**SVKM’S NMIMS**

**MUKESH PATEL SCHOOL OF TECHNOLOGY MANAGEMENT & ENGINEERING**

**HR ANALYTICS: A PARAMETRIC STUDY FOR STATISTICAL ANALYSIS AND PREDICTION OF ATTRITION**

**A Research Project submitted in partial fulfilment**

**Of the requirements for the degree of**

**MBA (Tech)**

**By:**

**Akash Chattopadhyay J006**

**Rishu Chourey J007**

**Simran Jain J010**

**Komal Mistry J017**

**Under the supervision**

**of**

**Prof. Sudarshana Sarkar**

****

**Abstract**

The human resources industry has long been enslaved by traditional methods and biased judgements led by intuition. Since human capital is one of the most important assets of an organization, structured and well formulated processes are required to retain and maintain resources profitably. Increasingly HR analytics is being used to aid in decision making. HR or People analytics also helps in predicting various workforce parameters which can be of great help while forming the strategy for the organization. Various aspects of Human Resource Management can be streamlined by the help of HR analytics i.e. Talent Acquisition, Absenteeism, Turnover, Employee relations, training etc. This report analyses the workforce parameter of employee churn or attrition. Using statistical techniques and various machine learning models the prediction of attrition will be analysed and compared.

**ACKNOWLEDGEMENT**

We would like to express a deep sense of gratitude to the dean of NMIMS MPSTME, Dr. N.T. Rao for providing us with opportunity to work on this research project.

We take this opportunity to express our regards to our faculty guide Ms. Sudarshana Sarkar for her exemplary guidance, monitoring and constant encouragement.

We’d also like to thank all the faculties for their cordial support, valuable information and guidance, which helped us in completing this project through various stages.

**INDEX**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  | | --- | --- | | **Topic** | **Page no.** | | Abstract | 2 | | Acknowledgement | 3 | | Literature Review/Secondary Research | 5 | | Research Objective | 4 | | Methodology | 4 | | Data Analysis | 5 | | Correlational Analysis | 6 | | Overall Results | 9 | | Descriptive Analysis | 10 | | Predictive Analysis | 13 | | Conclusion | 14 | | Limitations of the study | 17 | | Scope for further research | 17 | | References | 18 | | Appendix | 19 | |  |

**Literature Review / Secondary Research**

Human resource is one the most critical capital asset owned by an organization. Employee salaries make up close to half of many organizations’ operating expenses and can be even higher in some industries such as financial services, so the contribution of the workforce to organization success is perhaps the most important aspect of competitive advantage. (Oracle whitepaper 2011). The performance and training of this resource is thus crucial to the organization’s growth. Busy HR managers often rely upon traditional practices and intuition to make critical choices affecting the future of their firms. According to Rousseau, Evidence based HR decision making is crucial to avoid faulty practices. EBHR (Evidence Based HR) is a decision-making process which combines critical thinking with the use of available scientific evidence and business information. The research points out four sources of information that need to be tapped to make unbiased decisions: use of available scientific evidence, reliable and valid organisational facts, metrics and assessments, practitioner reflection and opinion and the perspective and interests of affected stakeholders (Rousseau and Barends, 2011).

EBHR as theorized by Rousseau, exhorts that the essence of such data-based decision making is to approach risk and uncertainty in a controlled fashion instead of relying on pure intuition. The HRM team thus need to develop a critical thinking approach necessary for the development of a questioning mindset. EBHR also encourages managers to make more explicit decisions which consider the opportunities available and evaluate each before concluding. Thus, we see that research lays emphasis on having a more data based decision making approach over the traditional HR practices that may be considerably tampered by inherent and unavoidable biases. The need for critical thinking and scientific analysis of available evidence is the need of the day.

Evidence based HRM is thought to be a family of HR practices which primarily combine research evidence with contextual information. Evidence based management provides tools and a common language that helps both the analytics-people and the people-people make better use of data in decision making. (Reddy and Lakshmikeerthi, 2017). In their research Reddy defines HR analytics as an EBHRM tool, which helps communicate data from disparate sources such as surveys, records, and operations to help make a cohesive background to aid in decision making.

It is important to note that HR analytics and HR metrics are not the same things. HR managers use several metrics to measure HR data such as turnover, absenteeism, sick leaves etc. HR analytics is used to gauge employee engagement and retention. Where HR metrics reflect past data, HR analytics helps predict future data. The reasons why data must be gathered are to: Describe, Predict, Explain, Optimise HR metrics and processes. These reasons must be kept in mind before the laying down the structure for HR analytics and implementation (Reddy and Lakshmikeerthi,2017).

HR metrics is more of an inside-outside perspective used for data reporting, while HR analytics is more of an outside-inside approach which studies the insights obtained from the data (Johannink,2015). Different levels of analytics have been identified: Descriptive, Predictive and Prescriptive. Descriptive analytics only gathers and puts forth the data over past trends. Predictive analytics focuses on the factors causing these trends, while Prescriptive analytics is used mainly for fixing issues; using prescriptive analytic techniques models can be created to understand alternative investments in employee training can affect firm’s revenue (King, 2016).

Predictive analytics finds applications in recruitment screening, selection testing, job performance, turnover management and so on. Talent analytics thus is the most interesting avenue for HR analytics. It can help organizations determine the skill sets and personal traits to look for in candidate to hire effectively. Employers need to need to identify and prioritize the characteristics of top performers and conduct psychometric tests which assess these traits. Companies need to apply analytics to assess the links between demographic info, performance ratings, educational background, career progression etc to determine the behaviour and characteristics of their human resource to manage it more effectively (Kaur and Raj,2017). According to Davenport, Harris, and Shapiro (2010), studies of attrition are one of the six ways that HR analytics is most frequently applied. Studies of attrition are popular as much of the information needed for these analyses is already held by the organization in a HR database (e.g., information about an employee’s date of hire, performance reviews, promotion history, etc.). This abundance of organizational data is further complemented by a number of studies on the topic of attrition that address theories about why employees choose to stay or leave.

The sheer volume, speed and availability of data has intensified the need for analytical decision making. The rising trend in the global marketplace is the use of talent analytics and big data to achieve competitive growth. The 3 factors which influence the use of HR analytics and big data in organizations today are: Silos, Skills and Suspicion.

Silos refers to the structural issues related to business unit hierarchy and system obstacles which arise due to incompatible technology and other technical shortcomings. Skills, this a crucial part of the puzzle, the right set of skills can galvanize the growth of insights by analytics and at the same time analysis performed by incompetent and unskilled employees can result in total wastage of HR budget and can seriously affect HR policies. Thus, the top management must first determine whether they would rely on in-house teams for analytics or take the help of management consultants. Suspicion, signifies the current perspective that most HR personnel are sceptic about decision making based primarily upon data and not human perception. This approach essentially dehumanizes the sanctum of traditional HR practice and poses a severe threat to the existing workforce which is still heavily reliant on these practices (Oracle, CIPD, 2013).

HR analytics needs to follow the right direction and must be focused upon improving efficiency and providing insights. It is suggested that HR analytics must start with a business challenge in mind rather than a definite dataset and key problem areas and issues. The beauty of analytics is that it helps find hidden knowledge from the existing dataset which can generate valuable insights and help establish relations between many independent factors. This relational linkage information can thus help an organisation to better understand the organization culture and the employee behaviour. This knowledge in turn helps effectively manage teams and set goals. The final goal of HR analytics is to discover the linkages that the changes and behaviour of human capital has upon the major business functions and the overall revenue of the organisation. The goal of HR analytics should then be to transcend from the HR mindset and to help cross functional teams and units unify their goal to achieve strategic harmony (Rasmussen and Ulrich,2015).

Numerous organizations are jumping onto the HR analytics bandwagon but only a handful have achieved a significant level of proficiency enough to help establish themselves amongst their competitors. A Harvard Business Review article by Davenport mentions that companies like Amazon, Red Sox, Google and Marriot have tactfully leveraged their HR analytics competency and established themselves superior amongst their competitors. Thus, it can be inferred that HR analytics if used and leveraged well can help an organization develop a competitive edge in the market.

**Research Objective**

1. To determine from the data collected the significant factors causing attrition.
2. To predict the attrition phenomenon of individuals from machine learning algorithm predictions.

**Methodology**

The research methodology will rely heavily on the use of statistical analysis for preliminary parametric evaluation. The predictions for attrition will then be made by various machine learning models. The respondents chosen here are the workforce of IBM Ltd. The company has provided a sample of its database for scientific analysis and research. The data provided by IBM will be the baseline for the research and various parameters collected from it, will be analysed for the course of the research process. IBM under its WATSON analytics initiative has provided a large dataset which has been curated and made open source by the in-house data analysts. The data set provided has 1470 unique entries curated upon 20 distinct employee parameters.

The research methodology will be to list down the method that can be used to acquire this data first hand, i.e. by the way for an employee survey. The information can be collected by a questionnaire that collects information related to every individual’s job design, role, salary structure, and satisfaction with the work culture.

The data collected will first be structured into normalized numeric and codified parameters. The resulting dataset is used for statistical testing. The parameters established will first be evaluated by a correlational analysis, to determine the significance of each attribute on the probability of attrition.

Initially descriptive analytics along with the results of correlational analysis will help to visualise data and in understanding the apparent behaviour of both dependent and independent variables.

For the predictive analysis, the dataset will be sectioned into training and testing sets. The training set will be used exclusively for building data model and the smaller test set will be used to evaluate accuracy of prediction. Since the principles of supervised machine learning are to be employed, the attrition variable will be modelled upon. The attrition variable is a simple classification outcome for whether the employee has resigned or is still retained in the organization.

The data is then modelled using various classification machine learning algorithms like Logistic Regression, Support Vector Machines, Decision Trees, Random Forest and Naïve Bayes algorithms. The prediction results obtained from each model will be compared and evaluated based on the following parameters:

1. Accuracy
2. Classification Error Percentage
3. Recall / Sensitivity
4. Precision of prediction
5. False Positive Rate / Specificity

The key problem areas will be highlighted by the mathematical evaluation of model parameter weights. This information will later be used to build recommendations on how the organisation can improve employee retention and in turn reduce the attrition rate.

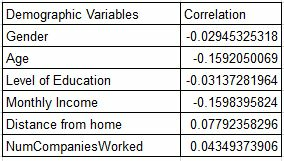
**Data Analysis**

The data collected from the database contains information upon the following independent variables:

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Demographic Variables** | **Job Description Variables** | **Job Satisfaction Variables** |
| 1. | Gender | Job Level | Environment Satisfaction |
| 2. | Level of Education | Overtime | Job Involvement |
| 3. | Monthly Income | Percentage salary hike | Job Satisfaction |
| 4. | Distance from home | Stock Option Level | Relationship Satisfaction |
| 5. | Number of companies worked for | Years at company | Work Life Balance |
| 6. |  | Years in current role |  |
| 7. |  | Years since last promotion |  |
| 8. |  | Years with current manager |  |

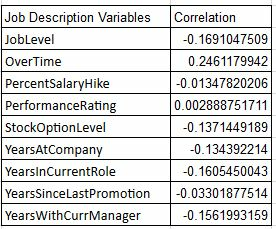
**Correlational Analysis**

The following tables show the results of correlational analysis of each independent variable with respect to dependent variable attrition. Since we have categorised attrition variable, the correlation will determine the effect of independent variable leading to employee turnover.



The overall correlation coefficient of demographic variables is -0.043.

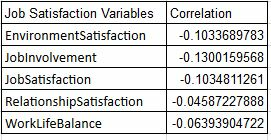
* Age of employee is strongly negatively correlated with attrition i.e. increase in age of employee is accompanied by decrease in the attrition phenomenon. Thus, older employees are more stable while their younger colleagues are comparatively difficult to retain.
* Monthly income is also strongly negatively correlated with attrition, thus employees with higher salaries are less likely to leave the organisation.
* Level of education is negatively correlated but has a less significant impact on the attrition phenomenon, indicating that a higher level of education doesn’t significantly reduce the chances of attrition.
* Distance from home has a positive correlation, therefore greater the distance higher the chances of attrition. But since the correlation constant is comparatively smaller its overall impact is weak.
* Number of companies that employee has worked for also has positive correlation, indicating that the employees who have a history of working for multiple companies are at a higher risk of attrition as compared to fresh recruitments. But since the correlation coefficient is small, the impact as compared to other factors is weak.



The overall correlation coefficient of job description variables is: -0.061

* The Overtime variable is positively correlated, it has the strongest correlation to attrition as compared to all other variables in the entire dataset. This indicates that overtime is one of the major factors that is causing attrition in the organisation.
* The performance rating variable is also positively correlated to the attrition variable, but it is a very weak relationship. This indicates that although higher performance rating decreases the chances of attrition of employee, it barely has any significance. Thus, it can be estimated that the performance rating doesn’t positively reinforce employees and company stands a risk of losing high performers.
* Job level and stock option level are negatively correlated to attrition, thus employees with high job level and high stock options are less likely to leave the organisation. This indicates that stock options provided to high job level employees make them stable in organisation to some extent.
* The years at company, current role, since last promotion and with current manager show overall development of the employee’s job role. All the variables are negatively correlated showing that increase in the no. of years a person has spent in company reduces the risk of attrition.

At the same time comparing the factors we see that no. of years employees has spent in the same role contributes to lower attrition risk as compared to the no. of years with current manager and with the company. The years since last promotion has the weakest correlation.



The overall correlation of job satisfaction variables is: -0.0872

* The job involvement variable has the strongest correlation with attrition. Indicating that higher levels of job involvement reduce the risk of attrition as compared to other variables in the set.
* Environment satisfaction and overall job satisfaction come second, indicating that after job involvement, these two factors are the contributors to reducing risk of attrition.
* Work life balance is also negatively correlated, indicating that employees who perceive themselves to have high work life balance levels are less likely to leave.
* The last factor is relationship satisfaction, it is negatively correlated so increased levels of relationship satisfaction with colleagues leads to lower risk of attrition. But the weak correlation indicates that relationship satisfaction contributes the least to minimizing the risk of attrition.

**Overall results:**

* The Job satisfaction variables have a higher collective impact, followed by job description variables and demographic variables have the least impact out of all.
* The increases in the following variables are significantly decreasing risk of attrition:
* Job level (-0.169)
* Years in current role (-0.160)
* Years with current manager (-0.159)
* Monthly Income (-0.159)
* The increase in the following variables are significantly increasing the risk of attrition:
* Overtime (+0.246)
* The following variables have little or no significance as compared to other variables on the risk of attrition:
* Performance rating (+0.002)

**Descriptive Analysis**

The following graphical representations give a descriptive view of the entire dataset.

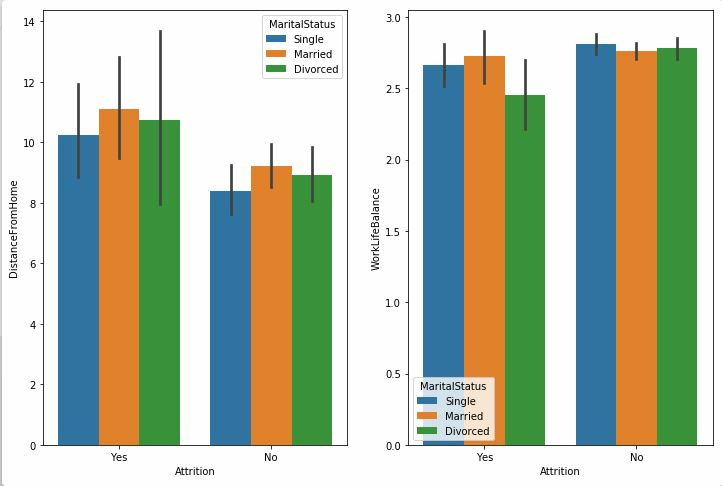


Fig 1.a Histogram showing relationship between marital status and distance from home wrt attrition.

b. Histogram showing relationship between marital status and work life balance wrt attrition.

The graph (a) above shows that married employees are the most sensitive to increase in distance from home as compared to divorced employees. It also shows that single employees have a strong tendency of attrition when distance from home increases above 9 as compared to their other counterparts.

The graph (b) shows that married employees with high work life balance tend to increase their risk of attrition as compared to their single and divorced counterparts.

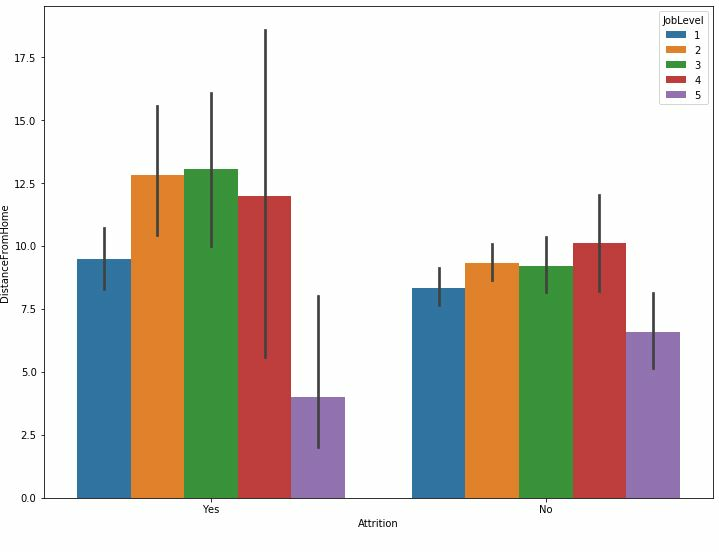


Fig 2 Relationship between job level and distance from home wrt attrition

The above graph shows the relationship between job level and distance from home, we can see that job levels 2,3,4 are most sensitive to distance from home as compared to job level 1 and 5.

Also job level 1 employees are less sensitive to increase in distance from home only till the maximum distance of 10. The job level 5 shows that even at higher distances attrition is low, this indicated that distance has very little significance to job level 5 employees and thus distance alone is not significant .

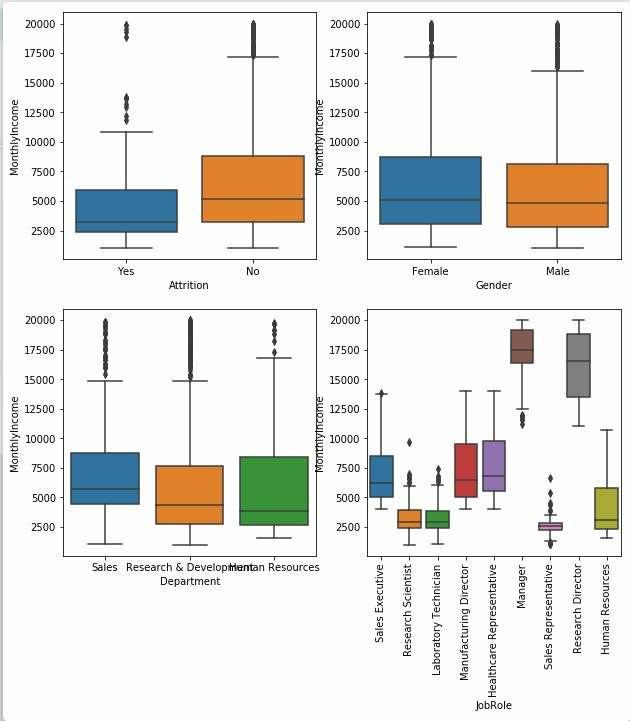


Fig 3 a. Relationship of monthly income wrt attrition.

b. Relationship of monthly income and gender

c. Relationship of monthly income and department

d. Relationship of monthly income and job role

The above box and whisker graphs indicate the mean and range of monthly income distribution.

**Predictive analysis:**

The following section deals with the prediction of attrition phenomenon using machine learning algorithms. The machine learning models used are:

1. Logistic Regression
2. Support Vector Machine
3. Decision Tree
4. Naïve Bayes
5. Random Forest

The data was first pre-processed using factor normalization and by accounting for missing values. The entire dataset was then partitioned into a training and testing dataset in the ratio of 80:20. This was done to build the model upon the data in training set and then to evaluate the model performance by comparing predictions with the test set data. The computation was done using the R statistical software.

The objective is to correctly predict the employees which are likely to leave the organization. The prediction results will thus focus more upon the identification of true positives i.e. employees who left the organization and were correctly predicted by the algorithm.

The following metrics were evaluated from the confusion matrices derived from the model performance on test set:

1. Accuracy: measures the percentage of correct predictions
2. Classification Error: measures the percentage of incorrect predictions
3. Recall / Sensitivity: measures the percentage of true positives wrt total number of actual positives in dataset.
4. Precision: measures the percentage of true positives wrt total number of positives predicted by the algorithm
5. False Positive Rate / Specificity: measures the percentage of false positives wrt actual total negatives in dataset.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Accuracy | Classification Error | Recall /Sensitivity | Precision | False Positive Rate |
| Logistic Regression | 90.13% | 9.86% | 44.68% | 87.5% | 1.21% |
| Support Vector Machine | 89.45% | 10.55% | 38.29% | 50% | 6.84% |
| Decision Tree | 85.71% | 14.28% | 27.65% | 61.90% | 3.23% |
| Naïve Bayes | 84.69% | 15.30% | 61.70% | 51.78% | 10.93% |
| Random Forest | 87.07% | 12.92% | 21.27% | 90.90% | 0.40% |

**Conclusion:**

The data analysis done by the correlational analysis shows that the following factors affect attrition the most. By identifying the relationship of these factors upon attrition , valuable insight can be generated which can then help make a more efficient and customised HR policy for the individuals at a high risk of attrition.

**Years in Current Role**:

In some jobs there is a lack of opportunity and growth. If the job is basically a dead-end proposition, this should be explained before hiring so as not to mislead the employee. If the employee has been working at the same level for many years and is not getting promoted despite deserving it then it is likely that the employee will quit and look somewhere else for better job growth potential (Fursso, 2015).

**Monthly Income**:

One of the most common reasons given for leaving the job is the availability of higher paying jobs. Inequity in pay structures or low pay is great causes of dissatisfaction and can drive some employees to quit. When they realize that they are being underpaid they tend to seek for better opportunities, thus affecting the turnover ratio of their present organization (Fursso, 2015).

**Overtime**:

Overtime and extended work hours increase with more resignations, additional workload, completion of certain tasks within a time horizon etc. If we consider a short time horizon, there is a possibility that an employee’s morale will initially rise slightly, as the employee tends to take advantage of the extra hours and extra money. However, when the bigger picture is taken into consideration, chronic and sustained staffing shortages require employees to work overtime at a regular basis, eventually diminishing the organizational capabilities. So with the increasing inefficiency within the work force, the rate of attrition increases significantly (Nandi and Mittal, 2012).

**Performance Evaluation and Rating**:

In previous research papers, it can be seen that attrition level is negatively correlated with performance appraisal. Thus an increase in the process of performance evaluation can lead to a reduced attrition within an organization. Certain companies have created provision for periodically reviewing and rating the performance of such employees. But in our research, based on the data gathered, we can conclude that performance rating doesn’t affect attrition significantly (Shanmugam and Babu, 2016).

**Job Level**:

In the research studies made earlier, it was noticed that the job level was high when the job satisfaction was high. High job level resulted in less attrition rate. High job level employees are less likely to quit. Organisation can take more advantage of the employees with high performance. High job level provides better working environment to the employees and they tend to feel satisfied with the salary they receive. They believe that their job security has increased and hence the retention rates are less. Entry level employees are mostly fresh graduates who tend to work to gain experience and may or may not settle down in the organisation. Thus, high job level increases the productivity and decreases attrition, as shown in our research. (Jitendra Kumar Singh & Dr. Mini Jain, 2013)

**Years with Current Manager:**

Employees may or may not be happy with the job because of their current manager. Previous researches show that relationship between employees and managers become stronger task by task, position to position and year by year. An important role of the manager is to connect with the employees. If he fails to do, job dissatisfaction comes into the picture. A good manager can work well by taking the team forward together. Our studies show a higher attrition due to increasing number of years with current manager. Thus, we can conclude that number of years of current manager doesn’t affect the attrition rate. (Dr. Wael M. EL Nabawy Saleh Dewydar, 2015)

The various machine learning models which were evaluated show that the Naïve Bayes model is the most appropriate choice for this specific use case. As this model gives the highest recall (61.70%) and an accuracy of (84.69%). The reason behind this selection is that in this case predicting correctly the employees that are most likely to leave is the greater focus, thus sensitivity / recall must be as high as possible.

Thus, by carrying out this analysis we have tried to address the research objectives which were to identify the factors causing attrition and predicting the attrition phenomenon. The implications of this study are that such statistical techniques and machine learning models can be used by HR practitioners to get a more information and evidence based approach to policy and decision making. Using the prediction algorithm various scenarios can also be analysed to see first-hand the results of changes in HR policy. Most importantly such techniques can help serve as an early indicator for alerting the HR team whenever a crucial or high performing employee is at a greater risk of attrition. Then the HR team can suitably compensate the employee or address the grievances faced by him before the probability becomes a reality. This can help minimise the cost of recruitment and training for new employees which will be required when the employee attrition rises.

**Limitations of the study:**

The limitations faced during the research process were the lack of larger dataset. By increasing the size of the training data, the model sensitivity can be considerably increased. Also, the presence of more features regarding each employee could also have made the model predictions more accurate in the real life setting where bias and environmental factors can change and affect the attrition phenomenon of workforce.

**Scope for further research:**

More advanced unsupervised machine learning models can be applied to the database to first make new statistically significant features and then a supervised learning model could be used to cluster the employees into different categories depending on probability of attrition.

**References**

[1] Dr. P. Raghunadha Reddy, P. L. (2017). HR Analytics - An Effective Evidence Based HRM Tool . *International Journal of Business and Management Invention* .

[2] Davenport, T. H. (2006, January). Competing on Analytics. *Harvard Business Review*.

[3] Denise M. Rousseau, E. G. (2011). Becoming an evidence-based HR practitioner. *Human Resource Management Journal, Vol 21*.

[4] Johannink, R. (2015). The future of HR Analytics: A Delphi method study . *5th IBA Bachelor Thesis Conference.* Enschede, The Netherlands: University of Twente, The Faculty of Behavioural, Management and Social sciences.

[5] King, K. G. (2016). Data Analytics in Human Resources: A Case Study and Critical Review. *Human Resource Development Review , Sage Journal*.

[6] Momin, W. Y. (2015). HR analytics transforming human resource management . *International Journal of Applied Research* .

[7] Shivinder Kaur, D. S. (2017). ANALYTICS IN MANAGEMENT WITH SPECIAL EMPHASIS ON USE OF ANN AS PREDICTIVE ANALYTICAL TOOL IN HRM PROCESSES . *International Journal In Applied Studies And Production Management* .

[8] Thomas Rasmussen, D. U. (2015). How HR analytics avoids being a fad. *Elsevier , science direct* .

[9]https://www.ibm.com/communities/analytics/watson-analytics-blog/hr-employee-attrition/

[10] Davenport, Harris, Shapiro (2010) *Competing on Talent Analytics. Harvard Business Review*.

[11] Aman Racho Fursso, (2015). *The Causes and Effects of Employees’ Turnover.*

[12] Veena Tewari Nandi, Harish Mittal (2012, January). Attrition in Knowledge Industry- Issues and Challenges. *International Journal of Knowledge and Research in Management & E-Commerce Vol.2, Issue 1*

[13] R. Shanmugam, Dr. N. Giri Babu (2016). Assessment of Employee Attrition among IT Employees. *International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 5.*

[14] Jitendra Kumar Singh, Dr. Mini Jain (2013) A STUDY OF EMPLOYEES’ JOB SATISFACTION AND ITS IMPACT ON THEIR PERFORMANCE Journal of Indian Research (ISSN: 2321-4155) Vol.1, No.4, October-December, 2013, 105- 111.

[15] Dr. Wael M. EL Nabawy Saleh Dewydar (2015) The Optimum Relationship between Managers and Employees International Journal of Business and Social Science Vol. 6, No. 8; August 2015

**Appendix:**

The training set had 1,176 data points (80% )

The test set had 294 data points (20%)

Both the training and test set data was randomly shuffled to avoid biases of consecutive data.

The confusion matrices for the machine learning models are as follows:

1. Logistic Regression

|  |  |  |
| --- | --- | --- |
|  | Predicted 1 | Predicted 0 |
| Real 1 | 21 | 26 |
| Real 0 | 3 | 244 |

1. Support Vector Machine

|  |  |  |
| --- | --- | --- |
|  | Predicted 1 | Predicted 0 |
| Real 1 | 18 | 29 |
| Real 0 | 2 | 245 |

1. Decision Trees

|  |  |  |
| --- | --- | --- |
|  | Predicted 1 | Predicted 0 |
| Real 1 | 13 | 34 |
| Real 0 | 8 | 239 |

1. Naïve Bayes

|  |  |  |
| --- | --- | --- |
|  | Predicted 1 | Predicted 0 |
| Real 1 | 29 | 18 |
| Real 0 | 27 | 220 |

1. Random Forest

|  |  |  |
| --- | --- | --- |
|  | Predicted 1 | Predicted 0 |
| Real 1 | 10 | 37 |
| Real 0 | 1 | 246 |