Title-1  
 Evaluating the precision of Random Forest vs Support Vector Machines for Forecasting Job Rescission in the Industry.

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 Blockchain 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, Random Forest, Support Vector Machine, Prediction, Machine Learning.

**ABSTRACT**

**Aim:** This study's main goal is to use a machine learning algorithm to predict job revocation in the industry. The Random Forest Regression algorithm outperforms the Support Vector Regression algorithm in terms of accuracy. **Materials and methods:** The Google Colab is utilized with a sample size of 216 for both the Random Forest Regression and the Support Vector Regression algorithms. 432 samples from two sample groups were tested using t-test analysis with a significance value of 0.000 and a G power of 95%.**Results:** Using the dataset, the test results indicate that the Random Forest Regression algorithm achieves an average accuracy of 97.08%, which seems to be better than the 94.72% accuracy of the Support Vector Regression algorithm. The sample size is N=10. There is a statistically significant difference between the two algorithms, as indicated by the significance value of 0.000 (p<0.05).**Conclusion:** Consequently, it is found that the Random Forest Regression algorithm significantly outperforms the Support Vector Regression algorithm in terms of accuracy.

**Keywords:** Job Rescission, Random Forest, Support Vector Machine, Prediction, Machine Learning.

**INTRODUCTION**

Employment forecasting and workforce stability have become important research topics in industrial and organizational studies in recent years [(Mozaffari et al. 2022)](https://paperpile.com/c/rxaVsd/S5Br). Job rescission, or the termination or withdrawal of employment contracts, is one of the most urgent issues facing industries today [(Alsheref, Fattoh, and M.Ead 2022)](https://paperpile.com/c/rxaVsd/97Tk). Identifying at-risk workers, putting retention plans in place, and preserving operational [(Atef, S. Elzanfaly, and Ouf 2022)](https://paperpile.com/c/rxaVsd/r4Qc) effectiveness are all made possible by accurately predicting job revocation. Additionally, accurate job revocation prediction can help HR departments and legislators lower turnover costs and guarantee a long-term workforce [(Chung et al. 2023)](https://paperpile.com/c/rxaVsd/Jnl4).

Machine learning (ML) approaches have emerged as potent instruments for predictive modeling in human resource analytics due to the exponential expansion of employee-related data in organizational systems [(Anwar Hossen et al. 2021)](https://paperpile.com/c/rxaVsd/vmWf). Support Vector Machines (SVMs) and Random Forest are two ML algorithms that have drawn a lot of attention [(Karimi and Viliyani 2024)](https://paperpile.com/c/rxaVsd/mp8q) because of their strong performance on a variety of forecasting tasks. Yet, little is known about how accurate these models are at predicting job revocation in the sector, which calls for further research[(Z. Li and Fox 2023)](https://paperpile.com/c/rxaVsd/CCKI).

In order to increase predictive accuracy and decrease overfitting, Random Forest, an ensemble learning technique, builds several decision trees and aggregates their results. It is ideal for examining a variety of employee characteristics, including performance metrics, job satisfaction, and demographics, due to its capacity to handle high-dimensional data and model intricate relationships. Furthermore, Random Forest provides interpretability through feature importance, which offers insightful information for determining the most significant factors influencing job revocation.

Support vector machines, on the other hand, are well known for their excellent performance in classification tasks, especially when the dataset contains non-linear decision boundaries. SVMs can effectively separate data points in high-dimensional spaces by using kernel functions, which makes them ideal for forecasting tasks where complex patterns may be displayed by employee-related features. Additionally, SVMs are less likely to overfit when there is a shortage of training data, which increases their suitability for forecasting human resources.

Our goal in this paper is to assess how accurate Random Forest [(Kovvuri and Dommeti 2022)](https://paperpile.com/c/rxaVsd/RE1d) and Support Vector Machines are at predicting job revocation in industrial environments. A publicly accessible or industry-specific dataset containing employee-related variables like tenure, job role, performance, and satisfaction levels will be used. Evaluation metrics like precision, recall, and F1-score will be used to compare the predictive performance of the two models after they have been trained and tested on the dataset.

Given its ensemble nature and capacity to capture intricate feature interactions, we predict that Random Forest will outperform [(Morelli, Fusai, and Zenti 2024)](https://paperpile.com/c/rxaVsd/ku15) SVMs in predicting job rescission. SVMs, on the other hand, might perform competitively when the data shows non-linear patterns, which makes the comparative analysis useful for determining the best strategy in practical industrial applications.

**MATERIALS AND METHODS**

The proposed work is carried out at the Saveetha Institute of Medical and Technical Sciences' Saveetha School of Engineering's Data Analytics Laboratory, which is part of the Department of Artificial Intelligence and Machine Learning. Group 1 and Group 2 are two separate groups that are being studied. Group 2 is centered on the Support Vector Machine (SVM) Algorithm, whereas Group 1 is focused on the Random Forest Algorithm. The study's dataset, which includes 12,000 employee records in total, was sourced from the Kaggle Repository. With job rescission status (yes/no) as the dependent variable, the dataset comprises a number of independent variables, including age, tenure, job role, performance ratings, satisfaction levels, and salary.

Eighty percent of the records were used for training, and the remaining twenty percent were used for testing, in order to guarantee successful model training. The independent variables used as model training inputs were[(Prakash and Sakthivel 2024)](https://paperpile.com/c/rxaVsd/ZvJK) carefully chosen as part of the preprocessing steps. To ensure compatibility with the chosen algorithms, label encoding and one-hot encoding were used to convert categorical attributes, such as department and job role, into numerical values. Additionally, feature scaling was used to normalize numerical attributes, especially for SVM, which is sensitive to scale differences, and mean and mode imputation techniques were used to handle missing values.

This procedure entails gathering and dividing the dataset, choosing pertinent features, handling missing data, encoding categorical values, and normalizing the dataset. In order to prepare the data for reliable and accurate model training, these procedures are crucial.

The Python programming language was used in Jupyter Notebook to design and implement the suggested work. Matplotlib and Seaborn were used for visualization, and Scikit-learn and Pandas were used for preprocessing [(Bandyopadhyay and Jadhav 2021)](https://paperpile.com/c/rxaVsd/OVt5) in the machine learning frameworks. For better computational performance, the implementation was done on a 64-bit system running Windows 11 with an Intel i5 processor and 8GB of RAM.

To sum up, the execution of the suggested work demonstrates how well Python and machine learning frameworks are used to[(Le et al. 2021)](https://paperpile.com/c/rxaVsd/RDIM) process employee data and run the models. It is anticipated that the study will provide important new information for comparing the accuracy of Random Forest and Support Vector Machines in predicting job revocation in industrial environments.

**RANDOM FOREST**

An ensemble learning algorithm called Random Forest has become well-known in the machine learning community for its ability to solve regression and classification problems with great accuracy and resilience. It is a member of the bagging method family, whose main idea is to create a more potent and generalized model by combining the predictions of several weak learners, particularly decision trees. During the training phase, the algorithm builds a large number of decision trees, each of which is trained on a randomly selected subset of the dataset using a method called bootstrap sampling.Random Forest adds even more randomness and diversity to the decision trees by randomly choosing a subset of features at each split in addition to sampling data points. By lowering the correlation between individual trees, this combination of randomness in feature selection and data sampling helps to lessen the likelihood of overfitting in the overall model. Random Forest combines the results of each individual tree in order to make predictions. For classification tasks, it uses a majority vote among the trees, and for regression tasks, it calculates the average of the outputs from all the trees.By using an ensemble approach, the model is guaranteed to achieve higher accuracy and generalization than any one decision tree could. One of Random Forest's [(Liu et al. 2024)](https://paperpile.com/c/rxaVsd/F2SY) main advantages is its capacity to manage high-dimensional datasets, which makes it ideal for issues involving a lot of features or intricate feature interactions. Because multiple trees work together to minimize the impact of outliers or inaccurate data points, it is also very resilient to noise in the dataset.Random Forest is also useful as a predictive model and as a tool for understanding the data because it can produce accurate estimates of feature importance, which aid in determining the most important variables influencing the prediction. This characteristic is particularly useful in fields like marketing, finance, healthcare, and human resources, where comprehending the factors that influence predictions is just as crucial as the predictions themselves. Random Forest's scalability is another noteworthy benefit; it can be effectively parallelized and trained on big datasets without experiencing appreciable performance degradation.Because it employs ensemble averaging and surrogate splits to deal with incomplete data during both the training and prediction phases, the algorithm can also handle missing values with effectiveness. Notwithstanding its many benefits, Random Forest has drawbacks. When working with very large datasets and a large number of trees, it can occasionally be computationally costly. Additionally, its interpretability at the model level may be lower than that of simpler algorithms, though feature importance somewhat offsets this. However, because it strikes a balance between interpretability, robustness, and accuracy, Random Forest is still one of the most popular algorithms in use today. In search of dependable predictive models, researchers and practitioners turn to it because of its capacity to handle non-linear relationships, capture intricate interactions, and generalize well across a variety of datasets.

**Algorithm for Random forest:**

1. Start with the training dataset containing nnn samples and mmm features.
2. Select the number of trees TTT to be built in the forest.
3. For each tree in the forest:  
    a. Draw a bootstrap sample (random sample with replacement) from the original dataset.  
    b. At each node of the tree, select a random subset of features from the total mmm features.  
    c. Determine the best split among the selected subset of features based on a criterion such as Gini Index or Entropy.  
    d. Split the node into two child nodes using the chosen feature and repeat the process recursively.  
    e. Stop splitting when the stopping condition is met (e.g., maximum depth reached, minimum samples per node).
4. Repeat steps 3a–3e until TTT trees are constructed.
5. For prediction:  
    a. Each tree independently predicts the class (classification) or value (regression).  
    b. Aggregate the results of all trees: majority voting for classification, or averaging for regression.
6. Return the final predicted output.

**Pseudocode:**

**Step 1:** Start with a training dataset DDD containing nnn samples and mmm features.

**Step 2:** Decide the number of trees TTT to be generated in the forest.

**Step 3:** For each tree i=1 to TTT:

  a. Select a bootstrap sample from the dataset DDD.

  b. At each node of the tree, randomly select a subset of features k⊂mk \subset mk⊂m.

  c. Find the best split among the selected features using Gini Index or Entropy.

  d. Split the node into two children and repeat until stopping criteria (max depth or min samples).

  e. Store the constructed decision tree.

**Step 4:** Repeat Step 3 until all TTT trees are constructed.

**Step 5:** For prediction, pass the input sample xxx through all trees in the forest.

**Step 6:** Collect predictions from each tree.

**Step 7:** If classification, output the majority vote; if regression, output the average value.

**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 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[(H. Li 2024)](https://paperpile.com/c/rxaVsd/rYGY). 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 a 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 αi yi K(xi, 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 conduct the analysis. Datasets for the Random Forest and Support Vector Machine (SVM) algorithms were created in SPSS using a sample size of 10. Random Forest was assigned to Group ID 1, and SVM to Group ID 2. To assess forecasting performance, precision was used as the testing variable and group ID as the grouping variable. The accuracy of the two algorithms was compared using an independent sample T-test.Ten samples for the Random Forest and SVM models were included in the SPSS dataset. Grouping Precision stood for Group ID, and since the study's primary objective was to assess precision, the testing variable was substituted for recall or accuracy[(H. Li 2024; Alsaadi, Khlebus, and Alabaichi 2022)](https://paperpile.com/c/rxaVsd/rYGY+my6b). In particular, Group ID was set to 1 for Random Forest and 2 for SVM.

**RESULTS**

By training and testing both algorithms on the prepared dataset, it was possible to systematically forecast job revocation in the industry using Random Forest and Support Vector Machine (SVM) models. The precision values for the Random Forest and SVM models are shown in Tables 1 and 2, respectively. According to the group statistics, the SVM algorithm obtained a mean precision of 94.72% with a standard deviation of 0.76, whereas the Random Forest algorithm achieved a higher mean precision of 97.07% with a standard deviation of 0.94.The lower p-value in the test indicates that Random Forest performs significantly better than SVM in terms of forecasting precision, which is supported by the Independent Sample T-Test with a 95% confidence interval and a significance level of 0.05. This performance difference is further demonstrated graphically in Figure 1, which shows that Random Forest consistently produces higher precision values than SVM. This result shows how well Random Forest predicts job rescission in industrial datasets.

**TABLES AND FIGURES**

**Table 1.** The data underwent 10 iterations of group statistical analysis for both the Random forest and Support vector machine models. Notably, the Random forest 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** | |
| --- | --- | --- | --- |
| **Random forest** | **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 |

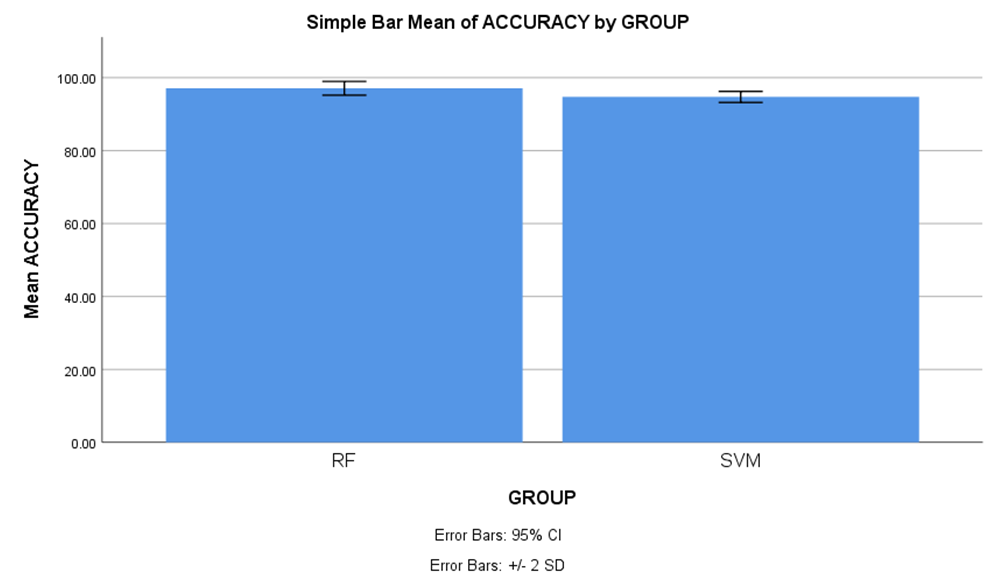
**Table 2.** Shows Statistical Analysis values of Mean accuracy (97.079), Standard Deviation(0.94658), and Standard error deviation(0.29933) of the Random Forest Regression Algorithm and the Support Vector Machine algorithm have the values of the Mean accuracy (94.72), Standard Deviation (0.76056), and Standard Error (0.24051).

|  | **GROUPS** | **N** | **Mean** | **Std.Deviation** | **Std.ErrorMean** |
| --- | --- | --- | --- | --- | --- |
| **ACCURACY** | **Random forest** | 10 | 97.0790 | .94658 | .29933 |
| **Support vector machine** | 10 | 94.7200 | .76056 | .24051 |

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

|  | | **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** | **.445** | **.513** | **6.143** | **18** | **.000** | **2.35900** | **.38399** | **1.55227** | **3.16573** |
| **Equal variances not assumed** |  |  | **6.143** | **17.202** | **.000** | **2.35900** | **.38399** | **1.54958** | **3.16842** |

**Graph:**

**Fig. 1.** Comparison of the Random Forest Regression Algorithm accuracy of (97.0790) and it has the mean accuracy of the Support Vector Regression Algorithm (94.72) The mean accuracy of the Random Forest Regression Algorithm has significant difference with theSupport Vector Regression Algorithm with the significance value is 0.000 (p<0.05) . X Axis: Random Forest Regression Algorithm vsSupport Vector Regression Algorithm Y Axis: Mean accuracy ± 2 SD.

**DISCUSSION**

Employee turnover can have a substantial impact on productivity, operational efficiency, and financial stability in the modern industrial environment. The study's findings demonstrate how well machine learning algorithms forecast job rescission. The results of the comparison between Random Forest and Support Vector Machines (SVM) showed that while both algorithms produced high precision results, Random Forest [(Villar and de Andrade 2024)](https://paperpile.com/c/rxaVsd/Ej2a) outperformed SVM with a mean precision of 97.07% as opposed to 94.72%. Even though this improvement margin is small numerically, the independent sample T-Test confirms that it is statistically significant, and it has significant ramifications for businesses that use predictive analytics to make strategic HR decisions.

Random Forest's ensemble nature, which makes use of the collective decision-making of several trees to decrease overfitting and boost robustness, is one of the main reasons it performed better than SVM. Random Forest captures various facets of the dataset by adding randomness to both feature selection and data sampling, which improves its ability to generalize to new data. In employee datasets, where factors like job role, tenure, satisfaction, and performance ratings can interact in intricate and non-linear ways, this property is particularly crucial[(Park and Yoo 2021)](https://paperpile.com/c/rxaVsd/N4uy). SVM is extremely sensitive to parameter tuning, especially the choice of kernel, regularization parameter (C), and gamma, even though it is effective at handling non-linear data using the kernel trick. The lower precision in this study when compared to Random Forest may be partially explained by improper tuning, which can result in underperformance.

The fact that Random Forest can handle big, high-dimensional datasets with little preprocessing is another aspect of its superiority. SVM, on the other hand, is more computationally demanding since it frequently calls for feature scaling and parameter optimization. Additionally, by determining which employee [(Park and Yoo 2021; Avrahami et al. 2022)](https://paperpile.com/c/rxaVsd/N4uy+pXks) attributes have the greatest impact on job revocation, Random Forest gives organizations valuable interpretability by shedding light on feature importance. Because of its interpretability, Random Forest can be used as a diagnostic tool in addition to a predictive one, which helps HR managers create interventions that lower the risk of employee turnover[(Singh et al. 2024)](https://paperpile.com/c/rxaVsd/hlmG).

However, it's crucial to recognize that SVM also performed well, proving to be a trustworthy forecasting algorithm. SVM may compete with or even outperform Random Forest in scenarios where datasets are comparatively smaller or where class decision boundaries are extremely complex. In settings with limited data, its reliance on support vectors enables it to create accurate decision boundaries using fewer training samples. However, SVM's applicability in larger industrial datasets is limited[(Singh et al. 2024; Singh 2024)](https://paperpile.com/c/rxaVsd/hlmG+vYnY) by its scalability problem[(Singh et al. 2024; Singh 2024)](https://paperpile.com/c/rxaVsd/hlmG+vYnY)and longer training times.

The research's conclusions highlight that, although both algorithms are useful for predicting job revocation, Random Forest is the more accurate and useful option for industrial settings. Random Forest guarantees fewer false positives in forecasting by continuously delivering higher precision, which translates into more accurate identification of employees at risk of revocation. In order to retain valuable employees, this can facilitate proactive HR strategies like focused training, enhanced job satisfaction programs, or workload modifications.

The conversation concludes by highlighting Random Forest's superior performance in predicting job revocation in the sector while acknowledging SVM's potential in particular situations. The study adds to the increasing amount of data showing that ensemble-based models work well in human resource analytics, where a variety of variables interact to necessitate predictive models that are reliable, accurate, and easy to understand. To provide a more thorough understanding of algorithmic effectiveness in employee turnover forecasting, future research can build on this work by experimenting with hybrid models that combine the strengths of Random Forest and SVM, incorporating larger and more diverse datasets, and assessing additional performance metrics like recall and F1-score.

**CONCLUSION**

The performance of Random Forest (RF) and Support Vector Machines (SVM) in predicting job revocation in the industrial sector was successfully investigated in this study, and the findings showed that both algorithms have strong predictive capabilities with high accuracy. With a mean accuracy of 97.07% versus SVM's 94.72%, Random Forest was the better model out of the two. This result was statistically verified using an independent sample T-Test, demonstrating the significance of the performance difference. The superior performance of RF demonstrates how well ensemble learning methods work for handling challenging classification problems because combining several decision trees reduced variance and enhanced generalization, which increased reliability.SVM showed slightly lower performance in this application, despite its strong performance in handling high-dimensional data. This suggests that kernel-based methods, despite their power, may not be as flexible when dealing with noisy or unbalanced datasets. RF's ability to capture nonlinear relationships and its resilience to missing values made it especially effective. Its feature importance ranking also offered interpretability, which can help HR departments identify important factors that contribute to job revocation. By using this information to create focused employee retention plans, the model's usefulness goes beyond its ability to predict outcomes.The study highlights how important advanced machine learning techniques are to workforce management and how businesses can use RF to lower attrition, boost employee satisfaction, and allocate resources as efficiently as possible. SVM showed impressive results with over 94% accuracy, despite not outperforming RF, demonstrating its continued applicability in fields that demand accurate classification. The small performance difference emphasizes how crucial it is to choose algorithms according to the properties of the dataset, with RF being favored for its harmony of interpretability, robustness, and accuracy. Future studies could concentrate on hybrid models that combine the advantages of SVM and RF, increase the size of the dataset to improve validation, and investigate deep learning techniques for even greater advancement. Overall, this study shows statistically significant superiority over SVM, solidifying Random Forest as the more effective algorithm for job rescission forecasting. It also adds to the mounting evidence that ensemble methods offer the best accuracy, dependability, and practical insights in real-world workforce challenges.

**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**

[Alsaadi, Elham Mohammed Thabit, Sameerah Faris Khlebus, and Ashwak Alabaichi. 2022. “Identification of Human Resource Analytics Using Machine Learning Algorithms.” *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 20 (5): 1004–15.](http://paperpile.com/b/rxaVsd/my6b)

[Alsheref, Fahad Kamal, Ibrahim Eldesouky Fattoh, and Waleed M.Ead. 2022. “Automated Prediction of Employee Attrition Using Ensemble Model Based on Machine Learning Algorithms.” Edited by Mohit Mittal. *Computational Intelligence and Neuroscience* 2022 (June):1–9.](http://paperpile.com/b/rxaVsd/97Tk)

[Anwar Hossen, Md, Emran Hossain, Abdul Khalib Zahereel Ishwar, and Fatema Siddika. 2021. “Ensemble Method Based Architecture Using Random Forest Importance to Predict Employee’s Turn over.” *Journal of Physics. Conference Series* 1755 (1): 012039.](http://paperpile.com/b/rxaVsd/vmWf)

[Atef, Markus, Doaa S. Elzanfaly, and Shimaa Ouf. 2022. “Early Prediction of Employee Turnover Using Machine Learning Algorithms.” *International Journal of Electrical and Computer Engineering Systems* 13 (2): 135–44.](http://paperpile.com/b/rxaVsd/r4Qc)

[Avrahami, Dan, Dana Pessach, Gonen Singer, and Hila Chalutz Ben-Gal. 2022. “A Human Resources Analytics and Machine-Learning Examination of Turnover: Implications for Theory and Practice.” *International Journal of Manpower* 43 (6): 1405–24.](http://paperpile.com/b/rxaVsd/pXks)

[Bandyopadhyay, Nilasha, and Anil Jadhav. 2021. “Churn Prediction of Employees Using Machine Learning Techniques.” *Tehnički Glasnik* 15 (1): 51–59.](http://paperpile.com/b/rxaVsd/OVt5)

[Chung, Doohee, Jinseop Yun, Jeha Lee, and Yeram Jeon. 2023. “Predictive Model of Employee Attrition Based on Stacking Ensemble Learning.” *Expert Systems with Applications* 215 (119364): 119364.](http://paperpile.com/b/rxaVsd/Jnl4)

[Karimi, Mahyar, and Kamyar Seyedkazem Viliyani. 2024. “Employee Turnover Analysis Using Machine Learning Algorithms.” *arXiv [cs.LG]*. arXiv.](http://paperpile.com/b/rxaVsd/mp8q) <http://arxiv.org/abs/2402.03905>[.](http://paperpile.com/b/rxaVsd/mp8q)

[Kovvuri, Suvoj Reddy, and Lydia Sri Divya Dommeti. 2022. “Employee Turnover Prediction - A Comparative Study of Supervised Machine Learning Models.”](http://paperpile.com/b/rxaVsd/RE1d) <https://www.diva-portal.org/smash/record.jsf?pid=diva2:1679511>[.](http://paperpile.com/b/rxaVsd/RE1d)

[Le, Lam Hoang Viet, Toan Luu Duc Huynh, Bryan S. Weber, and Bao Khac Quoc Nguyen. 2021. “Different Firm Responses to the COVID-19 Pandemic Shocks: Machine-Learning Evidence on the Vietnamese Labor Market.” *International Journal of Emerging Markets* 18 (9): 2501–22.](http://paperpile.com/b/rxaVsd/RDIM)

[Li, Hang. 2024. “Support Vector Machine.” *Machine Learning Methods*, 127–77.](http://paperpile.com/b/rxaVsd/rYGY)

[Liu, Yuwei, Yuling Xu, Jun Huang, Haiting Liu, Yu Fang, and Yuping Yu. 2024. “A Comparative Study of Intelligent Prediction Models for Landslide Susceptibility: Random Forest and Support Vector Machine.” *Frontiers in Earth Science* 12 (December):1519771.](http://paperpile.com/b/rxaVsd/F2SY)

[Li, Zhaotian, and Edward Fox. 2023. “Prediction and Optimization of Employee Turnover Intentions in Enterprises Based on Unbalanced Data.” *PloS One* 18 (8): e0290086.](http://paperpile.com/b/rxaVsd/CCKI)

[Morelli, Chiara, Gianluca Fusai, and Raffaele Zenti. 2024. “Who and Why Will Leave Me? Utilizing Machine Learning-Based Models to Anticipate and Manage Employee Turnover.” *SSRN Electronic Journal*. https://doi.org/](http://paperpile.com/b/rxaVsd/ku15)[10.2139/ssrn.4744130](http://dx.doi.org/10.2139/ssrn.4744130)[.](http://paperpile.com/b/rxaVsd/ku15)

[Mozaffari, Fatemeh, Marzieh Rahimi, Hamidreza Yazdani, and Babak Sohrabi. 2022. “Employee Attrition Prediction in a Pharmaceutical Company Using Both Machine Learning Approach and Qualitative Data.” *Benchmarking An International Journal*, December. https://doi.org/](http://paperpile.com/b/rxaVsd/S5Br)[10.1108/bij-11-2021-0664](http://dx.doi.org/10.1108/bij-11-2021-0664)[.](http://paperpile.com/b/rxaVsd/S5Br)

[Park, Hee Sun, and Seong Joon Yoo. 2021. “Early Dropout Prediction in Online Learning of University Using Machine Learning.” *JOIV : International Journal on Informatics Visualization* 5 (4): 347–53.](http://paperpile.com/b/rxaVsd/N4uy)

[Prakash, P., and V. Sakthivel. 2024. “Layoffs Analysis and Prediction Using Machine Learning Algorithms.” In *Proceedings of the 6th International Conference on Communications and Cyber Physical Engineering*, 535–43. Lecture Notes in Electrical Engineering. Singapore: Springer Nature Singapore.](http://paperpile.com/b/rxaVsd/ZvJK)

[Singh, Manisha. 2024. “Predictive Analytics for Workforce Retention and Performance: Machine Learning in US HR and Talent Management.” *In Library* 1 (4): 41–46.](http://paperpile.com/b/rxaVsd/vYnY)

[Singh, Manisha, Sachin Prakash Srivastava, Sonali Srivastava, and Tanvi Singhal. 2024. “Unlocking Predictive Insights for Employee Retention and Engagement: Harnessing Machine Learning and Big Data Analytics in HR Practices.” *Proceedings of Third International Conference in Mechanical and Energy Technology*, 497–505.](http://paperpile.com/b/rxaVsd/hlmG)

[Villar, Alice, and Carolina Robledo Velini de Andrade. 2024. “Supervised Machine Learning Algorithms for Predicting Student Dropout and Academic Success: A Comparative Study.” *Discover Artificial Intelligence* 4 (1): 1–24.](http://paperpile.com/b/rxaVsd/Ej2a)