Employee Attrition In Human Resource Using Machine Learning Techniques

Dr. T. S. Poornappriya[1] and Dr. R. Gopinath[2]

[1]Data Scientist, Tech Mahindra Bengaluru, Karnataka, India.

[2] D.Litt. (Business Administration) - Researcher, Madurai Kamaraj University, Madurai, Tamil Nadu, India.

ABSTRACT

Employee turnover is a concern that is serious knowledge-based organizations. When employees leave a company, they carry together with them tacit that is invaluable which will be usually the way to obtain competitive advantage when it comes to business. To help a company to continually have an increased advantage that is competitive its competition, it must ensure it is a duty to attenuate employee attrition. HR analytics is a multidisciplinary approach to put together methodology for enhancing the quality of people-related decisions in order to get better individual and organizational performance. There are transposable terms used for HR analytics are talent analytics, people analytics, and workforce analytics. HR analytics plays a role in every characteristic of the HR function, including recruiting, training and development, succession planning, retention, engagement, compensation, and benefits. In this research work, Neural Network based Regressor is used to predict the employee attrition. The performance of the proposed NNR is evaluated with their existing techniques using various evaluation metrics.

KEYWORDS: Employee Turnover, Employee Attrition, Human Resource Management, Machine Learning, Artificial Neural Network, Support Vector Regressor

1.   INTRODUCTION

The Human resource management [1] field includes various stages like planning, organizing, staffing, directing, controlling, recruiting, placement, performance appraisal, compensation and training etc. As a result, with the advancement of time, more and more data are available which is related to HR and obtained from various sources. Today, all the decisions are becoming more data driven than past experience and predictions. In all the fields, phases of HRM Big data analytics became the must to obtain higher, better quality and cost-effective outcome [2].

HR department also faces problem in the minority ways. The Amount of employee information is increasing quickly both in time and with the expansion of company. The previous employee records are also required for various purposes [23]. The skills, behavior and competencies at individual and organization level create the volume even more. All the information cum data is unless HR can store, access and analyze them for the organization speedy. Use of Big data analytics in HR gives improved speed, satisfaction and scale in an organization [3][4].

Big data is the collection of huge amounts of data which can't be processed. Data is essential to every field including management. Big data analytics and proper management of Big Data is very much important in various fields of management such as accessing and predicting customer behaviour, improved supply chain management, Workforce analysis, Human Resource management, Retention-Attrition, Talent analysis and management, Information Management, Knowledge management, HR planning, Business strategy planning, Management information System, Data Mining etc. Critical decision building systems are becoming more and more data driven [5][6][24].

2.   RELATED WORKS

Alduayj, Sarah S., and Phillip Smith [7] through this research work, neural networks and machine learning models, such as for example Long Short-Term Memory (LSTM), Logistic Regression, Naive Bayes and SVM, were taught to forecast the candidate's sentiments following a job interview. Every design was trained utilizing many data weighting and representations techniques, such as Term Frequency - Inverse Document Frequency (TF-IDF), Term Frequency, and Term Binary [25].

de Oliveira, Joao Marcos, et al [8] this paper provided to turnover literature by giving a novel logical attitude on important components of turnover models. The authors forecasting employee turnover may help companies to respond to these commonly negative events with, e.g., enhanced employee retention techniques.

Karande, Shubham, and L. Shyamala [9] dedicated to choosing the key attributes of voluntary employee turnover and just how they could be overcome ahead of when time. The thing is to forecast whether a member of staff will stay or leave predicated on some metrics [26]. The recommended work will utilize the application of ensemble understanding how to address the difficulty, in the place of centering on just one classifier algorithm. Each classification model should be assigned with a few weights predicated on the average person predicted accuracy. The ensemble model will determine the weightage average when it comes to probabilities regarding the specific classification and predicated on this weightage average, a member of staff may be classified.

Bindra, Harleen, Krishna Sehgal, and Rachna Jain [10] examined IBM Watson Human Resource Employee Attrition Dataset to forecast the employee attrition predicated on five selected attributes that are Gender, Education Field, Environment Satisfaction, Distance at home, and Work Life Balance away from 36 variables contained in the dataset. Association Rule Algorithm 'Apriori' alongside Decision Tree Algorithm 'C5.0' is employed.

Chourey, Anjali, Sunil Phulre, and Sadhna Mishra [11] mentioned the forecast of employee attrition utilizing different data mining strategies such as for example Gradient Boosted, Random Forest, Adaboost and Decision Tree Classifier.

Sehgal, Krishna, et al [12] used Particle Swarm Optimisation (PSO) and Grey Wolf Optimisation (GWO) Algorithm from the style of Decision Tree Algorithm "C5.0" which will be fed when you look at the inputs of Associated Rules, utilizing this optimised algorithm when it comes to forecast of employee attrition utilizing IBM Watson Human Resource Employee Attrition Data. The paper could be used to frame improve work guidelines which can help both the employee and employer.

Căplescu, Raluca-Dana, Miruna Ilie, and Vasile Alecsandru Strat [13] geared towards presenting alternatives for examining voluntary attrition in addition to situations if they can be utilized. The authors suggested both descriptive methods (cohort analysis, retention and turnover rates), so that you can introduce tools that any HR manager can very quickly employ, and prognostic techniques (survival analysis, logistic regression), that are more accurate and supply more actionable understanding to decreasing attrition, but require skills and data.

Maxwell, Sarita, and Ms Garima Singh [14] examined the HR procedures, and retention techniques that advertise employee retention in Indian IT sector. The turnover level of employees is quite saturated in this sector as a result of sufficient possibilities when it comes to new skilled employees in global and national, shortage of skilled employees in this industry.

Gulliford, Fred, and Amy Parker Dixon [15] focused to trace the original utilization of AI up to current day, predicated on Qlearsite's business knowledge. Utilizing actual activities, Qlearsite were assigned with representing how companies can improve realize their workforce, recognize efficiency obstacles and create ways of clear them, finally enhancing productivity.

Gulabbhai, Patel Parth kumar, and Manish Gangil [16] utilized the existing and previous employee's data to assess attrition behavior of employees also to give promotion/bonus to employee with non-attrition behaviour simply by using Fuzzy Rules and LSTM. The outcome indicated that the precision of model is improving pertaining to present methods by about 4%.

Immaneni, Kiran Mayi, and Naga Sailaja [17] used the straightforward random methodology to look for the sample size and obtained the principal data. The dwelling equation model happens to be applied to learn the end result of factors from the attrition of employees additionally the result mentioned that the long working hours are greatly affecting followed closely by the lower job profile. The factor analysis happens to be sent applications for the suggestive measures additionally the result stated that the management should reduce the working hours in addition they should provide the incentives with recognition when it comes to employees, so your service quality level should be improved when you look at the hotel industry.

Briker, Vitaly, et al [18] introduced comparative research on machine learning techniques since they are placed on predictions of customer churn, future purchase, and

product associations in after industry businesses. Association rules are employed make it possible to recognize designs across correlations and products in customer purchase behavior.

El-rayes, Nesreen, Michael Smith, and Stephen Michael Taylor [19] created tree-based models to determine the chances of a member of staff leaving a strong through a job change from a dataset of anonymously presented resumes via Glassdoor's on the internet portal. Dataset building and overview statistics are first described followed closely by an even more in-depth evaluation with four exploratory research.

Nair, Priyanka, Jaya Krishna, and Devesh Kumar Srivastava [20] focused to examine the effectiveness of visualization in predictive analytics applying rich R visuals aided by effective visualization tools. The principal tools which can be utilized for analytics want to provide the point on two ends, i.e., effective data visualization and efficient data analysis.

Liu, Liyuan, et al [21] The contribution for this research is the fact that designed a framework to aid a commercial aluminum firm to really make the decisions also to enhance system execution. The framework incorporated entity sentiment, predictive analysis and descriptive analysis. The authors utilized machine learning methods to forecast employees' turnover costs and locate risk aspects.

3.   EMPLOYEE ATTRITION IN HUMAN RESOURCE

Employee Attrition is merely one of several conditions that are major by any organization. In this chronological age of cut-throat competition, you'll find so many factors which lead to dissatisfaction in employee. long working hours, peer pressure, job location, job role, travelling time, a workplace, amenities on the job, perks and lots of other stuff reasons might be one factor for employee attrition. It is rather necessary for the HR department to learn employee satisfaction level [27]. Sometimes the employee many n't have any nagging problem once you glance at the company but others many offer a greater profile with better pay package. So, the employee could be thrilled to leave. Retaining one employee needs a total large amount of insight in many areas. In the last research work, this has found out factors that are very important end up in employee attrition. The results of the model can be used by HR department to plan an approach prior to the employee sends his resignation.

The program this is certainly systematic of methods on human resources (HR) related (big) info is referred to as HR analytics or people analytics. Typical problems in HR analytics is the estimation of churn rates, the identification of skill and knowledge in an organization or simply the prediction of success on a job [28]. HR analytics, as opposed to the use this is certainly simple of performance indicators, is an ever-growing field of good interest as a result of the rapid growth of volume, velocity and variety of HR data, driven due to the digitalization of work processes. Personnel files had previously held it's place in steel lockers in past times, these include now kept in company systems, alongside data from hiring processes, employee satisfaction surveys, emails, and process data.

4.   NEURAL NETWORK BASED REGRESSOR FOR THE FORECASTING OF EMPLOYEE ATTRITION

4.1   Artificial Neural Network

ANNs are characterised as a highly interconnected array of elementary neurons. ANN is inspired by the biological brain learning process. ANNs for functional applications such as design identification, gradation, feature approximation etc. have been commonly employed in recent years. Forest fire detection ANNs have also been effectively used and recent studies have used ANN for the diagnosis of forest fire areas. A typical ANN is a single input layer, one output layer and one or more layers that are hidden. The neuron structure is shown in Figure 1 schematically. There are multiple neurons in every layer, and neurons in layers have different relation weights associated with the neurons of the adjacent layer. Every network node in Figure 1 is represented by a neuron. The input layer neurons are supplied with input vectors. The signals obtained in each neuron of hidden layers and the output layers are multiplied by weights in the neuronal interconnection [29]. Then the neuron produces output through a transfer function via the summed signal. The conventional training of multi-layer neural networks is supervised methods. The network is supplied with the input vector set and the desired output from the network in a supervised learning template. Let inputs and goals for a neural network, for example, be:

$$\{p_1, t_1\}, \{p_2, t_2\}, \ldots, \{p_Q, t_Q\}$$

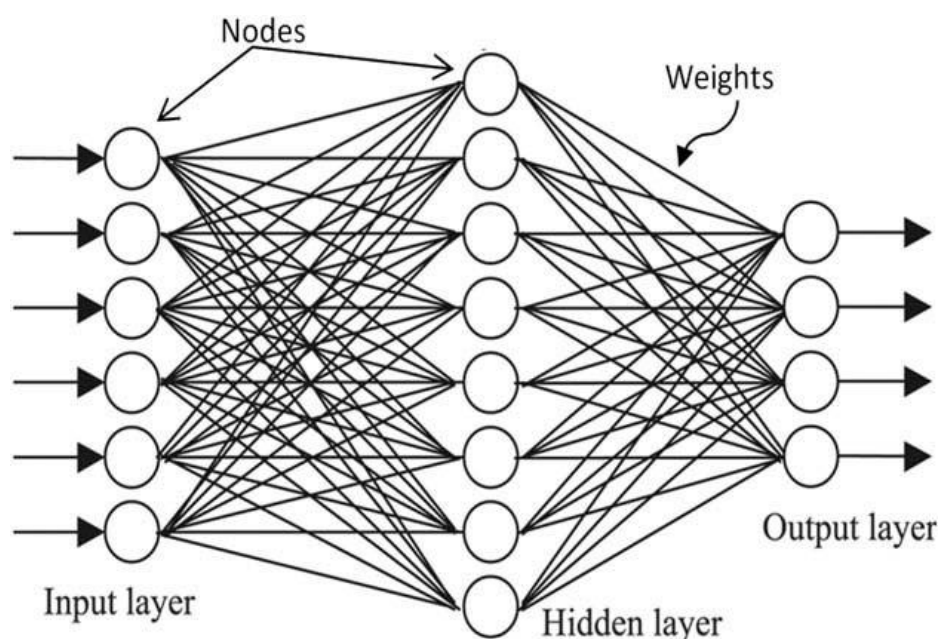where p_Q is an input to network and t_Q is its corresponding target.

Figure 1: Sample Artificial Neural Network

The network is trained iteratively. In each iteration, Mean Square Error (MSE) between target and network output is calculated. The MSE at the kth iteration is given by:

$$F(x) = (t_k - a_k)^2$$

where $F(x)$ is MSE function, t_k and a_k are the targets and output vectors at the kth iteration.

Weight and biases are modified to achieve network preparation, to reduce MSE $F(x)$ function. Return propagation (BP) is typically used for ANN trainings using Levenberg-Marquardt algorithms. BP is a descent method used in each of the layers to change the value of the weights to the neuron to decrease MSE. BP is an interconnecting weight. This adjustment process is repeated in the layer until the minimum error function is reached or MSE is reached to such a low value that the input vectors could be correctly classified. The weights at the metric level are determined by k+1st iteration:

As a downgrade method, BP is an effective and rapid algorithm for convergence. However, BP like any other gradient search technique, when implemented for complex non-linear optimisation problems such as ANNs, provides incompatible and unpredictable results. The error surfaces are extremely complicated due to the dynamic design of training ANNs. Since BP is to converge locally, solutions are highly dependent on the initial random weight draw, due to which BP algorithms may or may not be trapped by a local solution. This local convergence and inability to get out of local minima might pose serious problems while using ANNs.

4.2   Support Vector Regressor

Traditional regressors such as ANNs are very good regressors, but many training sets are required to prepare for good behaviour. In most actual implementations, this might not be possible. SVMs are also a tracked model of education, but they operate very well where only small sets can be found. Using small datasets SVM can be used to accurately classify. The Curse of Dimensionality does not affect SVMs as they are able to handle sparse data in large-scale datasets. Therefore, SVMs are far more generalised than ANNs and the solution given by SVM is much closer to the world's solution. The possibility that SVMs will get caught up in local minima is not threatened.

The fundamental concept of the SVM is that the optimum separation plane is formed under a linear separation. Figure 2 shows the fundamental theory (this shows classification of the data set containing two separate data groups, class I (circles) and class II (squares). This classification is shown. The SVM tries to set an optimal hyper plane (linear boundary) of both classes and orients it to maximise the margin (the distance from boundary to the closest point of each class). For the concept of the margin the closest data points to the segregating frontier are used and called support vectors.
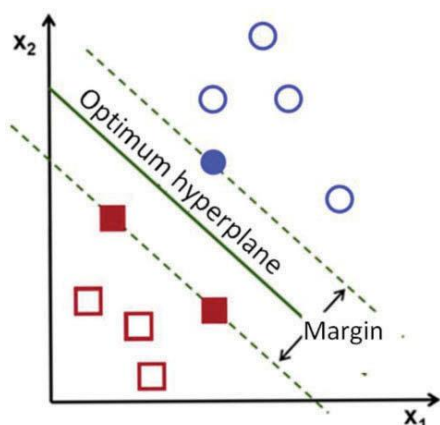
Figure 3.2: Sample SVM

4.3 Proposed Neural Network based Regressor

In this proposed ANN-SVM based Regressor, ANN Regressor is hybridized with SVM. The weights of the ANN are optimized with nearest point value of the classes from the SVM Regression technique.

Input: The benchmark dataset (HR Analytics Dataset)
Output: Number of Employee Attrition

Step 1: First, randomly initialize the weights of the ANN.
Step 2: Take the input data values from the given dataset for the training data.
Step 3: Propagate the input vector through the network to obtained an output.
Step 4: Determining an error rate by comparing the predicted output to the desired output.
Step 5: Adjust weights by means of nearest point values using SVM to minimize the overall error of ANN.
Step 6: Repeat steps 3-5 with next input vector until overall error lies in tolerance range.

5. RESULT AND DISCUSSION

5.1 Dataset Description

The performance of the proposed NN based regressor is evaluated for IBM HR analytics dataset [22]. The dataset is composed of 35 features includes attrition feature also.

Table 1: Description of the HR Analytics Dataset

Feature Number	Feature Name
1	Age
2	Attrition
3	Business Travel
4	Daily Rate
5	Department

6	Distance From Home
7	Education
8	Education Field
9	Employee Count
10	Employee Number
11	Environment Satisfaction
12	Gender
13	Hourly Rate
14	Job Involvement
15	Job Level
16	Job Role
17	Job Satisfaction
18	Martial Status
19	Monthly Income
20	Monthly Rate
21	Number of Companies worked
22	Over18
23	Overtime
24	Percentage Salary Hike
25	Performance Rating
26	Relationship Satisfaction
27	Standard Hours
28	Stock Option Level
29	Total Working Years
30	Training time last year
31	Work life balance
32	Years at Company
33	Years in current role
34	Years since last promotion
35	Years with current manager

5.2   Performance Analysis of the Proposed NN based Regressor

The evaluation of the Machine Learning algorithms like Support Vector Regressor (SVR), Decision Tree, Random Forest and Combined ANN-SVR method is done with Root Mean Square Error (RMSE). Table 6.1 depicts the performance analysis of the different type of regressor for predicting suspectability value of the employee attrition.

Table 2: RMSE obtained by the SVR, DTR, RFR, and combined SV-NN technique for predicting suspectability value

Regression Technique	RMSE value
Support Vector Regression	0.671

Decision Tree Regression	0.6241
Random Forest Regressor	0.614
Proposed NN based Regressor	0.468

Regression Error Characteristic (REC) Estimation

Receiver Operating Characteristic (ROC) curves provide a powerful tool for visualizing and comparing classification results. Regression Error Characteristic (REC) curves generalize ROC curves to regression. REC curves plot the error tolerance on the $x-axis$ versus the percentage of points predicted within the tolerance on the $y-axis$. The resulting curve estimates the cumulative distribution function of the error. The REC curve visually presents commonly-useds tatistics. The area-over-the-curve (AOC) is a biased estimate of the expected error. The R^2 value can be estimated using the ratio of the AOC for a given model to the AOC for the nul-model. Users can quickly assess the relative merits of many regression functions by examining the relative position of their REC curves. The shape of the curve reveals additional information that can be used to guide modeling. Figure 3 depicts the Regression Error Characteristic (REC) estimation obtained for the various models executed in this research work [32].
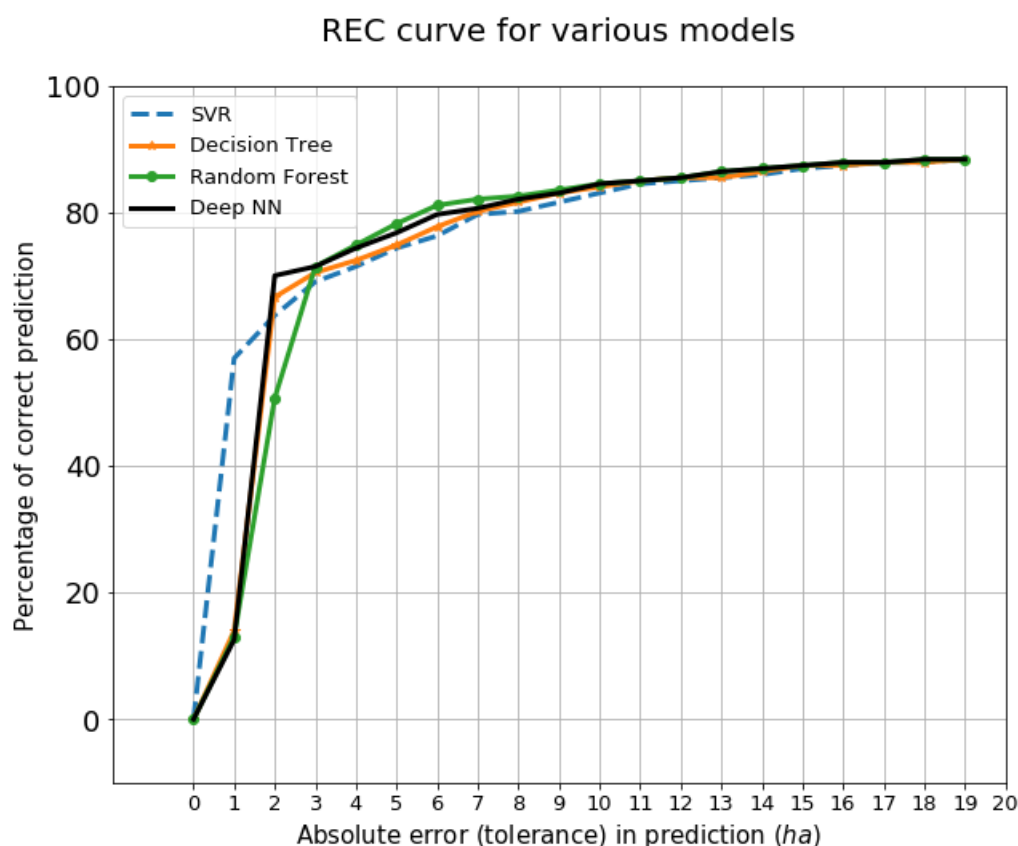
Figure 3: Regression Error Characteristic (REC) estimation obtained by the various Regression Models

The metrics like True Positive (TP), True Negative (TN), False Positive (FP), False negative (FN), Positive predictive value (PPV) (%), Negative predictive value (NPV) (%), Sensitivity (%), Specificity (%), Overall accuracy (%) (ACC), Kappa statistic, area under the ROC curve (AUC) are used to evaluate the models used in this research work. Table 4 depicts the SVR, DT, RF and ANN model performance during training phase for the given HR dataset.

Table 4: NN, DTR, RFR and SVR model performance analysis

Performance Metrics	Regression Techniques			
	NN	DTR	RFR	SVR
True Positive (TP)	175	165	169	171
True Negative (TN)	167	147	162	160
False positive (FP)	12	22	18	16
False negative (FN)	20	40	27	25
Positive predictive value (PPV) (%)	93.6	88.2	90.4	91.4
Negative predictive value (NPV) (%)	89.3	78.6	85.6	86.6
Sensitivity (%)	89.7	80.5	86.2	87.2
Specificity (%)	93.3	89.9	87.0	91.0
Accuracy (in %)	91.44	83.42	87.97	89.04
Kappa statistic	0.829	0.668	0.781	0.781
AUC	0.973	0.945	0.919	0.948

6. CONCLUSION

In virtually any organization's talent management is starting to become an method that is increasingly crucial of HR functions. Talent management can be explained as an outcome so that the right person when you look at the job that is right. Human capital is one of resource that is effective hiring the highly qualified personnel for enhancing the world economy and in addition for developing company's management [30]. Turnover of employee considers among the issue that is major every company faces. Especially, then the company faces great loss during that period if the employee has advance skills at his/her working field. Through this research work, Artificial Neural Network based Regressor is proposed using Support Vector Regressor [31]. This research work gives a prediction model for getting suspectability value of employee attrition by using Regression models. From the result and discussion, it is observed that the SVR, and NN performs good in terms of error rate and specificity, sensitivity, accuracy.

REFERENCES

[1] Mathis, R. L., & Jackson, J. H. (2008). Human resource management. Thomson/South-western.

[2]   Zang, S., & Ye, M. (2015). Human resource management in the era of big data.

[3]   Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., & Trichina, E. (2021). Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review. The International Journal of Human Resource Management, 1-30.

[4]   Alduayj, S. S., & Rajpoot, K. (2018, November). Predicting employee attrition using machine learning. In 2018 international conference on innovations in information technology (iit) (pp. 93-98). IEEE.

[5]   Jain, P. K., Jain, M., & Pamula, R. (2020). Explaining and predicting employees' attrition: a machine learning approach. SN Applied Sciences, 2(4), 1-11.

[6]   Mohbey, K. K. (2020). Employee's Attrition Prediction Using Machine Learning Approaches. In Machine Learning and Deep Learning in Real-Time Applications (pp. 121-128). IGI Global.

[7]   Alduayj, Sarah S., and Phillip Smith. "Sentiment Classification and Prediction of Job Interview Performance." 2019 2nd International Conference on Computer Applications   & Information Security (ICCAIS). IEEE, 2019.

[8]   de Oliveira, Joao Marcos, et al. "Mirror, Mirror on the wall, who is leaving of them all: Predictions   for   employee   turnover   with   gated   recurrent   neural networks." Collaborative innovation networks. Springer, Cham, 2019. 43-59.

[9]   Karande, Shubham, and L. Shyamala. "Prediction of Employee Turnover Using Ensemble Learning." Ambient Communications and Computer Systems. Springer, Singapore, 2019. 319-327.

[10]   Bindra, Harlieen, Krishna Sehgal, and Rachna Jain. "Optimisation of C5. 0 Using Association Rules and Prediction of Employee Attrition." International Conference on Innovative Computing and Communications. Springer, Singapore, 2019.

[11]   Chourey, Anjali, Sunil Phulre, and Sadhna Mishra. "Employee Attrition   Prediction Using Various Machine Learning Techniques."

[12]   Sehgal, Krishna, et al. "Prediction of Employee Attrition Using GWO and PSO Optimised Models of C5. 0 Used with Association Rules and Analysis of Optimisers." Innovations in Computer Science and Engineering. Springer, Singapore, 2019. 1-8.

[13]   Căplescu, Raluca-Dana, Miruna Ilie, and Vasile Alecsandru Strat. "Voluntary employee attrition. Descriptive and predictive analysis." Proceedings of the International Conference on Applied Statistics. Vol. 1. No. 1. Sciendo, 2019.

[14]   Maxwell, Sarita, and Ms Garima Singh. "Impact of Human Resource Practices on Employee Retention in the Indian IT Sector."

[15]   Gulliford, Fred, and Amy Parker Dixon. "AI: the HR revolution." Strategic HR Review (2019).

[16]   GULABBHAI, PATEL PARTHKUMAR, And MANISH GANGIL. "Employees Skills   Inventory using Deep Learning for Human Resource Management." Research Journal of Engineering Technology and Management (ISSN: 2582-0028) 2.04 (2019).

[17] Immaneni, Kiran Mayi, and Naga Sailaja. "A Study on Factors Effecting the Employees   Attrition in Hotel Industry with Reference Hyderabad." International Journal of   Management (IJM) 10.6 (2019).

[18] Briker, Vitaly, et al. "Identifying Customer Churn in After-market Operations using Machine Learning Algorithms." SMU Data Science Review 2.3 (2019): 6.

[19] El-rayes, Nesreen, Michael Smith, and Stephen Michael Taylor. "An Explicative and Predictive Study of Employee Attrition using Tree-based Models." Available at SSRN 3397445 (2019).

[20] Nair, Priyanka, Jaya Krishna, and Devesh Kumar Srivastava. "Visual Analytics Toward   Prediction of Employee Erosion Through Data Science Tools." Information and   Communication Technology for Sustainable Development. Springer, Singapore, 2020.   705-713.

[21] Liu, Liyuan, et al. "Using HR Analytics to Support Managerial Decisions: A Case Study." Proceedings of the 2020 ACM Southeast Conference. 2020.

[22] https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset

[23] Subhashini, M., & **Gopinath, R.** (2020). Employee Attrition Prediction in Industry using Machine Learning Techniques, International Journal of Advanced Research in Engineering and Technology, 11(12), 3329-3341.

[24] Upendran, V., & **Gopinath, R.** (2020). Feature Selection Based on Multi criteria Decision Making for Intrusion Detection System. International Journal of Electrical Engineering and Technology, 11(5), 217-226.

[25] Upendran, V., & **Gopinath, R.** (2020). Optimization Based Classification Technique for Intrusion Detection System. International Journal of Advanced Research in Engineering and Technology, 11(9), 1255-1262.

[26] Kalaiarasi, K., & **Gopinath, R.** (2020). Fuzzy Inventory EOQ Optimization Mathematical Model, International   Journal of Electrical Engineering and Technology, 11(8), 169-174.

[27] Kalaiarasi, K., & **Gopinath, R.** (2020). Stochastic Lead Time Reduction for Replenishment Python-Based Fuzzy Inventory Order EOQ Model with Machine Learning Support, International Journal of Advanced Research in Engineering and Technology, 11(10), 1982-1991.

[28] Shanmugavadivu, S. A., & **Gopinath, R.** (2020). On the Non homogeneous Ternary Five Degrees Equation with three unknowns $x^2 - xy + y^2 = 52z^5$, International Journal of Advanced Research in Engineering and Technology, 11(10), 1992-1996.

[29] Shanmugavadivu, S. A., & **Gopinath, R.** (2020). On the Homogeneous Five Degree Equation   with   five   unknowns $2(x^5 - y^5) + 2xy(x^3 - y^3) = 37(x + y)(z^2 - w^2)P^2$, International Journal of Advanced Research in Engineering and Technology, 11(11), 2399-2404.

[30] Priyadharshini, D., **Gopinath, R**., & Poornapriya, T.S. (2020), "A fuzzy MCDM approach for measuring the business impact of employee selection", International Journal of Management, 11(7), 1769-1775.

[31]   Poornapriya, T.S. & **Gopinath, R**. (2020), Rice Plant Disease Identification using Artificial Intelligence Approaches, International Journal of Electrical Engineering and Technology, 11 (10), 392-402.

[32]   Poornapriya, T.S. & **Gopinath, R**. (2020), Application of Machine Learning Techniques for Improving Learning Disabilities, International Journal of Electrical Engineering and Technology, 11 (10), 403-411.