DIONDRA STUBBS

IBM EMPLOYEE CHURN

Prediction & Analysis

Exploring and predicting IBM employee churn using Python.

YOUR ASSIGNMENT

The latest company performance report has been released and it looks like the employee retention rate has reduced to 84%.

Your assignment is to discover what the key drivers are for employees churning and predict the attrition of IBM employees. Others on your team suggest that a decision tree or Gaussian NB is best, but your task it to fit another classifier and compare it to the ones your team drafted. You are expected to use exploratory data analysis, visualizations and accuracy scores to make recommendations for IBM to increase employee retention and choose the best performing model.

MEASUREMENT PLANNING

What are the high-level goals of IBM?

Employee Retention

The company wants to keep their valuable employees. By addressing this goal we can look at ways that can be done.

Produce Revenue

IBM is best known for producing and selling computer hardware and software, as well as cloud computing and data analytics. In order to produce the best products they need their best workers. More reason to understand the employee churcn

Employee Satisfaction

Are employees satisfied at their job? Can it be improved? How?

IBM GOALS

Maintain Employee Base

How can we keep our current employee base

Focus Retention Efforts

How can we improve our retention percentage?

Predict employees at risk of churning

We need to predict churning so we can prepare in advance.

DATA PREP & PROCESS

We are professionals providing today's digital marketing and advertising services.

1. Collect the Data

IBM employee database & survey

2. Prep & Cleaning

Checking for and removing nulls

3. EDA

Why is there attrition?
Understanding the data.

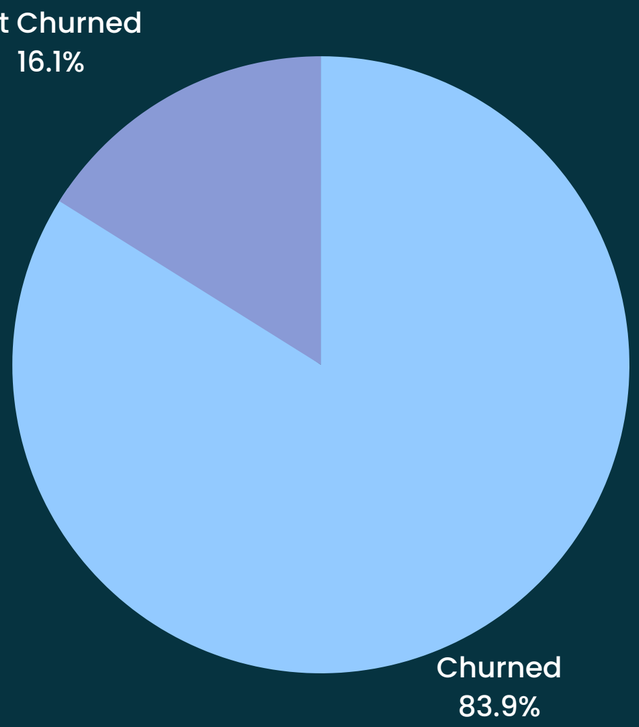
4. Modeling Builing

Coomparing 3 Classifier models

THE NUMBERS

How many employees have stayed?

83.9% (1233) of IBM's employees have stayed

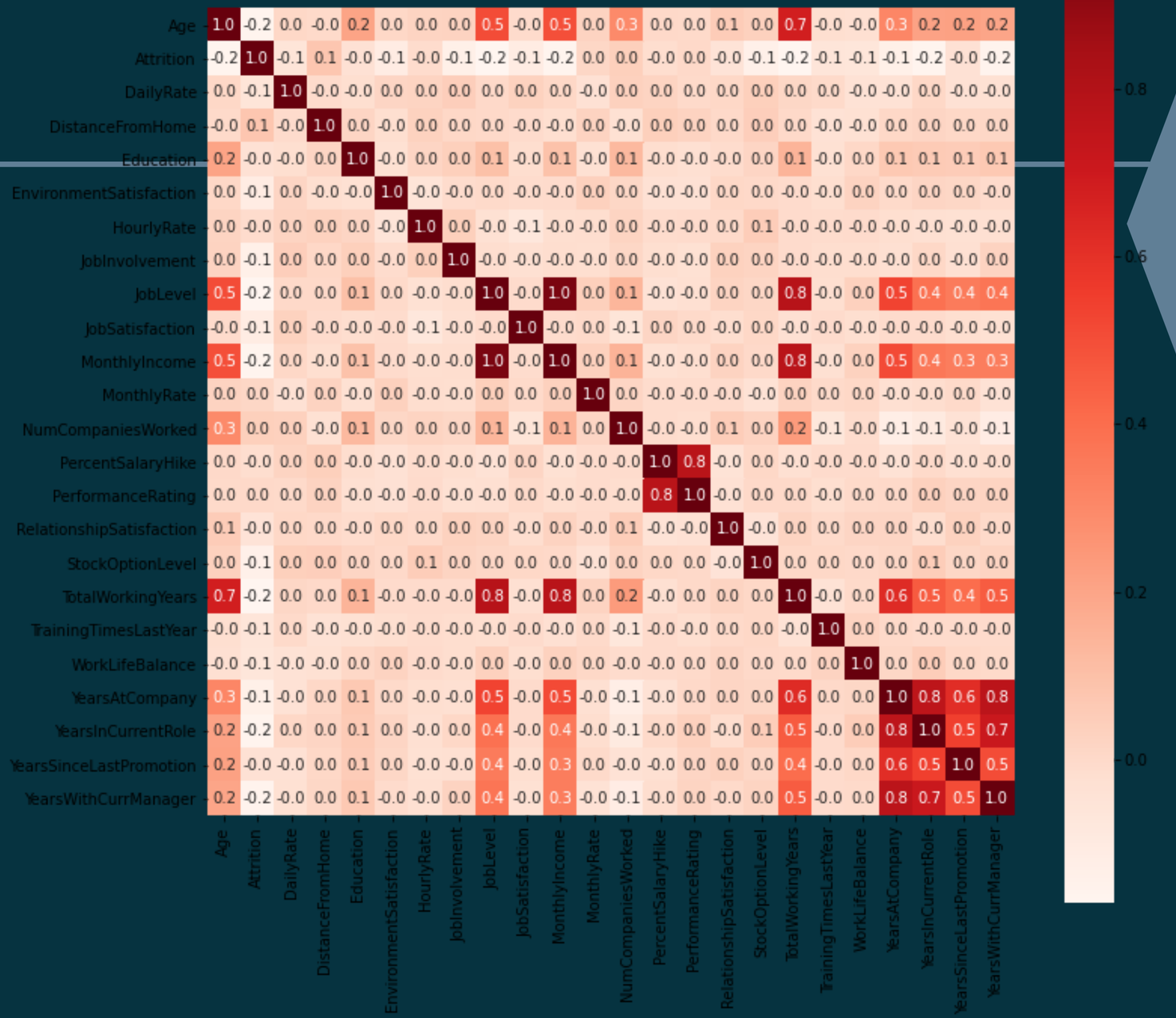


How many have left?

16.1% (237) have left.

FEATURES

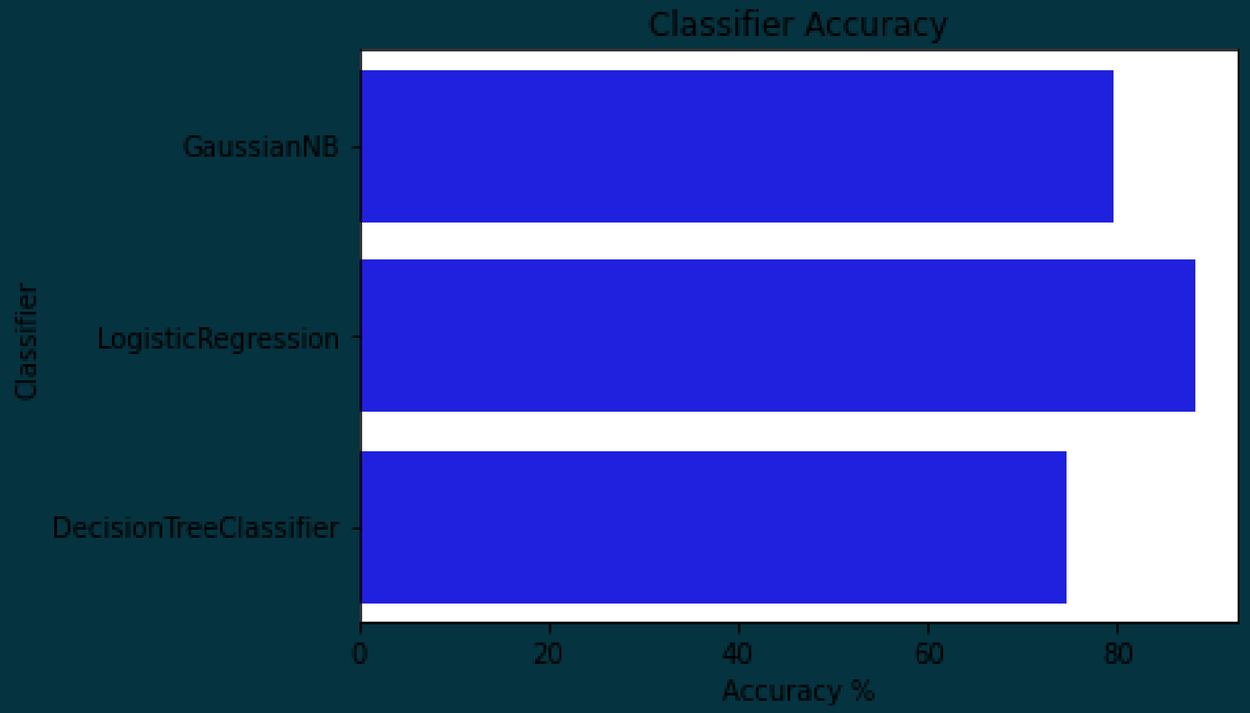
- ***Age and TotalWorkingYears are highly correlated. This would assume that older employees have worked longer.***
- ***MonthlyIncome and TotalWorkingYears are highly correlated insinuating that an employee's monthly income increases as they work longer***
- ***PerformanceRating and PercentSalaryHike are also highly correlated. Higher performance ratings leader to salary increases***
- ***YearsInCurrentRole is highly correlated with YearsAtCompany and YearsWithCurrManager and moderatly correlated with YearsSinceLastPromotion. This insinuates that many employees remain in their current role under the same manager overtime and there isn't much opportunity for promotion***



So what seems to be the key drivers of employee churn? Lack of promotion.

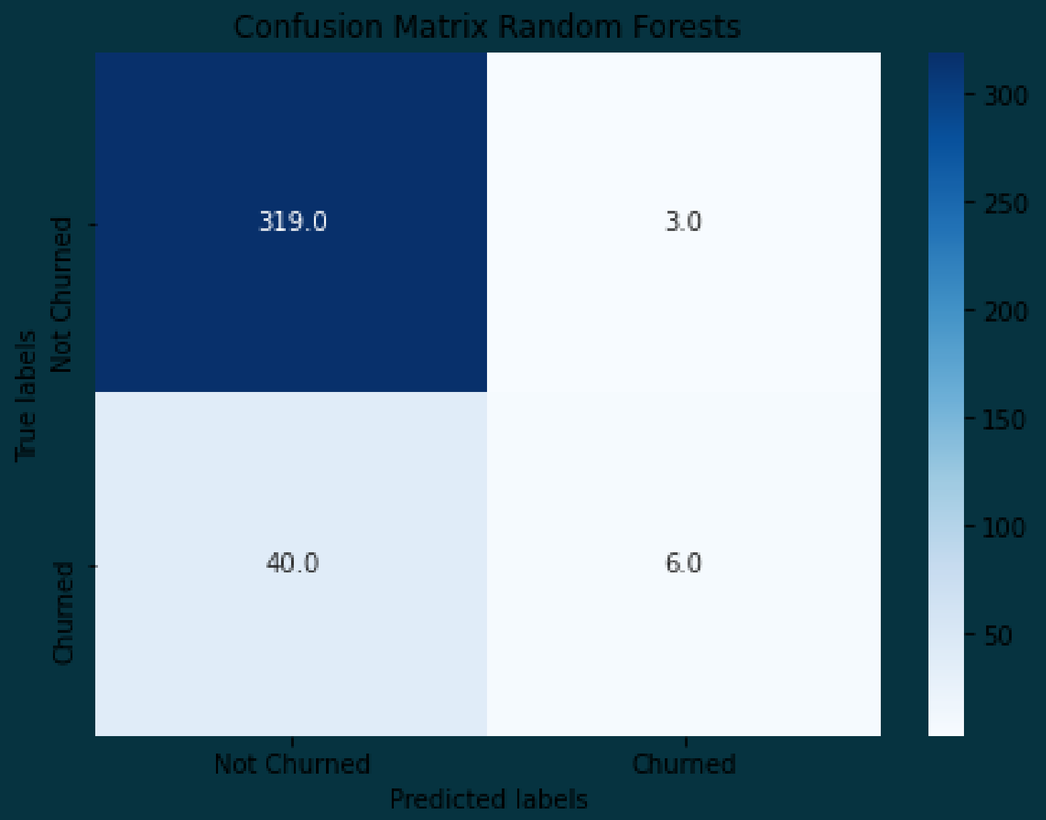
MODEL COMPARISON

So our results show that the Gaussian NB model gives ~80% accuracy, the Logistic Regression model gives ~88% accuracy and the Decision Tree model gives ~75% accuracy.



MODEL COMPARISON

he dimension of this matrix is 2*2 because this model is binary classification. The classes of the Attrition variable are 0 and 1. 88% (325/368) of the IBM employees were correctly classified as churned and not churned.



RESULTS & RECOMMENDATIONS

The Logistic Regression classifier is best for predicting employee churn

The key driver of employee churn is lack of promotions. IBM can improve retention among their employees by implementing more opportunities for promotion and growth within the company.