ATTRITION PROJECT

A PROJECT SUBMITTED TO AILEADTECH

SUBMITTED BY:

SOTAYO SEUN MUYIWA

NOVEMBER 2020.

**Objective:** To predict if an employee is going to attrite or not

**Introduction**

The project is carried out to determine the level of attrition among workers in the organization. The analysis would be carried out by determining:

1. Factors that drive attrition
2. Factors have no effect on attrition?
3. What is the accuracy of the prediction using the model?
4. Which algorithm gives the best accuracy?

Based on the findings of the analysis done, the factors that drive attrition from the data are:

* Age
* Total working years
* Years at company
* Years since last promotion
* Monthly rate
* Job level

The factors that have no effect on attrition are as follows:

* Daily rate
* Distance from home
* Education
* Environmental satisfaction
* Hourly rate
* Job involvement
* Job satisfaction
* Number of companies worked
* Present salary hike
* Performance rating
* Relationship satisfaction
* Stock option level
* Training times last year
* Work-life balance

The analysis shows that 2466 employees stayed in the company (i.e no attrition) while 474 employees left (attrition).

**Pictorial analysis of factors that affect attrition were taken and the results are given below**

1. Age

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Attrition | Max | Upper. Quartile | Mid Quartile | Lower Quartile | Minimum |
| No | 60 | 45 | 38 | 31 | 15 |
| Yes | 55 | 39 | 32 | 29 | 15 |

1. Total working years

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Attrition | Max | Upper. Quartile | Mid Quartile | Lower Quartile | Minimum |
| No | 32 | 17 | 10 | 6 | 0 |
| Yes | 20 | 10 | 7 | 4 | 0 |

1. Years at company

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Attrition | Max | Upper. Quartile | Mid Quartile | Lower Quartile | Minimum |
| No | 21 | 10 | 6 | 4 | 0 |
| Yes | 16 | 7 | 3 | 2 | 0 |

1. Years since last promotion

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Attrition | Max | Upper. Quartile | Mid Quartile | Lower Quartile | Minimum |
| No | 7 | 3 | 1 | 0 | 0 |
| Yes | 5 | 2 | 1 | 0 | 0 |

1. Monthly rate

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Attrition | Max | Upper. Quartile | Mid Quartile | Lower Quartile | Minimum |
| No | $25,000 | $20,000 | $14,000 | $8,000 | $2,000 |
| Yes | $25,000 | $22,000 | $15,000 | $9,000 | $5,000 |

1. Job level

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Attrition | Max | Upper. Quartile | Mid Quartile | Lower Quartile |
| No | 5.0 | 3.0 | 2.0 | 1.0 |
| Yes | 3.0 | 2.0 | - | 1.0 |

A histogram analysis was also carried out to show the effect of factors such as ‘total working years’, ‘Age’, ‘Standard hours’ and ‘Years at company’, on attrition.

1. Total Working Years: It was discovered that people who have worked for up to 10 years stayed while people with a little above a year at the company left.
2. Age: people between 34 and 35 years of age did not leave the company while people between 29 and 31 years of age have left the company.
3. Standard Hours: less than 500 people left the company based on the standard hours spent in doing their job while about 2434stayed.
4. Years at company: about 350 people who have been at the company for about 5-10 years stayed but 100 people who used about a year left.

This shows that the lesser the years spent working in the company, the higher the chances of leaving and the higher the years spent working at the company the lower the chances of leaving the company. Younger people between 29-31 years of age with lesser experience (about a year) at the company leave, while those with 5-10 years’ experience and about 34-35 years of age at the company tend to stay more.

Correlation of some factor was also carried and result showed that there is a moderately high correlation between

|  |  |  |
| --- | --- | --- |
| Factors | Correlation | Remarks |
| Age and Total working years | 0.68 | Moderately high |
| Job level and Total working years | 0.68 | Moderately high |
| Age and Job level | 0.5 | Moderate |
| Age and Monthly income | 0.5 | Moderate |
| Monthly income and Total working years | 0.68 | Moderately high |

**Methodology**

The analysis of the project employed models such as

* Logistic regression
* Confusion matrix
* Decision tree classifier
* Random forest Algorithm
* K-fold validation
* Leave one out validation

**Performance**

The analysis selected six features which are: Total working years, Years at company, Age, Years since last promotion, Monthly rate and Job level.

* The result of the logistic regression score show that the model is generally fit with the training score as 0.837 and the test score as 0.843. Therefore, the model performed well.
* Using the decision tree model classifier, the results are as follows:

Training showed 84.5% accuracy while test accuracy was 84.9%.

The above result showed that the model is generally fit.

* The random forest classifier showed an analysis on of training accuracy at 84.5% and test accuracy was estimated at 85.1% with the maximum depth set at 3 which means that the model is generally fit. Prior to this result, without the tuning of the data with maximum depth set at 3, the result of the analysis showed training accuracy at 100% and test accuracy at 95%
* The cross-validation score of the logistic regression was estimated at 97.3% with a maximum iteration of 1000 which is a high result.
* The data was standardized to give a training accuracy of 98.5% and 96.5% with ‘C’ set at 0.1which means the model over fits.
* The result of the K-fold scores with shuffle showed estimates at 94%, 88% and 94% respectively and the scores are all high
* The leave one out model was also adopted with 150 iterations and 96% mean score which is also high

**Summary**

From the data analysed based on age, 52.6% of people between 20-60years did not attrite while 43.4% of people between 18-55years attrite.

61.3% of people who have spent a total of about 0-32years did not leave while 38.7% of people who have spent between 0-20 years left the company.

59.4% of people who have spent 0-20 years at the company did not leave while 40.6% of people who have spent between 0-16years left.

57.9% of people who were last promoted between 0-7years stayed while 42.1% of people who were last promoted between 0-5yeaes left.

55.2% of those who earned between $2,000 to above $25,000 stayed while 44.8% of those who earned the same amount left.

64.7% of people with the job level of 0 - 5.0 stayed while 35.3% of people with job level of 0 – 3.0 left.

In summary, the algorithm with the highest prediction is the cross-validation algorithm with an estimated result of 97.3%.

The logistic regression, random forest algorithm, decision tree and K-fold all generally performed well.