5/20/2023

Employee Attrition for Healthcare

Yash Patkar

Yash.aj23syd006@spjain.org

MAIB-AJ23SYD006

**Content:**

|  |  |  |
| --- | --- | --- |
| **Sr no** | **Topic** | **Page** |
| 1 | Abstract | 2 |
| 2 | Introduction | 2 |
| 3 | Data Description | 4-5 |
| 4 | Research Questions | 6 |
| 5 | Research Findings | 7 |
| 6 | Analytical Findings | 15 |
| 7 | Recommendation | 16 |
| 8 | Conclusion | 16 |
| 9 | Reference | 16 |

1. **Abstract:**

The report is about the study of trends of employee attribution. In this report with the use of python graphs we visualize the trends in employee attribution and using ML algorithms create a model to predict the expectation of attrition based on the parameters in the data set.

1. **Introduction :**

Employee attrition, or employee turnover, is a critical challenge faced by organizations across various industries. It refers to the voluntary or involuntary departure of employees from an organization. High attrition rates can have detrimental effects on productivity, morale, and the overall success of a company. Understanding the factors contributing to employee attrition is essential for organizations to develop effective retention strategies and maintain a stable and engaged workforce.

Human Resources (HR) analytics provides valuable insights into employee behavior, performance, and engagement. By leveraging HR data and applying analytical techniques, organizations can identify patterns, trends, and potential predictors of attrition. This report aims to analyze employee attrition using HR analytics and explore the factors that influence employee turnover.

The report focuses on various variables that have been found to impact attrition rates. These variables include employee satisfaction, performance evaluations, project involvement, average monthly hours worked, tenure in the company, workplace accidents, promotions in the last five years, department, and salary. Analyzing these factors can provide a comprehensive understanding of the dynamics contributing to attrition.

Machine learning algorithms, such as Logistic Regression, Decision Trees, and Random Forests, are employed to build predictive models. These models utilize historical employee data to predict attrition and identify the most influential factors. Cross-validation techniques are used to evaluate the accuracy of the models and ensure their generalization capabilities.

The report also examines the issue of overfitting, which occurs when a model performs exceptionally well on the training data but fails to generalize to new data. By assessing the performance of the models on both the training and testing data, we can determine if overfitting is present and take steps to address it.

The findings of this report will provide organizations with valuable insights into the factors contributing to employee attrition. By understanding these factors, organizations can develop targeted strategies to improve employee satisfaction, engagement, and retention. This, in turn, can lead to higher productivity, reduced costs associated with turnover, and improved organizational performance.

In the following sections, we will discuss the methodology, data analysis, results, and recommendations based on the findings of the employee attrition analysis using HR analytics.

1. **Data Set:**  
   Link to dataset: - https://www.kaggle.com/datasets/jpmiller/employee-attrition-for-healthcare

This data set consists of **1676 rows and 35 features**. The Features in the dataset include:  
  
'EmployeeID', 'Age', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department', 'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount', 'EnvironmentSatisfaction', 'Gender', 'HourlyRate', 'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction','MaritalStatus', 'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked','Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating', 'RelationshipSatisfaction', 'StandardHours', 'Shift', 'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance', 'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion','YearsWithCurrManager'

Reason for selecting this dataset:

The above data set is concise and has quite a few numbers of features and all of them being on point. Even if the data is smaller compared to other data sets available this data set has no missing values and very few unique value columns making it perfect to attrition.

**Data Description:**  
The data is for the attrition of the employees in the health sector field

* EmployeeID – This is the unique identifier for each employee
* Age – This denotes the age of the employee
* Attrition – This denotes if the employee has left the sector or not
* BusinessTravel - This denotes if the employee has travelled for business or not
* DailyRate – This denotes the daily rate in terms of dollars
* Department – This denotes the Department the employees works in
* DistanceFromHome - This denotes the distance from the employees home
* Education – This denotes the level of education of the employee
* EducationField – This denotes the field of education
* EmployeeCount – This denotes the number of employees
* EnvironmentSatisfaction – This denotes the satiffaction with the work environment
* Gender – This denotes the gender of the employee
* HourlyRate - This denotes the hourly rate of the employee.
* JobInvolvement – This denotes the level of involvement with the job
* JobLevel – This denotes the level of job
* JobRole – This denotes the role of the employee
* JobSatisfaction – This denotes the satisfaction with the job
* MaritalStatus – This denotes the marital status
* MonthlyIncome – This denotes the monthly income
* MonthlyRate – This denotes the monthly rate in dollars
* NumCompaniesWorked – This denotes the Companies the employees has worked in
* Over18 – this denotes if the employee is above 18 or not
* OverTime – This denotes how much the employee has done overtime
* PercentSalaryHike – This denotes the salary hike in percentage
* PerformanceRating – This denotes the performance rating
* RelationshipSatisfaction – This denotes the relationship satisfaction
* StandardHours – This denotes the standard working hours
* Shift – This denotes the sift of the employee
* TotalWorkingYears – This denotes the total working hours
* TrainingTimesLastYear – This denotes the training time is the last year
* WorkLifeBalance – This denotes the work life balance
* YearsAtCompany – This denotes the years spent in the company.
* YearsInCurrentRole – This denotes the years spent in the current role
* YearsSinceLastPromotion – This Indicates the number of years since last promotion
* YearsWithCurrManager – This indicates the number of years since last change in the manager

1. **Research Questions** :

Some research questions that I have focused on in this project are :

1. Does the years spend at the company Make any Difference?
2. Which Department has the maximum number of Attrition.
3. Does the marital status have any significance to attrition?
4. **Research Findings:**

People that have spent more than 5 years with the company tend to stick to the company for a longer period and not leave.

A picture containing text, screenshot, line, plot

Description automatically generated

A picture containing screenshot, text, line, rectangle

Description automatically generated

A picture containing text, screenshot, line, rectangle

Description automatically generated

The maternity department has the most about of attritions

A picture containing text, screenshot, rectangle, diagram

Description automatically generated

A picture containing text, screenshot, diagram, number

Description automatically generated

The Data suggests that single people are more likely to leave

A picture containing screenshot, rectangle, text, plot

Description automatically generated

A picture containing screenshot, text, diagram, plot

Description automatically generated

Some other insights provided by the data are provided below:

The Spread of the Data per column

A picture containing text, square, screenshot, rectangle

Description automatically generated

Density Of the data

A picture containing colorfulness, screenshot, diagram, design

Description automatically generated

A picture containing colorfulness, screenshot, rectangle, diagram

Description automatically generated

A picture containing screenshot, diagram, plot, colorfulness

Description automatically generated

Scatter plot for all the data points

A group of white squares with blue dots

Description automatically generated with low confidence

A screenshot of a graph

Description automatically generated with low confidence

Attrition for the object type columns:

A picture containing screenshot, rectangle, colorfulness, diagram

Description automatically generated

A blue square with white text

Description automatically generated with low confidence

Columns with high Correlation

A picture containing text, screenshot, diagram

Description automatically generated

1. **Analytical Findings:**

Spliting the data into a 70 and 30 split we train multiple models to check the accuracy of the models. Random Forest model is the most effective model of all the models tested which means based on this models we can predict new data weather an employee is going to leave the company or not and if the likelihood of the leaving is high we can makes offer them a raise or make the work environment better depending on what is the highest contributor to the employee leaving.



1. **Recommendation:**• Implement retention tactics: Based on an examination of the factors driving employee attrition, it is advisable to develop and implement focused retention initiatives. Increasing employee engagement initiatives, offering opportunities for professional advancement and development, and building a positive and supportive work environment can all be part of this.

• Emphasis on at-risk populations: Develop retention techniques that are specifically targeted at high-risk groups for turnover, such as workers in a given department, with a certain amount of tenure, or earning a certain amount of money. Consider providing greater resources and support to staff members who have been with the company for a particular period of time or who work in departments with higher turnover rates.

• Work-life balance is a significant contributor to employee retention, so it is essential to pay attention to initiatives that promote a positive work-life balance.

1. **Conclusion**  
   This project helped me to think outside the box and able to come up with creative ways to explore the relations with the data . The exploration of data helped me better understand data and correlation it even help me understand the necessity of feature engineering in creation relationship between feature. Further this project gave me an opportunity as discovering a new and challenging concept of using regression models on real world data.
2. **Reference**Link to dataset:- https://www.kaggle.com/datasets/jpmiller/employee-attrition-for-healthcare

Information  
<https://www.w3schools.com/python/default.asp>  
<https://www.python.org/about/gettingstarted/>https://stackoverflow.com/  
<https://chat.openai.com/chat>  
<https://www.digitalocean.com/community/tutorials/exploratory-data-analysis-python>  
<https://towardsdatascience.com/exploratory-data-analysis-in-python-a-step-by-step-process-d0dfa6bf94ee>  
<https://www.analyticsvidhya.com/blog/2020/03/pivot-table-pandas-python/>

<https://pandas.pydata.org/docs/reference/api/pandas.pivot_table.html>

<https://stackoverflow.com/questions/37625334/convert-month-int-to-month-name-in-pandas>

https://sparkbyexamples.com/pandas/pandas-check-column-contains-a-value-in-dataframe/