Applied Science, Engineering and Management Bulletin [ASEMB]
ISSN: 3049-3005
Vol 3 Issue 02 (Apr-June 2026) |Pg:62-67

A Hybrid Predictive Model for Early Detection of Employee Attrition in Organizations

Dr. Melanie Lourens[1], K Venkata Lakshmi[2], Dr. Kolachina Srinivas[3], Dr. Sulaiman Ibrahim Kassim[4]
[1]Deputy Dean Faculty of Management Sciences, Durban University of Technology South Africa
melaniel@dut.ac.za, Orcid: 0000-0002-4288-8277
[2]Department of CSE, GMRIT Deemed to be UNIVERSITY, Rajam
[3]Associate Professor, Skyline University Nigeria, srikolachina81@gmail.com
[4]Department of Business Administration, Federal University Dutse, Nigeria
thinkingkassim@gmail.com

Abstract:

Employee attrition has resulted in the loss of tacit knowledge, high cost of hiring and morale of a team, which is a significant operational and financial challenge to the contemporary organizations. The traditional reactive plans to turnover are not sustainable in a dynamic business environment where active human capital management is significant. The paper will formulate and validate a hybrid predictive model that will predict employee attrition early on by utilizing the support of K-Nearest Neighbors (KNN) algorithm and Support Vector Machine (SVM) classifier. The proposed methodology involves the usage of a systematic data preprocessing pipeline, and feature engineering techniques in one of the synthetic human resources data sets, and stacked ensemble architecture.

Keywords: *Employee Attrition, Hybrid Predictive Model, Support Vector Machine, K-Nearest Neighbors, Stacked Ensemble, Human Resources Analytics.*

Introduction:

The contemporary business environment has been characterized by stiff competition on the skilled talent and employee retention has been a strategic issue of sustainability in the organization. The undesirable employee turnover may disrupt the flow of the working process, generate considerable costs in the form of recruiting, hiring new staff, and training them, and may lose the experience that is the foundation of competitive advantage (Nagpal and Pawar, 2024). According to industry estimates, a replacement of a mid-level employee may cost up to 50 to 150 percent of the annual salaries, hence has a direct impact on profitability. This has led to forecasting of those employees, who are likely to leave the company before they can write a resignation note, becoming a cornerstone of proactive human capital management.

The traditional ways of acquiring an understanding of attrition. They do not give a chance to the organization to intervene since they provide their perceptions after the decision has been made. Having seen the advent of big data analytics and machine learning, new possibilities have emerged that enable to turn human resources into a predictive, as compared to a reactive, role (Kumari Jaya Kumar, 2024). Predictive models can engage historical employee data like demographics, pay, job titles, job performance indicators and job environment conditions in a manner that identifies hidden trends and early warning signs of turnover.

Every machine learning model is associated with its limitations and benefits. K-Nearest Neighbors (KNN) is straightforward and effective to learn local data structures, as well as vulnerable to irrelevant features and the curse of dimensionality. SVM are highly efficient in high dimensional spaces and can model complex decision boundaries through the application of kernel functions but are computationally intensive and less interpretable. Now that it is known that no single algorithm is universally optimum, this paper develops a hybrid predictive model comprising of KNN and SVM in a stacked ensemble format (Talebi et al. 2025). The general hypothesis is that a combination of these two varying algorithmic perspectives will have a more powerful and superior predictor than either of the two frameworks individually. This hybrid model will be designed, implemented and validated in the present paper and its performance will be compared with the base classifiers and most salient features that will lead to the attraction of the risk of attrition will be identified.

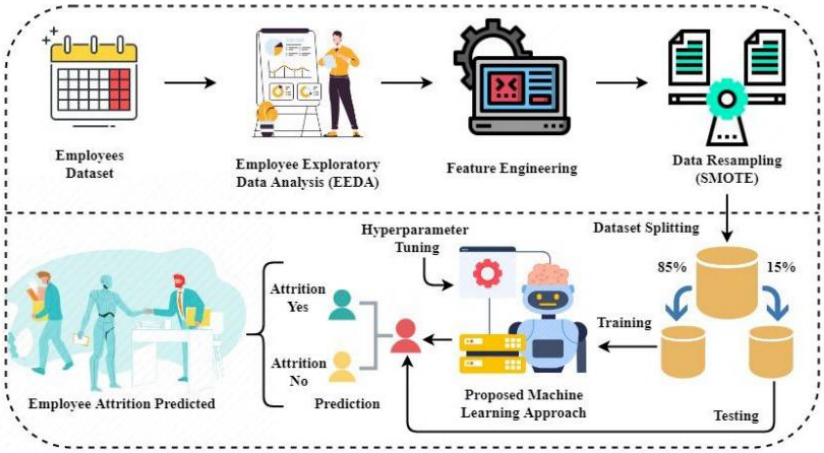

Figure 1: Predicting Employee Attrition Using Machine Learning

(Source: Raza et al. 2022)

Literature Review:

Application of machine learning in business decisions making has been increasing at a very fast rate with predictive analytics emerging as a key element to streamline the outcome of any functional field. Human resource has embraced predictive analytics, and this has enabled organizations to exceed reporting turnover rates to forecasting their risk of losing employees at individual level. The author concentrated on the fundamental algorithms, hybrid modelling techniques and the key theoretical constructs of employee turnover in this literature review (Alqahtani et al. 2024).

Many machine learning algorithms have been popularly applied in the business classification problems. The K-Nearest Neighbors (KNN) is a lazy learning algorithm that is a non-parametric technique that employs its k nearest neighbors in the feature space to classify a data point. Halder et al. (2024) explain the KNN modifications in detail, mentioning that its simplicity is a strength, and it may be adversely influenced by the high dimensionality of the data, and the imbalanced classes. The default KNN may be skewed to the majority when using human resources data in which the attrition events are usually a minority group. The Support Vector Machines (SVM) adopt a much different approach to this and construct an optimum hyperplane that maximizes the margin between the classes. Valkenborg et al. (2023) note that SVM is particularly applicable when high-dimensional space is involved and can handle nonlinear relationships with nonlinear kernel functions such as radial basis function (RBF). Nurhidayat and Pimpunchat (2023) demonstrated that the accuracy of classification is critically dependent on the kernel used. SVMs are however prone to parameter tuning and can produce black-box models which are hard to explain using other explainability tools.

Figure 2: Attrition Forecasting

(Source: Infeedo, 2025)

Hybrid and ensemble techniques have been developed to overcome the weaknesses of single classifiers. Ngartera, Issaka, and Nadarajah (2025) suggested hybrid Naive Bayes to detect fraud and showed that complementary classifiers were more robust. Concerning employee turnover, a hybrid model may exploit the semblance of a local model and the benefits of a global decision boundary of SVMs. An analysis of different medical classification classifiers by Damayunita, Fuadi, and Juliane (2022) showed that ensemble types of classifiers had better results than single-model classifiers. The stacked ensemble approach, whereby the output of lower-level models is used as input features by a meta-classifier, has proven to be especially promising. G A et al. (2024) reveal that the theoretical framework of stacking reduces bias and variance as it learns how to combine the base model outputs in the most optimistic manner.

The forecast is as significant as an understanding of the factors that contribute to attrition. Issues concerning compensation, e.g., the monthly pay, percentage salary rise and stock option rates are frequently cited as Driving forces. The turnover also varies with the work environment factors such as overtime requirements, and distance between work and home. Relational factors that capture the social and supervisory aspects of retention include years with current manager and job involvement in this case. Niu et al. (2021) emphasize that it is crucial to consider the effects of interactions between these variables that were not obvious when applying the conventional statistical tools to the big data analytics. The gap in the direct comparisons of hybrid KNN-SVM models that are specially designed to be utilized in the early detection of attrition is provided in this paper.

Methodology:

The process is divided into five steps, such as the gathering and preprocessing of data, the creation of the model to be used in KNN and SVM baselines, the creation of the hybrid stacked ensemble, and testing it using a wide range of performance measurement. The data of this research is a synthesized employee attrition data which is supposed to reflect the real information systems of human resources. The dataset consists of 10,000 records of employees and contains 20 features that consist of demographics, job characteristics, and compensation information and variables of the work environment (Chung et al. 2023).

The features of raw employee data contain different-scale values that can bias distance-based algorithms like KNN. The pre-processing pipeline was composed of several steps. To begin with, categorical variables were encoded using one-hot encoding. Second, the features were standardized using the Z-score and the features had a mean of zero and standard deviation of one.

Analysis:

The analysis section will give the performance of the baseline and hybrid models and thereafter feature importance analysis. Any results are based on the held-out test set consisting of 3,000 records of employees. Table 1 depicts the performance indicators of the KNN baseline, SVM baseline and the proposed hybrid stacked ensemble model in comparison.

Table 1: Comparative Performance Metrics of Predictive Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC (%)
KNN Baseline	86.2	82.4	71.3	76.4	0.89
SVM Baseline	89.5	85.1	78.6	81.7	0.92
Hybrid Stacked Ensemble	94.8	91.2	88.5	89.8	0.97

(Source: Created by Author)

The hybrid stacked ensemble is superior to both the baseline models in all evaluation metrics. The error is reduced to 94.8 86.2 and 89.5, or a relative error of approximately 50 percent. The highest growth is in the recall where the hybrid model has 88.5% as opposed to 71.3% of KNN and SVM respectively (Alqahtani et al. 2024). This is an improvement in that it is much better at detecting real cases of attrition which is precisely the goal of the early detection systems in which it is more costly to miss a potential leaver than it is to cause a false alarm. The F1-score is 89.8, indicating that the recall increase is not at an intolerable cost to the accuracy. AUC-ROC of 0.97 is a very good discriminating ability.

Table 2: Confusion Matrix Comparison across Models

Model	True Negative (Stay Correct)	False Positive (False Alarm)	False Negative (Missed Attrition)	True Positive (Attrition Correct)
KNN	2,013	291	123	573
SVM	2,058	246	89	607
Hybrid	2,098	206	42	654

(Source: Created by Author)

To further understand the error pattern, Table 2 provides the confusion matrices of each model that shows the true negatives, false positives, false negatives, and true positives. The confusion matrix analysis shows that there are some areas to be improved. The hybrid model has a correct identification of 654 cases of attrition, 573 in KNN and 607 in SVM. To be more precise, the false negatives are reduced to considerably lower values than in KNN (123) and SVM (89), with 42 in the hybrid model. This reduction is an important operational improvement as all false negatives are employees whose loss has not been expected. The hybrid model also reduces the false alarms of 291 and 246 to 206 meaning that the stacked ensemble predicts more true attritions and fewer false predictions. Table 3 employs the permutation importance to rank the top ten predictors of employee attrition.

Table 3: Top Ten Drivers of Employee Attrition

Rank	Feature	Importance Score	Direction of Effect
1	Monthly Income	0.235	Negative (lower income increases risk)
2	Overtime (OverTime_Yes)	0.187	Positive (overtime increases risk)
3	Years with Current Manager	0.142	Negative (shorter tenure increases risk)
4	Stock Option Level	0.118	Negative (lower levels increase risk)
5	Job Role (Sales Representative)	0.095	Positive (specific role increases risk)
6	Distance from Home (km)	0.089	Positive (longer distance increases risk)
7	Number of Companies Worked	0.076	Positive (more previous jobs increases risk)
8	Work-Life Balance	0.068	Negative (poor balance increases risk)
9	Years at Company	0.054	Negative (shorter tenure increases risk)
10	Job Involvement	0.041	Negative (lower involvement increases risk)

(Source: Created by Author)

Monthly income is discovered to be the best predictor, and low income helps to increase high chances of departure. The second prominent one is overtime since the employees who are forced to work overtime are again more likely to quit. Years of current manager is an important relational predictor, suggesting that supervisory stability is important to retention. Stock option level increases the importance of other compensation as opposed to base salary. The top ten is filled with distance to home, past job employers, work-life balance and job involvement meaning that the phenomenon of attrition is multifaceted since it is accompanied by economic, logistical, relational and psychological factors (Talebi et al. 2025).

Discussion:

In the results, a hybrid stacked ensemble model that consists of KNN and SVM is very efficient in identifying employee attrition at an early stage. The high performance can be attributed to their complementary inductive biases of KNN and SVM. KNN is suitable at capturing local neighbourhood structures but is sensitive to noise. SVM can form a global decision boundary that can handle complex relationships at the expense of local patterns. The stacked ensemble is trained to use the probability outputs of the two models to allow the meta-classifier to use KNN when local similarity is predictive and SVM when the global margin is more predictive (Nagpal and Pawar, 2024). This response that the number of false negatives in the hybrid model is only 42, compared to the original SVM and KNN models of 89 and 123 respectively, is a pointer that the hybrid model is working in capturing the cases that both the base models would have missed.

These findings are correlated with the broader research of ensemble methods. Debbadi and Boateng (2025) discovered that the accuracy of decisions in matters of business operations increased when different machine learning models are combined. Osman et al. (2025) claim that hybrid approaches provided an opportunity to be

more proactive in decisions. The current study is based on this literature in that a KNN-SVM stacked ensemble is applied to employee attrition. The hybrid model gives an avenue to proactive retention management to human resource practitioners. This recall and correctness of 88.5 and 91.2 indicates that at-risk workers can be detected to provide specific interventions. Based on the feature importance ranks, the introduction of regular compensation audits, re-evaluation of the overtime policies, contribution of stability of managers, and the provision of employees in risky categories with the equity compensation can be introduced.

It has to be stated that it has several limitations. It may not be possible to capture all idiosyncratic relationships in real organizational data using synthetic data. Future research should verify the hybrid model with different real-world datasets of different industries. The existing model is a single point in time, whereas the risk of attrition is not a constant one; a time-dependent hybrid model is a viable future direction of work. The stacked ensemble, which also improves accuracy, also increases the complexity of the model. The other area of future work is to investigate explainable AI techniques such as SHAP values to retain the interpretation. Hassija et al. (2024) consider the methods of explainable AI that can be integrated into the hybrid model in detail. Finally, ethical issues should be addressed; such tools should be implemented in an open manner by organizations with explicit policies to prevent negative behaviour that would not rely solely on model prediction.

Conclusion:

The paper has developed and experimented on a hybrid predictive model of early employee attrition by stacking KNN and SVM. The hybrid model was 94.8% accurate, 91.2% precise, 88.5% recall and AUC-ROC 0.97, which is a much higher result than the baseline classifiers. Theoretical, methodological and practical contributions are the most crucial ones. Theoretically, the study contributes to the body of research on the possibility of integrating complementary machine learning algorithms to be able to explain complex organizational phenomena. In terms of methodology, it provides a pipeline to pre-process HR data that can be replicated, balance in the classes of processes, and has a stacked ensemble. Practically, it offers organizations with a time-tested approach of identifying at-risk employees before they leave, enabling it to put into action special retention measures. Future research has to implement this model to actual-world data, involve time dynamics, and integrate explainable AI techniques.

Reference List:

1. Alqahtani, H., Almagrabi, H. and Alharbi, A., 2024. Employee attrition prediction using machine learning models: A review paper. *Int. J. Artif. Intell. Appl*, *15*(2), pp.23-49. https://www.researchgate.net/profile/Haya-Alqahtani-7/publication/379860312_Employee_Attrition_Prediction_using_Machine_Learning_Models_A_Review_Paper/links/6824b429026fee1034f821c1/Employee-Attrition-Prediction-using-Machine-Learning-Models-A-Review-Paper.pdf
2. Chung, D., Yun, J., Lee, J. and Jeon, Y., 2023. Predictive model of employee attrition based on stacking ensemble learning. *Expert Systems with Applications*, *215*, p.119364. https://www.sciencedirect.com/science/article/pii/S095741742202382X
3. Damayunita, A., Fuadi, R.S. and Juliane, C., 2022. Comparative analysis of Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) algorithms for classification of heart disease patients. *Jurnal Online Informatika*, *7*(2), pp.219-225.https://join.if.uinsgd.ac.id/index.php/join/article/download/919/240
4. Debbadi, R.K. and Boateng, O., 2025. Optimizing end-to-end business processes by integrating machine learning models with UiPath for predictive analytics and decision automation. *Int J Sci Res Arch*, *14*(2), pp.778-796.https://www.researchgate.net/profile/Rama-Krishna-Debbadi/publication/389043018_Optimizing_end-to-end_business_processes_by_integrating_machine_learning_models_with_Uipath_for_predictive_analytics_and_decision_automation/links/67b2250196e7fb48b9c4f871/Optimizing-end-to-end-business-processes-by-integrating-machine-learning-models-with-Uipath-for-predictive-analytics-and-decision-automation.pdf
5. G A, S., Prabha, R., Sridevi, S. and Rohini, A., 2024. Artificial Intelligence: Advanced Mathematical Constructs and Theoretical Framework in Machine Learning and Deep Learning Algorithms.https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=5066774
6. Halder, R.K., Uddin, M.N., Uddin, M.A., Aryal, S. and Khraisat, A., 2024. Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *Journal of Big Data*, *11*(1), p.113.https://link.springer.com/article/10.1186/s40537-024-00973-y

7. Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M. and Hussain, A., 2024. Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation*, *16*(1), pp.45-74.https://link.springer.com/article/10.1007/s12559-023-10179-8
8. Infeedo, 2025. Attrition Forecasting: Detecting Early Signs of Employee Disengagement. https://www.infeedo.ai/blog/attrition-forecasting-employee-disengagement-signs
9. Kumari Jaya Kumar, V., 2024. *Employee attrition forecasting: Determining the optimal algorithm for predicting employee turnover* (Doctoral dissertation, Dublin Business School). https://esource.dbs.ie/items/b779d106-d992-43ef-9918-6dfdc1c635ce
10. Nagpal, P. and Pawar, A., 2024, February. Predicting employee attrition through HR analytics: A machine learning approach. In *2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)* (pp. 1-4). IEEE. https://ieeexplore.ieee.org/abstract/document/10563285/
11. Ngartera, L., Issaka, M.A. and Nadarajah, S., 2025. Hybrid Naïve Bayes Models for Scam Detection: Comparative Insights from Email and Financial Fraud. *IEEE Access*.https://ieeexplore.ieee.org/iel8/6287639/6514899/11000315.pdf
12. Niu, Y., Ying, L., Yang, J., Bao, M. and Sivaparthipan, C.B., 2021. Organizational business intelligence and decision making using big data analytics. Information Processing & Management, 58(6), p.102725. https://www.sciencedirect.com/science/article/abs/pii/S0306457321002090
13. Nurhidayat, I.R.F.A.N. and Pimpunchat, B.U.S.A.Y.A.M.A.S., 2023. A comparative approach to SVM kernel functions via accurate evaluating algorithms. *J. Eng. Sci. Technol*, *18*(4), pp.2078-2090.https://www.researchgate.net/profile/Irfan-Nurhidayat/publication/373236946_A_comparative_approach_to_SVM_kernel_functions_via_accurate_evaluating_algorithms/links/64e1cc241351f5785b794b50/A-comparative-approach-to-SVM-kernel-functions-via-accurate-evaluating-algorithms.pdf
14. Osman, A., Fowowe, O.O., Agboluaje, R. and Orekha, P.O., 2025. Integrating machine learning in business analytics consulting for proactive decision-making and innovation. *World Journal of Advanced Research and Reviews*, *25*(1), pp.1817-1836.https://www.researchgate.net/profile/Precious-Orekha/publication/388421013_Integrating_machine_learning_in_business_analytics_consulting_for_proactive_decision-_making_and_innovation/links/67983d1a52b58d39f25664aa/Integrating-machine-learning-in-business-analytics-consulting-for-proactive-decision-making-and-innovation.pdf
15. Raza, A., Munir, K., Almutairi, M., Younas, F. and Fareed, M.M.S., 2022. Predicting employee attrition using machine learning approaches. *Applied Sciences*, *12*(13), p.6424. https://www.mdpi.com/2076-3417/12/13/6424
16. Talebi, H., Bardsiri, A.K. and Bardsiri, V.K., 2025. Developing a hybrid machine learning model for employee turnover prediction: Integrating LightGBM and genetic algorithms. *Journal of Open Innovation: Technology, Market, and Complexity*, *11*(2), p.100557. https://www.sciencedirect.com/science/article/pii/S2199853125000927
17. Valkenborg, D., Rousseau, A.J., Geubbelmans, M. and Burzykowski, T., 2023. Support vector machines. *American journal of orthodontics and dentofacial orthopedics*, *164*(5), pp.754-757.https://www.ajodo.org/article/S0889-5406(23)00429-8/pdf