Predicting the Perceived Employee Tendency of Leaving an Organization Using SVM and Naive Bayes Techniques

Ijeoma Lilian Emmanuel-Okereke*, Sylvanus Okwudili Anigbogu

Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria
Email: *ijayronk@yahoo.com

Abstract

There are several experienced and highly skilled employees considered to be assets in every organization; a good and flexible working environment is required to retain them. The perceived exit of well-skilled and highly experienced employees may result in financial losses, poor sales with customers' dissatisfaction and produced a low turnover. It also led to low production and output. The existing methods of operation lack the merit in producing accurate and reliable results. It could not generalize well with testing datasets and results in the problem of model over-fitting. Little or no work has been done in the area of predicting the perceived employee tendency of leaving an organization using Support Vector Machine (SVM) and Naive Bayes (NB) algorithm. The implemention was done using Python (Spyder IDE) in ANACONDA. In this paper, a model which is capable of predicting the perceived employee tendency of exiting an organization was developed using the support vector machine and the Naive Bayesian machine learning algorithm. The adopted techniques improved the prediction accuracy and generalized well with testing datasets in overcoming the problem of over-fitting. It also reduced the sudden occurrence of experienced and skilled employees leaving an organization. We adopted the SVM and NB to effectively handle overlapping and reduce data misclassification errors that can work well with a limited number of the dataset. The proposed NB model was trained, successfully tested and evaluated using the same dataset in comparison with the SVM technique. The experimental results of NB model produced 100% prediction accuracy with a 0.0000 RMSE error value in comparison with the SVM which gave a 97.00% success rate and 0.0258 RMSE value.

Subject Areas

Information Management

1. Introduction

Every business industry or organization that deals with skilled workers of human capital development focuses on profit maximization, turnover and cost minimization. There are profit-driven industries in the world today that suffer the backlog of financial losses caused by the exit of skilled and highly experienced workers [1]. The existing methods of operation are not accurate and reliable, and could not generalize well with testing datasets which results in the problem of model over-fitting.

Job exit is the process of an employee leaving an organization or industry which at times is beyond the control of that organization [2]. The act of highly experienced employees leaving an organization voluntarily and involuntarily can be controlled if we have a system that predicts the occurrence before it happens [3]. The voluntary exit of an employee from an organization is a crucial and important issue that results in a decline in human capital and financial loss when the best staff often leaves without prior knowledge. There are several factors responsible for or causes the perceived employee tendency of leaving an organization, namely: the organizational factors and individual factors [4]. The organizational factors include low salary, workload not commeasurable to salary and overtime pay, too many requirements for advancement, lack of appreciation for a job well done, etc [5]. The individual factors include frequent late-night meetings, family obligations, work conflicting with personal life responsibilities, personal relationships. There are other factors identified by [6] as new rules and organizational policies, lack of monetary benefits, extensive workload and stress, lack of leadership qualities, the relationship between manager and promotions, not being involved for staff training. Skilled and highly experienced employees are considered assets to any organization; therefore, a good and flexible working environment is recommended to retain them [7].

It is difficult and competitive to have qualified and highly experienced staff in fulfilling the needs of any organization around the globe. The success and efficiency of every organization depend on its capacity to retain skilled and well-experienced employees. The exit of well-skilled and highly experienced employees may result in losses in millions and billions of Naira, loss of revenue, poor sales and customers' dissatisfaction with a low turnover in meeting up with its goal and objectives in an organization. The unskilled employees are prone to make more errors which may give rise to low production, output and little or no work has been done in the area of predicting the perceived employee tendency of leaving an organization. And most of the existing systems in the application suffered

from data misclassification errors with overlapping patterns. It produced more False Positive (FP) cases than the True Positive (TP) and more False Positive (FP) classification than False Negative (FN) cases caused by model over-fitting.

The aim is to build a model capable of predicting the perceived Employee tendency of exiting an Organization using Support Vector Machine (SVM) and Naive Bayesian (NB) technique. This model is employed to help reduce overlapping and misclassification errors which affected the performance accuracy of the existing system techniques and could not generalize well with the testing dataset.

This paper is divided into different sections as followings: Section 1 contains the introduction, Section 2 presents a brief review of previous approaches relating to the study area and the gap in exploring the proposed model; Section 3 introduces materials and methods employed for developing the model; Section 4 focuses on the results and detailed discussion of results; Section 5 presents the conclusion to the paper.

2. Related Work

Saradhi & Palshikar [8] stated reasons for an employee exiting an organization to have a better offer or career growth relating to better salary, promotions, staff training and work environment. Ramamurthy *et al.* [9] developed a model in order to predict those employees qualified to be trained for a particular skill that suits their job function. Singh *et al.* [10] proposed a study on employee attrition using the C5.0 type of decision tree technique but suffers from model over-fitting and misclassification errors. Maharjan [11] developed a model to predict employee churn using SVM, Logistic regression and decision tree classifiers. An Extra tree class was invoked from Python SKlearn library to compute the score for all features. The information gain was used to filter relevant features, analyzed and compared without any order or rank. It uses internal mechanism in ranking the features without relying upon user calculated and ranked values. The dataset was imbalanced and employee attributes are less significant compared to non employee attributes. The dataset was divided into the ratio 70:30:70 of training and testing set to overcome the problem of over-fitting for the purpose of learning using a Stratified K-fold cross-validation test. A k value was defined based on what the entire dataset will get and divided in forming that number of K-folds. It provided a uniform data distribution with majority and minority across testing and training items. The LR performed better than the SVM. The accuracy rate was below average and could not be extended to work with handle cloud-based platform. Jayad *et al.* [12] proposed the use of Naive Bayes (NB) classification model as a technique in machine learning to predict employee performance drives the success of every organization. The implementation was done using WEKA and correctly classified and predicted target variable as required with 95.48% accuracy rate in 0.01 seconds and update performance score of 96.77% metrics of accuracy. The accuracy of NB increased along with the number of instances. The confusion matrix recorded more correctly predicted values (TP + TN) than wrongly

predicted values (FN + FP). The model flags up error message when instances are below ten (10) and accuracy level could not be computed.

Yahia *et al.* [13] adopted a deep machine leaning technique predict employee attrition support system. The deep driven machine learning approach was employed to detect key employee attributes with feature extraction technique with two different dataset that influences staff attrition. A small size of human resourced dataset of about 450 responses and a large sized kaggle HR dataset of 15,000 samples are used to train and test the model. The voting classifier performed better and produced 99% in terms of accuracy with real life dataset compared to other methods. The model was not suitable to work with imbalanced dataset with companies that have high turnover. Kamath *et al.* [14] employed the combination of Random Forest (RF), SVM, DT and Logistic Regression (LR) machine learning techniques for human resource attrition status, management and forecasting. The dataset was divided into training, testing and validation set in the ratio of 70%, 15% and 15% respectively. The results revels that employee attrition depends mainly on employee satisfaction as compared to other features and attributes. The RF was the best in performance with r-square value of 0.9773 while other like DT, SVM and LR recorded 0.8473, 0.8315 and 0.2299 respectively. The model could not work with large and unstructured dataset. Alshehhi *et al.* [15] combined DT and RF classifiers in machine learning to predict employee retention rates in an organization. The FR classifier was employed to forecast employee characteristics with retention rates using a training data of 13-years and testing dataset of 14-years.The dataset was divided into two to avoid model overfitting. It was trained to predict the occurrence of employee retention across each year, categories of department and training and used to determine if the organization losing an experienced Staff or not because of training and retraining. The RF classifier outperformed the DT technique in terms of accuracy and error rate. The RF and DT techniques could not work with large volume of dataset.

Senanayake *et al.* [16] employed the RF learning algorithm in ML to predict employee resignation in Swedish armed forces. The RF model was train to learn and predict employee that are due for resign and recommend possible recruitment policies that can be used to replace such retiring employees using a sizable dataset. The RF model produced 89.067% accuracy in comparison with the zero-guess that gave 84.533%. The dataset was quite small to achieve high accuracy rate as required.

3. Materials and Methods

In this paper, we are focusing on the use of SVM and Naive Bayes (NB) techniques to handle the problem of outliers efficiently with better accuracy rate and effectively handle overlapping classifications. The gamma, C set to 1.0 and random state variables are employed in the SVM class to have a better performance rate. We are adopting the Gaussian Naive bayes type of classifier because it is highly

scalable with number of data points, predictors and not sensitivity to irrelevant data features.

3.1. Data Source

The dataset (Table 1) used was obtained from a well-structured self study questionnaire distributed and collected through survey as a primary source containing five hundred and fourteen (514) items with attributes: timestamp, promoted, job satisfaction, work hour per day, training and working experience, job security, changed jobs, and employee exit as target. The dataset was divided into 80% training $\frac{80}{100}\% \times 514 = 412$ items and 20% testing $\frac{20}{100}\% \times 514 = 102$ set for predicting the perceived employee tendency of leaving an organization.

3.2. Data Preprocessing

The pre-processing stage is necessary for the training and reduces threshold value. It was adopted to help manipulate data and improve model performance because in gathering data sometimes poses difficulties and may result into out-of-range, missing, noisy and false data values. This involves data cleaning, instance selection, data normalization, transformation, feature extraction and selection. The preprocessing produces training data as output which can effectively be interpreted by models.

3.3. Classification

The classification system is adopted as a supervised learning process of determining or predicting data classes referred to as target, labels or categories. Classification is a predictive task or modeling of estimating a mapping function from

Table 1. Employee dataset.

	Gender	Experience	Job security	Working hours	- - - -	Target
0	Female	8	Yes	4	- - - -	Yes
1	Male	25	Yes	8	- - - -	No
2	Female	7	Yes	8	- - - -	No
3	Female	15	Yes	7	- - - -	No
4	Female	13	Yes	6	- - - -	No
- - -	- - - -	- - -	- - -		- - - -	- - -
510	Female	30	Yes	8	- - - -	No
511	Female	3	Yes	8	- - - -	No
512	Female	4	No	10	- - - -	Yes
513	Female	15	Yes	9	- - - -	No
514	Female	15	Yes	9	- - - -	No

input variables represented as "X" variable to a discrete output variable represented with "y". It depend mainly the area of application and the nature of available dataset [17]. The NB classifier is work based on Bayes theorem under simple assumption and the attributes are conditionally independent.

3.4. Feature Extraction

The feature selection process was adopted to determine the correlation between variable or attribute pars based on the level of correlation using a score value. The higher the score value the higher the correlation between attribute pairs [18].

3.5. Support Vector Machines (SVM) Classifier

The SVM is one of the simplest and more preferred machine learning techniques used by data professionals because it tendency of producing better and high accuracy with less computational error [19]. The SVM uses two main concepts namely: hypothesis space and the loss functioning finding an "optimal" hyper-plane as a solution to any learning problem [20]. The SVM is memory efficient and uses subsets of training data points in the support vectors called decision function. The simplest formulation of SVM is the linear one, where the hyper-plane lies on the space of the input data [21]. The SVM estimator was defined on the training dataset and tested to effectively predict the target variable. A SVM classifier was invoked from the sklearn.svm library in python and SVM model created. The gamma variable set to be scalable, c=1,0 and random states set to 101 with the Python script: svc=SVC(gamma='scale', C=1.0, rando_state=101). The model was trained with training dataset with svm.fit(X_train,y_test) and predicted using the testing dataset[svc.predict(X_test)]. The visualization was done using mat_plot_lib library in python. A SVM classifier was created with the pre-processed training data to make predictions about employees exit.

3.6. Naive Bayes Classifier

The Naive Bayes (NB) technique is one of the most popular known supervised machine learning algorithms that uses Bayes theorem. The Gaussian NB classification algorithm works with the principles of conditional probability as given by Bayes theorem. The Bayes theorem gives the conditional probability of an event "H" given whether event "D" has occurred. The Bayesian theorem basically computes the conditional probability of the occurrence of an event based on prior knowledge of conditions that might be related to the event [22]. It provides update to probability of hypothesis (H) for some given instance of data (D) which can be expressed in Equation (3.1) as follows:

$$P(H/D) = \frac{P(D/H)P(H)}{P(D)} \tag{3.1}$$

where P(D/H) is the probability of hypothesis and $P(D)$ dataset features/parameters.

The character or feature variables are encoded using label encoder at preprocessing stage and feature scaling technique employed for the training and testing dataset of the independent variables in producing better classification report.

The D is given as:

$$D = \left(d_1, d_2, d_3, \mathsf{L}\ d_n \right) \tag{3.2}$$

where d_1, d_2, d_3, ..., d_n represents the features mapped into the outlook.

Algorithm 1: Naive bayes (NB)

Step	Processes involved
1	Start
2	**Input:** Training_Dataset(T)
3	$F = (f_1, f_2, f_3, ..., f_n)$ // the predictor variables for testing items
4	**Output:** Class of testing items
5	**Compute** mean and standard deviation of predictor variables in each class
6	**Repeat** this step
	(a) Compute probabilities required for the Bayesian theorem for Exiting employees
	(b) Compute posterior probability of all those are not leaving the organization
7	**Compute** the likelihood of each class(first and second class)
8	**Get** the greatest likelihood
9	Return

Algorithm 2: Support vector machine (SVM)

Step	Processes involved
1	Start
2	Find candidate_SV with closest pair from classification (SV => support vector)
3	If there are violating points:
4	Find violating_points
5	Compute the candidate_SV= candidate_SV + voilating_points)
6	If there is any $a_p < 0$ due to the addition of c to S that gives negative:
7	Candidate_SV = candidate_SV
8	Repeat module to prune all data points
9	end_if
10	end_if

3.7. Performance Evaluation

The prediction accuracy, confusion matrix, classification report and ROC curve are employed to evaluate the performance of SVM and NB classifiers. The Classification accuracy is the ratio of correctly classified data points to the total no.

of points in the dataset which ranges from 0% - 100%.

$$\text{Classification accuracy} = \frac{\text{Number of correct classifications}}{\text{Total number of classifications}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3.3)$$

Precision: is a metrics used to measure the positive classifications represented as follows?

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3.4)$$

Recall is a metric used to measure the false negative classifications represented as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3.5)$$

F1-score: takes into consideration the true positive and false positive regardless of false negative and false positive classifications. The F1-score is sensitive to which class is positive and negative as given below in Equation (3.6):

$$\text{F1-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}} \quad (3.6)$$

The RMSE is a diagnostic tool employed to evaluate the quality of model predictions. It shows how far the model predictions fall from measured true values using the Euclidean distance. The RMSE computes residual and mean of each data point with the square of the same mean. The RMSE can be expressed in Equation (3.7) as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \left\| Y(i) - \widetilde{y}(i) \right\|^2}{N}} \quad (3.7)$$

where N is the number of points, $Y(i)$: the i-th measurement and $\widetilde{y}$ is the corresponding predictions

4. Results and Discussion

The results of SVM and NB classifiers are obtained through the use of seaborn heatmap, clustering graph, Charts and tables. The heatmap was employed in visualizing the correlations between target variable and other attributes or variables of the dataset. Clustering graph to group the nodes of exiting and not exiting employees into two different clusters represented with red and green colors. The Bar plots to show the categorical data with heights proportional to the value it represented and tables as a useful structural representation of organizing data into rows and columns. The design and implementation was done with some varying fine-tuned hyper-parameter values to have a better classification result. The prediction and classification accuracy of both model are visualized and discussed using confusion matrix, ROC and classification report as given bellow.

Figure 1 is the heat map or correlation matrix used to measure the relationship between variables. The matrix depicts a linear correlation between all possible
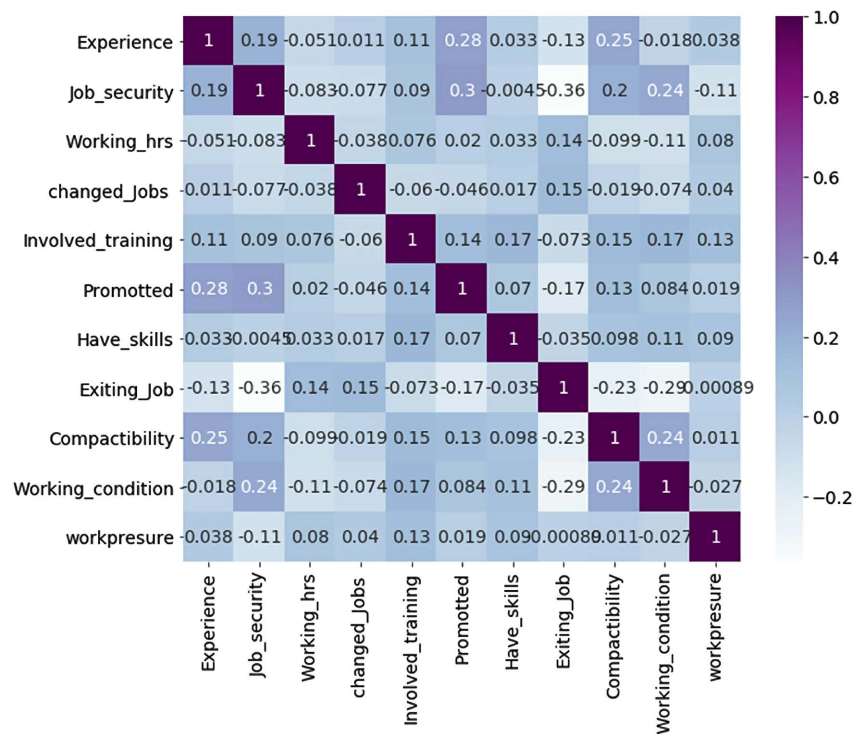
Figure 1. Correlation matrix of the proposed system dataset.

pairs of employee experience, job security, working hours, no. of changed jobs, promotions, job exit and etc. There is a strong relationship as shown in the main diagonal and other pairs.

Figure 2 depicts the number of those employees perceived to be leaving represented with red and those not exiting using blue color obtained from the proposed system dataset. The exiting employees as obtained from the dataset gave 245 items and those not leaving produced 269 items as visualized.

Figure 3 shows the two different clusters employees using the NB classifier obtained from the proposed system dataset. The employees perceived to exit the organization are grouped into one cluster as represented with red color and those staying with another cluster with blue color.

Figure 4 depicts the employee years of working experience as visualized and arranged in ascending order. It ranges from 1 to 39 as obtained from the proposed system dataset for decision making. The employees with 39-years of experience are very few and the least in number which requires good treatment in retaining them and compared to those with 10-years of experience with the highest number of staff as shown in the Bar chart.

Figure 5 shows the employee working hours per day that ranges from 3, 4, 5 hours to a maximum of 45 hours including overtime as obtained from the dataset. Those employees that work 8 hours per day are the highest compare to those spend 45, 25, 15, 14 hour and so as show in the Bar plot.

Figure 6 depicts the confusion matrix of proposed SVM classifier with leading diagonal elements or values showing the total number of correctly predicted values
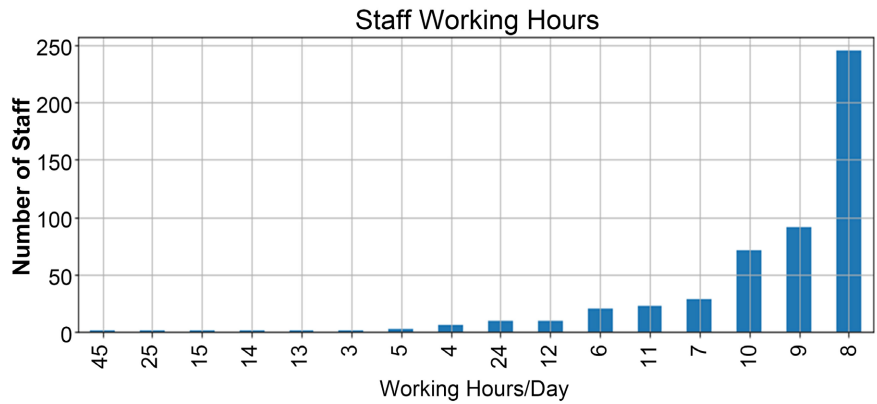
Figure 2. Classes of employee.

Figure 3. Clusters of employee by NB model.

Figure 4. Employee working experience.

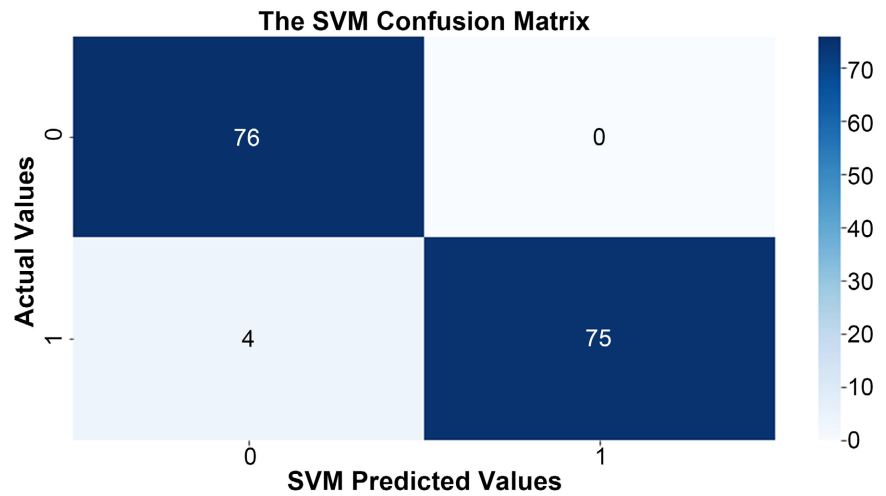**Figure 5.** Employee working hours per day.

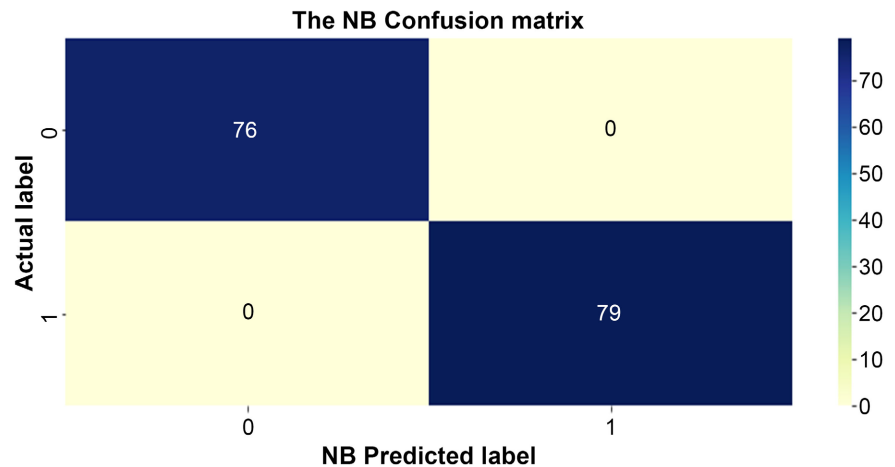**Figure 6.** The SVM confusion matrix.

that are equal to the actual or true values above and below the main diagonal cell values or off-diagonal elements shows the wrongly predicted values. The higher the diagonal values the better the prediction accuracy. From the confusion matrix: The total No. of correct predictions = TP + TN = 76 + 75 = 151 and wrong predictions = FP+ FN = 4 + 0 = 4.

Table 2 depicts the classification report of SVM containing the precession, recall and f1-score accuracy of exiting and not exiting employees. The precision accuracy score those employees not leaving the organization produced 0.95 and those exiting gave 1.00. The recall score for exiting employees recorded 0.95 and those not leaving to be 1.00 and f1-score 0.97 for both employees either exiting or not leaving.

Figure 7 shows the confusion matrix of the proposed NB classifier at testing stage with the correct predictions displayed at the secondary diagonal and wrongly predicted values recorded above and below the main diagonal called the off-diagonal elements. The total No. of correct predictions = TP + TN = 76 + 79 = 155 and wrongly predicted = FP + FN = 0 + 0 = 0 shown in Figure 7 where TP is true positive, FP false positive, FN false negative and TN true negative.

Table 2. The classification report of SVM.

	Precision	Recall	F1-score	Support
0	0.95	1.00	0.97	76
1	1.00	0.95	0.97	79
Accuracy		0.97	0.97	155
Macro avg	0.97	0.97	0.97	155
Weighted avg	0.98	0.97	0.97	155

The NB Confusion matrix

Figure 7. The NB confusion matrix.

Table 3. The classification report of NB.

	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	76
1	1.00	1.00	1.00	79
Accuracy			1.00	155
Macro avg	1.00	1.00	1.00	155
Weighted avg	1.00	1.00	1.00	155

Table 3 shows the classification report of NB classifier with precession, recall and f1-score classification accuracy of 1.00 for exiting and not exiting employees in an organization. The precision accuracy score gage 1.00, recall 1.00 and f1-score to be 1.00. There is a diplomatic tire as recorded for the precision, recall and f1-score values from the classification report.

Figure 8 is the Receiver Operating Characteristic (ROC) graph of SVM and NB classifiers showing the trade-off between sensitivity or true positive rate and specificity (1-FPR). The NB ROC curve is closer to top-left corner of the graph and is perfect and performed better than SVM model. The SVM curve is a bit away from the top X- and Y-axis with respect to the number of False Positive Rate (FPR) and True Positive Rate (TPR) as shown in the ROC graph. The proposed gave points lying along the diagonal (True Positive Rate = False Positive Rate) as expected.
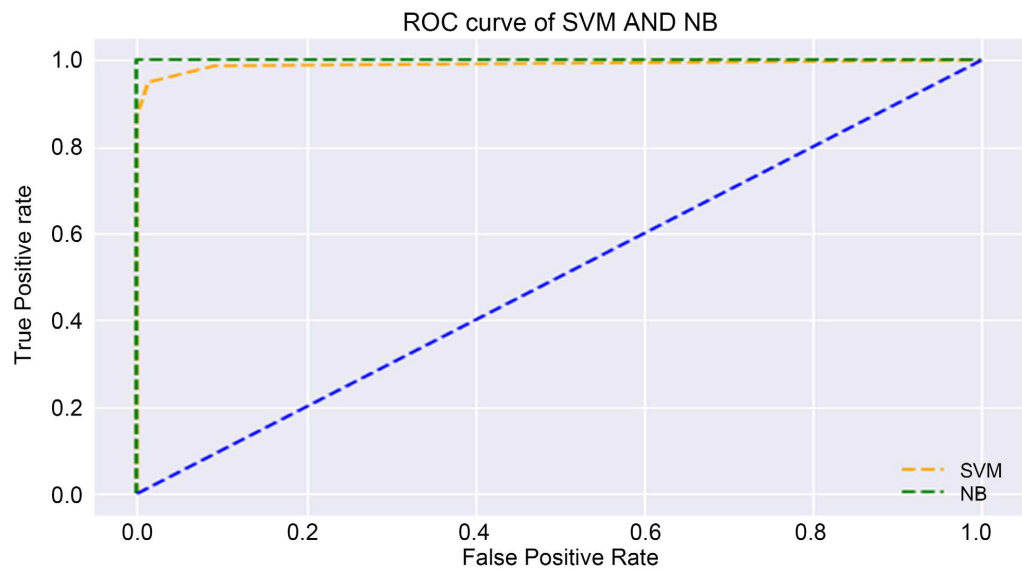
Figure 8. The ROC curve of SVM and NB.

Table 4. Training and validation test accuracy.

Model	Prediction Accuracy (%)	RMSE
SVM	97.0%	0.0258
NB	100%	0.0000

Table 4 shows the prediction accuracy measured in percentage and RMSE of SVM and NB classifiers. The result of NB classifier is recorded 100% to be faster with no RMSE value compared to the SVM techniques that produced 97.0% accuracy with 0.0258 RMSE value of testing dataset.

5. Conclusion

The prediction of the accuracy of NB was higher compared to SVM in terms of prediction accuracy and RMSE for all fine-tuned hyper-parameter values in determining the anticipated exit of skilled and highly experienced employees from an organization. This will help industries detect and prevent the possible occurrence of experienced employees' exit that may pose a danger to their throughput and can serve as a benchmark to other researchers because the model is scalable. The SVM prediction accuracy was high but recorded with a small error rate compared to the NB with a 100% accuracy rate with zero RMSE margin. The use of Python programming language simplified the implementation task because it has several machine learning inbuilt libraries and classes with deployable tools which can be achieved through a few lines of codes been optimized to achieve its best performance level.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Khera, S. and Divya, N. (2018) Predictive Modelling of Employee Turnover in Indian IT Industry Using Machine Learning Techniques. *Vision*, **23**, 12-21. https://doi.org/10.1177/0972262918821221

[2] Pandiyan, P., Kannadasan, K. and Vinoth, R. (2013) Prospective Control in an Organization through Two Grade Systems. 4*D International Journal of Multidisciplinary Research and Development*, **1**, 22-25.

[3] Kalaivani, J., Vinoth, R. and Elangovan, S.R. (2014) Survival Time to Trace the threshold Grade Level in an Organization. *International Journal of Multidisciplinary Research and Development*, **1**, 22-25.

[4] Morrell, K., Loan-Clarke, J. and Wilkingson, A. (2004) The Role of Shocks in Employee Turnover. *British Journal of Management*, **15**, 335-349. https://doi.org/10.1111/j.1467-8551.2004.00423.x

[5] Kannadasan, K., Pandiyan, P., Vinoth, R. and Saminathan, R. (2013) Time to Recruitment in an Organization through Three Parameter Generalized Exponential Model. *Journal of Reliability and Statistical Studies*, **6**, 21-28.

[6] Morrell, K. (2005) Towards a Typology of Nursing Turnover: The Role of Shocks in Nurses' Decision to Leave. *Journal of Advanced Nursing*, **49**, 315-322. https://doi.org/10.1111/j.1365-2648.2004.03290.x

[7] Kuwaiti, A.A., Raman, V., Subbarayalu, A.V., Palanivel, R.M. and Prabaharan, S. (2018) Predicting the Exit Time of Employees in an Organization Using Statistical Model. *International Journal of Scientific and Technology Research* (*IJSTR*), **5**, 213-217.

[8] Saradhi, V.V. and Palshikar, G.K. (2011) Employee Churn Prediction. *Expert Systems with Applications*, **38**, 19-30. https://doi.org/10.1016/j.eswa.2010.07.134

[9] Ramamurthy, K.N., Singh, M., Davis, M., Kevern, J.A., Klein, U. and Peran, M. (2015) Identifying Employees for Re-Skilling Using an Analytics-Based Approach. 2015 *IEEE International Conference on Data Mining Workshop* (*ICDMW*), Atlantic City, 14-17 November 2015, 345-354. https://doi.org/10.1109/ICDMW.2015.206

[10] Singh, M., Varshney, K.R., Wang, J., Mojsilovic, A., Gill, A.R., Faur, P.I. and Ezry, R. (2012) An Analytics Approach for Proactively Combating Voluntary Attrition of Employees. 2012 *IEEE* 12*th International Conference on Data Mining Workshops* (*ICDMW*), Brussels, 10-13 December 2012, 317-323. https://doi.org/10.1109/ICDMW.2012.136

[11] Maharjan, R. (2011) Employee Churn Prediction Using Logistic Regression and Support Vector Machine Support Vector Machine. SJSU Scholar Works: Master Degree Work, 1-72.

[12] Jayadi, R., Firmantyo, H.M., Dzaka, M.T.J., Sualdy, M. and Putra, A.M. (2019) Employee Performance Prediction Using Naive Bayes. *International Journal of Advanced Trends in Computer Science and Engineering*, **6**, 3031-3035. https://doi.org/10.30534/ijatcse/2019/59862019

[13] Yahia, N.B., Hlel, J. and Colomo-Palacios, R. (2017) From Big Data to Deep Data to Support People Analytics for Employee Attrition Prediction. *IEEE Access*, **34**, 1-12.

[14] Kamath, R.S., Jamsandekar, S.S. and Naik, P.G. (2019) Machine Learning Approach for Employee Attrition Analysis. *International Journal of Trend in Scientific Research and Development* (*IJTSRD*), **5**, 62-67. https://doi.org/10.31142/ijtsrd23065

[15] Alshehhi, K., Zawbaa, S.B. and Tariq, M.U. (2021) Employee Retention Prediction in Corporate Organizations Using Machine Learning Methods. *Academic of Entrepreneurship Journal*, **27**, 1-16.

[16] Senanayake, D., Muthugama, L., Mendis, L. and Madushanka, T. (2015) Customer Ch-

urn Prediction: A Cognitive Approach. *Internation Journal of Computer, Electrical, Automation, Control and Information Engineering*, **9**, 23-43.

[17] Foley, A.E. (2019) Using Machine Learning to Predict Employee Resignation in the Swedish Armed Forces. Kth Royal Institute of Technology School of Industrial Engineering and Management, Stockholm, 1-85.

[18] Panjasuchat, M. and Limpiyakorn, Y. (2020) Applying Reinforcement Learning for Customer Churn Prediction. *Journal of Physics: Conference Series*, **1619**, 12015.
https://doi.org/10.1088/1742-6596/1619/1/012016

[19] Zhang, X., Zhu, J., Xu, S. and Wan, Y. (2012) Predicting Customer Churn through Interpersonal Influence. *Knowledge-Based Systems*, **28**, 97-104.
https://doi.org/10.1016/j.knosys.2011.12.005

[20] Bryant, P.C. and Allen, D.G. (2013) Compensation, Benefits and Employee Turnover: HR Strategies for Retaining Top Talent. *Compensation and Benefits Review*, **45**, 171-175. https://doi.org/10.1177/0886368713494342

[21] Byun, H. and Lee, S.W. (2003) A Survey on Pattern Recognition Applications of Support Vector Machines. *International Journal of Pattern Recognition and Artificial Intelligence* (*IJPRAI*), **17**, 459-486. https://doi.org/10.1142/S0218001403002460

[22] Ramakrishnan, R., Bhattacharya, S. and Dhanya, P. (2018) Predict Employee Attrition by Using Predictive Analytics. *Benchmarking: An International Journal*, **26**, 2-18.
https://doi.org/10.1108/BIJ-03-2018-0083