Employee Attrition

Bill L. Andrews
Bellevue University

Abstract

Human resource analytics (HR analytics) is a subfield that refers to applying analytic processes to an organization's human resource department to improve employee performance and obtain a better return on investment. HR analytics is more than just collecting data on employee productivity. Instead, it seeks to provide insight into each process by collecting data and then analyzing it to make appropriate decisions about improving these processes.

Human resource attrition refers to the gradual loss of employees over time. In general, relatively high attrition is a problem for businesses. HR professionals frequently take the lead in developing company compensation programs, work cultures, and motivation systems that aid in the retention of top employees.

Business Problem

The cost of high employee attrition to an organization is a significant issue. The most common costs of losing and replacing employees are job postings, hiring processes, paperwork, and new hire training. Furthermore, high employee turnover prevents the company from expanding its collective knowledge and experience, especially concerning if the company serves customers, as customers prefer to interact with familiar faces. If the company constantly hires new employees, errors and problems are more likely.

Background/History

Employee attrition is the percentage of employees who leave a company. They are either replaced by new employees, or the position may remain vacant or be closed entirely.

Employee retention is critical and cannot be overstated. Indeed, the financial cost of losing top performers is enormous. According to the Work Institute's Retention Report, the replacement cost for a $50,000-a-year employee is a whopping $16,500.

With so much on the line, it is no surprise that employers are devoting more resources and time to retaining top talent. This is especially true in technology firms, where cutting-edge innovation is frequently driven by talented individuals working alone or in small groups.

Data Explanation

1. The dataset has 1470 rows for each employee and 35 attributes
2. Find the unique values for each attribute. Identify attributes that will not bring value to the model that will be removed.
    2.1. Employee Count
    2.2. Employee Number
    2.3. Over18
    2.4. Standard Hours
3. Review the data types
    3.1. Change categorical to numeric
4. Check for missing data

Methods

1. Define the Problem
2. Gather the Data
3. Prepare the Data (Data Wrangling)
    - 3.1. Import Python Libraries
    - 3.2. Load Libraries
    - 3.3. Data Cleaning
    - 3.4. Convert Formats
4. Exploratory Data Analysis
    - 4.1. Using descriptive and graphical statistics to search the dataset for potential problems, patterns, classifications, correlations, and comparisons is critical. Furthermore, data Employee Attrition 3 categorization (i.e., qualitative vs. quantitative) is essential for understanding and selecting the appropriate hypothesis test or data model.
    - 4.2. Identify Correlations
    - 4.3. Split data into train and test sets
    - 4.4. Data Visualizations
5. Model Data
    - 5.1. Data modeling can be used to summarize data or to forecast future outcomes. The dataset and expected results will determine the algorithms available for use. The incorrect model can result in poor performance and an incorrect conclusion (which is used as actionable intelligence) at worst.
    - 5.2. Encoding categorical variables with Ordinal Encoder
    - 5.3. Import Models
        - 5.3.1. Random Forest
        - 5.3.2. Gradient Boosting
        - 5.3.3. Decision Tree
6. Validate and Implement Data Model

Analysis

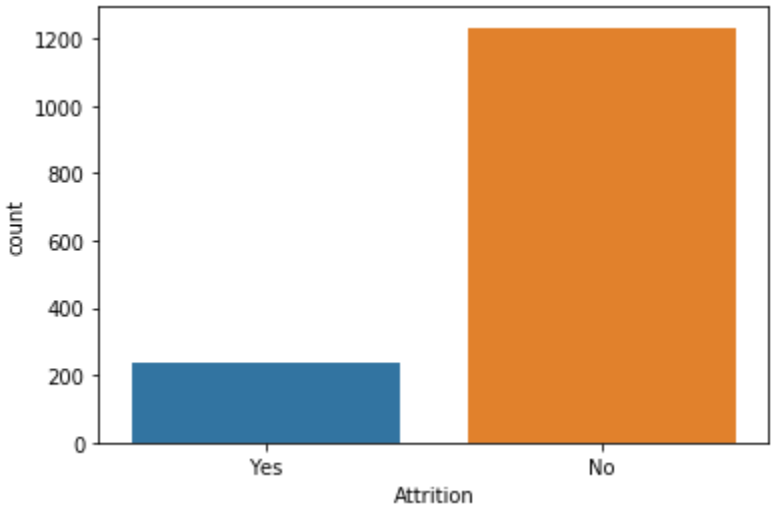
When categorical variables are examined, we can see that the majority of employees who left work in the Research & Development department, with the majority being laboratory technicians, sales executives, or research scientists. These employees received excellent performance ratings. It is never good to lose employees with such high levels of performance! Most of them held a Bachelor's degree and were educated in Life Sciences, Medicine, or Marketing. These employees reported high job involvement, high job satisfaction, and a better work-life balance. However, it is concerning that a large number of them reported low satisfaction with the work environment.

Looking at the attrition per age histogram, employees leave less when they get older. Most of the attrition is made up of employees aged 25 to 35. According to the data, the more working years, years at the company, and years in the current role employees have, the less likely they are to leave. In terms of income, most employees who left were among those with a lower monthly income, with monthly income.

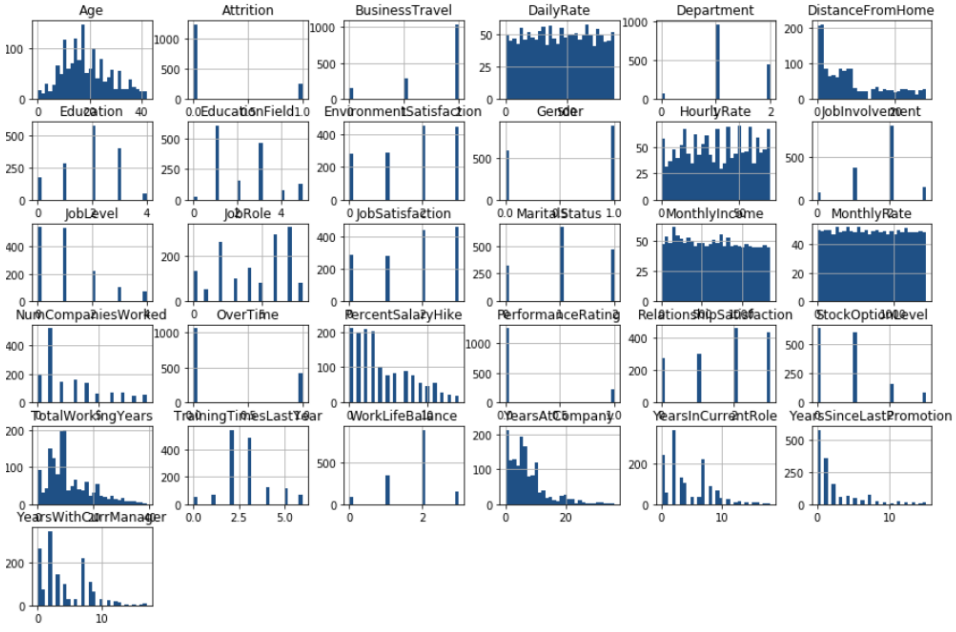
Those with a lower percentage salary increase tend to leave more than those with a higher percentage salary increase. Employees who leave tend to be young, with less time working for the company and at the start of their career because most of these employees worked for less than ten years in total.

Visualization

Attrition Count



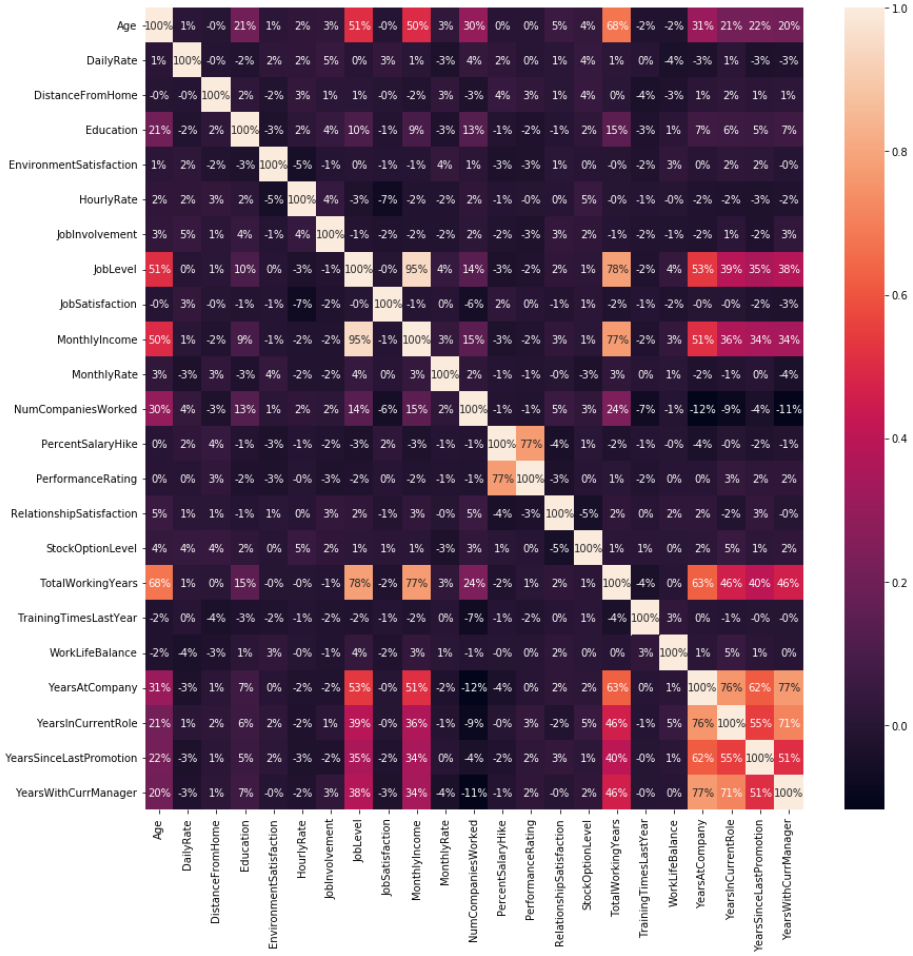
Histogram for all variables



Correlation

Correlation between Total Working Years with MonthlyIncome and JobLevel

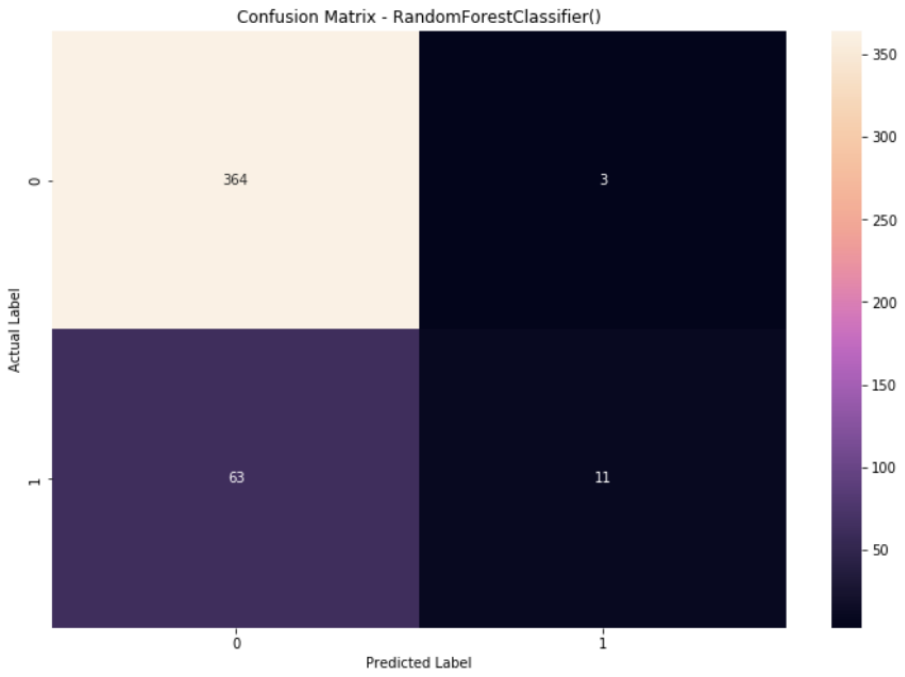
YearsAtCompany with YearsWithCurrManager and YearsInCurrentRole



Conclusion

The Random Forest received the highest score of 85.03% of all models.

```
Accuracy: 85.03%
Precision: 78.57%
Recall: 14.86%
F1_Score: 25.00%
```



Assumptions

Employee attrition is predictable under stable conditions, where a consistent pattern can be deduced from specific parameters that all times influence both the employee and the organization. Some of these parameters may be predictable, such as retirement age, while others may be unpredictable, such as company performance, external funding, and management shakeup.

Limitations

Data Shortage: For accurate predictions, several predictive data models require training on large datasets; otherwise, they may become overfitted. Massive datasets are only sometimes available or develop slowly. As a result, predictive data models must be constantly refined and updated, and their performance must be evaluated concurrently.

Data Labeling Errors: If the labels on the training data are incorrect, predictive data models can produce incorrect results. After all, a model is only as good as the data on which it is trained. A strong system of checks must be in place to reduce such errors.

Challenges

Why is it so difficult to find reliable and unbiased churn technical information? Because as with many other issues in data science, there are many methods for predicting which customers are likely to churn. Even the term "churn modeling" can refer to calculating the proportion of customers who are churning, forecasting a future churn rate, or predicting the risk of churning for specific individuals.

Because only some things discovered by this innovation are valuable, it is critical to keep predictive analytics focused on delivering helpful business insights. Some mined data is valid only to satisfy an inquisitive mind and has few or no business implications. Being diverted is an interruption that only

some businesses can afford. Furthermore, using more data in predictive modeling has some advantages. An excessive amount of data can skew the calculation, resulting in a functional or correct result.

Future Uses/Additional Applications

This model can help you make the following decisions:

- Employee requirements, as well as their strengths and weaknesses, are assessed.
- Reduce the cost of new talent acquisition based on employee profiling and company needs.
- Analysis and evaluation of the loss of expertise and skill sets
- Measurement of financial and productivity loss because of attrition Capability to plan for and minimize loss
- Provides a thorough understanding of labor supply and demand.
- Capable of developing contingency plans based on the model's insight and foresight

Recommendations

- Companies that treat their employees with opportunities and recognition will always have an advantage in attracting and retaining top talent, whether better at attracting top talent or keeping them.
- Companies that understand this concept invest in their employees, offering them career development opportunities such as training and education and more challenging work environments to keep people engaged.
- They also value good managers by providing additional support to critical performers willing to take on more team responsibility. This helps to retain key personnel for the long term.

Implementation Plan

People prefer to work for organizations that give back, not only financially but also through increased job security or more exciting work environments.

Organizations that implement this type of employee retention strategy will be able to reduce attrition rates while retaining high-performing employees who will ultimately contribute significantly to business growth and success.

Ethical Assessment

Data Bias: Racial minorities have historically been underrepresented in a variety of jobs. So, if a data model predicts a person's suitability for a role based on past data, racial minorities will continue to face discrimination. When training a model, it is critical to select the right features so that unrelated features, such as a person's race, do not influence the predictions. The point is that unless the factors contributing to historical biases are removed, the model's predictions will continue to reflect those biases.

References

[Predict Employee Attrition Using Machine Learning & Python | by randerson112358 | Analytics Vidhya | Medium]

Predictive Analytics Models: Everything to Know About It (analyticsinsight.net)

The Challenges of Building a Predictive Churn Model (oracle.com)

Working with Predictive Data Models: A Comprehensive Guide 101 - Learn | Hevo (hevodata.com)

People Analytics with Attrition Predictions | by Natasha Sharma | Medium (towardsdatascience.com)

Predicting Employee Attrition Using Machine Learning (akkio.com)

Predictive Attrition Model: Using Analytics to predict Employee Attrition (analyticsindiamag.com)

Predicting Employee Attrition Using Machine Learning (akkio.com)

EDA and Employee Attrition Prediction | Kaggle