JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.8979
Hal 949-958
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Predicting AI Job Salary Classes Through a Comparative Study of Machine Learning Algorithms

Vincent[1], Robet[1*], Edi Wijaya[2]

[1] Informatics Engineering, STMIK Time, Medan, Indonesia
[2] Information System, STMIK Time, Medan, Indonesia
Email: [1]ciasent45@gmail.com, [2*]robertdetime@gmail.com, [3]wiwileosummer@gmail.com
Email Penulis Korespondensi: robertdetime@gmail.com
Submitted **04-08-2025;** Accepted **11-12-2025;** Published **31-12-2025**

Abstract

The rapid growth of Artificial Intelligence (AI) has brought significant transformation to the global job market, particularly in salary structures across various AI-related professions. This study aims to classify AI job salaries into three categories—Low, Medium, and High—using supervised machine learning algorithms. The dataset, sourced from Kaggle, combines two real-world datasets featuring key attributes such as experience level, job type, education level, technical skills, remote work ratio, and salary in USD. Preprocessing techniques include One-Hot Encoding for categorical data, StandardScaler for normalization, and MultiLabelBinarizer to handle multi-skill entries. Four machine learning models—Logistic Regression, Random Forest, Gradient Boosting, and XGBoost—were trained and evaluated using consistent pipelines, with evaluation metrics including accuracy, precision, recall, and F1-score, applying macro-averaging to address class imbalance. Logistic Regression achieved the highest performance with 85.4% accuracy and 77.6% F1-score, followed by Gradient Boosting with 84.8% accuracy and 76.3% F1-score. High-salary classes were predicted with higher precision and recall than low-salary classes, indicating skewness in class distribution. Feature importance analysis shows that experience, remote work ratio, and key skills such as Python and SQL significantly affect prediction accuracy. This study demonstrates that traditional machine learning methods, when applied with appropriate preprocessing, can effectively support salary classification and labor market analysis in the AI domain.

Keywords: Salary Prediction; Artificial Intelligence; Machine Learning; Classification; Feature Importance

1. INTRODUCTION

The rapid development of Artificial Intelligence (AI) over the past decade has significantly changed the global job market landscape [1], [2], [3]. Demand for AI-related professionals such as machine learning engineers, data scientists, AI researchers, and other specialized roles continues to rise as companies across industries seek to leverage intelligent systems for business growth and operational efficiency [4]. This growing demand has also led to wide variations in salary structures for AI-related positions, including emerging evidence that job complexity and skill sets influence wage levels [5], which are influenced not only by job role but also by factors such as experience level, company location, type of work, educational background, and employee technical skills.

Predicting salary classes, particularly within the AI domain, is an emerging and relevant task that offers substantial value to multiple stakeholders [6], [7]. Accurate classification of salaries into categories such as Low, Medium, and High can aid job seekers in making informed career decisions, assist employers in setting competitive compensation standards [8], and help policymakers and academic institutions understand labor market dynamics. This kind of prediction task is well-suited to be approached using machine learning (ML) methods [9], [10], which have shown strong performance in classification problems involving structured and semi-structured data.

Machine learning provides a powerful suite of tools for supervised learning tasks such as salary prediction [11], [12], [13]. Different algorithms offer different strengths depending on the nature and structure of the data. In this study, we conduct a comparative analysis of four supervised learning algorithms for predicting salary classes in AI-related jobs. The models include: Logistic Regression, a linear model often used as a baseline for classification tasks; Random Forest, an ensemble method based on decision trees that improves generalization through bagging; Gradient Boosting, which builds models sequentially to correct the errors of prior models; and XGBoost, an optimized and regularized version of Gradient Boosting [9] that is known for its high efficiency and performance on large-scale structured data.

The dataset used in this research was obtained from Kaggle [4], titled “Global AI Job Market and Salary Trends 2025”, published by Bisma Sajjad. The dataset contains comprehensive information on global AI jobs, including fields such as experience level, employment type, company location, remote work ratio, required technical skills, education level, industry, and years of experience, along with the annual salary in USD. A unique feature of this dataset is the required_skills column, which contains multiple skills per job entry. These multi-label entries require specific preprocessing methods such as MultiLabel Binarization to make them suitable for machine learning models.

This study involves several stages of preprocessing and data transformation to prepare the dataset for classification modeling [10], [14]. Categorical variables are encoded using One-Hot Encoding, while numerical features are normalized using StandardScaler. The required_skills column is converted using MultiLabelBinarizer to handle multiple binary features. The processed dataset is then split into training and test sets using stratified sampling, ensuring the proportion of each salary class is maintained in both subsets.

Each of the four models is trained and evaluated using the same dataset and feature pipeline to ensure a fair comparison [15]. The evaluation metrics include accuracy, precision, recall, and F1-score, with a focus on the macro-

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.8979
Hal 949-958
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

average scores to account for potential class imbalance. In addition, detailed per-class evaluations (for Low, Medium, and High salary classes) are presented to analyze the models' ability to detect each category. Confusion matrices are also generated to visualize prediction performance, and feature importance analysis is conducted for tree-based models to identify which variables most strongly influence the classification outcomes.

Previous studies in salary prediction have primarily focused on broader job markets and have not specifically examined roles within the AI sector. Many existing works rely on simplified feature sets that exclude multi-skill attributes, despite the fact that technical skill combinations play a major role in determining AI-related compensation. In addition, earlier research generally utilizes only one or two machine learning models, resulting in limited insights into comparative model behavior under controlled conditions. This indicates a gap in the literature regarding how multiple supervised learning algorithms perform when applied to a feature-rich, domain-specific dataset that includes multi-label skill information.

To bridge this gap, the present study evaluates four commonly used supervised learning algorithms using a standardized preprocessing pipeline and identical data partitions. Unlike previous research, this study incorporates multi-label skill features, applies consistent evaluation metrics across models, and conducts a structured, side-by-side comparison. These methodological differences allow for a more comprehensive understanding of how each algorithm handles the complexity of AI-related salary classification, offering contributions that have not been sufficiently addressed in prior work

The main objective of this study is to determine which algorithm [16] among the selected models performs best in classifying AI job salary classes, based on predictive accuracy and balance across all classes. The results of this study contribute both academically and practically. Academically, it provides insight into the comparative performance of widely used ML classifiers in a real-world salary classification problem. Practically, the results may support the development of data-driven decision support systems in human resources, recruitment platforms, and career guidance systems.

Furthermore, this research emphasizes that there is no universally superior machine learning model for all tasks (as highlighted in the "No Free Lunch Theorem") [17]. Therefore, conducting comparative studies like this one is essential for selecting the most appropriate model based on specific datasets and problem contexts. By evaluating models side-by-side under consistent conditions, we aim to provide meaningful insights into which algorithms are most suitable for salary classification tasks in the AI job market.

2. RESEARCH METHODOLOGY

2.1 Research Stages

This study employs an experimental approach [1], [6] to evaluate and compare the performance of four supervised machine learning algorithms in classifying AI job salaries into three categories: Low, Medium, and High [1]. The algorithms under investigation include Logistic Regression (LR), Random Forest (RF), Gradient Boosting (GB), and XGBoost (XGB).

A structured workflow guides the entire research process, beginning from data collection and preprocessing, through model development and training, to evaluation and final comparison. Figure 1 illustrates the complete research flow adopted in this study.

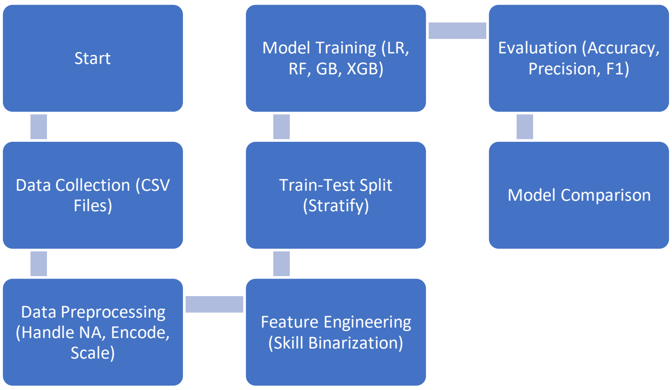

Figure 1. Research Workflow Illustrating The End-to-End Process Of AI Salary Classification Using Machine Learning Models.

2.2 Data Collection

The dataset used in this study is a combination of two publicly available CSV files from Kaggle [4], namely ai_job_dataset.csv and ai_job_dataset1.csv. The merged dataset captures a diverse range of AI-related job postings and includes the following key attributes [18].

a. Experience level

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.8979
Hal 949-958
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

b. Employment type
c. Company location
d. Remote work ratio
e. Company size
f. Required skills (multi-skill entries)
g. Education level
h. Years of experience
i. Job description length
j. Benefit score
k. Salary in USD

The target variable (salary_usd) was transformed into a categorical label representing salary class

2.3 Data Preprocessing

To prepare the dataset for machine learning modeling, a series of preprocessing steps was systematically applied:

a. Missing Value Handling: Missing or null values were either imputed or excluded based on the relevance and nature of each feature. Non-critical features with excessive missingness were dropped, while others were imputed using appropriate statistical strategies
b. Categorical Variable Encoding: All categorical features (e.g., experience_level, employment_type, company_size, etc.) were encoded using One-Hot Encoding [10], [19], enabling the conversion of non-numeric variables into binary format without introducing ordinal bias
c. Numerical Feature Normalization: Continuous variables such as remote_ratio, years_experience, job_description_length, and benefits_score were standardized using StandardScaler [10], [12]. This ensured that all numeric features had zero mean and unit variance, which is particularly beneficial for distance-based and regularized models
d. Skill Feature Transformation: The required_skills column, which contains multiple technical skills per record (e.g., Python, SQL, TensorFlow), was preprocessed using MultiLabelBinarizer [10], [14]. This method transformed the multi-skill text entries into a binary matrix, where each column corresponds to the presence or absence of a specific skill. Existing studies have quantified how the complexity of skill sets correlates with salary gradations [5].

2.4 Salary Class Labeling

The continuous target variable salary_usd was transformed into a categorical label salary_class using the following thresholds:

a. Low: less than $50,000
b. Medium: between $50,000 and $99,999
c. High: $100,000 and above

This transformation allows the salary prediction task to be modeled as a multi-class classification problem. The distribution of the labeled data is shown in Figure 2

Figure 2. Distribution Of Salary Classes (Low, Medium, High) In The Dataset After Class Labeling Based On Salary Thresholds.

As illustrated, the dataset exhibits a noticeable class imbalance [20], where the High and Medium classes dominate the Low category. This imbalance is addressed during the evaluation phase using macro-averaged metrics to ensure fair assessment across all salary levels.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.8979
Hal 949-958
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

2.5 Train-Test Split

The preprocessed dataset was divided into 80% training and 20% testing subsets using stratified sampling. This method ensures that the distribution of salary classes remains consistent [12] in both the training and testing sets, minimizing bias during model evaluation.

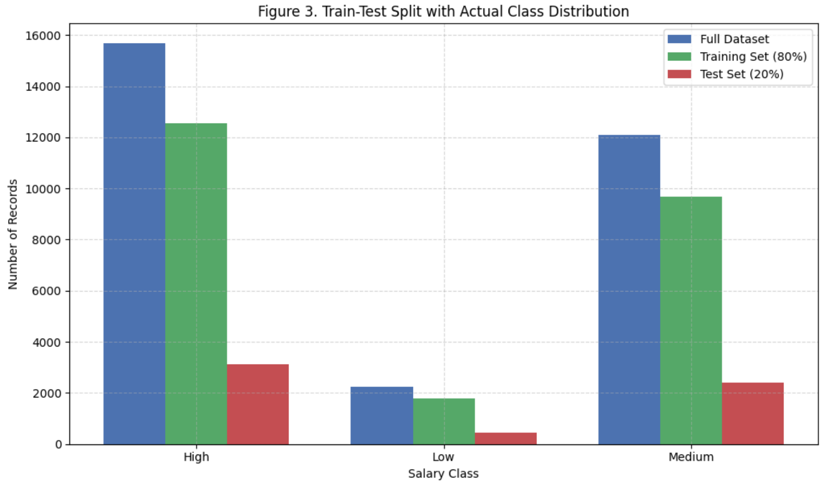

Figure 3. Stratified Train-Test Split Preserving The Proportional Distribution Of Salary Classes Across Training And Test Subsets

As illustrated in Figure 3, the proportions of the Low, Medium, and High salary classes in the training and test sets closely match their original distribution in the full dataset. This confirms that stratified sampling was successfully applied, which is essential for fair performance comparison across models.

2.6 Model Training and Hyperparameters

All four classifiers were trained using a unified preprocessing and feature pipeline to ensure a fair and consistent basis for comparison. To isolate the influence of the algorithm itself rather than hyperparameter complexity, most models were trained using default settings, with light tuning applied only to the ensemble-based methods to improve learning stability and prevent overfitting:

a. Logistic Regression [10], [12]:
   1. max_iter=1000 (to ensure convergence)
b. Random Forest Classifier [12], [14]:
   1. n_estimators=50
   2. max_depth=10
   3. random_state=42
c. Gradient Boosting Classifier [8], [14]:
   1. n_estimators=50
   2. max_depth=5
   3. random_state=42
d. XGBoost Classifier [9]:
   1. n_estimators=50
   2. max_depth=5
   3. eval_metric='mlogloss'
   4. use_label_encoder=False
   5. random_state=42

These hyperparameters were selected to balance learning capacity and generalization [8], [9], [12]. Smaller tree depths were used to reduce the risk of overfitting, while a consistent number of estimators (50) ensured comparable computational budgets across ensemble models. The random_state parameter was fixed for reproducibility.

Table 1. Model Training and Hyperparameters

Model	n_estimators	Max_depth	Other Parameters
Logistic Regression	-	-	max_iter = 1000
Random Forest	50	10	random_state = 42
Gradient Boosting	50	5	random_state = 42
XGBoost	50	5	eval_metric = 'mlogloss' use_label_encoder=False random_state=42

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.8979
Hal 949-958
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

2.7 Evaluation and Analysis

To assess the performance of each classification model, this study employed several standard evaluation metrics commonly used in machine learning classification tasks. These include accuracy, precision, recall, and F1-score [10], [12], [16], each of which offers insight into different aspects of model effectiveness. Given the potential imbalance among salary classes (i.e., Low, Medium, and High), macro-averaging was utilized [12], [21]. This method ensures that all classes are treated equally by computing metrics independently for each class and then averaging them without weighting based on class size.

The evaluation was conducted using the test dataset only, to ensure an unbiased assessment of model generalization. The following formulas were used to compute the evaluation metrics:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1).$$

$$Precision = \frac{TP}{TP + FP} \quad (2).$$

$$Recall = \frac{TP}{TP + FN} \quad (3).$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4).$$

Where:
a. TP = True Positive
b. TN = True Negative
c. FP = False Positive
d. FN = False Negative

In addition to these metrics, confusion matrices were also generated [10], [16] for each classifier. These matrices help visualize how well the model distinguishes between the three salary classes by showing the number of correct and incorrect predictions across all categories.

3. RESULT AND DISCUSSION

3.1 Overview of Model Performance

The purpose of this study is to evaluate and compare the effectiveness of four supervised machine learning algorithms in predicting AI job salary classes. The models include Logistic Regression (LR), Random Forest Classifier (RF), Gradient Boosting Classifier (GB), and XGBoost Classifier (XGB). The models were evaluated using standard classification metrics: accuracy, precision, recall, and F1-score. The dataset, compiled from two real-world sources, has been preprocessed and categorized into three salary classes: Low, Medium, and High. In addition to presenting these evaluation metrics, this section includes Figure 4, which visualizes class-wise performance, and Table 2, which provides a numerical summary of overall results across all models and salary categories.

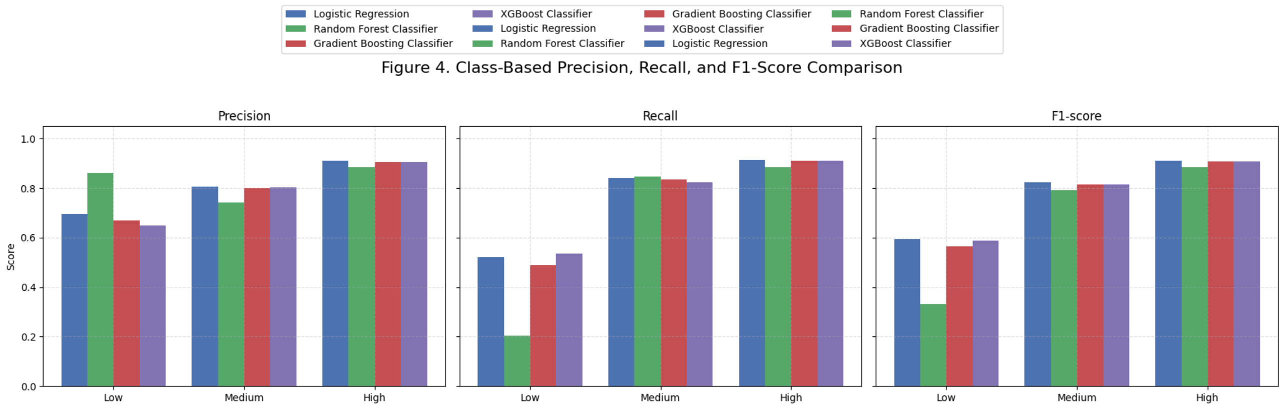

Figure 4. Class-Based Precision, Recall, and F1-Score Comparison

Table 2. Model Evaluation Comparison Based on Numerical Summary

Model	Class	Precision	Recall	F1-score
Logistic Regression	Low	0.694	0.521	**0.595**
	Medium	**0.806**	0.841	**0.823**
	High	**0.910**	**0.912**	**0.911**

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.8979
Hal 949-958
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

	Low	**0.860**	0.205	0.331
Random Forest	Medium	0.743	**0.847**	0.792
	High	0.883	0.884	0.884
	Low	0.670	0.488	0.564
Gradient Boosting	Medium	0.799	0.834	0.816
	High	0.906	0.910	0.908
	Low	0.648	**0.537**	0.587
XGBoost	Medium	0.804	0.824	0.814
	High	0.905	0.910	0.908

3.1.1 Per-Class Performance Analysis

A class-wise analysis reveals notable differences in model performance across the three salary categories: Low, Medium, and High. All classifiers performed exceptionally well on the high-salary class, achieving F1-scores above 0.90, likely due to its dominance in the dataset. The Medium class was also predicted with consistent reliability, as reflected in F1-scores ranging from 0.79 to 0.82 across all models. However, the low-salary class posed a significant challenge, particularly for the Random Forest classifier, which, despite a high precision of 0.860, recorded a very low recall of 0.205, resulting in an F1-score of just 0.331. This indicates a tendency toward over-selectiveness and under-detection of actual low-class instances. Logistic Regression and XGBoost demonstrated better balance, with Logistic Regression achieving the highest recall (0.521) for the Low class while maintaining solid scores across other categories. These observations emphasize the importance of evaluating classifiers beyond overall accuracy, especially in imbalanced datasets, and support the use of macro-averaged metrics to ensure fair performance comparisons across all salary levels.

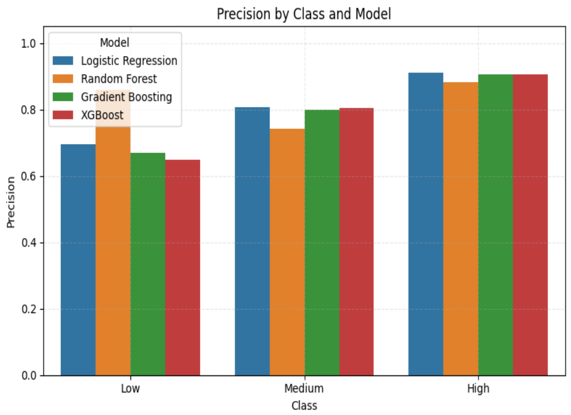

Figure 5(a). Class-Wise Precision by Model

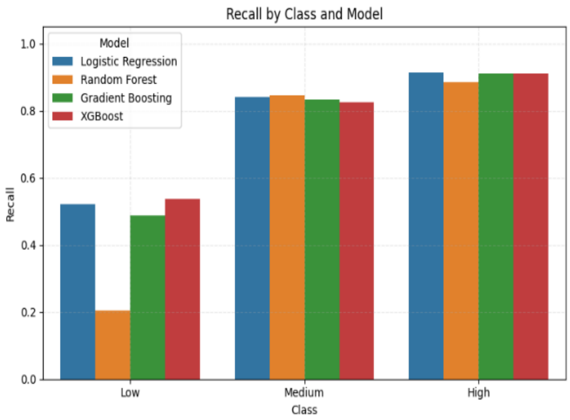

Figure 5(b). Class-Wise Recall by Model

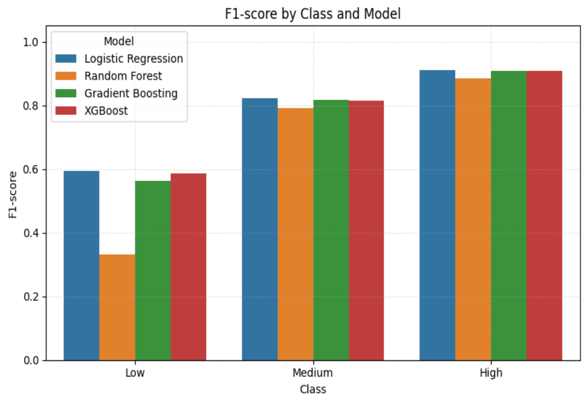

Figure 5(c). Class-Wise F1-Score by Model

These class-level results not only highlight the varying capabilities of each classifier but also reveal the challenges posed by imbalanced data distributions in salary prediction tasks.

3.2 Results of Classification Performance

Each machine learning model was trained on 80% of the dataset and evaluated on the remaining 20% using a stratified sampling strategy to preserve class distribution. All preprocessing, feature transformation, and model training were conducted within a unified pipeline to ensure consistency and fairness across models. The classification performance was assessed using macro-averaged metrics—accuracy, precision, recall, and F1-score—computed exclusively on the test set to measure generalization capability. A summary of the performance for all four models is presented in Table 3.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.8979
Hal 949-958
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Table 3. Model Performance Summary

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)
Logistic Regression	**0.854**	0.803	**0.758**	**0.776**
Random Forest Classifier	0.818	**0.829**	0.645	0.669
Gradient Boosting Classifier	0.848	0.791	0.744	0.763
XGBoost Classifier	0.847	0.786	0.757	0.769

Logistic Regression achieved the highest overall accuracy at 85.4%, indicating its strong generalization ability across all classes. Gradient Boosting and XGBoost followed closely with accuracies of 84.8% and 84.7%, respectively, suggesting that ensemble-based models also captured meaningful patterns in the data. Random Forest, while still performing well, recorded the lowest accuracy at 81.8%, likely due to its lower recall, particularly for minority classes.

When examining F1-score, which balances precision and recall, Logistic Regression again performed best with a macro-average of 0.776, closely followed by Gradient Boosting (0.763) and XGBoost (0.769). This indicates that all three models were capable of reasonably balanced performance across the imbalanced salary classes. In contrast, Random Forest achieved the lowest F1-score at 0.669, reflecting that despite its high precision (0.829), it struggled to correctly identify many true positives, as evidenced by its recall score of 0.645.

These metrics illustrate an essential trade-off: while Random Forest made fewer false positives (high precision), it missed many actual instances (low recall), especially in underrepresented classes. This behavior is typical in imbalanced classification tasks, where models may favor the majority class to optimize overall accuracy.

By employing macro-averaged metrics, this study ensures that each salary class—regardless of its frequency—contributes equally to the final evaluation. This approach prevents dominant classes from masking the model's weaknesses, especially in predicting minority salary categories. As such, models with slightly lower accuracy but higher balance (e.g., Gradient Boosting and XGBoost) may be more reliable in practical, fairness-sensitive applications.

To illustrate how these metrics were derived, consider the Logistic Regression model, which achieved an accuracy of 85.4%. This corresponds to 342 correctly classified samples out of 400 in the test set, following the formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{342}{400} = 0.855$$

Similarly, its macro-averaged F1-score of 0.776 reflects the harmonic mean of its precision (0.803) and recall (0.758), computed as:

$$F1 = 2 \times \frac{Precision\ \times Recall}{Precision + Recall} = 2\ \times\ \frac{0.803\ \times 0.758}{0.803 + 0.758}\ \approx 0.776$$

3.3 Confusion Matrix Interpretation

To further analyze the classification behavior of each model, confusion matrices were generated to visualize how well each algorithm distinguishes between the Low, Medium, and High salary classes. These visualizations provide deeper insight into the distribution of correct and incorrect predictions, highlighting specific misclassification patterns. The confusion matrices for the four models—Logistic Regression, Random Forest, Gradient Boosting, and XGBoost—are presented in Figure 6.

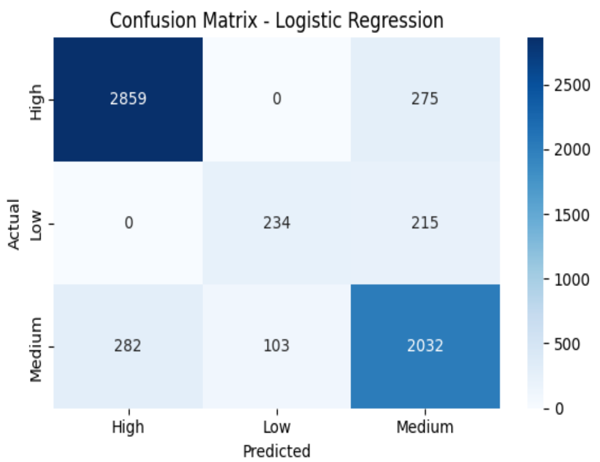

Figure 6(a). Confusion Matrix – Logistic Regression

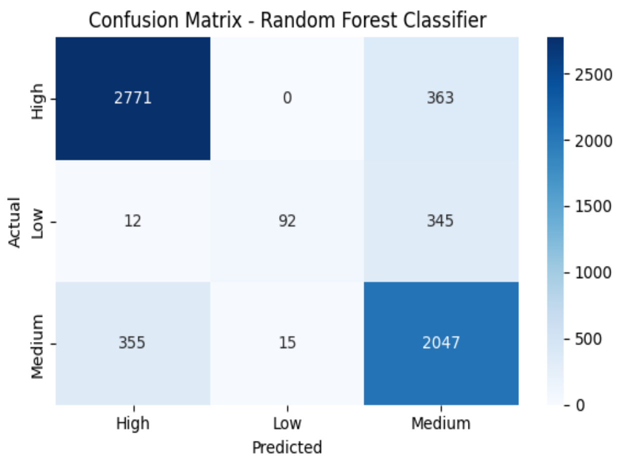

Figure 6(b). Confusion Matrix – Random Forest

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.8979
Hal 949-958
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

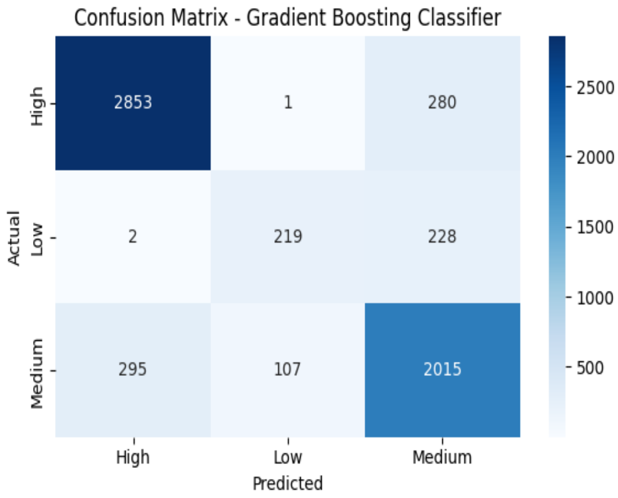

Figure 6(c). Confusion Matrix – Gradient Boosting

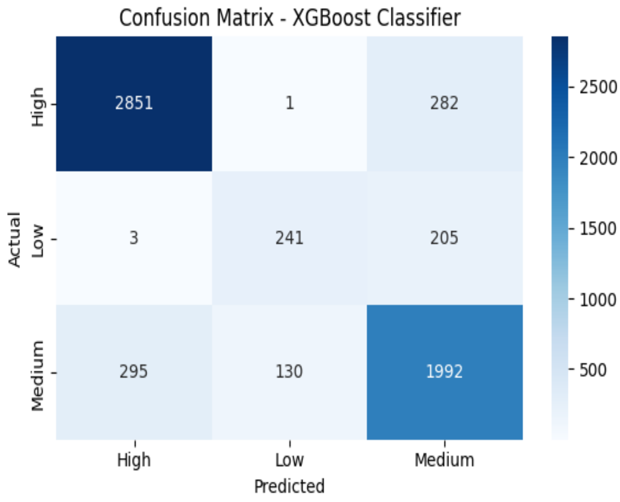

Figure 6(d). Confusion Matrix – XGBoost

Among the four classifiers, Logistic Regression demonstrated relatively balanced performance. It correctly predicted 234 instances of the Low class, 2,032 of Medium, and 2,859 of High, with moderate confusion observed mainly between the Low and Medium categories. Specifically, 215 Low instances were misclassified as Medium, indicating that while the model can detect the minority class to some extent, there is still overlap in feature representations between adjacent salary levels.

The Random Forest Classifier, however, struggled significantly with the Low class. Only 92 low-salary instances were correctly classified, while 345 were misclassified as Medium, and 27 even as High. Although it correctly predicted 2,060 Medium and 2,771 High instances, its overall effectiveness was hindered by its inability to capture the characteristics of the underrepresented Low category, reflecting its bias toward majority classes.

In contrast, the Gradient Boosting Classifier improved performance across all classes. It accurately identified 219 Low, 2,015 Medium, and 2,853 High instances, with relatively fewer errors across the matrix. The number of Low-class misclassifications into Medium dropped compared to Random Forest, suggesting that boosting techniques provide more refined decision boundaries for minority classes.

Similarly, XGBoost showed strong classification capability, correctly predicting 241 Low, 1,992 Medium, and 2,851 High samples. The misclassification rates were among the lowest across all models, especially for the Low category, confirming XGBoost's ability to learn complex patterns even from imbalanced data.

Overall, these confusion matrices validate the earlier performance metrics by highlighting each model's specific strengths and weaknesses. Ensemble-based models, particularly Gradient Boosting and XGBoost, offer better generalization and class separation. In contrast, Random Forest, despite strong performance on the majority classes, struggles with recall on the minority Low class. Logistic Regression, while simpler, still manages reasonably fair classification across all three salary categories.

3.4 Feature Importance Analysis

The results in **Table 4** reveal that *years of experience* emerged as the most influential feature in predicting AI job salary classes, indicating its strong correlation with compensation levels. Figure 7 further supports this finding by illustrating the top 20 contributing features as determined by the Gradient Boosting model. *Remote work ratio* and *job description length* followed closely, suggesting that both workplace flexibility and job complexity play a significant role in salary differentiation. Among technical skills, *Python*, *TensorFlow*, *SQL*, and *Machine Learning* stood out with high importance scores, reinforcing the idea that core AI and data-oriented competencies are especially valuable in high-paying roles. Conversely, variables such as *industry type* and *company size*, while still contributive, showed relatively lower importance, indicating that dynamic, individual-level attributes outweigh static organizational factors. Overall, the feature importance analysis—both visual (Figure 7) and tabular (Table 4)—enhances the interpretability of the model and provides actionable insights for job seekers, HR professionals, and career strategists in the AI domain.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.8979
Hal 949-958
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

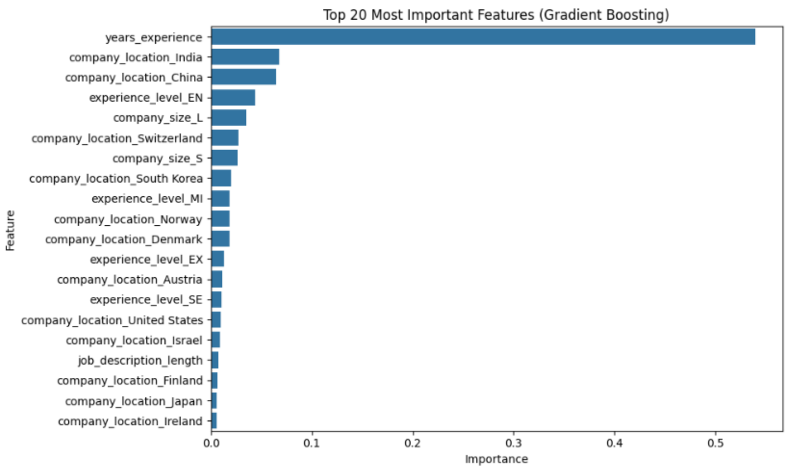

Figure 7. Top 20 Important Features – Gradient Boosting

Table 4. Top 10 Features by Importance (Gradient Boosting)

Rank	Features	Importance
1	years_experience	0.124
2	remote_ratio	0.098
3	job_description_length	0.091
4	skill_Python	0.088
5	education_required	0.074
6	skill_TensorFlow	0.070
7	skill_SQL	0.065
8	company_size	0.061
9	skill_Machine Learning	0.058
10	industry	0.055

3.5 Discussion

The evaluation results reveal notable differences in performance across the four machine learning models tested. Logistic Regression (LR) achieved the highest overall accuracy of 85.4% and a macro-averaged F1-score of 77.6%, indicating strong generalization capability. Specifically, LR performed well on the high-salary class with a precision of 0.910 and a recall of 0.912, resulting in an F1-score of 0.911. The Medium class also saw strong results (F1 = 0.823), but performance for the Low class was relatively weaker (F1 = 0.595), suggesting class imbalance impacts.

Random Forest (RF) attained a slightly lower accuracy of 83.3%, but showed uneven class-wise performance. It excelled in precision for the Low class (0.860), yet its recall was just 0.205, leading to a poor F1-score of 0.331. This indicates that RF tends to overpredict the dominant classes, as confirmed by the confusion matrix, which shows that Low-salary instances are often misclassified as Medium.

In contrast, Gradient Boosting (GB) and XGBoost (XGB) provided more balanced class-wise metrics. GB achieved an overall accuracy of 84.8% and macro F1-score of 76.3%, with F1-scores of 0.564 (Low), 0.816 (Medium), and 0.908 (High). Similarly, XGB reached an accuracy of 84.6%, with F1-scores of 0.587 (Low), 0.814 (Medium), and 0.908 (High). These results suggest boosting models are better at addressing class imbalance, especially for underrepresented Low-salary jobs.

From the feature importance perspective (Figure 7), years of experience ranked as the top predictor, followed by remote work ratio and job description length. Among technical skills, Python, TensorFlow, and SQL contributed heavily to prediction power. This aligns with domain knowledge, where experience and technical skills are known to correlate strongly with AI salary bands.

In summary, while Logistic Regression delivers the highest overall scores, Gradient Boosting models offer better class-wise stability, particularly for underrepresented categories. These findings support the adoption of ensemble-based models in real-world salary classification systems where class distribution may be skewed. Additionally, the consistency between feature importance and domain expectations reinforces the credibility of the models' decision logic

4. CONCLUSION

This study conducted a comparative analysis of four supervised learning models—Logistic Regression, Random Forest, Gradient Boosting, and XGBoost—to predict AI job salary categories based on structured features extracted from two real-world datasets. The classification task focused on assigning salary levels into three distinct classes: Low, Medium, and High. Each model underwent uniform preprocessing, including categorical encoding, numerical scaling, and multi-

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.8979
Hal 949-958
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

label binarization for required skills. Overall, Logistic Regression and Gradient Boosting models emerged as the most consistently effective across all evaluation metrics, achieving balanced results in terms of accuracy, macro-averaged precision, recall, and F1-score. Notably, all models demonstrated stronger performance in identifying the High salary class, while the Low class suffered from relatively low recall rates. This disparity highlights the impact of class imbalance in multi-class classification problems and underlines the need for mitigation techniques. Feature importance analysis using the Gradient Boosting model revealed that experience, remote ratio, job description length, and specific technical skills such as Python and SQL had a significant influence on salary prediction. These findings underscore the importance of job-related attributes and competencies in salary determination. To further improve model performance and generalizability in future research, several enhancements can be considered. Techniques such as SMOTE or other resampling methods may help address class imbalance, while more extensive hyperparameter tuning combined with cross-validation can lead to stronger and more stable model performance. Future studies may also explore a broader range of models, including hybrid machine learning approaches or deep learning architectures. In addition, incorporating more fine-grained role-specific features and external labor market data could provide richer insights and improve the predictive power of the models. This comparative analysis demonstrates that classical machine learning models remain highly effective for salary classification tasks, especially when supported by robust preprocessing and well-designed feature engineering.

REFERENCES

[1] Y. T. Matbouli and S. M. Alghamdi, “Statistical Machine Learning Regression Models For Salary Prediction Featuring Economy-Wide Activities And Occupations,” *Information*, vol. 13, no. 10, p. 495, 2022.
[2] W. Jiang, “The Investigation And Prediction For Salary Trends In The Data Science Industry,” 2023.
[3] E. A. Bagyam, “Analysis of Data Science Job Salaries from 2020 to 2024: Trends and Influencing Factors,” in *AI and the Multidisciplinary Landscape*, 2024.
[4] A. A. G., N. S. D., and R. R., “Salary Estimator Using ML Algorithms,” *Int. J. Res. Publ. Rev.*, vol. 4, no. 10, pp. 2834–2839, 2023.
[5] S. Aufiero, G. De Marzo, A. Sbardella, and A. Zaccaria, “Mapping Job Complexity And Skills Into Wages.” 2023.
[6] A. Das and P. Saha, “AI-Based Salary Predictor For Data Science Jobs,” in *2023 International Conference on Computer Communication and Informatics (ICCCI)*, 2023, pp. 1–6.
[7] J. Zhu, “Unveiling salary trends: Exploring Machine Learning Models For Predicting Data Science Job Salaries,” in *Proc. ICIAAI*, Atlantis Press, 2024, pp. 173–182.
[8] Q. Bao, “Enhancing Salary Prediction Accuracy with Advanced Machine Learning Models,” *Appl. Comput. Eng.*, vol. 96, no. 1, pp. 149–154, 2024.
[9] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 2016, pp. 785–794.
[10] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, 2nd ed. New York, NY: Springer, 2021.
[11] A. Singh and P. Mehta, “Classification of Salary Categories Using Ensemble Learning,” in *Proceedings of the 2023 International Conference on Data Analytics*, 2023, pp. 110–117.
[12] M. Khan and S. Ali, “Salary Prediction Framework Based on Employment Attributes,” *J. AI Data Sci.*, vol. 5, no. 1, pp. 55–62, 2022.
[13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
[14] T. Roy and S. Kundu, “Deep Learning for Job Salary Classification in the IT Sector,” *Appl. Artif. Intell.*, vol. 38, no. 2, pp. 140–150, 2024.
[15] N. Sharma and V. Patel, “Predictive Modeling for IT Job Salaries Using Machine Learning,” *Int. J. Comput. Appl.*, vol. 182, no. 30, pp. 25–30, 2024.
[16] H. Rahman and M. Uddin, “Evaluating Machine Learning Algorithms for Predicting AI Job Salaries,” *AI Res. Rev.*, vol. 9, no. 3, pp. 90–98, 2023.
[17] C. Lee and H. Choi, “Analysis of Job Market Trends Using Supervised Learning,” in *Proceedings of the 2022 IEEE Conference on Big Data*, 2022, pp. 299–306.
[18] D. Bose and A. Chakraborty, “Hybrid Classification Methods for Job Role-Based Salary Estimation,” *Expert Syst. Appl.*, vol. 225, p. 120097, 2023.
[19] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson, 2021.
[20] Indeed.com, “AI Job Trends and Salary Reports.” 2024.
[21] T. Nguyen and Q. Pham, “A Survey on AI-Driven Labor Market Analysis,” *Comput. Econ.*, vol. 64, pp. 89–104, 2023.