# Decoding Employee Attrition: A Deep Dive into HR Analytics Using Machine Learning

Introduction:

Employee attrition, or the rate at which employees leave a company, presents a significant challenge for businesses today. It's a costly issue that affects productivity and overall organizational performance. However, thanks to advancements in machine learning and HR analytic, we now have the tools to predict and mitigate employee attrition. In this blog post, we'll explore how machine learning can be applied to HR analytic to effectively address the attrition rate of employees.

Understanding Employee Attrition:

Before delving into the application of machine learning, let's first understand what employee attrition means. Employee attrition refers to the number of employees who leave a company within a specific time frame. This can be influenced by various factors such as job dissatisfaction, limited growth opportunities, poor work-life balance, and inadequate compensation. By analyzing historical data and identifying patterns, machine learning models can help organizations gain insights into these factors and predict the likelihood of attrition.

Data Collection:

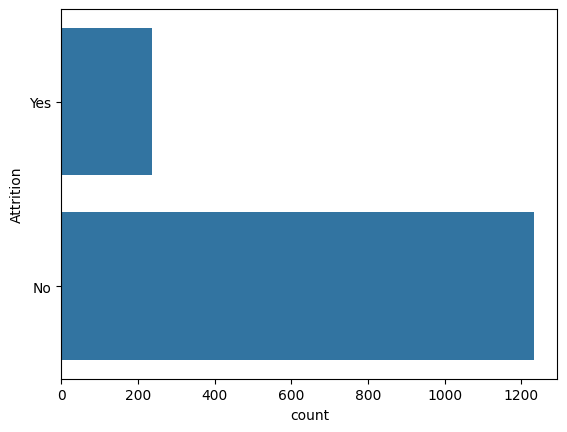
The Data is secondary and is taken from the link below:

https://github.com/dsrscientist/IBM\_HR\_Attrition\_Rate\_Analytics

Exploratory Data Analysis:

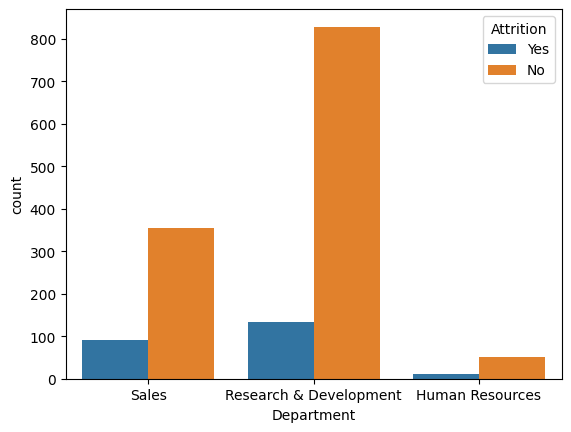
Initial Observations:

1. No null or duplicate values present in the data set
2. A classic binary classification problem with attrition column being the target
3. Raw Data has shape (1470,35)



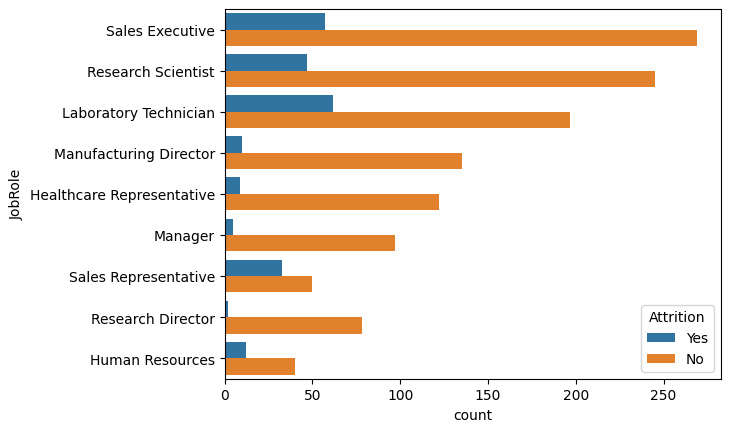
We have more data of employee attrition than non attrition which makes our data imbalanced.

**Question : Which department has highest attrition rate?**



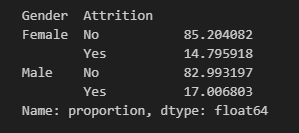
The Sales Department has the highest attrition rate, which could be due to intense competition and employees seeking better opportunities. On the other hand, the Research & Development Department has the lowest attrition rate.

**Question : Which job role has more attrition rate?**



Lab Technicians followed by Sales Executive job roles has highest rate of attrition which is understandable as these are entry level jobs and after certain serving period the employee would seek for increase in pay scale which is either matched by the organization or doesn’t. In the later case one tends to leave from the current company.

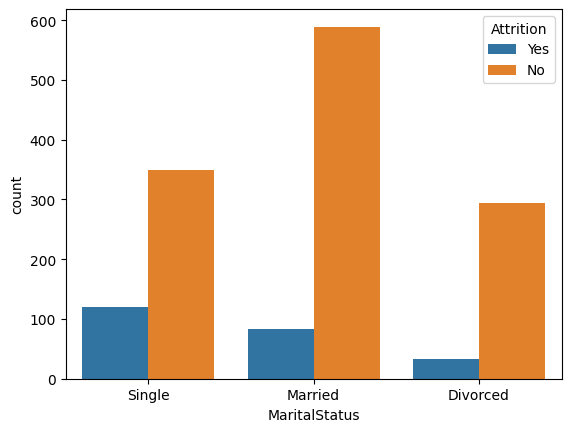
**Question : Does gender play a role in attrition rate?**



Apparently not!

There is not much difference in attrition rate when genders are taken into consideration.

**Question : Check whether Martial Status status affect the attrition rate?**



It is understandable that employee who are single has higher percentage chance of switching the company as most of these employees are young adults and can afford to take risk in their careers opposed to people who are married who seeks for stability before anything else.

Data Prepossessing:

Data has categorical and nominal features which needs to be encoded first.

**Categorical:**

1. BusinessTravel
2. Department
3. EducationField
4. JobRole
5. MaritalStatus

**Nominal:**

1. Gender
2. Overtime

Now as our dataset is imbalanced we will use a sampling technique to even out the odds. In this case we used **Random over sampler to achieve out goal.**

**Why Oversampling?**

As our dataset is a small one reducing data points can lead to significant lose of data which would be essential when getting into model training.

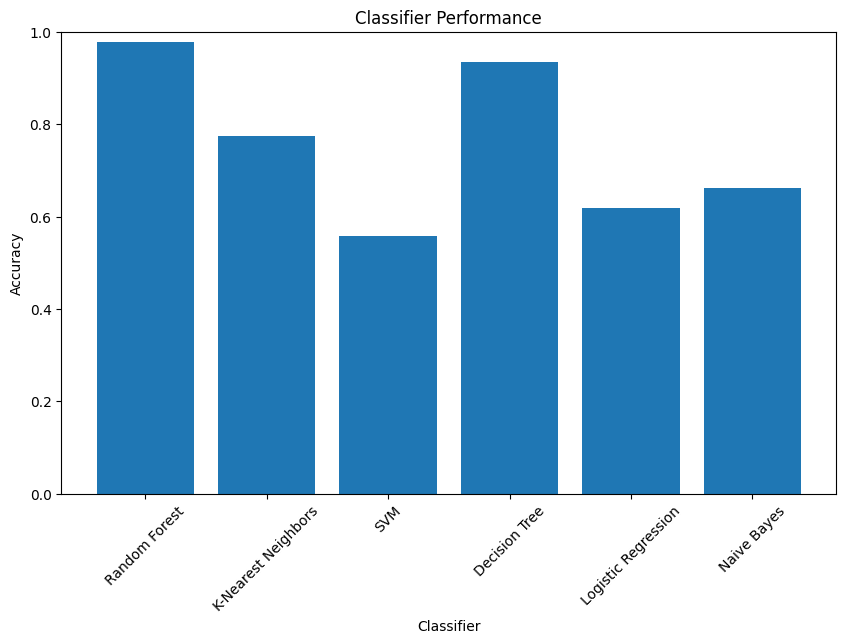
After Oversampling:

Train data shape: (1972, 33)

Test data shape: (494, 33)

**Model Selection and Training:**

With the preprocessed dataset it's time to select an appropriate machine learning algorithm which fits the best. There are several algorithms to consider, such as logistic regression, decision trees, random forests, and gradient boosting and more.



Looking at the accuracy Random forest classifier works the best in our case.

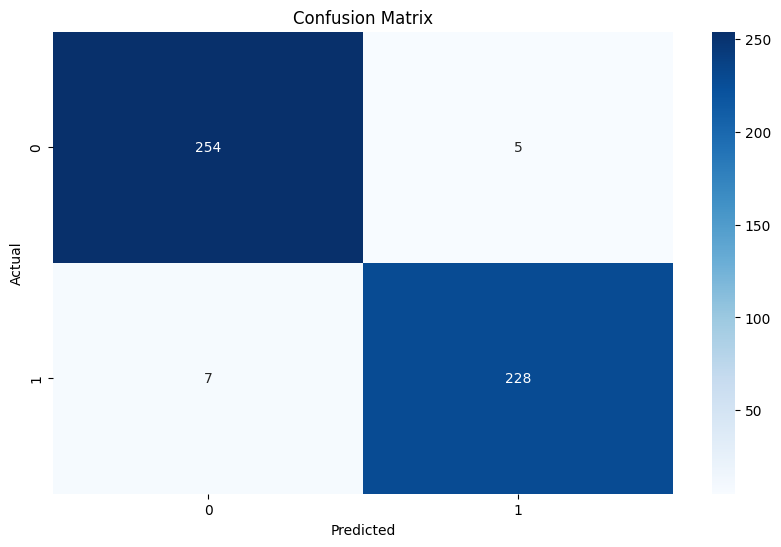
**Now Hyper parameter Tuning random forest classifier for best model creation.**

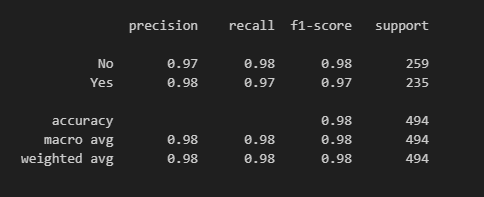
Best Parameters: {'max\_depth': None, 'n\_estimators': 100}

Best Score: 0.9644939921608943

Test Accuracy: 0.9757085020242915

**Confusion matrix and Classification Report of the best working model**





**Outcome:**

The classification model demonstrated impressive results in accurately predicting whether employees would leave or stay in the company. It exhibited a strong ability to classify employees correctly, with a precision score of 0.97 for the "No" class and 0.98 for the "Yes" class. Additionally, the model achieved high recall scores of 0.98 for the "No" class and 0.97 for the "Yes" class, indicating its effectiveness in identifying the majority of instances accurately. The F1-score, which considers both precision and recall, yielded excellent results of 0.98 for both classes. Collectively, these findings indicate that the model performed exceptionally well in accurately predicting employee attrition.

We can use this model to predict the attrition rate for new and unseen employee data. By inputting relevant employee information into the model, organizations can obtain attrition predictions with associated probabilities. These predictions help HR professionals identify employees at high risk of attrition and take proactive measures to retain them.

To conclude, machine learning algorithms have the potential to transform HR practices by enabling organizations to predict and reduce employee attrition rates. By leveraging HR analytics and data-driven insights, companies can address attrition-related challenges and implement effective retention strategies.

However, it is important to remember that machine learning techniques should be used alongside human expertise and experience to achieve the best results. By continuously improving and using data-driven insights, businesses can create a positive workplace environment that encourages employee loyalty and contributes to long-term success.