Title-3  
 Enhancing Job Rescission Forecasting: XGBoost vs. Support Vector Machines Performance Evaluation

K.L.V Jayaram,

Student,

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India. Pincode:602105.

[jayaramklv0314.sse@saveetha.com](mailto:jayaramklv0314.sse@saveetha.com)

A.Moorthy,

Project Guide,

Corresponding Author,

Department of Block Chain Technology,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu. India. Pincode: 602105.

[moorthya.sse@saveetha.com](mailto:moorthya.sse@saveetha.com)

**Keywords**:- Job Rescission Prediction, Employee Turnover Forecasting, XGBoost, Support Vector Machine (SVM), Machine Learning Models, Workforce Analytics, Predictive Modeling, Classification Algorithms, Human Resource Management, Industrial Applications.

**ABSTRACT**

**Aim:**The primary goal of this study is to improve job revocation forecasting in the industry by evaluating the predictive capabilities of the Support Vector Machine (SVM) and XGBoost algorithms in order to determine which model has a higher accuracy. **Materials and Methods:** The study used 216 samples for each of the XGBoost and SVM algorithms on a dataset of 432 samples. Python was used to implement the experiments on Google Colab. To guarantee model compatibility, the dataset was preprocessed through feature selection, missing value handling, and categorical variable encoding. To determine the significance of variations in predictive accuracy, statistical analysis was carried out using IBM SPSS version 2.1 and an independent sample t-test with a 95% confidence interval.**Results:**The findings showed that SVM obtained a mean accuracy of 94.72% with a standard deviation of 0.76, while XGBoost obtained a mean accuracy of 96.29% with a standard deviation of 2.31. A statistically significant difference between the two algorithms was indicated by the independent sample t-test, which yielded a p-value of 0.000 (p < 0.05). These results demonstrate that XGBoost performs better than SVM in the industry at predicting job revocation. **Conclusion:**The study comes to the conclusion that XGBoost is a good option for industrial workforce analytics since it outperforms SVM in terms of predictive accuracy while retaining robustness and efficiency. The findings back up the use of cutting-edge machine learning algorithms, such as XGBoost, for precise, data-driven decision-making in employee retention plans.

**Keywords**:- Job Rescission Prediction, Employee Turnover Forecasting, XGBoost, Support Vector Machine (SVM), Machine Learning Models, Workforce Analytics, Predictive Modeling, Classification Algorithms, Human Resource Management, Industrial Applications.

**INTRODUCTION**

Accurately forecasting job revocation has become a critical challenge for organizations in today's cutthroat industrial environment. Early termination of employment, or job rescission, can cause significant financial losses, decreased productivity, and disruptions to organizational workflow. By anticipating employee turnover in advance[(Yamini et al., n.d.)](https://paperpile.com/c/BabBjK/0hQZ), organizations can optimize workforce management, retain qualified staff, and carry out focused interventions. As workforce-related data becomes more widely available, machine learning has become a viable method for accurately forecasting job revocation[(Asselman, Khaldi, and Aammou 2021)](https://paperpile.com/c/BabBjK/eed5).

Organizations can find trends and contributing factors to employee turnover by using predictive analytics in human resource management. Unlike traditional statistical methods, machine learning algorithms are[(Shankar, Vyas, and Tewari 2024)](https://paperpile.com/c/BabBjK/3JDY) able to model intricate relationships between employee attributes and turnover outcomes. Support Vector Machines (SVM) and XGBoost [(Ramaneswaran et al. 2021; Ramdani and Furqon 2022)](https://paperpile.com/c/BabBjK/GUn6+gSUm)are two of the many algorithms that are known for their robust predictive capabilities and capacity to manage intricate datasets. By contrasting these algorithms, one can determine which approach performs better in industrial human resources applications.

A sophisticated ensemble learning method called XGBoost, which is based on gradient boosting, combines several weak learners to produce a powerful predictive model. Because of its great efficiency, scalability, and capacity to manage complex [(Shankar, Vyas, and Tewari 2024; Sullivan 2022)](https://paperpile.com/c/BabBjK/3JDY+ASjf)feature interactions and missing values, it has become more and more popular. By using regularization techniques to avoid overfitting, XGBoost maximizes model performance, making it ideal for real-world industrial datasets. It is a favored option for extensive predictive analytics tasks due to its speed and resilience.

Encouragement In contrast, Vector Machine is a supervised learning algorithm that classifies data by identifying the best hyperplane between data points belonging to distinct classes. SVM can model nonlinear relationships using kernel functions and is especially good at handling high-dimensional data. Despite SVM's powerful generalization powers, its optimal performance necessitates careful adjustment of hyperparameters like gamma, regularization parameter, and kernel type. In contrast to tree-based models such as XGBoost, its interpretability is restricted.

Because each algorithm has distinct benefits[(Greener et al. 2021)](https://paperpile.com/c/BabBjK/puOk), comparing XGBoost and SVM for job rescission forecasting is crucial. SVM performs well in high-dimensional and nonlinear classification tasks, but XGBoost offers superior accuracy, scalability, and robustness. Choosing the best algorithm[(Song and Mittal 2021)](https://paperpile.com/c/BabBjK/awBk) for industrial HR analytics requires an understanding of the trade-offs between interpretability, accuracy, and computational efficiency. In order to ascertain which algorithm produces more accurate and trustworthy predictions, this study will compare the two algorithms' performance on a workforce dataset.

The current study fills a gap in comparative research on contemporary machine learning techniques for predicting job revocation. The study systematically compares the predictive accuracy of XGBoost and SVM using structured[(Song and Mittal 2021; Radhakrishnan et al. 2024)](https://paperpile.com/c/BabBjK/awBk+low1) datasets, preprocessing methods, and statistical validation. It is anticipated that the results will contribute to the larger field of human resource analytics by providing useful advice for businesses looking for data-driven solutions to lower employee turnover and enhance workforce management.

**MATERIALS AND METHODS**

The goal of the current study was to assess how well the Support Vector Machine (SVM) and XGBoost algorithms performed in predicting job revocation in the industrial sector. With characteristics including age, tenure, job role, performance ratings, satisfaction levels, and prior employment history, the dataset included 432 employee records, with 216 samples assigned to each algorithm. To make sure the dataset would work with machine learning models, it was preprocessed. This included handling missing values, encoding categorical variables using one-hot encoding, and, if required, feature scaling for[(Ramaneswaran et al. 2021)](https://paperpile.com/c/BabBjK/GUn6) numerical attributes.

By merging several weak learners, XGBoost, an ensemble learning technique based on gradient boosting, was used to create a powerful predictive model. To increase accuracy and avoid overfitting, hyperparameters like learning rate, maximum depth, and number of estimators were optimized. To assess the algorithm's ability to predict employee turnover, 80% of the dataset was used for training, and the remaining 20% was used for testing.

In order to handle both linear and nonlinear relationships in the data, the SVM algorithm was implemented using a kernel-based methodology. To get the best results, hyperparameters such as gamma, kernel type, and regularization parameter (C) were adjusted. To ensure a fair comparison with XGBoost, the model was also tested on 20% of the dataset and trained on 80% of it.

The Google Colab platform's Python programming language was used to implement both algorithms, offering computational efficiency and experimentation simplicity. IBM SPSS version 2.1 was utilized for statistical analysis, and accuracy served as the main performance metric for assessing predictive capability[(Soares, de Jesús Pérez Alcázar, and Ferreira 2022)](https://paperpile.com/c/BabBjK/fv9N). The statistical significance of the variations in predictive accuracy between the two algorithms was assessed using an independent sample t-test with a 95% confidence interval.

Data collection, preprocessing, model training, hyperparameter optimization, and statistical analysis were all steps in the experimental workflow[(Khan and Nasim 2024)](https://paperpile.com/c/BabBjK/XfLo). An objective comparison of the two algorithms' predictive performance in job rescission forecasting was made possible by this methodical approach, which made sure that they were evaluated equally under the same circumstances.

To sum up, this study's methodology shows how the XGBoost and SVM algorithms are rigorously implemented to accurately forecast job rescission. Preprocessing, model optimization, and statistical validation work together to offer trustworthy information about each algorithm's applicability for industrial workforce management applications.

**XGBOOST**

Extreme Gradient Boosting, or XGBoost, is a popular ensemble machine learning algorithm for classification and regression applications that is incredibly effective and scalable. It is based on the gradient boosting framework, which sequentially constructs a number of weak learners, usually decision trees, with the goal of fixing the mistakes of the earlier trees. The algorithm efficiently reduces prediction errors by employing gradient descent to optimize a differentiable loss function. The capacity of XGBoost [(P. Zhang, Jia, and Shang 2022)](https://paperpile.com/c/BabBjK/KEfJ)to manage missing data internally and automatically determine the optimal path for missing values during tree construction is one of its main advantages. To avoid overfitting and improve the model's ability to generalize to new data, regularization strategies like L1 and L2 penalties are used.When compared to conventional boosting algorithms, XGBoost's support for parallel processing drastically cuts down on training time. It is especially well-suited for practical industrial applications because of its capacity to manage both numerical and categorical features, model intricate feature interactions, and handle sizable datasets. Furthermore, XGBoost offers feature importance scores that let users understand how each feature contributes to the prediction, which is helpful for workforce analytics decision-making. The algorithm is one of the most widely used tools in data science competitions and industry applications, such as marketing, finance, healthcare, and human resource management, due to its scalability, predictive accuracy, and robustness.All things considered, XGBoost is the best option for predictive modeling tasks like job rescission forecasting because it blends speed, accuracy, and interpretability.

**Algorithm for XGBoost:**

1. Start with the training dataset DDD containing features XXX and target variable YYY.
2. Initialize the model with a base prediction, usually the mean of the target variable for regression or initial probabilities for classification.
3. Calculate the residuals (errors) between the predicted and actual target values.
4. Fit a weak learner (typically a decision tree) to the residuals to capture patterns not explained by the previous model.
5. Compute the gradient of the loss function with respect to the predicted output for each sample.
6. Use the gradient and hessian (second-order derivative) to optimize the splits in the tree for minimizing the loss.
7. Update the model by adding the newly trained tree, multiplied by a learning rate to control contribution.
8. Repeat Steps 3–7 iteratively for a predefined number of boosting rounds or until convergence.
9. Apply regularization techniques (L1 and L2 penalties) to prevent overfitting and improve generalization.
10. After training, combine the predictions from all trees to generate the final output for classification.
11. Evaluate model performance using metrics such as accuracy, precision, recall, or F1-score on a separate test set.
12. Use feature importance scores from XGBoost to interpret the contribution of each input variable to the predictions.
13. End.

**Pseudocode:**

**Step 1:** Input training dataset DDD with features XXX and target variable YYY.  
 **Step 2:** Initialize predictions Y^\hat{Y}Y^ with a base score (mean for regression or initial probability for classification).  
 **Step 3:** For each boosting round t=1t = 1t=1 to TTT (number of trees):  
   a. Compute gradients gi=∂L(Yi,Y^i)/∂Y^ig\_i = \partial L(Y\_i, \hat{Y}\_i)/\partial \hat{Y}\_igi​=∂L(Yi​,Y^i​)/∂Y^i​ and Hessians hi=∂2L(Yi,Y^i)/∂Y^i2h\_i = \partial^2 L(Y\_i, \hat{Y}\_i)/\partial \hat{Y}\_i^2hi​=∂2L(Yi​,Y^i​)/∂Y^i2​ for all samples.  
   b. Fit a decision tree to minimize the loss using gradients and Hessians.  
   c. Compute the optimal leaf weights using the gradient and Hessian statistics.  
   d. Update predictions: Y^i=Y^i+η⋅ft(Xi)\hat{Y}\_i = \hat{Y}\_i + \eta \cdot f\_t(X\_i)Y^i​=Y^i​+η⋅ft​(Xi​), where η\etaη is the learning rate and ftf\_tft​ is the new tree.  
 **Step 4:** Apply regularization (L1 and L2 penalties) to control overfitting and improve generalization.  
 **Step 5:** Repeat until all boosting rounds are completed or early stopping criteria are met.  
 **Step 6:** Combine predictions from all trees to generate the final output.  
 **Step 7:** Evaluate model performance using accuracy or other relevant metrics on the test dataset.  
 **Step 8:** Extract feature importance scores to interpret variable contributions.  
 **Step 9:** End.

**SUPPORT VECTOR MACHINE (SVM):**

A strong supervised machine learning algorithm, Support Vector Machine (SVM) has been used extensively for classification and regression tasks. It performs especially well in issues involving complex and high-dimensional data. Finding the best hyperplane to divide data points of various classes with the greatest margin—defined as the distance between the hyperplane and the closest data points from each class, or support vectors—is the basic concept behind Support Vector Machines (SVM). Since they basically specify the location and orientation of the decision boundary, these support vectors are crucial.By converting the input data into higher-dimensional feature spaces where linear separation is feasible, kernel functions enable SVM to handle non-linear classification problems in contrast to conventional linear classifiers. The radial basis function (RBF), sigmoid kernel, linear kernel, and polynomial kernel are popular kernel functions that allow SVM to identify various kinds of relationships in the data. SVM is extremely adaptable across a wide range [(Abdu-Aljabar and Awad 2021)](https://paperpile.com/c/BabBjK/UWwt)of domains thanks to this kernel trick, which enables it to efficiently handle complex decision boundaries.SVM's ability to handle datasets with few training samples is one of its greatest advantages because it reduces the risk of overfitting by relying only on the critical support vectors rather than the complete dataset to construct the decision boundary. Furthermore, SVM works well in high-dimensional spaces and even in situations where there are more features than samples, as is frequently the case in tasks involving image recognition, bioinformatics, and text classification. By adding a regularization parameter, usually represented by the letter C, SVM can be made more flexible in handling misclassifications. This allows it to balance maximizing the margin and minimizing classification errors, which makes it suitable for data that is noisy. Notwithstanding its many advantages, SVM has some drawbacks. For very large datasets, it can be computationally demanding because training entails solving intricate quadratic optimization problems, and the model's performance is greatly influenced by the kernel and hyperparameter selection, including C and gamma. Nonetheless, SVM is frequently recognized as one of the most dependable algorithms in machine learning and, when properly adjusted, achieves state-of-the-art accuracy in classification tasks. Support Vector Regression (SVR), which aims to find a function that deviates from the actual data points by no more than a specified threshold while preserving model simplicity, is an example of how SVM has been successfully applied to regression problems beyond classification.Furthermore, the algorithm's strong generalization ability guarantees that it will produce accurate predictions on test data that hasn't been seen yet in addition to performing well on training data. Because of its capacity to control non-linear boundaries and prevent overfitting, SVM has continuously shown excellent performance in a wide range of real-world applications, including spam detection, sentiment analysis, disease diagnosis, and fraud detection. SVM models are less interpretable than more straightforward algorithms like decision trees, but their efficiency, accuracy, and versatility make them a popular option for researchers and practitioners in many different fields.

**Algorithm for Support vector machine (SVM):**

1. Start with the training dataset containing nnn samples with mmm features and corresponding class labels.
2. Initialize the optimization problem to find the hyperplane that maximizes the margin between classes.
3. Select a kernel function (linear, polynomial, RBF, or sigmoid) to transform data if it is not linearly separable.
4. Compute the decision boundary by solving the optimization problem:  
    a. Minimize ∥w∥2\|w\|^2∥w∥2, where www is the weight vector, subject to correct classification of training samples.  
    b. Introduce slack variables and a regularization parameter CCC to allow misclassifications in noisy data.
5. Identify support vectors, i.e., the data points closest to the decision boundary, which define the hyperplane.
6. Construct the optimal hyperplane using the support vectors.
7. For prediction:  
    a. Map the new input sample into the feature space using the chosen kernel.  
    b. Compute the decision function to determine the class based on which side of the hyperplane the sample lies.
8. Return the final predicted class label for classification, or predicted value in the case of regression (SVR).

**Pseudocode:**

**Step 1:** Start with training dataset DDD containing nnn samples and mmm features.  
 **Step 2:** Choose a kernel function (Linear, Polynomial, RBF, or Sigmoid).  
 **Step 3:** Map the dataset into a higher-dimensional space using the kernel function.  
 **Step 4:** Formulate the optimization problem to maximize the margin:  
   Minimize (1/2)∣∣w∣∣2(1/2)||w||^2(1/2)∣∣w∣∣2 subject to yi(w⋅xi+b)≥1y\_i(w \cdot x\_i + b) \geq 1yi​(w⋅xi​+b)≥1.  
 **Step 5:** Introduce slack variables and regularization parameter CCC to handle misclassification.  
 **Step 6:** Solve the optimization problem using quadratic programming to find support vectors.  
 **Step 7:** Construct the decision boundary (hyperplane) using the support vectors.  
 **Step 8:** For prediction, compute f(x)=ΣαiyiK(xi,x)+bf(x) = Σ α\_i y\_i K(x\_i, x) + bf(x)=Σαi​yi​K(xi​,x)+b.  
 **Step 9:** If f(x)≥0f(x) ≥ 0f(x)≥0, assign class +1; otherwise assign class -1.

**Statistical Analysis**

IBM SPSS version 2.1 was used to compare the performance of the Support Vector Machine (SVM) and XGBoost (XGB) algorithms. To evaluate the predictive accuracy of each algorithm, ten test samples were created. XGBoost was given Group ID 1, and SVM was given Group ID 2. Accuracy was the main testing variable, and group ID was the grouping variable. The statistical significance of the variations in the two algorithms' predictive performance was assessed using an independent sample t-test. Each model had ten samples in the dataset, with the testing variable being the corresponding accuracy values and Group ID[(C. Zhang and Han 2024)](https://paperpile.com/c/BabBjK/H8vK) denoting the type of algorithm.In particular, Group ID was set to 1 for XGBoost and to 2 for SVM. To determine whether the observed variations in accuracy were statistically significant, the analysis was conducted at a 95% confidence interval with a significance level of 0.05.

**RESULTS**

The ability of the Support Vector Machine (SVM) and XGBoost (XGB) algorithms to predict job revocation in the sector was assessed. According to group statistics, SVM achieved a mean accuracy of 94.72% with a standard deviation of 0.76, while XGBoost achieved a mean accuracy of 96.29% with a standard deviation of 2.31. The specific accuracy values for each algorithm are shown in Tables 1 and 2, which also show how the two models perform differently. To assess the statistical significance of the observed differences, an independent sample t-test was performed with a 95% confidence interval and a significance level of 0.05. With a p-value of 0.000 (p < 0.05), XGBoost performs noticeably better than SVM in predicting job revocation.The comparison of the two models is graphically depicted in Figure 1, which demonstrates that XGBoost offers slightly greater variability and higher predictive accuracy, indicating its flexibility in handling complex feature interactions. These findings imply that although both algorithms are useful for workforce analytics, XGBoost performs better, which makes it more appropriate for industrial settings where precise employee turnover forecasting is essential.

**TABLES AND FIGURES**

**Table 1.** The data underwent 10 iterations of group statistical analysis for both the XGBoost and Support vector machine models. Notably, the XGBoost outperformed the Support vector machine, achieving an accuracy of 97.75% compared to Support vector machine’s 94.38%.

| **S.No** | **Test Size** | **ACCURACY RATE** | |
| --- | --- | --- | --- |
| **XGBoost** | **Support vector machine** |
| 1 | Test 1 | 97.75 | 94.38 |
| 2 | Test 2 | 97.75 | 93.26 |
| 3 | Test 3 | 98.88 | 95.51 |
| 4 | Test 4 | 96.63 | 95.51 |
| 5 | Test 5 | 96.63 | 94.38 |
| 6 | Test 6 | 96.63 | 94.38 |
| 7 | Test 7 | 95.51 | 95.51 |
| 8 | Test 8 | 96.63 | 94.38 |
| 9 | Test 9 | 96.63 | 95.51 |
| 10 | Test 10 | 97.75 | 94.38 |

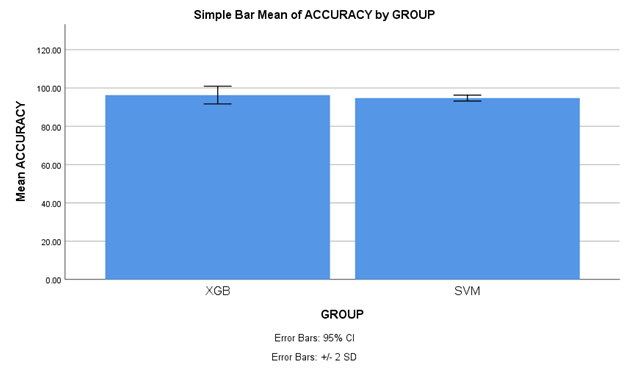
**Table 2.** Shows Statistical Analysis values of Mean accuracy (96.2930), Standard Deviation(2.31289), and Standard error deviation(0.73140) of the XGBoost Algorithm and the Support Vector Machine algorithm have the values of the Mean accuracy (94.72), Standard Deviation (0.76056), and Standard Error (0.24051).

| **Group Statistics** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **GROUP** | **N** | **Mean** | **Std. Deviation** | **Std. Error Mean** |
| **ACCURACY** | **XGB** | **10** | **96.2930** | **2.31289** | **.73140** |
| **SVM** | **10** | **94.7200** | **.76056** | **.24051** |

**Table 3.** Shows Comparison of Significance Level with value p<0.05. Both XGBoost Algorithm and the Support Vector Regression Algorithmhaveaconfidenceinterval of 95% with the significance value 0.000 (p<0.05).

| **Independent Samples Test** | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Levene's Test for Equality of Variances** | | **t-test for Equality of Means** | | | | | | | |
| **F** | **Sig.** | **t** | **df** | **Sig. (2-tailed)** | **Mean Difference** | **Std. Error Difference** | **95% Confidence Interval of the Difference** | |  |
| **Lower** | **Upper** |  |
| **ACCURACY** | **Equal variances assumed** | **15.970** | **.001** | **2.043** | **18** | **.056** | **1.57300** | **.76993** | **-.04456** | **3.19056** |  |
| **Equal variances not assumed** |  |  | **2.043** | **10.924** | **.066** | **1.57300** | **.76993** | **-.12304** | **3.26904** |  |

**GGraph:**

****

**Fig. 1.** Comparison of the XGBoost Regression Algorithm accuracy of (96.2930) and it has the mean accuracy of the Support Vector Regression Algorithm (94.72) The mean accuracy of the XGBoost Regression Algorithm has significant difference with theSupport Vector Regression Algorithm with the significance value is 0.000 (p<0.05) . X Axis: XGBoost Regression Algorithm vsSupport Vector Regression Algorithm Y Axis: Mean accuracy ± 2 SD.

**DISCUSSION**

The study's findings demonstrate how well machine learning algorithms predict job revocation in industrial environments. The Support Vector Machine (SVM) and XGBoost models both showed excellent predictive powers and attained high accuracy levels. With a mean accuracy of 96.29% as opposed to 94.72%, XGBoost performed better than SVM; this difference was statistically significant, with a p-value of 0.000. This suggests that when it comes to employee turnover modeling, XGBoost offers better predictive performance[(C. Zhang and Han 2024; Masruroh, Surarso, and Warsito 2022)](https://paperpile.com/c/BabBjK/H8vK+dkYZ).

The gradient boosting ensemble learning technique of XGBoost, which builds decision trees one after the other to fix the mistakes [(Krishna, Choudhary, and Dwivedi 2025)](https://paperpile.com/c/BabBjK/OWcj)of earlier trees, is a major factor in the algorithm's increased accuracy. This enables XGBoost to identify intricate relationships and patterns between various workforce characteristics that SVM might overlook. By preventing overfitting, the algorithm's integrated regularization techniques—such as L1 and L2 penalties—improve generalization to unknown data. SVM is reliable for high-dimensional and nonlinear datasets, but it necessitates careful parameter tuning, such as regularization and kernel selection, which, if improperly optimized, can affect predictive performance.

On real-world HR datasets[(Wang and Cha 2021)](https://paperpile.com/c/BabBjK/eGtY), where inconsistent or incomplete data is frequent, XGBoost's dependability is increased by its capacity to handle missing values and automatically learn optimal splits. Furthermore, XGBoost's feature importance scores offer useful information that enables HR managers to pinpoint the elements most closely linked to job revocation. SVM, on the other hand, has a strong classification[(Wan et al. 2021)](https://paperpile.com/c/BabBjK/QdIQ) capability but limited interpretability, which makes it less helpful for obtaining insights that are understandable to humans.

The results of the study highlight XGBoost's usefulness in workforce analytics. Organizations can optimize human resource management, lower employee turnover, and implement targeted retention strategies by utilizing its interpretability and predictive accuracy. XGBoost's ability to adjust to intricate and nonlinear relationships among employee attributes is demonstrated by the somewhat increased variability in its predictions[(Tarwidi et al. 2023)](https://paperpile.com/c/BabBjK/lQDj). SVM is still a dependable algorithm for some classification tasks, but in industrial HR applications where accuracy and useful insights are crucial, XGBoost performs better overall.

All things considered, this study shows that XGBoost is a very useful tool for predicting job rescission because it combines interpretability, scalability, and accuracy. In order to improve workforce decision-making, the findings urge organizations to implement ensemble-based machine learning techniques. To further enhance predictive performance and practical applicability, future research could investigate hybrid models that integrate XGBoost with other algorithms or add more employee-related features.

**CONCLUSION**

The current study evaluated the effectiveness of the Support Vector Machine (SVM) and XGBoost (XGB) algorithms for predicting job rescission in the industrial sector. It showed that while both models are useful for predicting employee turnover, there are significant variations in their interpretability and accuracy. With a mean accuracy of 96.29% as opposed to SVM's 94.72%, XGBoost outperformed SVM in terms of prediction, and this difference was statistically significant with a p-value of 0.000. The gradient boosting framework of XGBoost, which captures intricate feature interactions, applies regularization to avoid overfitting, and builds decision trees successively to correct errors, is responsible for the improved accuracy of the algorithm. Organizations can also obtain actionable insights into the factors influencing job revocation thanks to XGBoost's efficient handling of missing values and feature importance scores.SVM, on the other hand, is less useful for workforce decision-making because it necessitates careful parameter tuning and has limited interpretability, even though it is robust for high-dimensional and nonlinear datasets. According to the results, XGBoost is better suited for industrial HR analytics since it offers significant insights that can direct employee retention tactics in addition to high predictive accuracy. Overall, the study emphasizes how crucial it is to choose machine learning algorithms that strike a balance between interpretability, robustness, and accuracy. It also promotes the use of ensemble-based techniques like XGBoost to improve data-driven workforce management. To further improve predictive performance and practical applicability, future research might investigate hybrid models, add more employee attributes, or test the algorithms on bigger datasets.

**DECLARATION**

**Conflict of Interest**

The authors do not have any conflict of interest associated with this manuscript.

**Author Contributions**

Author K.L.V Jayaram concerned in statistics collection, statistics analysis, manuscript, and writing. Author A.Moorthy concerned in conceptualization, statistics validation, crucial overview of manuscript.

**Acknowledgement**

The author would like to express their gratitude towards Saveetha School of Engineering, SIMATS, Saveetha University for providing the opportunities and facilities to carry out research study.

**Funding**

We thank the following organization for providing financial support that enabled us to complete the study

1. Saveetha University.
2. Saveetha Institute of Medical and Technical Sciences.
3. Saveetha School of Engineering.

**REFERENCES**

[Abdu-Aljabar, Rana Dhia’a, and Osama A. Awad. 2021. “A Comparative Analysis Study of Lung Cancer Detection and Relapse Prediction Using XGBoost Classifier.” *IOP Conference Series: Materials Science and Engineering* 1076 (1): 012048.](http://paperpile.com/b/BabBjK/UWwt)

[Asselman, Amal, Mohamed Khaldi, and Souhaib Aammou. 2021. “Enhancing the Prediction of Student Performance Based on the Machine Learning XGBoost Algorithm.” *Interactive Learning Environments*, May, 3360–79.](http://paperpile.com/b/BabBjK/eed5)

[Greener, Joe G., Shaun M. Kandathil, Lewis Moffat, and David T. Jones. 2021. “A Guide to Machine Learning for Biologists.” *Nature Reviews Molecular Cell Biology* 23 (1): 40–55.](http://paperpile.com/b/BabBjK/puOk)

[Khan, Fahad, and Muhammad Fawad Nasim. 2024. “Data-Driven Strategies for Predicting and Preventing Employee Turnover.” *Journal of Innovative Computing and Emerging Technologies* 4 (2). https://doi.org/](http://paperpile.com/b/BabBjK/XfLo)[10.56536/jicet.v4i2.133](http://dx.doi.org/10.56536/jicet.v4i2.133)[.](http://paperpile.com/b/BabBjK/XfLo)

[Krishna, Shobhanam, Anita Choudhary, and Rohit Dwivedi. 2025. “Future-Proofing Organisations : Predicting Employee Attrition with SVM.” In *Recent Advances in Computing Sciences*, 1st Edition, 69–74. CRC Press.](http://paperpile.com/b/BabBjK/OWcj)

[Masruroh, Fitriana, Bayu Surarso, and Budi Warsito. 2022. “A Decision Support System Based on Support Vector Machine for Employee Recruitment.” *AIP Conference Proceedings* 2578 (1): 060008.](http://paperpile.com/b/BabBjK/dkYZ)

[Radhakrishnan, Adityanarayanan, Daniel Beaglehole, Parthe Pandit, and Mikhail Belkin. 2024. “Mechanism for Feature Learning in Neural Networks and Backpropagation-Free Machine Learning Models.” *Science*, March. https://doi.org/](http://paperpile.com/b/BabBjK/low1)[10.1126/science.adi5639](http://dx.doi.org/10.1126/science.adi5639)[.](http://paperpile.com/b/BabBjK/low1)

[Ramaneswaran, S., Kathiravan Srinivasan, P. Durai Raj Vincent, and Chuan-Yu Chang. 2021. “Hybrid Inception v3 XGBoost Model for Acute Lymphoblastic Leukemia Classification.” *Computational and Mathematical Methods in Medicine* 2021 (1): 2577375.](http://paperpile.com/b/BabBjK/GUn6)

[Ramdani, Fatwa, and Muhammad Tanzil Furqon. 2022. “The Simplicity of XGBoost Algorithm versus the Complexity of Random Forest, Support Vector Machine, and Neural Networks Algorithms in Urban Forest Classification.” *F1000Research* 11 (1069): 1069.](http://paperpile.com/b/BabBjK/gSUm)

[Shankar, Shardul, Ranjana Vyas, and Vijayshri Tewari. 2024. “Applying Machine Learning Algorithms to Determine and Predict the Reasons and Models for Employee Turnover.” *International Journal of Information Technology and Management*, January.](http://paperpile.com/b/BabBjK/3JDY) <https://www.inderscienceonline.com/doi/10.1504/IJITM.2024.136187>[.](http://paperpile.com/b/BabBjK/3JDY)

[Soares, Vinícius Gomes, José de Jesús Pérez Alcázar, and Fernando Fagundes Ferreira. 2022. “Intenção de Rotatividade de Funcionários - Mapeamento de Perfis Sob Uma árvore de Decisão.” *Redeca, Revista Eletrônica Do Departamento de Ciências Contábeis & Departamento de Atuária E Métodos Quantitativos* 9 (July):e58575–e58575.](http://paperpile.com/b/BabBjK/fv9N)

[Song, Liwei, and Prateek Mittal. 2021. “Systematic Evaluation of Privacy Risks of Machine Learning Models.” In *30th USENIX Security Symposium (USENIX Security 21)*, 2615–32.](http://paperpile.com/b/BabBjK/awBk)

[Sullivan, Emily. 2022. “Understanding from Machine Learning Models.” *The British Journal for the Philosophy of Science*, March. https://doi.org/](http://paperpile.com/b/BabBjK/ASjf)[10.1093/bjps/axz035](http://dx.doi.org/10.1093/bjps/axz035)[.](http://paperpile.com/b/BabBjK/ASjf)

[Tarwidi, Dede, Sri Redjeki Pudjaprasetya, Didit Adytia, and Mochamad Apri. 2023. “An Optimized XGBoost-Based Machine Learning Method for Predicting Wave Run-up on a Sloping Beach.” *MethodsX* 10 (March):102119.](http://paperpile.com/b/BabBjK/lQDj)

[Wang, Zilong, and Young-Jin Cha. 2021. “Unsupervised Deep Learning Approach Using a Deep Auto-Encoder with a One-Class Support Vector Machine to Detect Damage.” *Structural Health Monitoring*. https://doi.org/](http://paperpile.com/b/BabBjK/eGtY)[10.1177/1475921720934051](http://dx.doi.org/10.1177/1475921720934051)[.](http://paperpile.com/b/BabBjK/eGtY)

[Wan, Zhibo, Youqiang Dong, Zengchen Yu, Haibin Lv, and Zhihan Lv. 2021. “Semi-Supervised Support Vector Machine for Digital Twins Based Brain Image Fusion.” *Frontiers in Neuroscience* 15 (July):705323.](http://paperpile.com/b/BabBjK/QdIQ)

[Yamini, B., K. Venkata Ramana, M. Nalini, D. Chitra Devi, B. Maheswari, and Subramanian R. Siva. n.d. “Customer Churn Prediction Model in Enterprises Using Machine Learning.” Accessed September 10, 2025.](http://paperpile.com/b/BabBjK/0hQZ) <https://search.proquest.com/openview/e58098ade9ececd57139ac85633c5e07/1?pq-origsite=gscholar&cbl=2037694>[.](http://paperpile.com/b/BabBjK/0hQZ)

[Zhang, Chunyang, and Wenjing Han. 2024. “Ensembles of Decision Trees and Gradient-Based Learning for Employee Turnover Rate Prediction.” *PeerJ Computer Science* 10 (October):e2387.](http://paperpile.com/b/BabBjK/H8vK)

[Zhang, Ping, Yiqiao Jia, and Youlin Shang. 2022. “Research and Application of XGBoost in Imbalanced Data.” *International Journal of Distributed Sensor Networks*. https://doi.org/](http://paperpile.com/b/BabBjK/KEfJ)[10.1177/15501329221106935](http://dx.doi.org/10.1177/15501329221106935)[.](http://paperpile.com/b/BabBjK/KEfJ)