ISSN PRINT 2319 1775 Online 2320 7876
 Volume 11, Sp.Iss 7 , 2022

DEVELOPING A MODEL TO ESTIMATE RETENTION PERIOD OF EMPLOYEE USING PREDICTIVE ANALYTICS

Dr. Sundari Dadhabai[1]K L Amrutha Varshini [2]

1. Associate Professor, Department of MBA, Koneru Lakshmaiah Education Foundation (KLEF), Deemed to be University, Vaddeswaram, Green fields, Guntur, Andhra Pradesh, India -522302. Mail ID: sundaridadhabai@gmail.com.Ph. 9059849965.
2. MBA Student,,Department of MBA,,Koneru Lakshmaiah Education Foundation (KLEF), Deemed to be University, Vaddeswaram, Green fields, Guntur, Andhra Pradesh, India - 522302, Mail ID: amruthakota99@gmail.com. Ph: 9346661555

DOI : 10.48047/IJFANS/11/S7/023

ABSTRACT

People analytics is the practise of analysing human behaviour in the workplace using data and analytical methodologies. Employee retention is one area where people analytics may be quite valuable. Organisations may acquire a better understanding of why employees leave, which individuals are at risk of leaving, and what variables contribute to high employee turnover by analysing data on employee retention. This data can assist organisations in making better educated decisions on how to boost retention and minimise turnover.

In this study, researches made an attempt to develop a methodology for estimating employee retention periods using predictive analysis. The model predicts the probability of an employee adhering with the organisation for a set amount of time based on historical data on numerous employee-related characteristics such as job role, performance evaluations, previous working years, and job satisfaction. In order to develop the retention estimation model, we use machine learning techniques such as support vector machine and random forest algorithms. The model is trained using employee data from an IBM dataset.Further research could possibly be performed to develop and evaluate the model utilising real-world data from organisations to validate its efficacy in various circumstances.

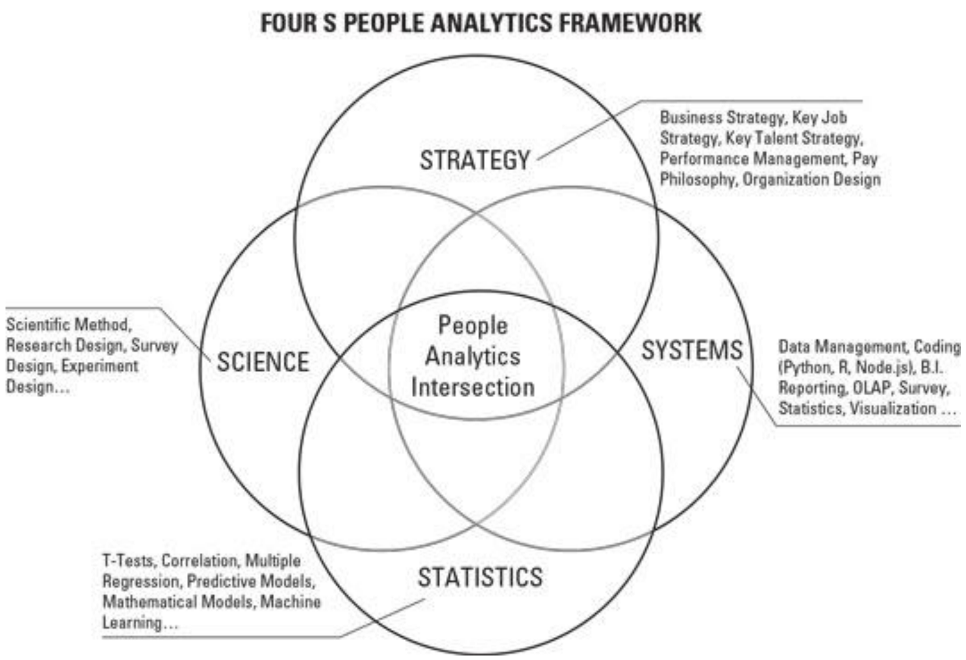
Key Words:

Employee Retention, Predictive analytic, Model Validation, Random Forest, Support Vector Machine

INTRODUCTION

HR Analytics:

People analytics have evolved from the use of basic human resource metrics in the 1980s and 1990s to the exposure of sophisticated human resource information systems in the 2000s. In the early 2010s, the term "people analytics" emerged in order to describe the application of sophisticated analytics and data science methodologies to support HR decision-making.In recent years, there has been a rising emphasis on employing predictive analytics to forecast future employees trends and detect potential challenges before they take place People analytics is increasingly viewed as a vital component of company strategy, affecting larger decisions about organisational design, talent management, and workforce planning. Advances in technology, a rising realisation of the relevance of data-driven decision-making, and a focus on aligning HR practises with larger business goals have all propelled the rise of people analytics.



By analysing multiple data points linked to employee behaviour, work performance, and job satisfaction, People Analytics may be used to forecast and increase employee retention. People Analytics can discover the characteristics that contribute to employee retention and anticipate which employees are likely to depart using data analytic techniques such as regression analysis, machine learning, and predictive modelling.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 7 , 2022**

This data may assist organisations in taking proactive actions to retain employees, such as providing training and development opportunities, enhancing salary and benefits, and responding to employee issues and comments. Organisations may develop a more stable and engaged workforce by using People Analytics to boost employee retention, leading to higher productivity and overall company success.

Machine Learning using Regression Analysis:

Machine learning using regression analysis involves building models to predict continuous numerical values based on the relationships between independent variables (predictors) and a dependent variable (target).Once the model is trained, it can be used to make predictions on new, unseen data. The performance of the regression model can be evaluated using metrics such as mean squared error (MSE), mean absolute error (MAE), or R-squared, which quantify the accuracy and goodness of fit.

For this research, the researcher has used 4 types of regression analysis as follows:

1. **Support Vector Machine (SVM):** SVM handles non-linear relationships through the use of kernel functions. It is effective in high-dimensional spaces and can handle large feature sets. It is less prone to overfitting when appropriate regularization is applied.
2. **Random Forest (RF):** RF is an ensemble method that combines multiple decision trees. It can handle non-linear relationships, interactions, and feature importance assessment. RF is robust against overfitting, requires minimal data preprocessing, and can handle high-dimensional datasets.
3. **k-Nearest Neighbors (kNN):** kNN is a non-parametric algorithm that can capture complex relationships without making strong assumptions. It is simple to understand and implement, and it can handle both regression and classification tasks. kNN is effective when there is a clear local structure or clusters in the data.
4. **Linear Regression (LR):** Linear regression is a simple and interpretable model that provides insight into the relationships between predictors and the target variable. It is computationally efficient and can handle large datasets. Linear regression is well-suited when there is a linear or near-linear relationship between predictors and the target.

LITERATURE REVIEW

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
 Volume 11, Sp.Iss 7 , 2022

İrem ERSÖZ KAYA (2021)resulting in balanced prediction performance. The accuracy level achieved in the blind test was 80% using RBF and SVM techniques. The author **Semu Bacha (2016)**examine several research and identify the elements that lead to employee turnover, investigate the repercussions of staff attrition and grasp of the complexities of employee attrition and its possible consequences.**Ravishankar S Ulle (2018)** suggests that factors contributing to attrition include job dissatisfaction, poor management, and inadequate recognition or appreciation. **Kotresh Patil and Dr. T P RenukaMurthy**develop effective strategies for employee retention to improve employee retention rates and reduce attrition. **SaswatBarpanda and Athira S**used a triangulation approach that combined data from surveys, interviews, and secondary sources to provide a comprehensive understanding of the causes of attrition.**Rahul Yedida (2018)** says employee turnover can have a significant impact on organizational performance, and predicting employee attrition through various machine learning and statistical modeling techniques that can help organizations take proactive measures to prevent it.**Guru Vignesh Sridharand Sarojini Venugopal**, the challenges associated with employee attrition and retention, several strategies to improve employee retention.**Dr B.Latha Lavanya (2017)**finds the effective HRM practices and importance of understanding the underlying causes of attrition to develop targeted interventions that address the specific needs and preferences of employees.**Kishori Singh, Reetu Singh (2019)** the study identifies several factors contributing to attrition and these factors through effective HRM practices.**Salah Al-Darraji, Dhafer G. Honi (2021)**develop a model that DNNs can effectively predict employee turnover by considering various factors and can also provide valuable insights to organizations.**Krishna Kumar Mohbey (2020)** various ML algorithms have been used to predict employee turnover by considering various factors such as employee demographics, performance metrics, and job-related factors.**Alao D. & Adeyemo A. B. (2013)**explores the importance of considering the unique needs and preferences of employees when developing retention strategies based on decision tree analysis.**I Setiawan, S Suprihanto (2020)** utilizes logistic regression to analyze various employee-related variablesand develops a predictive model for identifying employees at risk of leaving.**Dr. R. S. Kamath and Dr. P. G. Naik (2019)** explores various ML algorithms to predict and understand factors contributing to employee turnover and aims to develop accurate predictive models for identifying employees at risk of attrition.**Ali Raza, Kashif Munir, Mubarak Almutairi (2022)**examines the effectiveness of various ML approaches in forecasting employee turnover and aims to identify the most accurate and reliable algorithm for predicting employee attrition.

IJFANS International Journal of Food And Nutritional Sciences
Official Publication of International Association of Food and Nutrition Scientists

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 7 , 2022**

MOTIVATION BEHIND THE STUDY

Various factors for Employee Attrition and challenges faced by organizations for retaining the employees and their strategies implemented to retain the employees. Antecedents and consequences of Employee Attrition, in addition predicting the employee turnover based on signification factor.Existing literature focused on only factors which are leading for attrition in different sectors and few articles focused on challenges and causes.Every few research works made on attempt to predict attrition with the help of most influencing factors but there's no research exists on prediction of Employee Attrition based on algorithm and equation.

OBJECTIVES

The present research was carried out with the following objectives

- To know / identify influencing factors at IBM for Attrition
- To identify which test is more accurate in prediction.
- To develop a model based on selected factor for predicting Employee Attrition

RESEARCH METHODOLOGY

The methodology for analysing the variables to predict the employee retention in years at the organization.IBM HR dataset which is secondary data with 1470 observations and 35 variables.

DATA ANALYSIS:

Among 35 variables, the researcherhave considered 13 variables which the researcher considered as relevant for retention were selected and verified the literature support. Those variables are as follows:

1. Age
2. Department
3. Education Field
4. Gender
5. Job Role
6. Job Satisfaction
7. Monthly Income
8. Num Companies Worked

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 7 , 2022**

9. Percent Salary Hike
10. Performance Rating
11. Relationship Satisfaction
12. Total Working Years
13. Years At Company

From the above mentioned variables, taken Years At Company as Dependent variable and others are Independent variables

1. Feature Selection: In order to develop a proper model, the researcher used Recursive Feature Elimination (RFE) method, based on which 5 variables were identified as the most significant variables that influences and predicts the employee retention in years at the company as follows:

1. Monthly Income
2. Past Working Years
3. Age
4. Relationship Satisfaction
5. Job Satisfaction

2. Performed different tests on the selected variables such as Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbours (KNN), Linear Regression (LR): As dependent variable (YearsAtCompany) is continous, the researcher performs regression analysis on these tests used to know the employee retaining at the company with the independent factors like monthly income of the employee, Age, Past Working Years, Relationship satisfaction and job satisfaction.

Support Vector Machine: The SVM model's performance based on the provided metrics is moderate. The MAE and RMSE values indicate that the average prediction error is around 2.37 units, and the R2 value suggests that the model explains approximately 55.45% of the variance in the target variable.

Random Forest: The random forest model achieved moderate performance. It demonstrates a relatively low mean squared error and explains a significant portion of the variance.

KNN: The model seems to have decent performance based on the provided metrics. The MSE and RMSE values indicate that the average prediction error is around 2.95 units, while the R2

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **© 2012 IJFANS. All Rights Reserved**, **Volume 11, Sp.Iss 7 , 2022**

value of 0.6808788 suggests that the model explains approximately 68.09% of the variance in the target variable.

Decision Tree: As R2 value should be always in positive value, but the received value got after predicting the decision tree model it is negative. Thus it says the decision tree test cannot be performed for thisdataset.

Linear Regression: The linear regression model has limited predictive power in explaining the "YearsAtCompany" variable. The R-squared value suggests that only about 32% of the variability, the MAPE value of approximately 70.22% suggests that the model's predictions have a high percentage error.

Identification of test: Among the above mentioned tests for the variables considered, the accuracy results are identified based on the MAPE (Mean Absolute Percentage Error) value. For this purpose, a comparative table of all the tests with their MAPE, RMSE, R2 values etc., was generated as given below.

Table showing comparative values of all the tests conducted:

Test	Train dataset				Test dataset			
	MSE	RMSE	MAPE	R2	MSE	RMSE	MAPE	R2
SVM	NA	3.357788	46.92808	0.554525	11.27474	3.357788	46.92808	0.554525
RF	9.800518	3.130578	37.26063	0.2688832	9.800518	3.130578	37.26063	0.268883
KNN	8.736958	2.955835	38.84792	0.6808788	8.736958	2.955835	38.84792	0.680879
LR	15.1316	3.889936	70.21972	0.3237932	15.1316	3.889936	70.21972	0.323793

Based on the MAPE value, Random Forest gives the more accuracy compared to others models performed. The less MAPE value, better the model in terms of accuracy.

3.Developing a model on selected factors: Developed a model with the considered factors such as Age, Monthly Income, Job Satisfaction, Relationship Satisfaction and Past Working Years taken, Random Forest test gives high accuracy and the model as follows

YearsAtCompany = Age+,JobSatisfaction+MI+RelationshipSatisfaction+PastWorkingYears.

- Testing using the actual data from dataset with the equation above mentioned as, Age is 45, JobSatisfaction is 2, MI (Monthly Income) is 5.13, RelationshipSatisfaction is 4 and 0 for PastworkingYears. Therefore, the model predicts the employee can retain for 9.5008 years where the actual value in the dataset is 10 years.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Sp.Iss 7 , 2022

As the result is very close to the actual data, it can be concluded as that this model is able to predict with accuracy.

CONCLUSION:

Developing a predictive analytical model for estimating employee retention in years at the organization offers a valuable tool for effective workforce management. By considering factors such as age, monthly income, job satisfaction, past working years, and relationship satisfaction with the company, the model can accurately predict how long an employee is likely to stay with the organization. This information can help organizations identify employees at risk of leaving and develop targeted retention strategies to mitigate turnover. Additionally, the model can provide insights into the factors that drive employee engagement and satisfaction, enabling organizations to make data-driven decisions to improve the overall employee experience. Thus, it helps in reduced costs, creates more stable and productive work environment.

REFERENCES:

1. Alao D, A. A. (2013). ANALYZING EMPLOYEE ATTRITION USING DECISION TREE ALGORITHMS. *Computing, Information Systems & Development Informatics* .
2. Bacha, S. (2016). Antecedents and Consequences of Employee Attrition: A Review of Literature. *Reserch Gate* .
3. Dr. R. S. Kamath, D. S. (2019). Machine Learning Approach for Employee Attrition Analysis. *IJTSRD* .
4. Guru Vignesh Sridhar, S. V. (2018). Employee Attrition and Employee Retention-Challenges & Suggestions. *Research gate* .
5. I Setiawan, S. S. (2020). HR analytics: Employee attrition analysis using logistic regression. *IOP Publishing Ltd* .
6. İrem ERSÖZ KAYA, O. K. (2021). Machine Learning Approach for Predicting Employee Attrition and Factors Leading to Attrition. *Cukurova University Journal of the Faculty of Engineering* .
7. Kishori Singh, R. S. (2019). A Study on Employee Attrition: Effects and Causes. *International Journal of Research in Engineering, Science and Management* .
8. Lavanya, D. (2017). A Study on Employee Attrition: Inevitable yet Manageable. *International Journal of Business and Management Invention* .

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **© 2012 IJFANS. All Rights Reserved**, **Volume 11, Sp.Iss 7 , 2022**

9. Mohbey, K. K. (2020). Employee's Attrition Prediction Using Machine Learning Approaches. *Research gate* .
10. Norsuhada Mansor, N. S. (2021). Machine Learning for Predicting Employee Attrition. *IJACSA*.
11. Rahul Yedida, R. V. (2018). Employee Attrition Prediction. *Research gate* .
12. Rupesh Khare, D. K. (2011). Employee Attrition Risk Assessment using Logistic Regression Analysis. *Academia* .
13. Salah Al-Darraji, D. G. (2021). Employee Attrition Prediction Using Deep Neural Networks. *MDPI* .
14. Ulle, R. S. (2018). Rational of Employee Attrition and Strategies to Employee Retention. *Reserch Gate* .