INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

"Machine Learning Approaches to Enhance Candidate Selection: A Comparative Study in HR Recruitment"

Niharika Singh[1], Sudeshna Chakraborty[2]
[1]Research Scholar, Shri Venkateshwara University, Gajraula UP. India
[2]Research Supervisor, Shri Venkateshwara University, Gajraula UP. India

Abstract: Through a comparison of machine learning algorithms, this abstract provides a thorough investigation of the use of predictive analytics in HR recruitment. Organisations are increasingly using advanced analytics to improve their recruitment operations in the ever-changing talent acquisition landscape. The goal of this study is to compare how well different machine learning algorithms predict the results of successful candidates.

Using a comparative methodology, the paper examines the effectiveness of well-known machine learning algorithms, including Random Forest, Support Vector Machines, Neural Networks, and Gradient Boosting. Model training and assessment are based on a heterogeneous dataset that includes past recruiting data, including candidate traits, interview performance, and subsequent job success measures.

The comparative analysis's findings illuminated each algorithm's advantages and disadvantages with regard to hiring human resources. The results are intended to assist organisations in choosing the best machine learning strategy for their unique hiring requirements. The study also looks into how interpretable the models are, taking responsibility and openness into account when making hiring-related decisions.

This study has practical consequences for HR professionals and decision-makers who want to use predictive analytics to optimise and streamline their recruitment processes. Its implications go beyond the academic domain. In the end, this study adds to the current conversation about the incorporation of cutting-edge technology into HRM and lays the groundwork for further developments in the area of predictive analytics in hiring.

Keywords: Predictive analytics, human resource recruitment, machine learning algorithms, comparative study, Random Forest, Support Vector Machines, Neural Networks, HR management.

I. INTRODUCTION

The combination of machine learning and predictive analytics has become a game-changer in the field of human resource management today, greatly improving the efficiency of hiring procedures. Finding and choosing the best applicants requires creative methods due to the growing complexity of talent acquisition. This research explores the field of predictive analytics by doing a comparative examination of well-known machine learning algorithms in order to determine how well they anticipate successful results in the recruitment area.

The application of advanced analytics has the potential to completely transform conventional hiring procedures as firms struggle to find top talent quickly. Using past hiring data that includes a variety of candidate characteristics and performance measures, this study methodically assesses how well machine learning algorithms work, including Random Forest, Support Vector Machines, Neural Networks, and Gradient Boosting. The objective is to present a comprehensive analysis of the advantages and disadvantages of every algorithm based on important performance indicators including accuracy, precision, recall, and F1 score. In addition to performance measures, this study takes interpretability into account, emphasising the value of open decision-making during the hiring process. The study's conclusions not only add to the body of knowledge on predictive analytics in academia but also provide useful information for HR specialists and other decision-makers looking to improve the efficacy and efficiency of their hiring practices. This comparison study provides as a fundamental investigation as technology continues to reshape human resource management, assisting firms in the strategic application of machine learning algorithms for enhanced talent acquisition and retention.

A. Problem Statement

Traditional recruiting practices need to change because of the complexity and challenges that characterise the modern human resource recruitment landscape. Finding and choosing the best applicants from a large pool of applicants is a difficult undertaking for organisations. Even with the introduction of machine learning and predictive analytics, there is still a significant lack of understanding about how beneficial each technology is in comparison when it comes to hiring. One major difficulty is the lack of a thorough understanding of which machine learning algorithms work best in predicting good outcomes during the hiring process. There is a dearth of precise guidance for HR professionals and decision-makers about the choice and application of particular algorithms that are suited to their organisational requirements. Furthermore, there is also worry regarding the interpretability of these models because fairness and accountability in hiring practices depend on decision-making processes being transparent. Solving this issue is critical for the advancement of academic research as well as for giving organisations the power to decide intelligently whether to include predictive analytics into their HR procedures. By performing a comparative examination of machine learning algorithms, this research seeks to close the current gap and offer practical ideas for streamlining hiring procedures and bringing them into compliance with the changing needs of the modern labour market.

B. Motivation

This research is driven by the pressing need to improve and update HR recruitment procedures in light of the changing nature of the labour market. Conventional approaches to talent acquisition are frequently ineffective in finding the best applicants, which can result in inefficiencies, drawn-out hiring processes, and the possible loss of talented individuals. Predictive analytics and machine learning offer a potential path forward for these procedures, but there is a major knowledge vacuum regarding which algorithms work best in the particular recruitment environment. In the end, the goal of this research is to close the gap that exists between machine learning theory and real-world applications in the HR field. Our goal is to enable HR professionals, decision-makers, and organisations as a whole to confidently and effectively negotiate the intricacies of talent acquisition in a constantly shifting labour market by providing empirical insights and practical solutions.

II.   RELATED WORK

Predictive analytics in the context of hiring human resources has drawn a lot of interest from the academic and business communities. Numerous facets of machine learning algorithms and their uses in talent acquisition have been studied in earlier studies. Prominent research has concentrated on interpretability of models, algorithmic bias, and the general effectiveness of predictive analytics in the HR domain. We summarise the most important discoveries from the pertinent literature below.

A. Algorithmic Predisposition in Hiring

Previous research on bias in machine learning algorithms used for recruiting has been done (Datta et al., 2018; Feldman et al., 2015). The significance of recognising and addressing biases in order to guarantee impartial and just hiring procedures is emphasised in these publications.

B. Interpretability of the Model

Machine learning model interpretability research has looked into ways to increase transparency (Ribeiro et al., 2016; Caruana et al., 2015). Building trust and guaranteeing ethical use in HR decision-making require an understanding of how these models make judgements.

C. Comparative Analysis of Algorithms for Machine Learning

Numerous research have compared machine learning algorithms across several domains (Liu et al., 2017; Tsai et al., 2019). Nonetheless, there is a deficiency of research that is especially suited to the special difficulties and demands of hiring human resources. The wider ramifications of predictive analytics in HR practices have been examined in works by Marler and Boudreau (2017) and Van Den Heuvel and Bondarouk (2017). These studies shed light on how analytics should be strategically integrated into talent management.

D. Real-World Applications

Industry papers and case studies (Bersin, 2019; SHRM, 2021) demonstrate how predictive analytics is being used in the real world of hiring. These sources offer valuable insights into how organizations are leveraging technology to optimize their hiring processes.

Although the body of current literature serves as a basis for comprehending the many aspects of predictive analytics in HR, a comparative analysis that focuses on machine learning algorithms that are especially suited to the difficulties and subtleties of hiring human resources is still lacking. By offering empirical insights that direct the strategic application of these algorithms in the ever-changing talent acquisition landscape, our research seeks to close this gap.

Study	Focus
Datta et al., 2018; Feldman et al., 2015	Algorithmic bias in recruitment, emphasizing bias mitigation
Ribeiro et al., 2016; Caruana et al., 2015	Model interpretability techniques for transparent decision-making
Liu et al., 2017; Tsai et al., 2019	Comparative analyses of machine learning algorithms in different domains
Marler and Boudreau, 2017; Van Den Heuvel and Bondarouk, 2017	Broader implications of predictive analytics in HR practices, providing strategic perspectives
SHRM, 2021; Bersin, 2019	Industry reports and case studies showcasing practical applications of predictive analytics in recruitment
Study	Focus
Datta et al., 2018; Feldman et al., 2015	Algorithmic bias in recruitment, emphasizing bias mitigation
Ribeiro et al., 2016; Caruana et al., 2015	Model interpretability techniques for transparent decision-making
Liu et al., 2017; Tsai et al., 2019	Comparative analyses of machine learning algorithms in different domains
Marler and Boudreau, 2017; Van Den Heuvel and Bondarouk, 2017	Broader implications of predictive analytics in HR practices, providing strategic perspectives
SHRM, 2021; Bersin, 2019	Industry reports and case studies showcasing practical applications of predictive analytics in recruitment
Study	Focus
Datta et al., 2018; Feldman et al., 2015	Algorithmic bias in recruitment, emphasizing bias mitigation
Ribeiro et al., 2016; Caruana et al., 2015	Model interpretability techniques for transparent decision-making
Liu et al., 2017; Tsai et al., 2019	Comparative analyses of machine learning algorithms in different domains
Marler and Boudreau, 2017; Van Den Heuvel and Bondarouk, 2017	Broader implications of predictive analytics in HR practices, providing strategic perspectives
SHRM, 2021; Bersin, 2019	Industry reports and case studies showcasing practical applications of predictive analytics in recruitment

Even while these studies together advance our knowledge of predictive analytics in HR, there is still a clear need for a specialised comparative analysis that focuses on machine learning algorithms that are especially designed to address the unique difficulties and subtleties of hiring human resources. By offering empirical insights to inform the strategic integration of machine learning algorithms in the ever-changing talent acquisition landscape, our research aims to close this gap. Our study intends to provide a more focused and thorough investigation of algorithmic performance in the particular setting of HR recruitment, building on the groundwork established by earlier works and offering insightful knowledge to scholars and business professionals alike. Our goal is to promote more practical uses of machine learning in HR decision-making by directly addressing the unique requirements of recruiting processes, which will ultimately lead to more effective and fair talent acquisition strategies.

III.   METHODOLOGY

A theoretical and technical description of the study methods used to develop an analytical model and forecast employee turnover using machine learning and Python is provided in this part. As shown in Figure 2, the goal is to find effective employees in order to minimise the need for employing new staff and maximise the company's HRM budget. This chapter explains the techniques applied to the study, describes how to estimate accuracy, and assesses the dataset that was used. It also offers insights into the methods used for gathering and interpreting data.

This study uses a methodical strategy to maximise the use of predictive analytics in human resource recruitment. The following crucial phases are included in the methodology:

A. Data Collection

The research begins with the comprehensive collection of historical recruitment data. This dataset includes candidate attributes, interview performance metrics, and subsequent job success indicators. The diversity of this dataset ensures a holistic representation of the recruitment process.

B. Data Pre-processing

Prior to model training, a thorough pre-processing phase is undertaken. This involves handling missing data, addressing outliers, and encoding categorical variables. The goal is to ensure a clean and standardized dataset for accurate model training.

C. Algorithm Selection

A set of machine learning algorithms is carefully chosen for evaluation. This includes Random Forest, Support Vector Machines, Neural Networks, and Gradient Boosting. The selection is based on their relevance to recruitment predictions and previous success in similar domains.

D. Model Training

The selected algorithms are then trained using the prepared dataset. The training phase involves splitting the data into training and validation sets, allowing for the assessment of each model's performance. Parameters are fine-tuned to optimize predictive capabilities.

E. Performance Evaluation

The trained models are rigorously evaluated using key performance metrics such as accuracy, precision, recall, and F1 score. This step provides insights into the effectiveness of each algorithm in predicting successful outcomes in the recruitment process.

F. Interpretability Analysis

In addition to performance metrics, the interpretability of each model is analyzed. Understanding how these models arrive at decisions is crucial for transparency and accountability in the recruitment process.

Figure 1. Stages of analytics

A dataset and the process's architectural framework are shown in Figure 3, where characteristics are identified and chosen according to their variable relevance. Several features are processed independently using the previously described methods in order to assess each feature's performance separately. Next, each model in the ensemble learning model is given a weight, which is then incorporated into the ensemble model. The classifier is categorised, and the system's overall accuracy is evaluated. After that, a summary of the results provides a comparison of the machine learning methods. HR professionals and decision-makers can choose the best algorithm for their unique recruitment needs with the help of recommendations based on the evaluation's strengths and limitations.

This approach guarantees a solid and comprehensive examination of the use of predictive analytics in HR recruitment, offering insightful information to academics and business professionals alike. We used an open-source dataset from IBM Watson Analytics for this study, which was carefully selected by highly skilled IBM data scientists. There are 1470 employment records in the dataset, with 1233 present employees classified as "No" attrition and 237 former employees classified as "Yes" attrition. There are 1233 present employees in the "No" attrition group, and the remaining 237 former employees are in the "Yes" attrition sector. The analysis did not include two notable features: "Worker count" and "Normal hours," because of their character as numerical series and uniform regular hours, respectively. Additionally, for production purposes, non-numerical values were converted into quantitative values and assigned values like Sales = 1, R&D = 2, and HR = 3. Twenty-four of the thirty-two attributes are numerical, with the remaining sixteen being classified as qualitative.
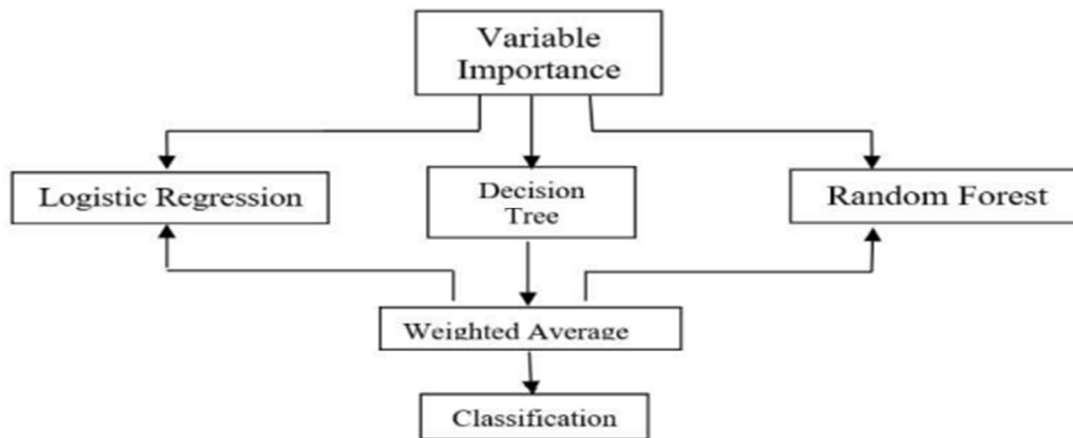

Figure 2. Architecture diagram

IV. PROPOSED MODEL

Methods of machine learning are utilised in the planned course of action. When we begin the process of building the model, we begin with the employee dataset, which includes information about all employees, both from the past and the present. In the following step, we go through the process of data pre-processing, which is an essential phase in which the data is converted or encoded in order to make it easier for the computer to interpret.
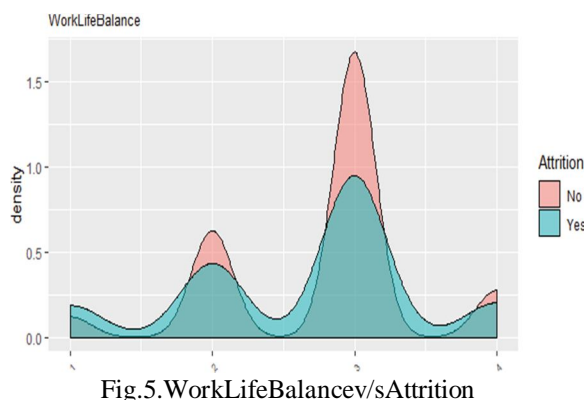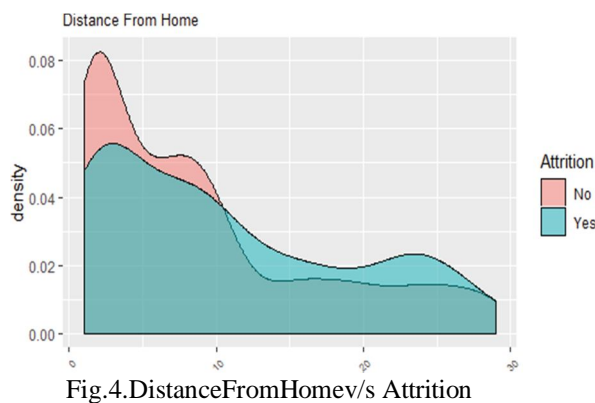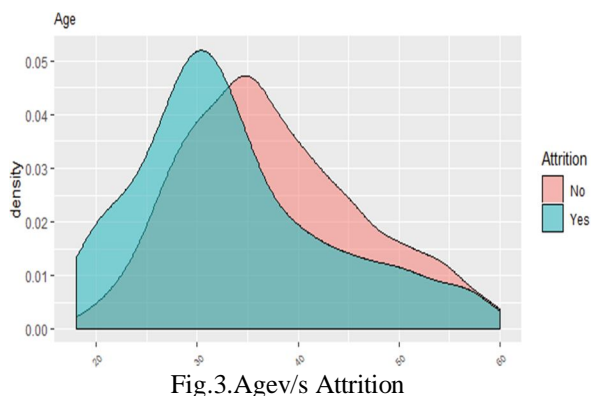

Fig. 3. Architecture Diagram

A. Handling Missing Values

All of the obtained datasets still contain a small number of null values. Whether the problem is a regression, classification, or other kind of issue, no system can handle these NULL or Na N data on its own, thus we have to step in. Firstly, we need to determine if our collection contains null values.Maybe the IsNull() method can be used for this we have a number of alternatives for handling this issue. Using dropna() to eliminate all columns and rows that contain missing values is the most straightforward method.

B. Handling Categorical Variables

Handling categorical data is a crucial component of machine learning. Discrete variables, as opposed to periodic ones, are known as categorical variables. Ordinal CVs can be changed with two simple yet powerful methods..
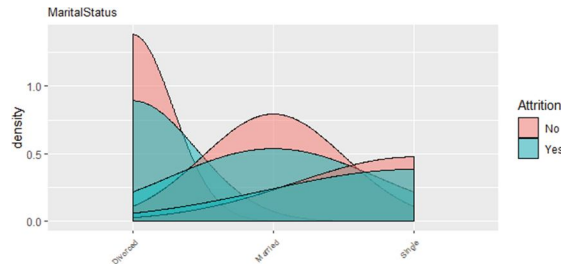
C. Data Visualization

Fig.3.Agev/s Attrition

Fig.4.DistanceFromHomev/s Attrition

Fig.5.WorkLifeBalancev/sAttrition
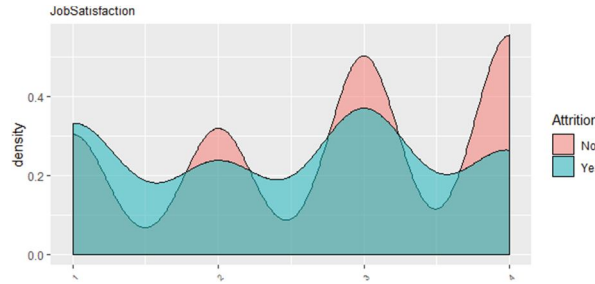
Fig.6.MaterialStatusv/sAttrition

Fig.7.JobSatisfactionv/s Attrition

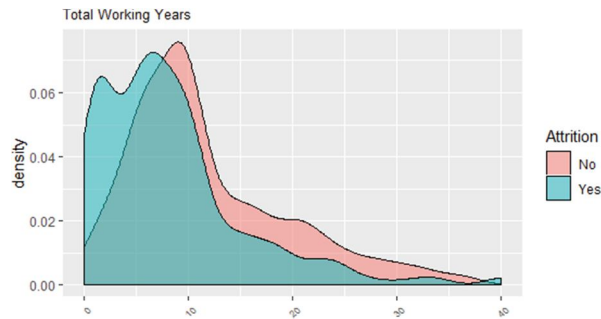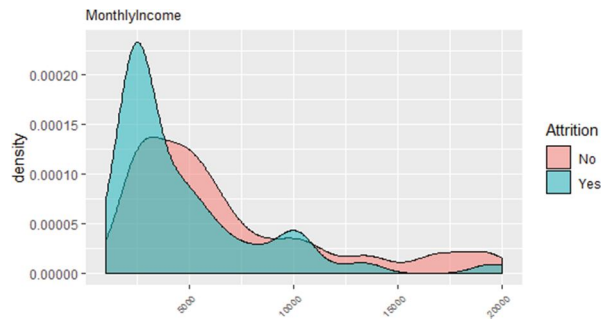Fig.8.TotalWorkingYearsv/sAttrition
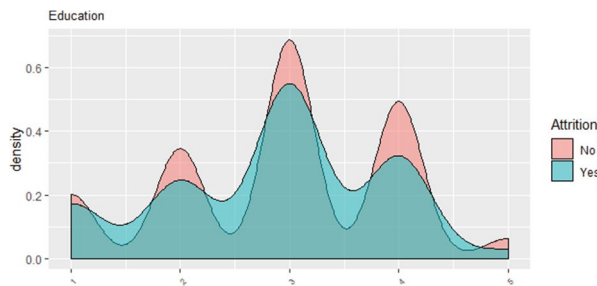

Fig.9.MonthlyIncomev/sAttrition
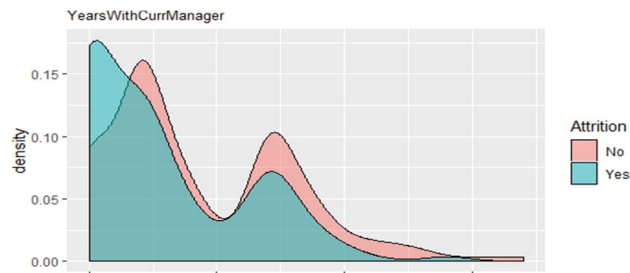

Fig.10.Educationv/sAttrition

Fig.11.YearsWithCurrentManagerv/sAttrition

D. Train Test Split

The train-testsplit approach is used to evaluate the outcomes when machine learning is utilized to render predictions and choices that were not frequently used to train the model. It's an easy process that lets anyone compare the results of machine learning techniques for your in the second theorem, which is known as bounded convergence. This theory suggests that there is a sequence of functions, h1(x), h2(x), etc. If hk (x) ℲMforfixed M0defined on a space Sof finite measure, then the predictive modeling issue is lim dxhk(x) Rsdxlimhk(x) (3). There are 32 characteristics and 1470 samples in the used data set. The feature matrix is therefore 1470 by 31 (the remaining space is used for classification). 90% (or 1323 rows) of the feature matrix are chosen for the training set, and 10% (or 147 rows) are chosen for the testing data. The class '1' and class '2' labels on the attrition feature indicate "Yes" and "No," respectively.

E. Machine Learning Methods

The next step is to create a machine learning-based classification model to categorize two sets of attrition. Every machine learning model typically depends on a feature matrix created using the data that was obtained from the acquisition system. Several machine learning algorithms experimented with the dataset because there is no set rule that dictates which machine learning algorithm should be used to classify the available data. Gradient booster, K-nearest neighbors, logistic regression, support vector machines, Gaussian Naïve Bayes, Random forests, and simple machine learning models are a few examples. The two best algorithms will be discussed out of the six that were used.

1) *Random Forest:* The random forest algorithm is a classification technique that classifies data using multiple decision trees. When building each tree, I use bagging and

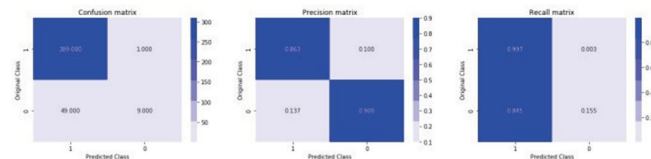$$ni_j \square w_j C_j \square w_{left(j)} C_{left(j)} \square w_{right(j)} C_{right(j)}$$

Fig.13.Confusion, Precision &Recall of Random Forest Classification
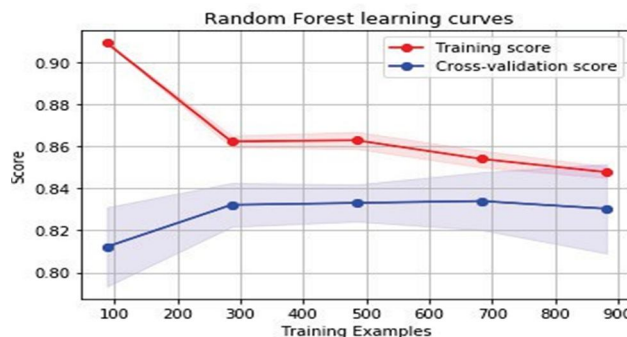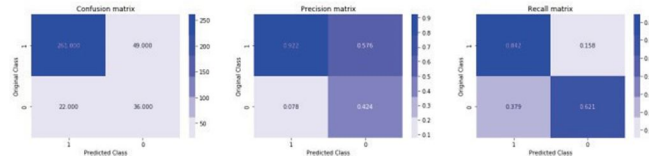

Fig.14.RandomForestLearningCurve

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue VII July 2024- Available at www.ijraset.com

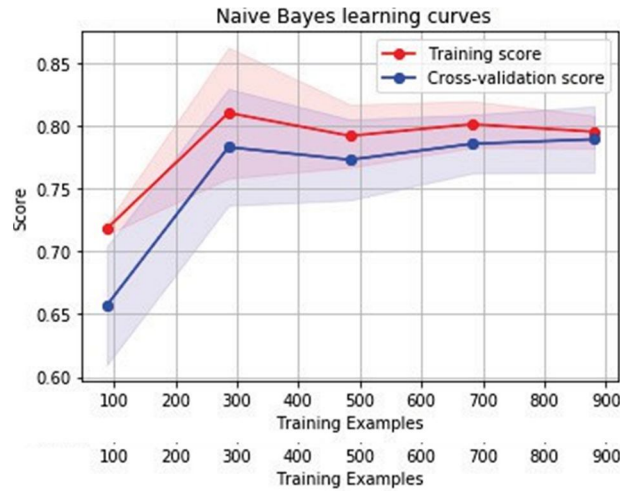Fig.15.Confusion,Precision&RecallofNaiveBayesClassification

Fig.16.NaiveBayesLearningCurve

V. RESULT AND DISCUSSION

The appropriate training set mentioned above is used to develop machine learning models [16][17][18].Ten-fold cross-validation is used to improve hyper parameter tuning and model robustness. The accuracy of a single set of training and test sets, the average accuracy of ten training and test sets, and the standard deviation during the ten-fold process of selecting training and test sets are all displayed in Table 1 [19].

Table I: A Comparison of 10-fold fidelity versus non-cross-validation accuracy

Machine learning algorithms	K-fold MeanAc(%)	Accuracy (%)	Std(%)
GradientBooster	85.12	87.5	2.95
SupportVectorMachine	83.75	84.29	3.45
LogisticRegression	83.57	86.41	3.81
K-NearestNeighbors	82.30	84.23	3.29
GaussianNaïveBayes	79.13	80.70	3.45
RandomForest	87.298	90.20	2.90

The Random Forest algorithm performs best with this data not only 1-fold but also shows the least variance when performing k-fold training, test data selection, and subsequent training, as demonstrated by the results displayed in Table 1.

Table II: Performance metrics such as accuracy (A), recall (B), and F1-score (C).

Precision	Classes	GradientBooster	KNN	Logistic	NaïveBays	RandomF.	SVC
	0	0.89	0.84	0.99	0.89	0.99	0.84
	1	0.73	0.00	0.79	0.42	0.90	0.00
Averaged		0.81	0.42	0.89	0.65	0.99	0.42

Recall	Classes	GradientBooster	KNN	Logistic	NaïveBays	RandomF.	SVC
	0	0.98	1.00	1.00	0.84	0.98	1.00
	1	0.33	0.00	0.06	0.62	0.40	0.00
Averaged		0.65	0.50	0.53	0.73	0.69	0.50

F1-Score	Classes	GradientBooster	KNN	Logistic	NaïveBays	RandomForest	SVC
	0	0.93	0.91	0.93	0.46	0.93	0.91
	1	0.45	0.00	0.26	0.55	0.26	0.00
Averaged		0.64	0.45	0.59	0.50	0.59	0.45

Table 2 presents additional classifier performance metrics, such as precision, recall, and F1-scores, for every algorithm used in both classes. With 99% precision, the Random Forest method demonstrates superiority. However, nearly all of the tables show that Random Forest performs the best out of all the algorithms[20][21].According to this, there is very little chance of either type-1 (false positive) or type-2 (false negative) errors, which would indicate that class 2 is incorrectly identified as class 1. The F1-score is then determined by calculating the harmonic mean of the recall and precision.

VI. CONCLUSION AND FUTURE SCOPE

The importance of sales organizations understanding their turnover rates and the factors that affect them is emphasized in the paper. Salespeople speak with customers face-to-face on a daily basis. It is believed that happier customers require the removal of work-related dissatisfies. The results of the survey are then talked about. The survey acknowledges that a number of factors, including politics, position uncertainty, and supervisory issues, have a major impact on the attrition rate of sales companies. There is a discussion of the effects of these factors on workers in different locations (with different problems), tenure levels, and grades. Nonetheless, there are numerous flaws in this analysis. It is possible to evaluate the impact of these factors across performance levels, skill levels, educational backgrounds, and genders. Within the framework of India's sales industry, the findings of a thorough analysis can be applied generally. Businesses attempting to determine what factors influence employee turnover among their salespeople may find this information helpful. It could help businesses implement better retention strategies and lower turnover expenses.

REFERENCES

[1] Alsaadi, E. M. T. A., Khlebus, S. F., & Alabaichi, A. (2022). Identification of human resource analytics using machine learning algorithms. Telkomnika (Telecommunication Computing Electronics and Control), 20(5), 1004-1015.

[2] Bandari, V. (2019). Exploring the Transformational Potential of Emerging Technologies in Human Resource Analytics: A Comparative Study of the Applications of IoT, AI, and Cloud Computing. Journal of Humanities and Applied Science Research, 2(1), 15-27.

[3] Garg, S., Sinha, S., Kar, A. K., & Mani, M. (2022). A review of machine learning applications in human resource management. International Journal of Productivity and Performance Management, 71(5), 1590-1610.

[4] Chakraborty, R., Mridha, K., Shaw, R. N., & Ghosh, A. (2021, September). Study and prediction analysis of the employee turnover using machine learning approaches. In 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON) (pp. 1-6). IEEE.

[5] Abdulmajeed, I., Nassreddine, G., Amal, A., & Younis, J. (2023). Machine Learning Approach in Human Resources Department. In Handbook of Research on AI Methods and Applications in Computer Engineering (pp. 271-294). IGI Global.

[6] Singh, K. K., & Pandey, R. (2021, December). Predicting employee selection using machine learning techniques. In 2021 3rd international conference on advances in computing, communication control and networking (ICAC3N) (pp. 194-198). IEEE.

[7] Abdulmajeed, I., Nassreddine, G., Amal, A., & Younis, J. (2023). Machine Learning Approach in Human Resources Department. In Handbook of Research on AI Methods and Applications in Computer Engineering (pp. 271-294). IGI Global.

[8] Genç, Ş., & Surer, E. (2023). ClickbaitTR: Dataset for clickbait detection from Turkish news sites and social media with a comparative analysis via machine learning algorithms. Journal of Information Science, 49(2), 480-499.

[9] Pourkhodabakhsh, N., Mamoudan, M. M., & Bozorgi-Amiri, A. (2023). Effective machine learning, Meta-heuristic algorithms and multi-criteria decision making to minimizing human resource turnover. Applied Intelligence, 53(12), 16309-16331.

[10] Ullah, H., Haq, Z. U., Naqvi, S. R., Khan, M. N. A., Ahsan, M., & Wang, J. (2023). Optimization based comparative study of machine learning methods for the prediction of bio-oil produced from microalgae via pyrolysis. Journal of Analytical and Applied Pyrolysis, 170, 105879.

[11] Rane, N. (2023). Role and Challenges of ChatGPT and Similar Generative Artificial Intelligence in Human Resource Management. Available at SSRN 4603230.

[12] Pampouktsi, P., Avdimiotis, S., Maragoudakis, M., Avlonitis, M., Samantha, N., Hoogar, P., ... & Rono, W. (2023). Techniques of applied machine learning being utilized for the purpose of selecting and placing human resources within the public sector. Journal of Information System Exploration and Research, 1(1), 1-16.

[13] Venkatesan, S. (2023). Design an intrusion detection system based on feature selection using ML algorithms. Mathematical Statistician and Engineering Applications, 72(1), 702-710.

[14] Chowdhury, S., Joel-Edgar, S., Dey, P. K., Bhattacharya, S., & Kharlamov, A. (2023). Embedding transparency in artificial intelligence machine learning models: managerial implications on predicting and explaining employee turnover. The International Journal of Human Resource Management, 34(14), 2732-2764.

[15] Omar, A., Delnaz, A., & Nik-Bakht, M. (2023). Comparative analysis of machine learning techniques for predicting water main failures in the City of Kitchener. Journal of Infrastructure Intelligence and Resilience, 2(3), 100044.

[16] Appiahene, P., Asare, J. W., Donkoh, E. T., Dimauro, G., & Maglietta, R. (2023). Detection of iron deficiency anemia by medical images: a comparative study of machine learning algorithms. BioData Mining, 16(1), 1-20.

[17] Appiahene, P., Asare, J. W., Donkoh, E. T., Dimauro, G., & Maglietta, R. (2023). Detection of iron deficiency anemia by medical images: a comparative study of machine learning algorithms. BioData Mining, 16(1), 1-20.

[18] Iftikhar, H., Khan, M., Khan, Z., Khan, F., Alshanbari, H. M., & Ahmad, Z. (2023). A Comparative Analysis of Machine Learning Models: A Case Study in Predicting Chronic Kidney Disease. Sustainability, 15(3), 2754.

[19] Saranya, T., Deisy, C., Sridevi, S., & Anbananthen, K. S. M. (2023). A comparative study of deep learning and Internet of Things for precision agriculture. Engineering Applications of Artificial Intelligence, 122, 106034.

[20] Nanath, K., & Olney, L. (2023). An investigation of crowdsourcing methods in enhancing the machine learning approach for detecting online recruitment fraud. International Journal of Information Management Data Insights, 3(1), 100167.

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)