International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 06 | June -2020 ISSN: 2582-3930

Employee Attrition Prediction Model

Satyam Gupta , Anand Kuma Shrivastava

Computer Science & Engineering Department

ABES Engineering College, Ghaziabad, U.P., 201009, India

Abstract

The research paper entitled 'Employee Attrition Prediction Model' is an attempt to understand the thinking and working of the various categories of employees in IT sector towards the reduction of employee attrition in the organizations. Employee attrition is important especially in today's marketplace where employees are the most important human capital assets, attrition directly or indirectly impacts on organization's competitive advantage. The tangible costs of employee attrition would be the cost of training new employees, the recruitment and selection costs, adjustment time, possible product or service quality problems, costs of temporary staff, the cost of training, the cost of loss productivity, the cost of lost knowledge and the cost of the position remaining vacant till a suitable replacement is found. The intangible costs, which may be even more significant than the tangible cost, this involves the effect of attrition on organizational culture, employee morale, social capital or organizational memory. All these costs would significantly take away the profitability and the competitive advantage of the firm.

Keywords: Employee Attrition, Employee Satisfaction , Employee Appraisal and Training, Attrition rate.

Introduction

Attrition is an alarming concern for all the organizations across varying sectors. Anything that involves manpower would be affected by employee attrition. Whenever employees leave an organization, the knowledge and intellect also leaves. Human Resource practitioners can try and capture the knowledge and intellect of the talent, however, this is easy said than done. Nowadays, the importance and relevance of the employees is well understood by the organizations and hence the focus is on

understanding the key factors that affects employee attrition. The aim of this research is to understand and identify various researches conducted on employee attrition across sectors. For the same, the literature starting from 1955-2014 is studied. The research has highlighted the factors like dissatisfaction with compensation offered , payment below prevailing market rate and inadequacy in the internal and external equity, Inappropriate Reward and recognition, toxic work environment &incompatible work culture, employee attit ude, Insufficient support, unsatisfactory relationships with superior, colleagues and subordinates (work relationship), and inadequate opportunities for growth, hiring practices, and managerial style, which affect employee attrition. Understanding attrition is imperative for every organization. Mere recognition of factors would not do any better, both for employees as well as employer. This study focuses on compiling the factors that affect attrition and has a further scope where these factors can be empirically tested in different sectors and recommendations can be incorporated s as to observe the difference in the attrition rate.

a. Objective:

- First objective is to identify the underling features which mostly influent the employee attrition in the organization.
- Second objective to analyze the behavior and working of machine learning algorithm on this dataset.
- The last objective is to derive the best accuracy rate of different machine learning algorithm.

b. Problem Identification & Definition

The HR head wants to understand what key drivers are for attrition in their organization. Also want to ensure that they attract and retain our top talent and take remodel action if it looks like they are going to loss good people.

Employee Turnover (attrition) is a major cost to an organization and predicting turnover is at the forefront of the needs of Human resources in many organizations

Related Work

a. EMPLOYEE ATTRITION ANALYSIS

Abstract:

This paper discusses on the analysis of factors affecting employee attrition and predicting it beforehand so as to take the necessary measures to retain a skilled and valuable employee. It gives a detailed account of the factors affecting an employee's decision to leave the company,predicted probabilities of their leaving the company the variation of a factor's influence on them.

Conclusion:-

Though employee attrition may seem like a trivial issue but it is one of the costliest problem that every industry faces irrespective of their field of work. Companies spend a large share of their budget on employee training, hence it becomes very difficult to train another employee again, once a skilled and trained employee leaves the company. It is also difficult to find skilled individuals in the market these days in the current economic scenarios. Hence through this project, we are able to understand and tackle the issues afflicting the employees as a preventive measure. Data analytics provides some of the most accurate and intrinsic insights and aids in places where human judgment might error, thus providing a way to solve human issues with a little aide and insight from analysis.

Early Prediction of Employee Attrition using Data Mining Techniques

Abstract:-

Bill Gates was once quoted as saying, "You take away our top 20 employees and we [Microsoft] become a mediocre company". This statement by Bill Gates took our attention to one of the major problems of employee attrition at workplaces. Employee attrition (turnover) causes a significant cost to any organization which may later on effect its overall efficiency. As per CompData Surveys, over the past five years, total turnover has increased from 15.1 percent to 18.5 percent. For any organization, finding a well trained and experienced employee is a complex task, but it's even more complex to replace such employees. This not only increases the significant Human Resource (HR) cost, but also impacts the market value of an organization. Despite these facts and ground reality, there is little attention to the literature, which has been seeded to many

misconceptions between HR and Employees. Therefore, the aim of this paper is to provide a framework for predicting the employee churn by analyzing the employee's precise behaviors and attributes using classification techniques.

Conclusion:-

Employee attrition can affect an organization in many ways like goodwill, revenues and cost in terms of both time and money. The predictive attrition model helps in not only taking preventive measure, but also making better hiring decisions. In this study implementation of various classifications method helps in predicting whether a particular employee might leave the organization in the near future by deriving trends in the employee's past data. It was intuited that salary or other financial aspect like promotions are not the sole reasons behind the attrition of employees. These models can help us in prioritizing the features with higher impact in attrition of an employee and the possible reasons behind it so that HR can take appropriate decision for the retention process. The main purpose of this research is to build reliable and accurate models which can optimize the hiring and retention cost of quality employees. This could be done by determining the attrition status of employee under consideration by using the appropriate data mining techniques.

Proposed Methodology

Model : A mathematical equation that establishes a relationship between the x variables and the Y variable. Models are created in 3 Stages:

Part I : Pre Modelling:

1. Understanding objective and deriving the Business problem
2. Validate the Problem
3. Convert Business Problem to Stats problem by defining possible X and Y, type of stat problem
4. Understand the data is required
5. Get the different data and create Data Audit Report at File
6. Consolidate the required data at objective level (e.g. Customer 360) and attain data dictionary
7. Create Data Audit Report on the consolidated file

Part II : Implementing Model (Modeling)

- Select algorithm based on the Business/Stat Problem
- Data Preparation Level I
- Data Preparation Level II
- Data Preparation Level III

- Data Preparation Level IV
- Model Implementation
- Hyper Parameters Selection and optimization (Machine Learning)

Part III : Post Modeling

1. Model Validation
2. Calculate Metrics (Model Evaluation)
3. Exporting The Model Object
4. Implementation

Implementation and results

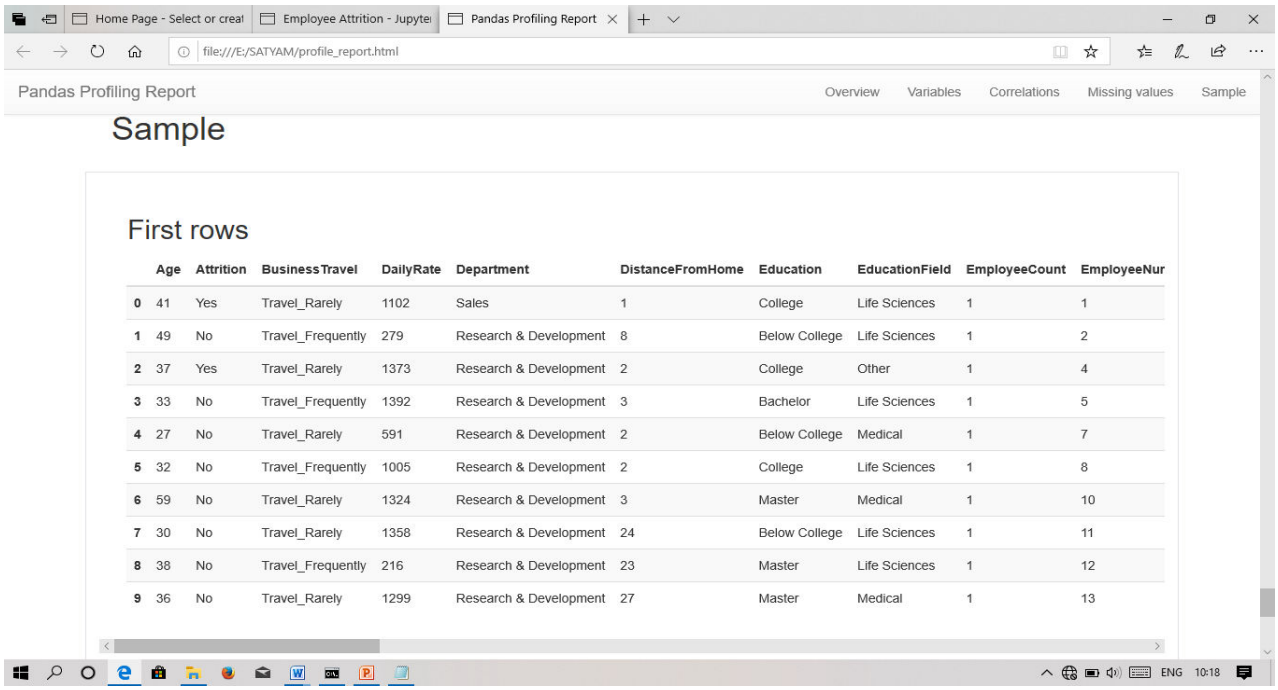
POPULATION / Sample Size

Population The population size 1500 employees of an organization that is related to the IBM HR Analytics team.

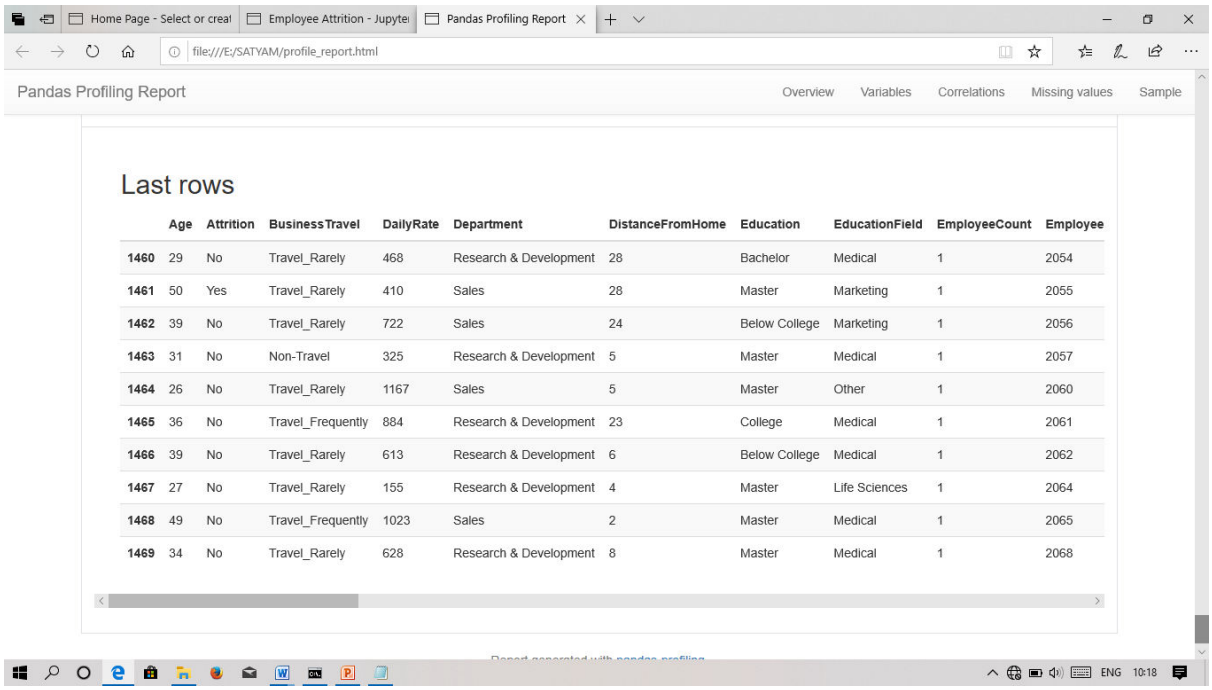
Sample Size:- For this research paper sample of 70% data is used for training the model and rest is for test the model.

DATA COLLECTION The dataset can be available on many online forums like Kaggle, Github etc.

DATASET We Have Been Given A Data Set About All Of Past And Current Employees, So Here we Have Access To Some Demographic Information Like Age, Gender And Marital Status Along With Their Compensation And Performance Related Details. Here are some view of Dataset……



	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNur
0	41	Yes	Travel_Rarely	1102	Sales	1	College	Life Sciences	1	1
1	49	No	Travel_Frequently	279	Research & Development	8	Below College	Life Sciences	1	2
2	37	Yes	Travel_Rarely	1373	Research & Development	2	College	Other	1	4
3	33	No	Travel_Frequently	1392	Research & Development	3	Bachelor	Life Sciences	1	5
4	27	No	Travel_Rarely	591	Research & Development	2	Below College	Medical	1	7
5	32	No	Travel_Frequently	1005	Research & Development	2	College	Life Sciences	1	8
6	59	No	Travel_Rarely	1324	Research & Development	3	Master	Medical	1	10
7	30	No	Travel_Rarely	1358	Research & Development	24	Below College	Life Sciences	1	11
8	38	No	Travel_Frequently	216	Research & Development	23	Master	Life Sciences	1	12
9	36	No	Travel_Rarely	1299	Research & Development	27	Master	Medical	1	13



	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Employee
1460	29	No	Travel_Rarely	468	Research & Development	28	Bachelor	Medical	1	2054
1461	50	Yes	Travel_Rarely	410	Sales	28	Master	Marketing	1	2055
1462	39	No	Travel_Rarely	722	Sales	24	Below College	Marketing	1	2056
1463	31	No	Non-Travel	325	Research & Development	5	Master	Medical	1	2057
1464	26	No	Travel_Rarely	1167	Sales	5	Master	Other	1	2060
1465	36	No	Travel_Frequently	884	Research & Development	23	College	Medical	1	2061
1466	39	No	Travel_Rarely	613	Research & Development	6	Below College	Medical	1	2062
1467	27	No	Travel_Rarely	155	Research & Development	4	Master	Life Sciences	1	2064
1468	49	No	Travel_Frequently	1023	Sales	2	Master	Medical	1	2065
1469	34	No	Travel_Rarely	628	Research & Development	8	Master	Medical	1	2068

Pre Modelling:

The objective of this business problem is to identify some important features that are responsible for the attrition in the company.

In this problem y (Independent variable) is the Attrition column and x(dependent variable) is other columns like age,gender,satisfaction level,YearAtCompany etc.

X=Attrition

Y= (['Age', 'BusinessTravel', 'DailyRate', 'Department', 'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount', 'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender', 'HourlyRate', 'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction', 'MaritalStatus', 'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked', 'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating', 'RelationshipSatisfaction', 'StandardHours', 'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance', 'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion', 'YearsWithCurrManager'].

TOOLS USED FOR Feature Selection:-

a) Simple percentage method

b) Chi-square method

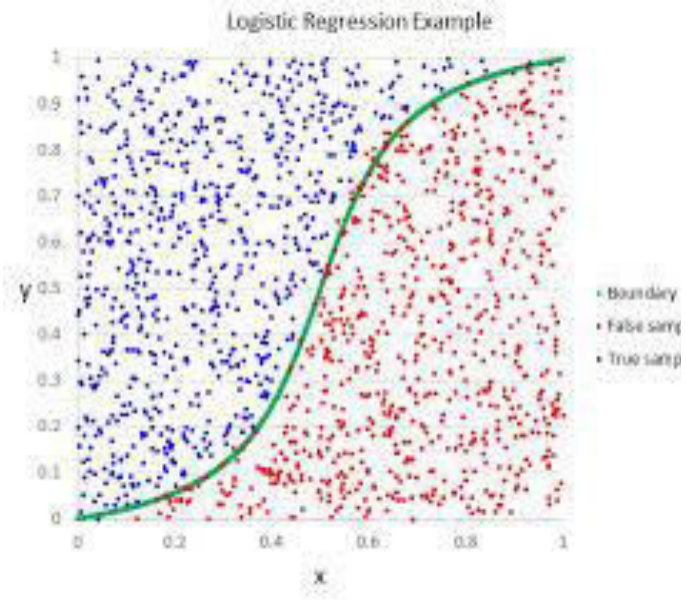
c) Correlation & Coefficient method

In this project we have used Correlation & Coefficient method

Algorithm:-

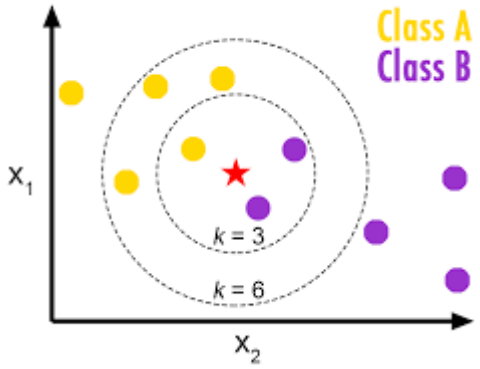
1. **Logistic Regression:**

A classification method built on the same concept as linear regression. In this method the target variable is categorical it can be two category or multi category variable. With the help of logigtic regression you cam identify weather the email i spam or not and so many thing.



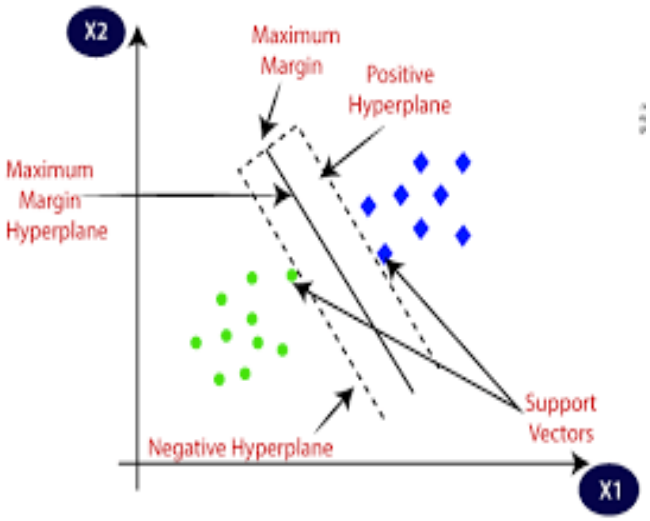
2. **KNN Algorithm :**

KNN algorithm is used for both classification and prediction . It is based in the most nearing neighbors so it is called K-nearest-Neighbors algorithm. It looks k closest data record in reference data.



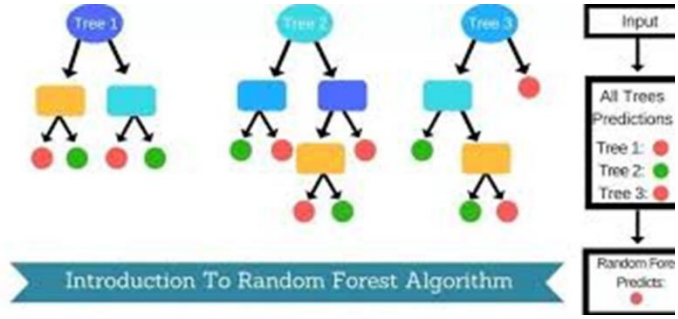
3. **Support Vector Machine:-**

Support vector machine is a discriminate classifier. It works on various dataset and used for classification.



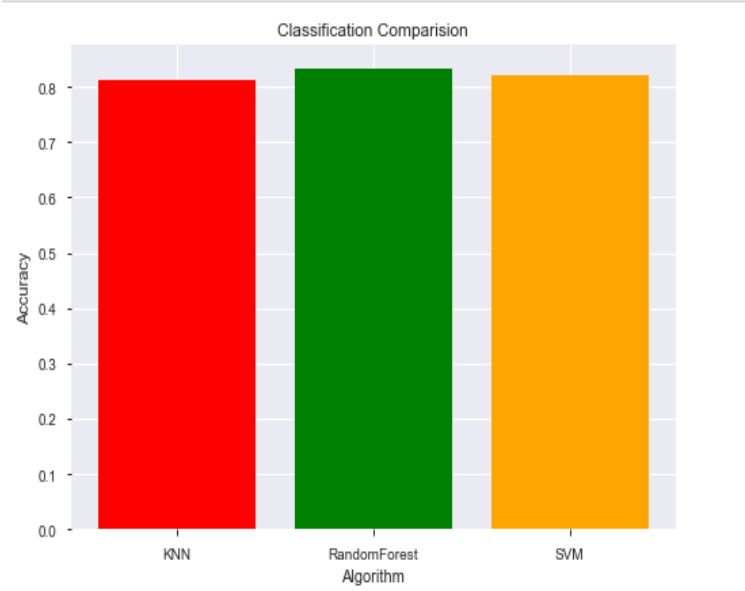
4. Random Forest Algorithm:-

Random Forest machine learning is a both classification and prediction algorithm. It works as decision tree. It makes various decision tree and analyse all trees.

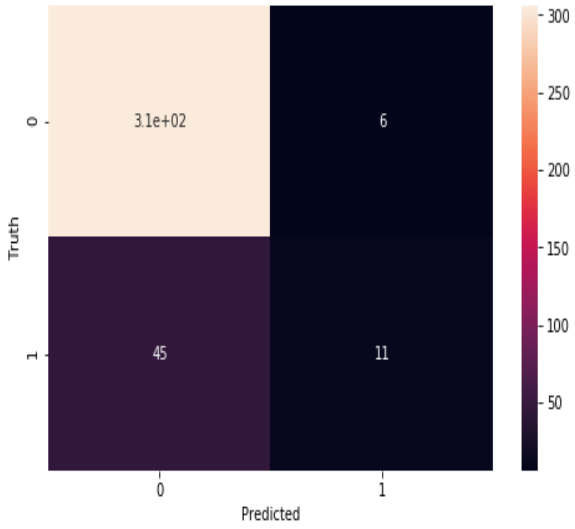


Performance Analysis:-

Amongst the three algorithms, Random Forest is said to have the highest accuracy (0.8614) as seen while comparing and choosing the algorithms in below Figure.

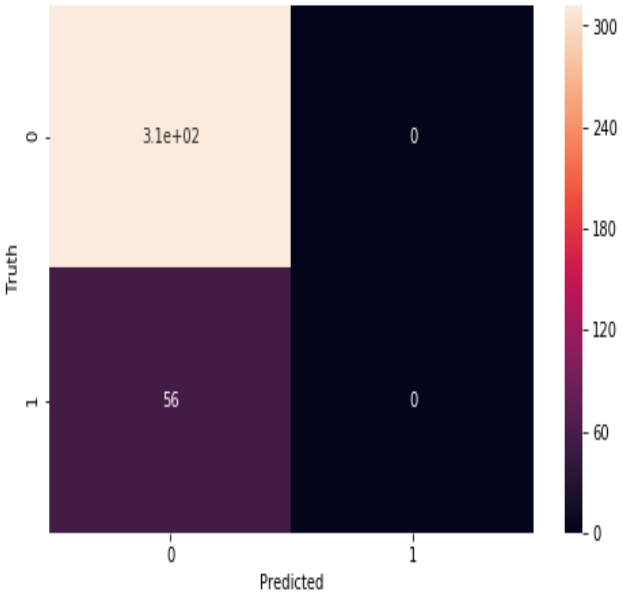


It also has the highest number of cases for which the predicted probabilities and the true probabilities are accurate as shown in the confusion matrices.



This confusion matrices have 306 records which are 0 as truth and predicted correctly as 0 and 11 I's which are predicted as 1.

In SVM method it has the accuracy rate is 0.847814% and the confusion matrix is below in which 312 0's are predicted 0 correctly and no single 1's are predicted correctly.



Conclusion:-

In this project I have used various classification techniques which gives different accuracy rate and on the bases of his precision value and confusion matrix Random Forest algorithm is finded best.

Reference:-

1. "Analytix Labs" large summary for all the steps of data preprocessing and model implementation.

2. YOUTUBE TUTORIALS (Codebasics channel) for algorithm and data processing.

3. "Kaggle" for Dataset.

4. Github

5. Fechete, Nedelcu, "Analysis of the Economic Performance of an Organisation using Multiple Regression", AFASES, May 2014.

6. Olton[Online].Available: "Solving Employee Attrition with Data", TowardsDataScience.com; April 2017

7. Natasha Sharma[Online]. Available: "People Analytics with Attrition Prediction", May 2018