Prediction Employee Performance Using Weighted Average Ensemble Method

*B. Kalaiselvi, S. Saranya, T. Suresh Kumar, K. Shanmugapriya

Mahendra Engineering College, Namakkal, Tamil Nadu, India.
*Corresponding Author Email: kalaiselvib@mahendra.info

Abstract: *Employees are the most crucial resources in a firm, but they can leave their jobs unexpectedly, resulting in significant costs. Generally, attitude and effort of employee are driven by personality qualities, although job satisfaction that may be the consequence of particular observations from an organization with respect to environmental factors. Meanwhile, acquiring a new staff may be time-consuming and costly. Similarly, a newly employees hired may required for considering some efforts to make the particular organization productive. Job satisfaction is an important component in leaving the organization. Employee performance prediction and reasons to leave the organization have undertaken by Human Resource Management (HRM). This type of prediction must advance from HRM to examine the most talented and most experienced employees' reasons for leaving the organization through different data mining techniques, but a precise prediction is not accomplished. This can be determined by observing some of the most experienced and talented employees departing their firm. As a result, this paper aimed to construct an ensemble model that can accurately predict employee performance and job satisfaction using an HR analytics dataset. The present research paper focused on examining the employee performance and job satisfaction mentioned by employees in the Employee Attrition dataset by prediction by Weighed Average (WA) as an ensemble approach with Logistic Regression (LR). Furthermore, the suggested ensemble approach performance evaluation has been contrasted to the traditional method for analyzing better employee performance and job satisfaction predictions.*

Keywords: *weighted average mechanism, employee performance, ensemble method, Job satisfaction, organization.*

1. INTRODUCTION

Nowadays constantly evolving workplace, wherein global talent intersects and organizations compete towards excellence, HRM serves as a critical conductor, coordinating worker dynamics to accomplish success. When individuals recognize the recent era of human social and economic growth generated by advanced technical innovation, it have to recognize the "creative destruction" caused by emerging technologies with iterative renewal [1]. The questionnaires about development of technologies have overcome the manual work and effect of relationship between human and employee because of future Artificial Intelligence (AI) technology improves the productivity may resulted in large scale with long-term unemployment are being extensively discussed. China began a deep collaboration with recent phase and growth of the "new technology cluster" which involves internet technologies and the actual world economy. At present, biological, Physical and cyber space are completely unified as well as novel businesses, models, as well as forms of commerce have impended. As digital technology continues to grow rapidly, its employment qualities, like excellent retention capacity, diverse job demands and flexible form become prominently, and several recent vocations arise. The sharing economy, platform economy, gig economy, and full-time economy have adopting in lead with and the economy's, innovative transformation and development are resulted for essential shifts in employment carriers, occupational skills need and employment forms [2-4]. In addition to the widespread use of IT-based business systems as well as social media platforms, the automobile industry can generate massive amounts of databases and unstructured data on a wide range of business activities. The ability of a car manufacturer's human resources department for monitoring and evaluating employee performance is crucial to the business's overall achievement. The capacity of organization is utilized for an individual resource in gaining an aggressive advantage has critical to its success. Organizations can maximize their HRM by integrating them and their operational skills. The fourth industrial revolution has resulted in a paradigm shift in company operations. It requires human resources exist incorporated into operations as well as up-skilled [5]. HR Analytics employs data mining techniques on employee as well as productivity-related data to increase its capacity to predict the efficiently employees performance. In today's environment of dynamic technological, the inclusion of AI into various areas is emerged as a innovative significant driver. AI is making a big impact in Human Resources (HR) [6]. A set of technologies known as AI enables devices and computers to acquire information obtained from sensors,

mobile devices, and storage, interpret and analyze the data employing Natural Language Processing (NLP), generate intelligent choices or recommend events determined from ML, and respond to situational demands. Although HRM continues to be an essential profession, it has evolved dramatically in both structure and operation during the last 30 years. Business executives are quickly embraced with the digital world, and AI empowered using ML has ability in transforming the HR department in multiple fronts. The AI impact employment is ambiguous, and also brunt in human activity cannot be categorized into a simple "utopian" or "dystopian" scenario, but instead results in a hybrid of "utopia" and "dystopia" [7]. At the enterprise level, robotics' job-creation impacts coexist with the formation of new employment as a result of technical change [8]. Adopting an appropriate AI operation module can compensate the market, firms, and people' inadequate allocation of resources for labor-intensive jobs, reverse the nondirectional robots deployment over labor sector as well as stimulate their reorganization in the industrial as well as service sectors. The influence extents in employment across the entire society are unknown [9-11]. The management of employee performance has developed significantly in current years, migrating farther conventional annual assessments as well as into a high dynamic, data driven approach. Instead of focusing solely on recent achievement, the goal is for correlating person's goal with company objectives, encourage continuous improvement, and optimize employee potential. The ability of AI to scan and analyze massive volumes of data, identify trends, and deliver practical recommendations is changing the way organizations manage and develop talent.

2. LITERATURE REVIEW

AI can help to automate repetitive administrative tasks that allow HR employees in focusing their other job features. HR personnel can develop into high productive as a result, allow it for focusing on performing tasks that require their specific knowledge. Ismail et al. have reviewed the research, which acknowledges the vital need for early detection of COVID-19, especially for medical contexts, as well as underlines the importance of Deep Transfer Learning (DTL) in automation of this process. Additionally, it investigates the ensemble DTL models used in chest X-ray comprehension, with a particular emphasis on detection of COVID-19, and highlights the real-time importance about Internet of Things (IoT) method. Comparative evaluations with various algorithms have provided, emphasized the study's sensible utility in supporting radiologists in generating rapid and accurate diagnosis of COVID-19 as well as the combining IoT with deep TL importance for the earlier detection cases [12]. Dutta et al. have underlined the lack of knowledge in hospitality and tourism, as well as the increasing significance of technological advances, especially in the metaverse to gain competing benefits. The research gap has been indicated in a metaverse application for workplace learning. Objective of the project's are discussed with a focus on the way metaverse enabled HRM procedures might promote engagement of learning in remote employees. The research used a technique of qualitative case study informed using practice theory. Based on the finding in the study, the metaverse serves an augmented tool, promoting both synchronous and asynchronous instruction in the hotel and tourism industries [13]. Johan has emphasized the quick growth of financial industry as the consequence of FinTech companies development and also the requirement in established financial firms for integrating technology to remain competitive. It describes the research goal of identifying how technology influences the employment processes which recognized financial organizations, with a particular emphasis on business structure comparisons. The qualitative research approach used includes multipurpose financial organization. Chatbots, AI, automation, as well as face recognition are some technological opportunity being investigated to optimize work processes, resulting in digital transformation initiatives. To overcome the high costs involved, the report recommends collaboration as a viable approach to establish financial institutions in implementing technology, includes collaborating with FinTech companies [14]. Chowdhury et al. have offered an abstract that acknowledges the expanding usage of AI in HRM as well as the immense value that may give. Although huge investments, they shows the challenges that businesses face in achieving their assured AI advantages. Accessible HRM research is reviewed, with an emphasis on AI application as well as implication. The research purpose has performed a multidisciplinary literature in systematic review with an objective to better comprehend the organizational resources breadth needed to achieve AI capability over HRM, with a focus on non-technical resources as well as technical resources. The study's contributions includes the development of an AI capability framework as well as recommendations to promote AI scholarship in HRM, which will be valuable for managers to evaluate preparation and establishing AI-enabled HRM methods [15]. Hooi explained the study's goal is to look at the Management Development Index (MDI) effects, HRM, and Employee Engagement (EE) in performance of organization. It covers the research methodology, involving Stratified Random (SR) sampling of multinational corporate management and analysis using structural equation modeling. The report also discusses the key findings, which reveal linkages among Management Development, HRS, and an organization performance in EE acting as a limited intermediary in the performance of HRS-firm performance associated [16]. Singh et al. used the Bayesian Classification algorithm to predict employee performance and leave management using a dataset of 430 employees that they collected themselves. Revenue, pay rate, attendance, expenditure, and employee feedback were among the data points [17]. Hooi explained the study's goal, which is to look at the impact of the MDI, HRS, and EE on company performance which covers the research methodology, involving SR sampling of multinational corporate analysis as well as management using modeling of structural equation. The report has discusses the key result that reveal linkages among MDI, HRS, and an organization performance from EE acting as a limited intermediary in the HRS organization's performance association [18]. Singh et al. used the Bayesian Classification algorithm for forecasting employee performance as well as manage leave using a dataset of 430 employees that they collected themselves. Revenue, pay rate, attendance, expenditure, and employee feedback

were among the data points [19]. Duan and Zhang have illustrated about the employees who provide digital actual time service for employers located remote from their residences, as well as acquire convenient employment data and enhance their digital abilities through the use of digital assets, leading to task economy, sharing economy, crowd sourcing economy, and various kinds of economic activity. At last, the web's virtual space may support practically any number of firms at the same time. In digital trade with commercial context although all firm can get all intermediate provide from the internet market, its ultimate output can instantaneously becomes an intermediary source of other organizations. As a result, firms' supply of commodities as well as product sales is dependent on the overall market. The market scale impact caused by intermediate ingredients can now be indefinitely increased because it is not anymore limited by regional aggregation [20].

3. RESEARCH METHODOLOGY

The research investigation attempted to determine the most efficient patterns for predicting job dissatisfaction between individuals in this database. As a result, the target category in this database is job satisfaction. Data have gather from practitioners belongs to many job functions from services as well as manufacturing organizations located over same association gets collected with "employee attrition" dataset.
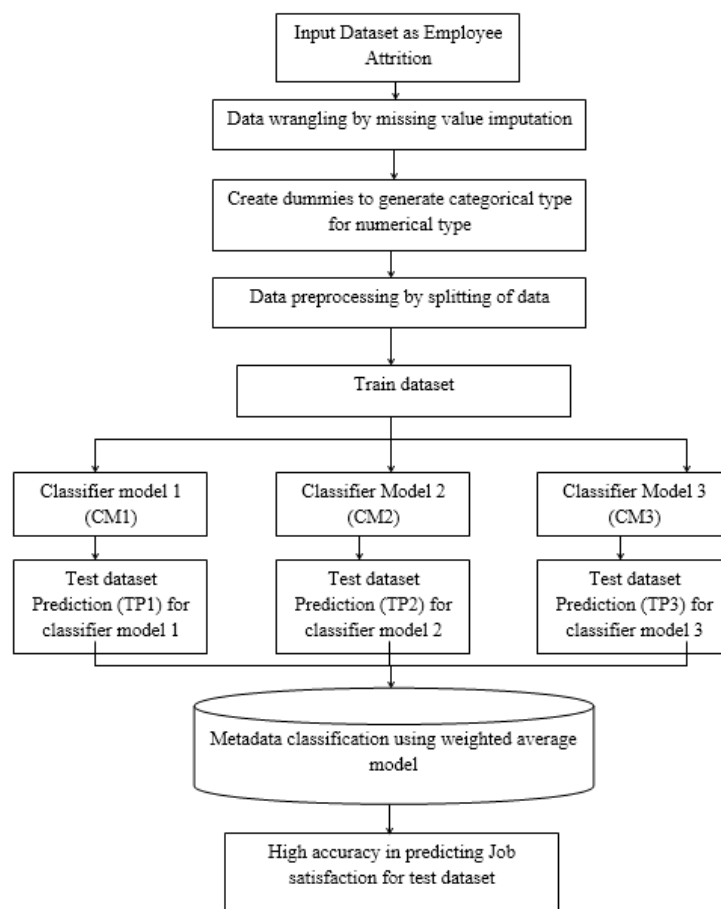

FIGURE 1. Weighted Average Mechanism (WAM) architecture for job satisfaction prediction

Figure 1 illustrates the collected data is used to develop an accurate prediction of employee satisfaction. At first, the data wrangling procedure was used to discover and input missing values, followed by the generation of dummies to convert categorical values into numerical values.

3.1 Ensemble Model with WAM
Whenever the dummies are formed, the dataset is quite simple to analyze the variable. The dataset has been separated into two parts namely the train dataset with 70% and the test dataset with 30%. The proposed WA based ensemble with LR method consists of three classifiers are chosen as classification methods namely CM1 as Extra Tree Classifier (ETC), CM2 as Gradient Boosting Machine (GBM), and CM3 as eXtreme Gradient Boosting (XGB). Moreover, these predictions of classifier are utilized to ensemble by WA algorithm that aims for minimizing the relationship among estimators over an ensemble model through classifier training on feature random samples instead of the complete set. Thus, the WA algorithm considers the weights to specific attributes determined by the importance of the target variable as well as obtains the final outcome are shown in figure 2.
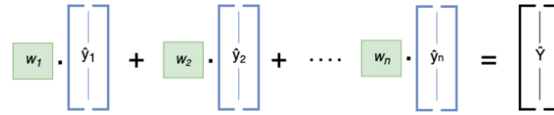
FIGURE 2. WA ensemble model working framework

Stacking is an ensemble learning strategy that integrates multiple machine learning (ML) algorithms through another ML model acting as a meta-learner. The base-level models are trained on the complete training dataset, whereas the meta-model is trained using the combined predictions of all base models as input features. This study incorporates both bagging and boosting approaches to mitigate variance and bias, which assists in training the stacking ensemble and enhances its predictive accuracy.

Furthermore, stacking is implemented via a logistic regression (LR) model, which serves as a mixing layer to determine the optimal aggregation of individual predictions. Stacking using the WA (Weighted Average) algorithm is particularly beneficial for specific tasks. However, its performance may vary significantly depending on the regression complexity. The WA algorithm is defined in Equation 1.

$$A_W = TP_1 \cdot w_1 + TP_2 \cdot w_2 + TP_3 \cdot w_3 + \cdots + TP_n \cdot w_n \tag{1}$$

where:

- A_W is the weighted average for the meta-classifier,

- TP_n is the prediction output of the n-th individual classifier,

- w_n is the specific feature weight assigned to the n-th classifier.

The WA-based ensemble LR model is shown to predict employees' job satisfaction status with high accuracy. The algorithmic process is detailed below.

WA-based Ensemble Algorithm

1. The employee attrition dataset is split into training and testing subsets during preprocessing.

2. The training data is input into three different classifiers, named `CM1`, `CM2`, and `CM3`.

3. Each classifier is executed independently using distinct hypotheses and algorithms. These act as sub-models and generate predictions denoted by TP_1, TP_2, and TP_3 respectively.

4. The outputs of the classifiers are assigned weights w_1, w_2, and w_3.

5. The WA stacking mechanism combines the three classifiers using Equation 1.

6. The final ensemble prediction is made by passing the weighted average output to a logistic regression model acting as the meta-classifier.

This ensemble approach effectively combines the strengths of multiple classifiers, leading to improved prediction accuracy. By selecting and weighting features from different classifiers, the model enhances the reliability of forecasting employee job satisfaction within an organisational context.

4. RESULT AND DISCUSSION

According to the experimental study, the employee attrition dataset contains 500 records with 70% of the dataset used as training dataset as well as 30% used for test dataset. To select the best prediction, utilize three other classifier models as distinct classifiers. Individual classifier testing predictions are recognized, and the parameters of confusion matrix values for ETC, GBM, and XGB are displayed in Table 1. Figure 3 shows the confusion matrix for ETC whereas the implementation has been carried out using Python language, and the findings obtained to perform evaluation of the proposed WA with ensemble model are compared to current classifier prediction value of confusion matrix parameter.

TABLE 1. Confusion matrix attributes value for WA based ensemble LR and existing method

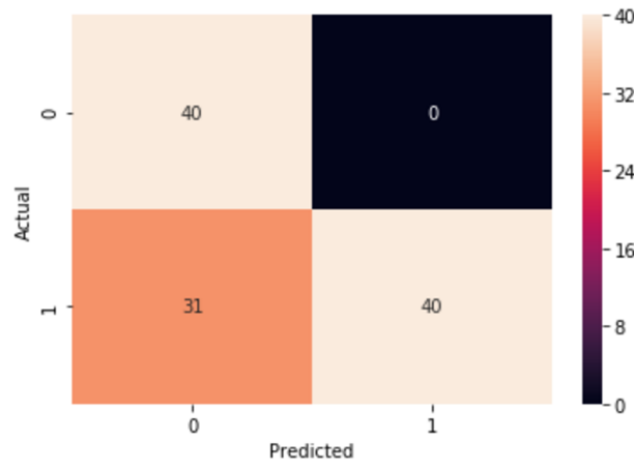| Classification method	True Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)
ETC	40	0	40	31
GBM	65	0	40	6
XGB	68	3	39	0
Proposed WA ensemble LR	71	2	38	0

FIGURE 3. Confusion matrix for ETC

The confusion matrix measures helped to evaluate the model's performance efficiency in accurately forecasting employee job satisfaction. Accuracy is a statistic that is calculated by dividing the number of right predictions by the total number of patients evaluated for observation. The ensemble LR model has an accuracy score of 98.23%, indicating that it predicts work satisfaction the most accurately when compared to other existing models. Table 2 shows ETC, GBM, and XGB.

TABLE 2. Performance evaluation in Job satisfaction prediction

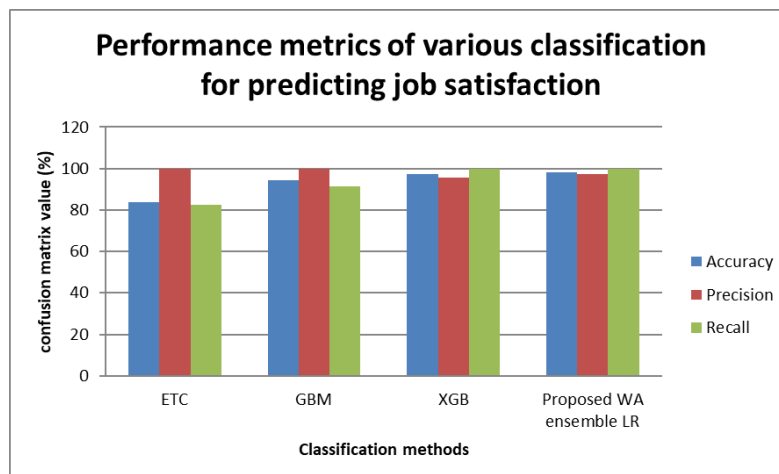| Classification method	Accuracy	Precision	Recall
ETC	83.95	99.89	82.62
GBM	94.47	99.91	91.57
XGB	97.48	95.82	99.86
Proposed WA ensemble LR	98.21	97.34	99.88

FIGURE 4. Performance Evaluation of WA ensemble model

Figure 4 shows that the suggested WA with ensemble LR model has a precision of 97.34%, which is lower than ETC and GBM but higher than XGB based on the number of TP predictions divided by the overall positive predictions (TP+FP). As a result, the suggested WA algorithm assists in improving feature extraction, resulting in datasets with relevant variables linked to the dataset's aim. The WA based ensemble LR model aggregate the significant features from the three independent classifiers such as ETC, GBD, and XGB. Thus, the accumulated WA ensemble LR classifier prediction improved the prediction of employee job satisfaction and the performance of the employee with 98.21%.

5. CONCLUSION

Most of the research focused on the analysis of salary for the "employee job satisfaction" source but this research concentrated in attempting to measure the salary received to the task upon completion of the employment

representative's. However, the integrated study model highlighted the association between employees and job satisfaction by buffering the psychological endorsement effect and lowering the impact of proactive behavior. Therefore, the extracted dataset is utilized with three classifiers are ETC, GBM, and XGB. It serves as the training input to an ensemble LR model, whilst the WA has utilized to collect relevant features as trained dataset that help in improving prediction results. Thus, the findings suggest that a lack of pride in working for one's employer can have a important impact in entire job discontent. Similarly, recent job might be directly related to job discontent, whereas a stressful work environment has had a significant impact on job unhappiness categorization. The ensemble learning technique has the advantage of combining the output of various data mining models such as ETC, GBM, and XGB for classification and ensemble LR is utilized for predicting job satisfaction using those classifier models. The proposed WA ensemble LR has been implemented and its performance accuracy as 98.21% is examined by contrasting ensemble models using three classifier methods which are higher than the existing classifier methods. As a result, the experimental findings showed that the suggested method has a greater prediction accuracy and sensitivity than other current methods, outperforming bagging ensemble results.

REFERENCES

[1]. Forsythe E, Kahn LB, Lange F, Wiczer D (2022) Where have all the workers gone? Recalls, retirements, and reallocation in the COVID recovery. *Labour Econ* 78:102251. https://doi.org/10.1016/j.labeco.2022.102251

[2]. Dunn M (2020) Making gigs work: digital platforms, job quality and worker motivations. *New Technol Work Employ* 35(2):232–249. https://doi.org/10.1111/ntwe.12167

[3]. Li L, Mo Y, Zhou G (2022) Platform economy and China's labor market: structural transformation and policy challenges. *China Econ J* 15(2):139–152. https://doi.org/10.1080/17538963.2022.2067685

[4]. Wong SI, Fieseler C, Kost D (2020) Digital labourers' proactivity and the venture for meaningful work: Fruitful or fruitless? *J Occup Organ Psychol* 93(4):887–911. https://doi.org/10.1111/joop.12317

[5]. Riaz and Ghanghas (2024) Artificial Intelligence in Employee Performance Evaluation and Its Managerial Implication. *J Inform Educ Res* 4(1):299–307.

[6]. Chukwuka EJ, Igweh FN (2024) Strategic Impact of Artificial Intelligence (AI) on Entrepreneurial Creativity and Management. *Int J Bus Entrep Res* 14(11):33–46.

[7]. Kolade O, Owoseni A (2022) Employment 5.0: The work of the future and the future of work. *Technol Soc* 71:102086. https://doi.org/10.1016/j.techsoc.2022.102086

[8]. Ni B, Obashi A (2021) Robotics technology and firm-level employment adjustment in Japan. *Jpn World Econ* 57:101054. https://doi.org/10.1016/j.japwor.2021.101054

[9]. Berkers H, Smids J, Nyholm SR, Le Blanc PM (2020) Robotization and meaningful work in logistic warehouses: threats and opportunities. *Gedrag Organisatie* 33(4):324–347.

[10]. Tschang FT, Almirall E (2021) Artificial intelligence as augmenting automation: Implications for employment. *Acad Manag Perspect* 35(4):642–659. https://doi.org/10.5465/amp.2019.0062

[11]. Reljic J, Evangelista R, Pianta M (2021) Digital technologies, employment, and skills. *Ind Corp Change* dtab059. https://doi.org/10.1093/icc/dtab059

[12]. Ismail M, Binti Mustaffa SN, Abed M (2023) AI-Enabled Deep Learning Model for COVID-19 Identification Leveraging Internet of Things. *Wasit J Comput Math Sci* 2(2):33–39. https://doi.org/10.31185/wjcms.146

[13]. Dutta D, Srivastava Y, Singh E (2023) Metaverse in the tourism sector for talent management: A technology in practice lens. *Inf Technol Tour* 25(3):331–365. https://doi.org/10.1007/s40558-023-00258-9

[14]. Johan S (2021) Enhanced Financial Business Competitiveness by Leveraging Technology and Innovation. *CommIT J* 15(2):79–89. https://doi.org/10.21512/commit.v15i2.6968

[15]. Chowdhury S, Dey P, Joel-Edgar S, Bhattacharya S, Rodriguez-Espindola O, Abadie A, Truong L (2023) Unlocking the value of artificial intelligence in human resource management through AI capability framework. *Hum Resour Manag Rev* 33(1):100899. https://doi.org/10.1016/j.hrmr.2022.100899

[16]. Hooi LW (2021) Leveraging human assets for MNCs performance: The role of management development, human resource system and employee engagement. *Int J Hum Resour Manag* 32(13):2729–2758. https://doi.org/10.1080/09585192.2019.1590443

[17]. Singh H, Vishnavat K, Srinivasan R (2018) Employee performance and leave management using data mining technique. *Int J Pure Appl Math* 118:2063.

[18]. Thakre G, Wankhede P, Patle S, Joshi S, Chauhan A (2021) Implementation of Machine Learning Model for Employee Retention Prediction. *Int J Sci Res Sci Technol* 8(3):503–508. https://doi.org/10.32628/IJSRST2183122

[19]. Lather S, Malhotra R, Saloni P, Singh P, Mittal S (2020) Prediction of employee performance using machine learning techniques. In: *Proc 1st Int Conf Adv Inf Sci Syst*, New York, NY, USA, pp. 1–6. https://doi.org/10.1145/3373477.3373696

[20]. Duan X, Zhang Q (2023) Industrial digitization, virtual agglomeration and total factor productivity. *J Northwest Norm Univ (Soc Sci)* 60(1):135–144. https://doi.org/10.16783/j.cnki.nwnus.2023.01.016

[21]. Employee Attrition Dataset. Kaggle. https://www.kaggle.com/colearninglounge/employee-attrition