**Predicting Employees who are likely to churn**

**Abstract:** Using the dataset about employee churn, I built a simple model using Random Forest Classifier to predict the employees who are likely to churn or leave the organization. The model had an accuracy score of 0.992.

**Keywords:** Employee, predict, churn

**Introduction:** Employee churn is an important topic for many senior leaders in organizations. For some employees, they join organizations for financial reward, access to high networks, professional growth among others. However, some employees also leave their company for reasons such as: pursuing further opportunities, lack of motivation, slow or no growth experiences, too much work among many others. For senior leaders, the lower the turnover rate, the better it is for the company’s image and its ability to attract high performing employees. Managers rely on the power of predictive analytics to help them predict employees who are likely to leave the company so that appropriate measures can be put in place to prevent the churn from happening. In this project, data on current employees and employees who churned were provided. In total, it has about 15000 data points. Each data points represented an employee.

**Project Objective:** The goal of this project is to use techniques in data science to investigate what type of employee is likely to churn or leave the company.

**Exploratory Data Analysis:** The dataset was loaded into a Jupyter Notebook and analysis were done using python and other libraries written in python. The following figures were plotted:

A picture containing drawing, room

Description automatically generated

Figure : Count plot of categorical columns

A picture containing room

Description automatically generated

Figure : Boxplot of Churned Employees and other columns

A picture containing clock

Description automatically generated

Figure : Satisfaction Level vs Number of Project

A screenshot of a cell phone

Description automatically generated

Figure : Satisfaction Level vs Promotion

A picture containing drawing

Description automatically generated

Figure : Number of years vs Promotion

**Figure 1** shows the count plot of all categorical variables in the dataset. One of the notable insights from these plots is that the count plot of *time\_spend\_company* is right skewed. This means that majority of the employees have spent four or a smaller number of years in the company.

**Figure 2** shows plots of various columns with the *churned* column. Some of the notable insights:

* Churned employees have lower satisfaction levels
* Churned employees usually work more hours per month
* Churned employees spend relatively lesser number of years in the firm than those who did not churn.
* Churned employees usually work on more projects than those who did not churn.

**Figure 3** shows the satisfaction level vs the number of projects. We observe that satisfaction levels are low for employees with 2 projects. Satisfaction levels are also low for employees with 6 or more projects. This means that optimal number of projects for employees are from 3 to 5. This is because employees who had number of projects from 3 to 5 had the most satisfaction levels.

**Figure 4:** In this figure, we observe that employees who are promoted have a median satisfaction level higher than those who were not promoted.

**Figure 5:** In this figure, we observe that employees who were promoted tend to stay longer in the company than those who were not promoted.

**Model Selection:** In figure 1, we see that the proportion of churned employees to non-churned employees is not balanced. 76% of our dataset were current employees and 24% were employees who churned. This poses a risk of our model being biased towards employees who did not churn. Due to this reason, I selected Random Forest Classifier algorithm because it performs very well in scenarios like this.

**Predictions:** I divided the dataset into 70% training and 30% test. After fitting the model, I got an accuracy of 99%. Following that I used the current employee dataset to predict which employees are likely to churn. I added the resulting column to the data frame and plotted the proportion of employees likely to churn. Figure below.

A picture containing drawing

Description automatically generated

Figure : Employees likely to churn or not

**Recommendations:** From the exploratory data analysis and the variables for prediction, we can see that the top two features that account for employee churn:

* Satisfaction – Employees who are less satisfied are more likely to churn
* Number of projects – Employees with 2, 6 or more project are more likely to churn

Based on this, I recommend that management:

* Put in policies and mechanisms to improve employee satisfaction in the workplace.
* Do not overload employees with too many projects as it might lead to burnout.
* Employees should be working on 3 to 5 projects for optimal satisfaction.
* Management should institute promotion policies that rewards high performing employees as employees who are promoted tend to stay in the company longer and are less likely to churn.

**Conclusion:** In this project, I built a model to predict the employees who are likely to churn in the company using Random Forest Classifier Algorithm. I also identified the type of employees who are likely to churn and advised management on ways to reduce employee churn.