Logistic and Linear Regression Classifier Based Increasing Accuracy of Non-Numerical Data for Prediction of Enhanced Employee Attrition

Dr. Ganesh Khehare,
Associate Professor,
Department of Computer Science and Engineering,
PIT, Parul University
Vadodara, Gujarat, India.
ganesh.khekare19325@paruluniversity.ac.in

Dr K Balaji,
Associate Professor and Chairperson,
Amrita School of Arts, Humanities and Commerce,
Dept of Commerce and Management,
Amrita Vishwa Vidyapeetham,
Bogadi, Mysuru campus, India
kbalaji@my.amrita.edu

Ms. Malika Arora,
Assistant Professor,
Department of Electronics & Communication Engineering,
Chandigarh Engineering College
Jhanjeri, India
malika.j540@cgc.ac.in

Raman R.Tirpude
Assistant professor
Department of managemnt
Maharashtra National Law University
Nagpur, India
ramantirpude100@gmail.com

Dr. Bhawna Chahar
Associate Professor
Department of Business Administration,
Manipal University Jaipur
Jaipur, India
drbhawnachahar@gmail.com

Aniruddha Bodhankar
Management, Assistant Professor
Dept of Management
Dr. Ambedkar Institute of Management Studies & Research
Nagpur
aniruddha.bodhankar16@gmail.com

Abstract—Nowadays attrition rate of employees in several organizations are very high. This higher attrition rate of employee depends on several factors. Logistic and linear regression classifier based increasing accuracy of non-numerical data for prediction of enhanced employee attrition is proposed in this paper.1470 samples of Employee attrition dataset have been used in this study for the categorization of Employee Attritionby Logisticas well as Linear Regression. Accuracy rate of both the approaches are compared. On the basis of accuracy rate supremacy of the Logistic regression over Linear regression is also presented in this paper.

Keywords-Attrition rate,Linear regression,Deviance test,Wald Chi Square Test, Logistic Regression,Likelihood-Ratio Test.

I. INTRODUCTION

The motivation behind human resource management is covering specialist execution and responsibility, examining laborers' joint effort models, researching employee pal and turnover, and making worker lifetime assessment. Human asset examination is a creating application space for examination of HRM targets [1]. HR Analytics is confined with logical proof examination of a particular employee[2]. Employee is a basic resource while considering the present information driven industry. Employee attrition can turn out to be exorbitant for an association. The extra charges as expensescalculated for an employee by the HR computation where some of lost information may occur, worker spirit, motivation to multiple workers is also defined for employees. In 2018 utilizing representatives' survey answers with respect to their administrator, performed an exploration to quantify individuals' administration abilities. The chief end is that individuals the board has a solid negative relationship with workers whittling down [3]. Employee attrition can be used in applications such as score boards, bonus point calculation etc. Also, retention of employees as an application helps to involve performance analysis.

Numerous analysts have demonstrated [4], [5] the value of all working resource, creation and the executives, also recognizing association for employee usefulness. 105 research papers have been published in IEEE Xplore while 230 articles in Google Scholar. In 2016. Workplace has many forms of control where there is no relative understanding between elements. Also, they are finding the position of representing various infrastructure around the workplace [6]. It examined people's Attrition information to foresee the worker's steady loss dependent on five classes. [7][8].The calculated relapse model is the most oftentimes utilized relapse model for the examination of this information [9]. Strategic Regression has been utilized to recognize the components that decide work fulfillment [10], balance between serious and fun activities [11], workplace and creation misfortune [12], stress, and disappointment on representatives [13], and impetus consequences for worker commitment [14 -24].

The old-stylescheme of employee attrition is in the physicalmethodcompanies have to check on the employees who are leaving the firmsince the physicalstructure attrition rate turn out to behigher. Theoutcomesare carried in relations of typical style capacities and intentions that have shaped the finestnon-Numericalinformationoutcomesaccessible for the Logistic Regression Classifier. The aim of this employee attrition forecastscheme is to diminish the attrition rate of the firms which are suffering from high degree of attrition rate.

II. RESOURCES AND PROCEDURES

The lab is called Data Analytics. The configuration of the systems is high in the lab for obtaining effective results once the experiments are carried out. Group count is two with each group comprising 62 samples [25-27]. The total size of samples used is 124 while considering the proposed with existing systems. Values of alpha as well as beta are 0.05 along with 0.2 respectively with 80% of G-power [3].

The information covers 261 days1470 staffs are recognized. The firm is separated into three workplaces and is tasked with the work of the project. The Research and Development section where 65% of the staffsare employed is considered as the key of this study. Working out rates for

institution staff are 31%. The field of over-training of the establishmentis Life Science and healthcare with 47% and 22%, correspondingly. The normal time hours of at work are 7-8 hours daily and recognized theft linkedevidence is 19.70 percent.

Jupyter Notebook is essential for compiling the idea of working module. Operating system with 64-bit and 11 update, which has power of 8GB RAM and 256GB Read Only Memory for working the non-Numericaldata, and Processor of speed with high i7.Jupyter notebook integrated with Python 3.7 has been used for applied for the projected technique.[28-30] Logistic regression algorithm is shown in Table 1.

In the preparation of group 2sample, Linear regression is used a classifier for newestimate, to appropriate the model: Objective flexibility $'b'$ is defined as an opportunity to leave the first manager and a_n means the same forms of assessment, pay, working duration and related control factors including employee business, job title and unique metropolitan region. There are severalunusablefeatures in this dataset. Its ideas are basic and helpful in learning. To create the capacity for self-determination through the progress of each job in which can present each employee as an obligation to stay or leave his or her current business; here the expected probability is combined at 1 and surprised at 0. Include this model for compelling reasons after the encounter with the previous models. Table 2 shows the pseudocode of Linear regression.[31-32]

TABLE I. PREDICTED ACCURACY OF EMPLOYEE ATTRITION

S. No.	LR Accuracy (%)	SVM Accuracy (%)	S. No.	LR Accuracy (%)	SVM Accuracy (%)
1	85	85	32	85	84
2	87	85	33	81	75
3	91	86	34	91	73
4	81	71	35	77	81
5	80	77	36	80	87
6	77	79	37	79	73
7	80	75	38	80	81
8	79	74	39	86	76
9	89	83	40	89	83
10	86	81	41	89	81
11	86	86	42	83	83
12	86	84	43	79	74
13	92	87	44	81	75
14	80	70	45	88	79
15	81	78	46	83	77
16	76	78	47	89	71
17	81	76	48	78	86
18	78	75	49	78	85
19	90	82	50	87	85
20	85	80	51	80	77
21	86	85	52	81	79
22	89	86	53	91	71
23	79	77	54	87	75
24	80	75	55	85	86
25	77	83	56	86	74
26	80	85	57	89	85
27	81	71	58	79	83
28	91	79	59	80	85
29	87	74	60	77	81
30	85	81	61	85	81
31	87	83			

Informationexisting in the dataset is transformed into metadata in directive for easy legibility and analyzing and encrypting the input characteristics. Here, creates an input

function for reading and analyzing the file, and to convert the features and labels into a variable named tf-data. Dataset for the purpose of training and validation. Lastly, the preprocessing and the input by planning the target label to an index took place.

A. ArithmeticalInvestigation:

IBM SPSS is used to conduct implementation work-based statistics. Computation of mean along with standard deviation is performed including graph plotting. Two groups are selected for conducting T-test independently as well as significance value calculation. Gender along its attrition is taken as independent variables whereas various parameters such as accuracy as well precision are considered to be dependent ones. The proposed work is also subjected to analysis of T-test independently.

III. RESULTS

Predicted accuracy of employee attrition is represented in Table 1 withan accuracy of 83.59 % forlogistic regression and of 79.69% for linear regression. The total sample size taken for logistic as well as linear regression is 62. In the presentedtechnique, logistic regression has relativelyimproved accuracy as compared to linear regression algorithm.

TABLE II. ALGORITHM FOR LOGISTIC REGRESSION

Logistic Regression Algorithm
Input:embedding_ecoder
Function Logistic regression (classification.model):
1. Input → df.describe () //Input layer
2. H0 → layers. Concatenate (encoded_features) //
3. H1→ layers. concatenate //
4. H= import LabelEncoder
5. X→sklearn.model_selection (classification.model):
6. X1=def classification report (model):
7. return Input, H
Output: Logistic regression (features)

Logistic regression algorithm is shown in Table 2.Primarily the input is the inserting encryption and the key functionality is the logistic regression that has the constraints as categorization of Model. Subsequent the variable statementoccurs in such a means that the variable H is demarcated to be layers. Concatenate prearrangedstructures and variable H1 is demarcated as layers. concatenate a completelyassociated layer and all the variables are demarcatedconsequently. Lastly, the outcome is derived by running the topographies model and forecasts the finest accuracy.

Linear regression algorithm is shown in Table 3. Firstly the input is the set in encryption and the foremost functionality is the Linear regression that has the constraints as administered Model. Subsequent the variable announcementoccurs in such a means that the variable H is

demarcated to be layers. Preprocessing (prearrangedtopographies) and variable H1 is demarcated as layers. preprocessing a entirelyassociated layer and all the variables are demarcatedconsequently. In conclusion, the yield is derived by running the topographies model and forecasts the finestaccurateness.

Computation of tested samples among logistic regression as well as linear regression are shown in Table 4. Logistic regression has mean accuracy 83.5900 and the linear regression is 79.6900.

Logistic regression has Standard deviation of 1.43948 and the linear regression is 2.09295. Logistic regression has mean loss as 5.6880 and the linear regression is 9.6581. Logistic regression has Standard error mean as 0.57509 as well as linear regression as 1.76508.

TABLE III.   ALGORITHM FOR LINEAR REGRESSION

Linear Regression Algorithm
Input:embedding_ecoder
Function Linear regression (supervised.model):
Input → df.describe () //Input layer
H0 → layers. Preprocessing (encoded_features)
H1→ layers. Preprocessing
H= import LabelEncoder
X0 → sklearn. ensemble import Linear regression (supervised.model):
X1→ sklearn.import confusion_matrix
X =def classification_report (model):
return Input, H
Output: Linear regression (features)

TABLE IV.   ARITHMETICAL COMPUTATION FOR LOGISTIC REGRESSION (LR)ANDLINEAR REGRESSION

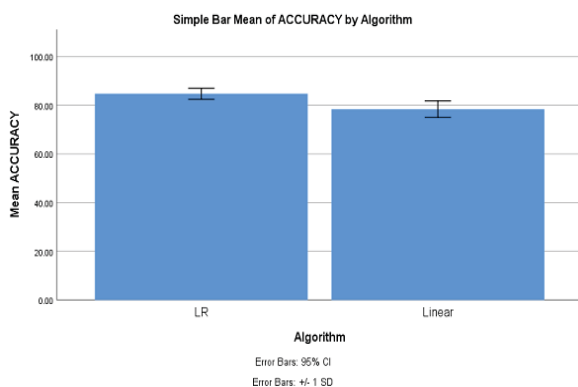| Algorithm	N	Mean	Std. Deviation	Std Error Mean
L R	62	83.590	1.43948	0.57509
Linear	62	79.690	2.09295	1.76508

Fig. 1.   Bar Graph of comparative Accuracy

Accuracy rate of logistic as well as linear regression classifier is shown in Figure 1. Accuracy rate of Logistic Regression gives better performance over linear regression. Logistic regression gives an accuracyof 83.59 % while linear regression's accuracy is 79.69%.

IV. DISCUSSION

Logistic regression obtains more accuracy in comparison with linear regression. Statistical mean accuracy is computed for logistic regression as 83.59% as well as for linear regression arithmetical mean accuracy value is 79.69%.

The choices for various employees understanding the HR comparison can be shown by focusing on individual parameters of employees. These calculations of individual features like age, salary, empid, etc. are highlighted with the relative results [26]. Presentation assessment, utilized the disarray framework to infer a progression of central measurements to quantitatively communicate the proficiency of every calculation: recording precision, accuracy, recall, and sensitivity [27]. In this manner, the Logistic Regression can be used for classifying objects and help in accomplishing the aim of investigation [28]. Prediction of employee attrition rate is presented in this paper using theutmostsignificant Logistic and Linear Regression Classifiers. Wherever it is assumed to stipulate the dimensiondescription, the projected model displays Logistic and Linear Regression. The Logistic Regression algorithm shows an accuracy rate of 83.59% whileLinear Regression algorithm has 79.69% of accuracy rate [29].

The restrictions of employee attrition forecast are that it can only forecast the attrition rate based on former employees' records. Besides, the application comprises the partition of the logistic and linear regression to regulate whether it is conceivable for a demonstrative to leave the association[30]. In which the Logistic Regression provides higher outcomes compared with Linear Regression Classifier in Employee Attrition Prediction. For future examination on attrition expectation, physical components in regards to employee attrition are proposed for investigation. The impact of the quantity of accessible opportunities for every business, thinking about his details and situational factors identifying with his/her wearing down likelihood, can likewise be examined in future works.

V. CONCLUSION

Due to highest attrition rate of employees in several organizations, Logistic and linear regression classifier based increasing accuracy of non-numerical data for prediction of enhanced employee attrition is presented in this paper. Results shows that the prediction accuracy of the Logistic Regression model is 83.59% which is higher than the Linear Regression model with 79.69%.

REFERENCES

[1]   S. N. Mishra, D. R. Lama, Y. Pal, and Others, "Human Resource Predictive Analytics (HRPA) for HR management in organizations," International Journal of Scientific & Technology Research, vol. 5, no. 5, pp. 33–35, 2016.

[2]   J. H. Marler and J. W. Boudreau, "An evidence-based review of HR Analytics," The International Journal of Human Resource Management, vol. 28, no. 1, pp. 3–26, Jan. 2017.

[3] M. Hoffman and S. Tadelis, "People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis," J. Polit. Econ., vol. 129, no. 1, pp. 243–285, Jan. 2021.

[4] M. Marchington, A. Wilkinson, R. Donnelly, and A. Kynighou, Human Resource Management at Work. Kogan Page Publishers, 2016.

[5] N. Bloom and J. Van Reenen, "Chapter 19 - Human Resource Management and Productivity," in Handbook of Labor Economics, vol. 4, D. Card and O. Ashenfelter, Eds. Elsevier, 2011, pp. 1697–1767.

[6] D. V. A. J. Raja and R. A. R. Kumar, "A study to reduce employee attrition in it industries," International Journal of Marketing and Human Resource Management, vol. 6, no. 3, pp. 01–14, 2015.

[7] H. Bindra, K. Sehgal, and R. Jain, "Optimization of C5.0 Using Association Rules and Prediction of Employee Attrition," International Conference on Innovative Computing and Communications. pp. 21–29, 2019. doi: 10.1007/978-981-13-2354-6_3.

[8] S. N. Mishra and D. R. Lama, "A decision making model for human resource management in organizations using data mining and predictive analytics," International Journal of Computer Science and Information Security, vol. 14, no. 5, p. 217, 2016.

[9] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression. John Wiley & Sons, 2013.

[10] E. Villar-Rubio, J. Delgado-Alaminos, and P. Barrilao-González, "Job satisfaction among Spanish tax administration employees: A logistic regression analysis," J. Labor Res., vol. 36, no. 2, pp. 210–223, Jun. 2015.

[11] K. Karhula et al., "Objective working hour characteristics and work-life conflict among hospital employees in the Finnish public sector study," Chronobiol. Int., vol. 34, no. 7, pp. 876–885, Jun. 2017.

[12] M. Lohela-Karlsson, J. Hagberg, and G. Bergström, "Production loss among employees perceiving work environment problems," Int. Arch. Occup. Environ. Health, vol. 88, no. 6, pp. 769–777, Aug. 2015.

[13] G. Halkos and D. Bousinakis, "The effect of stress and dissatisfaction on employees during crisis," Econ. Anal. Policy, vol. 55, pp. 25–34, Sep. 2017.

[14] K. K. Merriman, S. Sen, A. J. Felo, and B. E. Litzky, "Employees and sustainability: the role of incentives," &ctx_ver=Z, vol. 39, p. 88, May 2016.

[15] M. S. Samuel, J. Bhattacharya, S. Raj, N. Santhanam, H. Singh, and N. D. Pradeep Singh, "Efficient removal of Chromium(VI) from aqueous solution using chitosan grafted graphene oxide (CS-GO) nanocomposite," Int. J. Biol. Macromol., vol. 121, pp. 285–292, Jan. 2019.

[16] J. Johnson, G. Lakshmanan, B. M, V. R M, K. Kalimuthu, and D. Sekar, "Computational identification of MiRNA-7110 from pulmonary arterial hypertension (PAH) ESTs: a new microRNA that links diabetes and PAH," Hypertens. Res., vol. 43, no. 4, pp. 360–362, Apr. 2020.

[17] H. Venu, L. Subramani, and V. D. Raju, "Emission reduction in a DI diesel engine using exhaust gas recirculation (EGR) of palm biodiesel blended with TiO2 nano additives," Renewable Energy, vol. 140, pp. 245–263, Sep. 2019.

[18] B. Keerthana and M. S. Thenmozhi, "Occurrence of foramen of huschke and its clinical significance," J. Adv. Pharm. Technol. Res., vol. 9, no. 11, p. 1835, 2016.

[19] E. P. Thejeswar and M. S. Thenmozhi, "Educational research-iPad system vs textbook system," J. Adv. Pharm. Technol. Res., vol. 8, no. 8, p. 1158, 2015.

[20] R. N. Krishna and K. Y. Babu, "Estimation of stature from physiognomic facial length and morphological facial length," J. Adv. Pharm. Technol. Res., vol. 9, no. 11, p. 2071, 2016.

[21] A. Subashri and M. S. Thenmozhi, "Occipital emissary foramina in human adult skull and their clinical implications," J. Adv. Pharm. Technol. Res., vol. 9, no. 6, p. 716, 2016.

[22] N. Sriram, Thenmozhi, and S. Yuvaraj, "Effects of Mobile Phone Radiation on Brain: A questionnaire based study," J. Adv. Pharm. Technol. Res., vol. 8, no. 7, p. 867, 2015.

[23] R. K. Jain, S. P. Kumar, and W. S. Manjula, "Comparison of intrusion effects on maxillary incisors among mini implant anchorage, j-hook headgear and utility arch," J. Clin. Diagn. Res., vol. 8, no. 7, pp. ZC21–4, Jul. 2014.

[24] A. Menon and M. S. Thenmozhi, "Correlation between thyroid function and obesity," J. Adv. Pharm. Technol. Res., vol. 9, no. 10, p. 1568, 2016.

[25] S. Moguilner, S. P. Knight, J. R. C. Davis, A. M. O'Halloran, R. A. Kenny, and R. Romero-Ortuno, "The Importance of Age in the Prediction of Mortality by a Frailty Index: A Machine Learning Approach in the Irish Longitudinal Study on Ageing," Geriatrics, vol. 6, no. 3, Aug. 2021, doi: 10.3390/geriatrics6030084.

[26] K. Bhuva, K. Srivastava, and Others, "Comparative Study of the Machine Learning Techniques for Predicting the Employee Attrition," IJRAR-International Journal of Research and Analytical Reviews (IJRAR), vol. 5, no. 3, pp. 568–577, 2018.

[27] F. Fallucchi, M. Coladangelo, R. Giuliano, and E. William De Luca, "Predicting Employee Attrition Using Machine Learning Techniques," Computers, vol. 9, no. 4, p. 86, Nov. 2020.

[28] S. Ponnuru, G. Merugumala, S. Padigala, R. Vanga, and B. Kantapalli, "Employee attrition prediction using logistic regression," Int. J. Res. Appl. Sci. Eng. Technol, vol. 8, no. 5, pp. 2871–2875, 2020.

[29] D. C. Montgomery, E. A. Peck, and G. Geoffrey Vining, Introduction to Linear Regression Analysis. John Wiley & Sons, 2015.

[30] L. D. Schroeder, D. L. Sjoquist, and P. E. Stephan, Understanding Regression Analysis: An Introductory Guide. SAGE Publications, 2016.

[31] M S Madhu; Karthikeyan P. R "Detection of Liver disorder using Quadratic Support Vector Machine in comparison with RBF SVM to measure the accuracy, Precision, sensitivity and specificity" 15-16 July 2022

[32] S. Venkata Manoj Kumar Reddy, K. Malathi "an innovative method to enhance the accuracy in the classification of spam detection for youtube comments with using logistic regression over k– nearest neighbor" 15.03.2023