
Subsidiary or Business Segment	28
Government	22
College / University	21
Self-employed	6
Private Practice / Firm	5
Unknown	4
Contract	4

Type of ownership	count
Private	623
Public	488
Nonprofit Organization	70
Other Organization	55
Hospital	43
Subsidiary or Business Segment	28
Government	22
College / University	21
Self-employed	6

Figure 6: Type of ownership—before and after trimming.

and San Francisco (CA, USA) can be combined to form one entry—CA. With this, the number of unique job locations reduces to 57. Figure 7 shows the top 15 Job Locations. Similarly, Jobs in various sectors are shown in Figure 8, Number of companies with different sizes (i.e., number of employees working in it) are shown in Figure 9.

a.v. Feature selection

This stage involves selecting the significant features for final prediction as per requirements. In our research we have estimated the salary by using the attributes rating, founded, sector, ownership, job title, and job skills (extracted from job description).

e. Model development and testing

A range of ML techniques were applied to construct prediction models of employment focused outcomes, including salary prediction and job title prediction. In the model development outline, Lasso regression [19] was used as a baseline pay prediction model. This method implements L1 regularization that imposes penalties on larger coefficients to select features and discourage overfitting. Random Forest and XGBoost-DMLC (Distributed Machine Learning Community), the algorithms that can be used for both classification

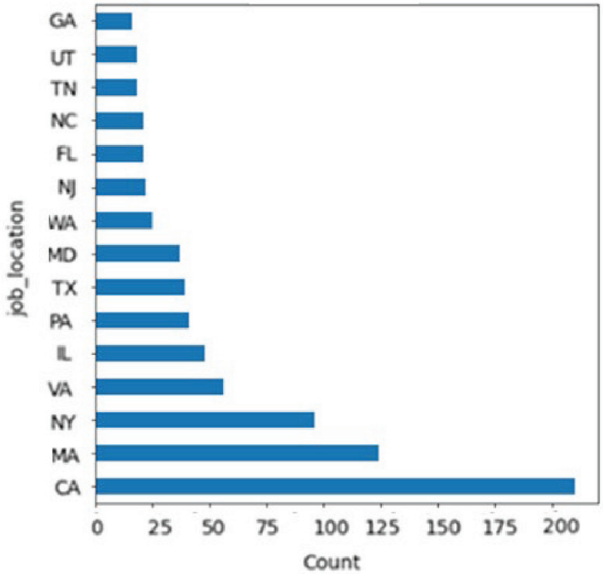

Figure 7: Top 15 job locations preferred by employees.

and regression are also implemented. Random Forest can handle large amounts of data and a large number of features effectively. XGBoost performs well with structured data, and is helpful in predicting continuous variables [20, 21]. LightGBM-Microsoft Corporation is a hyper parameter-tuning technique for training a

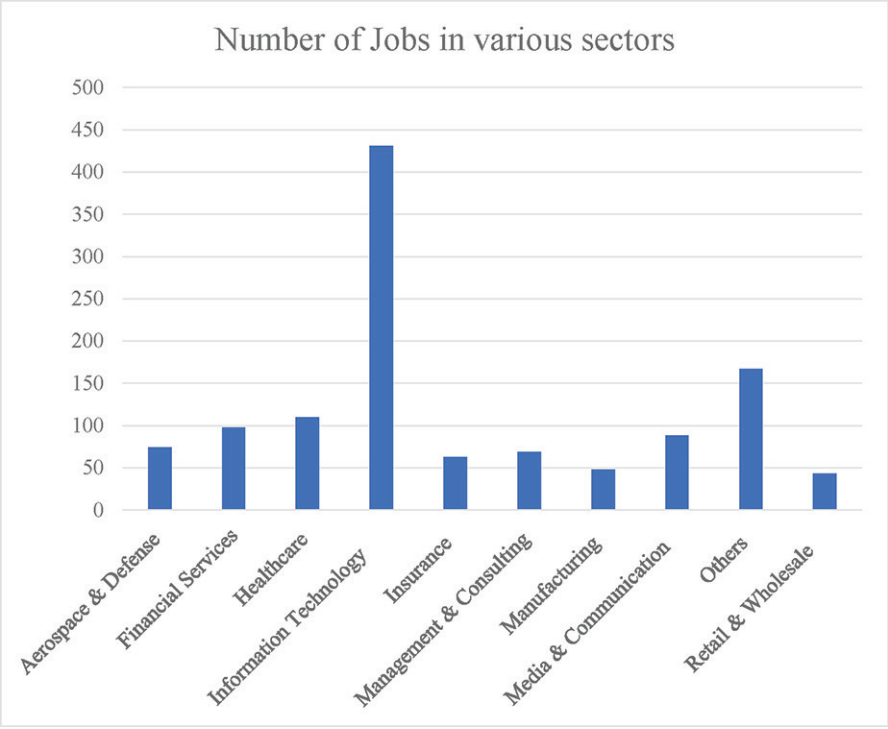

Figure 8: Number of Jobs in various sectors.

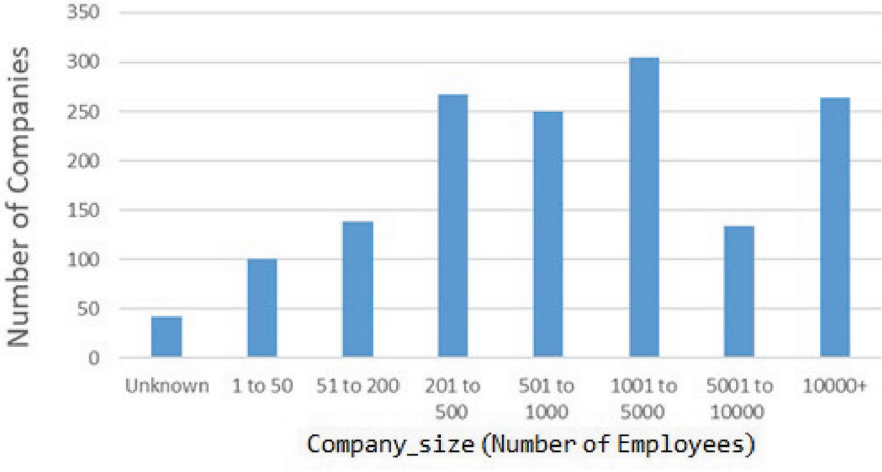

Figure 9: Plot based on Company_size.

gradient-boosted framework that estimates job salaries and is principally designed to be efficient and rapid, which is beneficial for lesson dataset sizes. An Ensemble methodology for prediction performance was additionally applied in the model outline, specifically using Voting Regressor for ensembles, including Random Forest and Gradient Boosting [22, 23]. The intent of the approach is to combine predictions from different algorithms while taking advantage of their strengths. To have better results and to enhance model performance, 10-fold cross-validation was done and the results evaluated with and without cross-validation. In the 10-Fold Cross-Validation method, the data are split into 10 equal parts (or folds). The model is trained on 9 folds and tested on the remaining fold, iterating this process 10 times, with each fold being used as the test set exactly once. This is proved to be more reliable and reduces risk of over fitting.

To measure the performance of these models, we used six performance metrics. Accuracy, the ratio of

correct predictions to total predictions is the most commonly used. It is used when the dataset is balanced. Low accuracy indicates a significant number of wrong predictions by the model, but sometimes it may be a result of over fitting, under fitting, or imbalanced data. Root mean squared error (RMSE) measures the difference between the actual values and the predicted values, with an emphasis on large errors by squaring them before taking the square root. Normalized root mean square error (NRMSE) adjusts the RMSE by dividing it by a measure of the scale of the data. This normalization makes NRMSE useful for comparing models across different datasets or situations where the scale of the target variable is different. Mean absolute error (MAE) measures the average magnitude of the errors in a set of predictions that tells how far off the predictions are from the actual values on average. A lower RMSE/NRMSE/MAE value indicates a better model, as it suggests that the model's predictions are closer to the actual values. R-squared (R^2), also known as the coefficient of determination, tells how well the model's predictions match the actual data and how much of the variance in the target variable is explained by the model. The value should be near to 1 for better performance. In some rare cases when the data are poorly fitted, it is negative. Standard deviation (SD) indicates the spread or dispersion of data points from the mean. It is used as a metric of prediction performance, with the implication and/or prolongation of predicting salary due to potential outcomes becoming negative based on deviations from real salaries [24]). A low SD means that the data values are close to the mean, while a high SD means that the data values are spread out over a wide range [25].

f. Prediction

After selection of the best performing model, an analysis of the model is done to assess the outcome [26]. A fresh set of attributes are fed to the selected model for generating a customized response as per the need. The outcome based on input fed to the model is shown in Figures 10 and 11, respectively.

g. Algorithm for methodology

Input: Raw Glassdoor Dataset
Output: Predicted Job Title and Salary

1. *Load the Glassdoor dataset*
2. *Perform Data Preprocessing:*
   a. *Handle missing values and outliers (e.g., drop or impute)*
   b. *Normalize/scale numerical features*
   c. *Encode categorical variables (e.g., one-hot or label encoding)*
   d. *Remove duplicates and irrelevant columns*

```
# Prediction
# Input sequence: 'company_rating', 'company_founded', 'company_sector','company_ownership', 'job_title', 'job_skills'
salary = predict_salary(3.0, 2000,'Health Care', 'Public', 'data analyst',['python', 'tableau'])
print('Estimated salary (range): {}(USD) to {}(USD) per annum.'.format(int(salary*1000)-9000, int(salary*1000)+9000))
```

```
Estimated salary (range): 84740(USD) to 102740(USD) per annum.
```

Figure 10: Sample result obtained for predicting salary.

```
Enter company rating (e.g., 4.2): 4
Enter company size (e.g., 500): 300
Enter company founded year (e.g., 1990): 2021
Enter company revenue (in millions, e.g., 100): 100
Enter expected salary (e.g., 75.0 for $75,000): 70
Enter job skills (comma-separated, e.g., python,sql): sql
Enter company ownership (e.g., Private, Public, Nonprofit Organization): private
Enter company sector (e.g., Information Technology, Finance): Information Technology

Predicted job title: job_title_data scientist
```

Figure 11: Sample result obtained for predicting job title.

3. *Conduct Exploratory Data Analysis (EDA):*
   a. *Visualize distributions, outliers, and correlations*
   b. *Identify feature importance or selection criteria*
4. *Apply Feature Engineering:*
   a. *Extract relevant features (e.g., company rating, location, skills)*
   b. *Create derived features (e.g., job level from title)*
   c. *Handle ordinal and nominal categorial features*
   d. *Select Relevant Features*
5. *Split the dataset into training and testing sets (e.g., 80/20 split)*
6. *Select and train multiple machine learning models:*
   a. *Models = {Random Forest, Lasso, LightGBM, XGBoost, Voting(Gradient Boosting + Random Forest.}*
   b. *Train each model on training data.*
   c. *Metrics = {Accuracy, RMSE, NRMSE, R^2,MAE, SD}*
   d. *Select best-performing model(s)*
7. *Predict salary and job title for new/unseen data*

IV. Results, Analysis, and Discussion

A comparative analysis of the models is done before finalizing the selected model for predictions. Table 4 gives a tabular representation of model performance

Table 4: Comparison of model performance

Model performance	Accuracy	RMSE	NRMSE	R^2	MAE	SD
Random Forest	0.9853	0.0646	0.1966	0.8133	0.0166	0.0646
Lasso	0.8750	0.2103	0.6402	0.0061	0.0888	0.2103
LightGBM	0.9559	0.4441	1.3520	0.5373	0.1160	0.4430
XGBoost	0.9963	0.1819	1.3108	0.9224	0.1113	0.1816
Voting	0.9963	0.0646	0.5501	0.9234	0.0117	0.1803

MAE, mean absolute error; NRMSE, normalized root mean square error; RMSE, root mean squared error; SD, standard deviation.
The underlined values indicate the maximum values in the respective columns which in turn shows the best performing model.

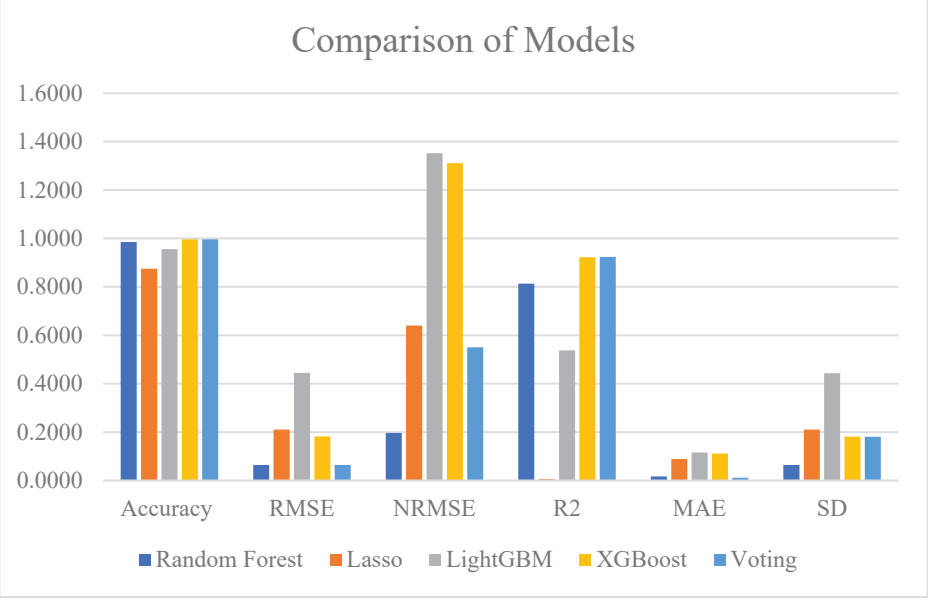

Figure 12: Representation of performance metrics. MAE, mean absolute error; NRMSE, normalized root mean square error; RMSE, root mean squared error; SD, standard deviation.

obtained after 10-fold cross-validation [27, 28]. The table shows that among the five ML models used in this study, the performance of Random Forest exceeds that of others in terms of RMSE and NRMSE, and it is the second performer with respect to MAE. The other measures can be avoided as the data are not balanced (as expected for better accuracy), and it is categorical (as not expected for SD). So Random Forest is the model proposed for final prediction of job title and salary for a candidate searching for a job. The ensemble model, the Voting Classifier with Gradient Boosting and Random Forest models, performs well for predicting Jobs for jobs seekers as well as for tentative candidates who are suitable for the company's requirements. The model can be customized as per the requirements of employer or employee [29]. A comparative analysis of work done in a similar domain is listed in Table 4. A graphical representation is shown in Figure 12 for better understanding.

V. Conclusion

The objective of this project is to predict job characteristics as required for employees based on salary and job title utilizing multiple ML algorithms from the job postings dataset. After engaging in extensive data preprocessing, feature engineering, and model evaluation, various models were applied such as Random Forest, Lasso regression, LightGBM, XGBoost, and subsequently a Voting Regressor aggregation involving Random Forest and Gradient Boosting. Results demonstrated that Random Forest produced the most accurate results based on its lowest average NRMSE and SD. Alternatively, the ensemble techniques, such as the Voting Regressor, successfully optimized accuracy while maintaining stability, indicating the value of aggregating models to improve prediction consistency. The proposed model is able to predict job title and salary with minimum error. It can also be customized as per the needs of the individual based on their specific skills, as well as for the requirements of the industry. Furthermore, there is need for integration of external labor market indicators (e.g., industry trends, regional demands) for more robust prediction, and the development of a real-time job recommendation system using the trained model for deployment in recruitment platforms. The findings support using newer ensemble techniques in the job outcome predictions.

VI. Future Scope

In the future, this work can be extended in several impactful directions to enhance its practical relevance and predictive performance. One promising avenue is the integration of external labor market indicators, such as regional employment trends, inflation rates, and sector-specific demands, to improve the contextual accuracy of salary and job title predictions. Additionally, the model can be adapted for real-time applications, such as intelligent job recommendation systems for recruitment platforms, enabling dynamic matching of candidates with suitable roles based on their skills and preferences. Further improvements can be achieved through advanced hyperparameter tuning and the application of deep learning models or transformer-based architectures for a richer semantic understanding of job descriptions. Expanding the dataset to include more diverse job domains and geographies will also enhance generalizability. Ultimately, these enhancements will help build a robust, adaptable system that responds effectively to the evolving landscape of the global job market.

References

[1] Nasser, I., & Alzaanin, A. H., "Machine learning and job posting classification: A comparative study", International Journal of Engineering and Information Systems (IJEAIS), Vol. 4, Issue 9, pp. 6–14, September 2020.

[2] Singh, A., Sharma, P., & Jain, M., "Predicting student placement success using ensemble machine learning techniques: A comparative study", Journal of Intelligent & Fuzzy Systems, Vol. 46, No. 2, pp. 2311–2322, 2024

[3] Makhi, M. V., Dhumal, P., Jagdhane, U., & Mane, S., "Skill Based Job Recommendation System", Vol. 11, Issue 3, 2023.

[4] Walecha, W., & Gupta, B., "Salary Estimator using Data Science", International Journal for Modern Trends in Science and Technology, Vol. 6, Issue 12, pp. 319–322, December 2020.

[5] Mohammed, A. K., Danlami, A. A., Saeed, D. I., Lawan, A. A., Hussaini, A., & Mohammed, R. K., "An Ensemble Machine Learning Approach for Classifying Job Positions", Academic Journal of Nawroz University, Vol. 12, Issue 3, pp. 547–555, August 2023.

[6] Tee, Z., & Raheem, M., "Salary prediction in data science field using specialized skills and job benefits–a literature", Journal of Applied Technology and Innovation, Vol. 6, Issue 3, pp. 70–74, July 2022.

[7] Kablaoui, R., & Salman, A., "Machine learning models for salary prediction dataset using python", International Conference on Electrical and Computing Technologies and Applications (ICECTA), Ras Al Khaimah, UAE, pp. 143–147, November 2022.

[8] Satpute, B. S., Yadav, R., & Yadav, P. K., "Machine Learning Approach for Prediction of Employee Salary using Demographic Information with Experience", IEEE Global Conference for Advancement in Technology (GCAT), Bangalore, India, pp. 1–5, December 2023.

[9] Wang, G., "Employee Salaries Analysis and Prediction with Machine Learning", International Conference on Machine Learning and Intelligent Systems Engineering (MLISE), Guangzhou, China, pp. 373–378, August 2022.

[10] Asaduzzaman, A., Uddin, M. R., Woldeyes, Y., & Sibai, F. N., "A Novel Salary Prediction System Using Machine Learning Techniques", Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON), Chiang-Mai, Thailand, pp. 38–43, February 2024.

[11] Kumar, G., Agrawal, R., Sharma, K., Gundalwar, P. R., Kazi, A., Agrawal, P., Tomar, M., & Salagrama, S., "Combining BERT and CNN for Sentiment Analysis: A Case Study on COVID-19", International Journal of Advanced Computer Science and Applications, Vol. 15, No. 10, 2024.

[12] Pawar, L., Saw, A. K., Tomar, A., & Kaur, N., "Optimized features-based machine learning model for adult salary prediction", IEEE International Conference on Data Science and Information System (ICDSIS), Hassan, India, pp. 1–5, July 2022.

[13] Yang, S., "Automated employee salary prediction algorithm based on machine learning", International Conference on Computer Vision, Application, and Algorithm (CVAA), Chongqing, China, pp. 243–249, April 2023.

[14] Nguyen, T., Zhang, H., & Lee, Y., "Context-Aware Job Matching and Salary Prediction Using BERT and Graph Neural Networks", Information Processing & Management, Vol. 62, No. 1, 103215, 2025.

[15] Ayua, S. I., Malgwi, Y. M., & Afrifa, J., "Salary Prediction Model for Non-Academic Staff Using Polynomial Regression Technique", Artificial Intelligence and Applications, Vol. 2, Issue 4, pp. 330–337, June 2023.

[16] Buditjahjanto, I. G. P. A., Pratama, A., & Samani, M., "The Intelligent Decision System Based on Hybrid Decision Tree to Determine The Level of Lecturer Performance", International Journal of Advances in Soft Computing and its Application, Vol. 16, Issue 1, pp. 219–232, March 2024.

[17] Saeed, A. K. M., Abdullah, P. Y., & Tahir, A. T., "Salary Prediction for Computer Engineering Positions in India", Journal of Applied Science and Technology Trends, Vol. 4, Issue 1, pp. 13–18, February 2023.

[18] Zaqaibeh, B., "Assessing and achieving intended learning outcomes against the NQF case of CS program at Jadara university", International Journal of Advances in Soft Computing and its Application, Vol. 14, Issue 2, pp. 139–151, June 2022.

[19] Loukili, M., Messaoudi, F., & El Ghazi, M., "Supervised Learning Algorithms for Predicting Customer Churn with Hyperparameter Optimization", International Journal of Advances in Soft Computing and its Application, Vol. 14, Issue 3, pp. 49–63, November 2022.

[20] Park, J., Feng, Y. and Jeong, S.P., 2024. Developing an advanced prediction model for new employee turnover intention utilizing machine learning techniques. Scientific Reports, 14(1), p. 1221.

[21] Parida, B., KumarPatra, P. and Mohanty, S., 2022. Prediction of recommendations for employment utilizing machine learning procedures and geo-area based recommender framework. Sustainable operations and computers, 3, pp. 83-92.

[22] Chen, J., Mao, S., & Yuan, Q., "Salary prediction using random forest with fundamental features", Third International Conference on Electronics and Communication; Network and Computer Technology (ECNCT), Harbin, China, Vol. 12167, pp. 491–498, March 2022.

[23] Ji, Y., Sun, Y., & Zhu, H. (2025, April 9). Enhancing Job Salary Prediction with Disentangled Composition Effect Modeling: A Neural Prototyping Approach. Arxiv. https://arxiv.org/html/2503.12978v1

[24] Ramachandran, R. (2023, October). *Salary Estimator Using ML Algorithms*. ijrpr. https://ijrpr.com/uploads/V4ISSUE10/IJRPR18476.pdf.

[25] Gopal, K., Singh, A., & Kumar, H. (2021, June). *Salary Prediction Using Machine Learning*. ijirt. https://ijirt.org/master/publishedpaper/IJIRT151548_PAPER.pdf.

[26] Malgwi, Y. M., & Afrifa, J., "Salary prediction model for non-academic staff using polynomial regression technique", Artificial Intelligence and Applications, Vol. 2, No. 4, pp. 330–337, 2024.

[27] Sankhe, V., Shah, J., Paranjape, T., & Shankarmani, R., "Skill based course recommendation system", International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, pp. 573–576, October 2020.

[28] Siswanto, J. V., Castilani, L. A., Winata, N. H., Nugraha, N. C., & Sagala, N. T. M., "Salary Classification & Prediction based on Job Field and Location using Ensemble Methods", International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), Jakarta, Indonesia, pp. 325–330, February 2023.

[29] Wang, Z., Sugaya, S., & Nguyen, D. P. T., "Salary Prediction using Bidirectional-GRU-CNN Model", Computer Science, Business, pp. 292–295, Assoc. Nat. Lang. Process., 2019.