

Content available at: <https://www.ipinnovative.com/open-access-journals>

Journal of Management Research and Analysis

Journal homepage: <https://www.jmra.in/>

Original Research Article

Various prediction methods for finding employee retention rate

Vishal Mahajan^{1,2,*}, P.S. Aithal³, Shailly Mahajan⁴

¹Post-Doctoral Fellow (Srinivas University), Mangaluru, Karnataka, India

²Sardar Beant Singh State University (Applied Chemistry, Humanities and Management), Gurdaspur, Punjab, India

³College of Management & Commerce, Srinivas University, Mangalore, Karnataka, India

⁴Shanti Devi Arya Mahila College, Punjab, India



ARTICLE INFO

Article history:

Received 12-05-2022

Accepted 30-07-2022

Available online 30-08-2022

Keywords:

Prediction

Forecasting

Future Process

Projection framework and Guess

analysis

ABSTRACT

Prediction method are the methods performed on a data set either to foresee the reaction variable worth in view of an indicator variable or to concentrate on the connection between the reaction variable and the indicator factors. With the help of prediction, it will assist any organisation whether it is corporate or academic for the future retention rate of the employees and the grey area on which the organisations has to focus on rectification of employees concerned problems to increase the loyalty rate. Purpose of this paper is to explain the various techniques which are available presently for predicting or forecasting. This paper also explains about various prediction techniques and also provide insights about these techniques which are available for finding employee retention.

This is an Open Access (OA) journal, and articles are distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/), which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprint@ipinnovative.com

1. Introduction

Modern research associated with retention of employee are emerging issues in contemporary corporate world. The chief issue lies in the management of essential workforce is broad and is presumably going to present troubles to different organisations in the near future. Corporate houses are presumably going to conform to the various designs of definitive approach to acting concerning the genuine elements of the continuous work environment, whose accomplishment additionally, life range relies upon creativity, innovativeness, and versatility. Employee retention approach is a complex way of thinking.

The human asset in each company is seen to be the alone parameter for upper hand advantage and the most important resource in each corporate houses. For each company, skilled workers are the spine. Hence, corporate houses

need to address current difficulties by giving preparation of training to the representatives so that they can help to achieve the goals of the corporates.

2. Literature Review

S. Lai, et. al.¹⁻⁸ have proposed the setting based data grouping; RCNN is exceptionally valuable. The execution is best in a few datasets especially on report level datasets. Contingent upon the words utilized in the sentences, loads are allotted to it and are pooled into least, normal and the maximum pools. Here, at the most pooling is applied to separate the catchphrases from the sentences which are generally significant. RNN, CNN and RCNN when contrasted and other conventional strategies, for example, LDA, Tree Kernel and calculated relapse creates high precise outcomes. A. Hassa, et. al.⁹ have proposed RNN for the construction sentence portrayal. This tree like construction catches the semantic of the sentences.

* Corresponding author.

E-mail address: vishal643.vm@gmail.com (V. M.).

The text is investigated word by word by utilizing RNN then the semantic of every one of the past texts are safeguarded in a decent size stowed away layer. For the proposed framework LSTM plays significant job, being a memory stockpiling it holds the characters which helps in foreseeing the following word. J. Y. Lee, et. al.⁷ have suggested that text arrangement is a significant assignment in normal language handling. Many methodologies have been created for arrangement like SVM (Support Vector Machine), Naïve Bayes etc. Normally short message shows up in grouping (sentences in the record) thus utilizing data from going before text might work on the characterization. This paper presented RNN (Recurrent Neural Network) and CNN (Convolutional Neural Network) based model for text characterization. V. Tran, et. al.⁵ have recommended that n-gram is a coterminous grouping of 'n' things from a given succession of text. Assuming the given sentence is 'S', we can build a rundown on n-grams from 'S', by tracking down sets of words that happens close to one another. The model is utilized to infer likelihood of sentences utilizing the chain rule of unqualified likelihood. Z. Shi, et. al.⁴ have characterized that repetitive brain network has information, yield and secret layer. The current secret layer is determined by current info layer and past secret layer. LSTM is a unique Recurrent Neural Network. The rehashing module of common RNN has a basic construction rather LSTM utilizes more complicated capability to trade it for more exact outcome. The vital component in the LSTM is the cell state which is moreover called as covered up layer state. J. Shin, et. al.¹⁰ have characterized that understanding the relevant parts of a sentence is vital while its characterization. This paper mostly centers around it. Different methodologies like SVM, T-LSTM, and CNN have been recently utilized for sentence characterization. However, the proposed C-CNN (Contextual-CNN) gives improved results for example the C-CNN accomplishes condition-of-the-art exactness 52.3% on the fine-grained opinion forecast task and 95.2% on the TREC question arrangement task. W. Yin, et. al.¹¹ have characterized different grouping undertakings are significant for Natural language handling applications. These days CNN are expanding utilized as they can show long reach conditions in sentence, the frameworks utilized are with fixed-sized channels. Yet, the proposed MVCNN approach breaks this boundary furthermore, yields improved results when applied to different datasets: twofold with 89.4%, Sentiment 140 with 88.2% and Subjectivity grouping dataset (Subj.) with 93.9% exactness. Multichannel instatement brings two benefits: 1) Frequent words can have c portrayals to start with (rather than only one), and that would not joke about this has more accessible data to use 2) An uncommon word missed in some implanting renditions can be "made up" by others (we refer to it as "to some extent known word"). I. Sutskever, et. al.¹² have characterized profound

getting the hang of being the freshest innovation in the period has progressed in many fields. One of the procedures called as Deep Neural Networks are extremely strong AI models what's more, have accomplished effective and astounding execution on numerous issues like discourse acknowledgment, visual article location and so on because of its capacity to perform equal calculation for the unassuming no of steps. Many endeavors have been made to resolve the issues with brain organization. The outcomes showed that an enormous profound LSTM with a restricted jargon can outflank a standard SMT-based framework.

2.1. Problem of the study

Finding the methods for prediction of employee retention and their suitability in organisation.

2.2. Objective of the study

Following are the objectives of the study.

1. To find the methods for employee retention.
2. To check the analytical part of each method.

2.3. Tools and techniques for study

There were various prediction techniques. Need of Employee Prediction Method can be explained by following points.

1. *Overseeing labour force:* If the bosses or HR came to be aware of a representative that they will intend to leave the organization then they could reach out to those workers which can assist them with remaining back or they can deal with the labour force by recruiting the new option of those representatives.
2. *Consistent Workforce:* If every one of the representatives in the ongoing task are working consistently on a venture then the pipeline of that undertaking will be smooth yet on the off chance that guess one proficient resource of the project (employee) unexpectedly leave that organization the work process will be not really smooth.
3. *Recruiting Management:* If HR of one specific task came to be familiar with the representative who will leave the organization then he/she can deal with the quantity of employing and they can get the significant resource at whatever point they need so for the proficient progression of work.

Some of the best techniques which has been used for study are:

3. Regression Analysis

Regression analysis is utilized to anticipate a ceaseless objective variable from one or different autonomous

factors. Commonly, regression investigation is utilized with normally happening factors, instead of factors that have been controlled through trial and error.

4. ANOVA

ANOVA, or investigation of fluctuation, is to be utilized when the objective variable is persistent and the reliant variables are category wise. The null hypothesis in this examination is that there is no huge distinction between the various groupings. The populace ought to be typically circulated, the example cases ought to be free of one another, and the variables ought to be roughly equivalent among the gatherings. Below Table 1 shows the example of ANOVA.

5. Linear Regression

Linear regression is to be utilized when the objective variable is nonstop and the dependent variable is constant or a combination of ceaseless and unmitigated, and the connection between the autonomous variable and dependent variables are straight. Besides, all the indicator factors ought to be regularly circulated with consistent fluctuation and ought to exhibit almost no multi collinearity nor autocorrelation with each other. Linear regression calculations of prediction is better explained in Tables 1 and 2 along with Figure 1.

Normal P-P Plot of Regression Standardized Residual

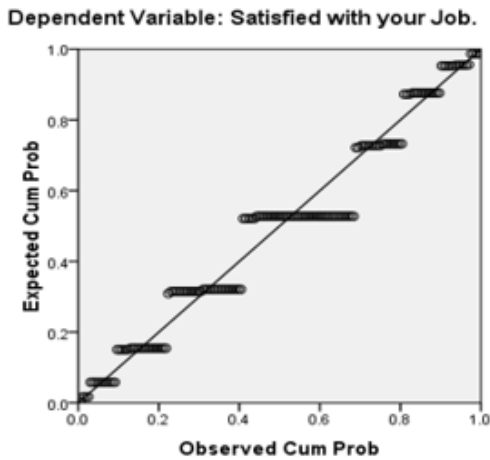


Fig. 1: Logistic regression.

Logistic regression doesn't need a straight connection between the objective and the dependent variables. The objective variable is paired (accepts a value of 0 or 1) or dichotomous. The residuals/errors of a logistic regression need not be regularly distributed and the change of the residuals needn't bother with to be consistent. Nonetheless,

the dependent variables are parallel, the perceptions should be autonomous of one another, there should be almost no multi collinearity nor autocorrelation in the information, and the sample size ought to be enormous. Finally, while this investigation doesn't need the dependent and independent variable(s) to be directly related, the independent variables should be straight connected with the log chances. Logistic regression calculations of prediction is better explained in Tables 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and 13 along with Figure 2.

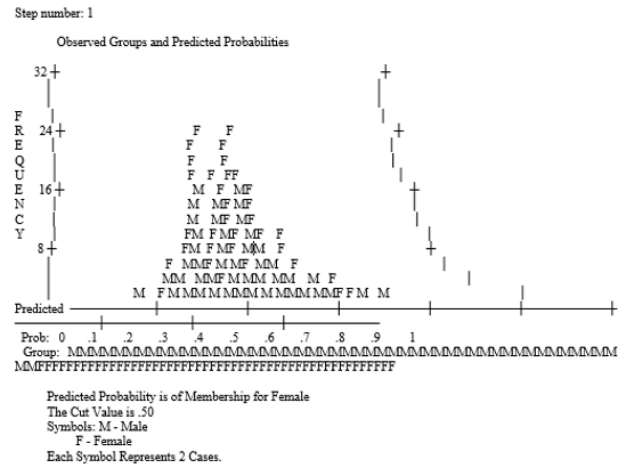


Fig. 2: Logistic regression plot of male and female.

6. Ridge Regression

Ridge regression is a method for breaking down multiple regression variables resultant in multi collinearity. Ridge regression adopts the common least squares strategy, and respects that the residuals experience high differences by adding a level of predisposition to the regression estimates to lessen the standard mistakes. The suspicions follow those of multiple regression, the dissipate plots should be straight, there should be consistent change without any anomalies, and the dependent variables should display freedom.

7. Time Series

Time-series relapse investigation is a strategy for anticipating future reactions in light of reaction history. The information for a period series ought to be a bunch of perceptions on the qualities that a variable requires some investment. The information is bivariate and the independent variable is time. The series should be fixed, they are normally appropriated: the mean and change of the series are consistent throughout extensive period of time. Besides, the residuals ought to likewise be ordinarily circulated with a steady mean and variance throughout a

Table 1: ANOVA for work culture in colleges in relation to anxiety level.

Anxiety Level					
Work culture	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	13.7	5	3.42	3.3	0.01
Within Groups	151.0	145	1.04		
Total	164.7	150			

* significance level ($\alpha \leq 0.05$).

Table 2: Linear regression analysis.

Residuals Statistics					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1.4287	3.4592	2.1800	.54766	150
Residual	-2.45918	2.06365	.00000	.91758	150
Std. Predicted Value	-1.372	2.336	.000	1.000	150
Std. Residual	-2.671	2.241	.000	.997	150

a. Dependent Variable: Satisfied with your Job.

Table 3: Model logistic regression case processing summary.

Unweighted Cases^a			N	Percent
	Included in Analysis	Missing Cases		
Selected Cases		Total	150	100.0
			0	.0
Unselected Cases		Total	150	100.0
			0	.0
Total			150	100.0

a. If weight is in effect, see classification table for the total number of cases.

Table 4: Depend entvariable in logistic regression.

Dependent variable encoding	Original Value	Internal Value
Male		0
Female		1

Table 5: Logistic regression classification table of prediction.

Classification Table^{a,b}		Observed		Predicted		Percentage Correct
		What is your Gender?		What is your Gender?		
Step 0	What is your Gender?	Male	Female	Male	Female	100.0
				82	68	
Overall Percentage						54.7

a. Constant is included in the model.
b. The cut value is .500

Table 6: Variables in logistic regression equation.

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.187	.164	1.303	1	.254	.829

Table 7: Variables not in logistic regression equation variables not in the equation.

		Score	df	Sig.
Step 0	Variables	1.243	1	.265
	Job Stress	.301	1	.583
Overall Statistics		2.223	2	.329

Table 8: Omnibus tests of model coefficients.

Omnibus Tests of Model Coefficients		Chi-square	df	Sig.
Step 1	Step	2.249	2	.325
	Block	2.249	2	.325
	Model	2.249	2	.325

Table 9: Nagelkerke R square.

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	204.386 ^a	.015	.020

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

Table 10: Hosmerand lemeshow test.

Step	Chi-square	df	Sig.
1	9.508	7	.218

Table 11: Contingency table for hosmer and lemeshow test.

Contingency table for hosmer and lemeshow test						
		What is your Gender? = Male		What is your Gender? = Female		Total
		Observed	Expected	Observed	Expected	
Step 1	1	9	9.915	6	5.085	15
	2	8	7.835	5	5.165	13
	3	13	11.216	6	7.784	19
	4	6	7.986	8	6.014	14
	5	13	9.884	5	8.116	18
	6	5	7.978	10	7.022	15
	7	15	11.665	8	11.335	23
	8	8	8.841	10	9.159	18
	9	5	6.679	10	8.321	15

Table 12: Classification table classification table^a

Observed	What is your Gender?	Predicted		Percentage Correct
		Male	Female	
Step 1	Male	69	13	84.1
	Female	48	20	29.4
Overall Percentage				59.3

a. The cut value is .500

Table 13: Variables in the equation.

Variables in the Equation							
			S. E.	Wald	df	Sig.	Exp (B)
Step 1a	Job	-.232	.168	1.902	1	.168	.793
	Stress	.168	.169	.987	1	.321	1.183
	Constant	-.069	.457	.023	1	.880	.933

a. Variable(s) entered on step 1: Job, Stress.

Table 14: Time series model Fit between job evaluation, personnel policies and job satisfaction model fit.

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile						
					5	10	25	50	75	90	95
Stationary R-squared	.207	.	.207	.207	.207	.207	.207	.207	.207	.207	.207
R-squared	.254	.	.254	.254	.254	.254	.254	.254	.254	.254	.254
RMSE	.933	.	.933	.933	.933	.933	.933	.933	.933	.933	.933
MAPE	41.891	.	41.891	41.891	41.891	41.891	41.891	41.891	41.891	41.891	41.891
MaxAPE	191.250	.	191.250	191.250	191.250	191.250	191.250	191.250	191.250	191.250	191.250
MAE	.716	.	.716	.716	.716	.716	.716	.716	.716	.716	.716
MaxAE	2.865	.	2.865	2.865	2.865	2.865	2.865	2.865	2.865	2.865	2.865
Normalized BIC	-	.	-.006	-.006	-.006	-.006	-.006	-.006	-.006	-.006	-.006

Table 15: Decision tree model of job satisfaction dependent and independent variables model summary.

Specifications	Growing Method	CHAID
		Dependent Variable Independent Variables
Results	Validation	None
	Maximum Tree Depth	3
	Minimum Cases in Parent Node	100
	Minimum Cases in Child Node	50
	Independent Variables Included	College/University following the UGC and AICTE norms.
	Number of Nodes	3
	Number of Terminal Nodes	2
	Depth	1

Table 16: Risk model of job satisfaction risk.

Estimate	Std. Error
.413	.040
Growing method: CHAID dependent variable: Satisfied with your Job.	

Table 17: Chaid growing method of satisfaction with job classification.

Observed	Predicted					Percent Correct
	Strongly Satisfied.	Satisfied Some what.	Neutral	Dissatisfied Some what.	Strongly Dissatisfied.	
Strongly Satisfied.	37	6	0	0	0	86.0%
Satisfied Somewhat.	11	51	0	0	0	82.3%
Neutral	2	24	0	0	0	.0%
Dissatisfied Somewhat.	3	10	0	0	0	.0%
Strongly Dissatisfied.	0	6	0	0	0	.0%
Overall Percentage	35.3%	64.7%	.0%	.0%	.0%	58.7%

Growing Method: CHAID Dependent Variable: Satisfied with your Job.

Table 18: Network information between job evaluation, personnel policies and job satisfaction network information.

Input Layer	Factors	1	Personnel policies of college.
	Covariates	1	Job Evaluation System of Faculty.
Hidden Layer(s)	Number of Units ^a	6	Standardized
	Rescaling Method for Covariates	1	
	Number of Hidden Layers	1	
Output Layer	Number of Units in Hidden Layer 1 ^a	3	Hyperbolic tangent
	Activation Function	1	Satisfied with your Job.
	Dependent Variables	5	Softmax
	Number of Units	1	Cross-entropy
	Activation Function		
	Error Function		

a. Excluding the bias unit

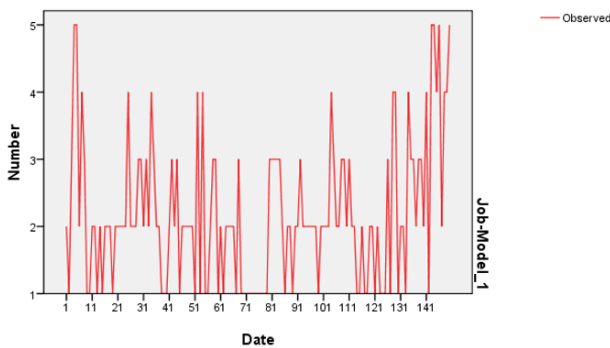


Fig. 3: Time series job satisfaction model.



Fig. 4: Decision tree model of job satisfaction.

significant period of time, as well as uncorrelated. The series shouldn't contain any exceptions. Assuming irregular shocks are available, they ought to be sure be randomly dispersed with a mean of 0 and a steady variance. Table 14 along with Figure 3 elaborate the prediction calculation of time series.

8. Decision Trees

Decision trees are a sort of oversight learning algorithm calculation which more than once parts the sample in view of specific inquiries concerning the samples.

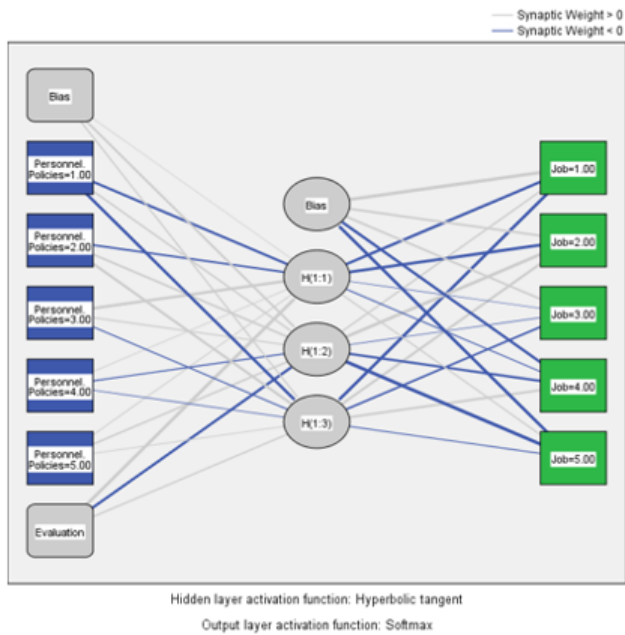


Fig. 5: Neural network between personnel policies and job satisfaction.

These are extremely helpful for issues related with classification problems. They are somewhat straightforward and exceptionally viable. Decision trees address a few choices followed by various possibilities of event. This procedure assists us with characterizing the main variables and the connection between at least two variables. Tables 15, 16 and 17 along with Figure 4 explains the calculations of decision tree model for prediction analysis.

9. Conclusion

Considering everything, these are just a little pack of the decisions of different insightful techniques that can be used to show data. It should be seen that making causal associations between factors when using judicious assessment procedures is very unsafe. We can't communicate that one variable caused one more in farsighted examination, rather, we can communicate that a variable impacted another and what that effect was.

10. Source of Funding

None.

11. Conflict of Interest

None.

References

- Lai S, Xu L, Liu K, Zhao J. Recurrent Convolutional Neural Networks for Text Classification. *Proceedings of the Twenty-Ninth AAAI Conference on AI*. 2015;p. 1137–55.
- Ongsulee P. Artificial Intelligence, Machine Learning and Deep Learning. *15th International Conference on ICT and Knowledge Engineering*. 2017;p. 1–7.
- Yin W, Kann K, Yu M, Schütze H. Comparative study of CNN and RNN for Natural Language Processing. *Comput Sci*. 2017;1:7.
- Shi Z, Shi M, Li C. The prediction of character based on Recurrent Neural network language model. *IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*. 2017;p. 1–4.
- Tran V, Nguyen K, Bui D. A Vietnamese Language Model Based on Recurrent Neural Network. *Eighth International Conference on Knowledge and Systems Engineering*. 2016;.
- Arnold KC, Gajos KZ, Kalai AT. On Suggesting Phrases vs. Predicting Words for Mobile Text Composition; 2016.
- Lee J, Derroncourt F. Sequential Short-Text Classification with Recurrent and Convolutional Neural Networks; 2016.
- Liang M, Hu X. Recurrent Convolutional Neural Network for Object Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015;.
- Hassan A, Mahmood A. Deep Learning for Sentence Classification. *Appl Technol Conf*. 2021; Available from: <https://towardsdatascience.com/deep-learning-techniques-for-text-classification-78d9dc40bf7c>.
- Shin J, Kim Y, Yoon S. Contextual CNN: A Novel Architecture Capturing Unified Meaning for Sentence Classification. *IEEE International Conference on Big Data and Smart Computing (BigComp)*. 2018;p. 15–7.
- Yin W, Schütze H. Multichannel Variable-Size Convolution for Sentence Classification. *19th Conference on Computational Language Learning*. 2015;1:1603–45.
- Sutskever I, Vinyals O, Le QV. A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification; 2016.

Author biography

Vishal Mahajan, Assistant Professor in (Sardar Beant Singh State University)

P.S. Aithal, Professor

Shailly Mahajan, Assistant Professor

Cite this article: VM, Aithal PS, Mahajan S. Various prediction methods for finding employee retention rate. *J Manag Res Anal* 2022;9(3):122-129.