

# Attrition Capstone

Avery Clark

January 1, 2020

## Executive Summary

In this analysis, I used machine learning methods to build prediction models designed to predict what whether an employee will stay with the company (IBM) or will leave.

In this section I'll describe the dataset and summarize the goal of the project and key steps that were performed.

The data was provided by IBM and can be found on Kaggle here: <https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>

My goal was to build a prediction model with a prediction accuracy 88%. I surpassed that goal.

I split the data into a training set (90% of data) to train the prediction models and a testing set (10% of data) to test the accuracy of the prediction model.

After running three prediction models, the highest accuracy obtained was 0.8911565 or 89.11565%. Surpassing my goal of 88% prediction accuracy.

The most effective prediction model was "Generalized Linear Model".

This report contains four sections: Executive Summary, Analysis, Results, and Conclusion.

Executive Summary describes the dataset and summarizes the goal of the project and key steps that were performed.

Analysis explains the process and techniques used, such as data cleaning, data exploration and visualization, any insights gained, and the modeling approach.

Results presents the modeling results and discusses the model performance.

Conclusion gives a brief summary of the report, its limitations and future work.

Thank you for taking the time to look at this report. I hope that you will run this code by stepping through (by pressing Ctrl + Enter) as I'm explaining it.

## Analysis

In this section, I'll explain the process and techniques used, such as data cleaning, data exploration and visualization, any insights gained, and the modeling approach. You'll see these models in action in the Results section.

90% of the data was designated for training the prediction model and 10% of the data was reserved for testing the accuracy of that model's predictions.

A simple way of thinking about this is that the model (or algorithm) will learn about the data by taking in different factors and will make a prediction of which employees will stay and which will leave. Different approaches will have the model/algorithm using the factors given to it in different ways to make predictions.

The model/algorithm decides to predict a review rating “Y” based on factors “A”, “B”, and “C” (or more). Then the model/algorithm is exposed to the testing dataset to see if what it predicts as the review rating “Y” (based on the factors in the new dataset “A”, “B”, and “C”) is actually that accurate or not.

I hope that you will step through the code with me as I explain it.

You can run all of the code by clicking Run. You can run it line by line by pressing Ctrl + Enter on your keyboard. You can also highlight a section of code and run just that by clicking Run or pressing Ctrl + Enter on your keyboard.

Let’s dig in!

These next lines will install what is needed to run the code and will skip what your system already has installed.

## Conclusion

In this section I’ll give a brief summary of the report, its limitations and future work.

I split the data into a training set (90% of data) to train the prediction models and a testing set (10% of data) to test the accuracy of the prediction model.

When I tried to reach a higher accuracy level by using only some columns that had proven to be significant in early tests, my accuracy actually decreased. So I let each type of analysis decide for itself which predictors to include from the entire list.

After running three prediction models, the highest accuracy obtained was 0.8911565 or 89.11565%. Surpassing my goal of 88% prediction accuracy.

The most effective prediction model was “Generalized Linear Model”.

I feel as though my report has some limitations. I could have taken more modeling approaches to potentially reach a higher prediction accuracy.

I would like to improve this analysis in the future by finding some prediction model approaches that will give me a prediction accuracy of greater than 93%.

Thank you for reading my report. I hope you enjoyed it. - Avery Clark