FA541 Project - HR attrition rate analysis and prediction

Jinxin Qu

1. **What is the data that will be used for your problem?**

The data set I am going to use is from Kaggle.

1. **Please provide a sample of a few rows and talk about the variables in the data set**

In the original data, there are 35 columns which include 34 independent variables and 1 dependent variable.

The first five rows are as following:

Graphical user interface, application

Description automatically generated

The dependent variable is “Attrition” and the independent variables include age, daily rate, distance from home…

Table

Description automatically generated

1. **Discuss what would be the variable(s) your model would predict**

In this project, I want to explore features and factors that could determine if a candidate will work for the company or leave the company. In other words, why workers quit?

1. **Discuss some hypothetical approaches to the original idea**

I will first do some exploratory analysis about the dataset and then I will use the following ML models to approach the problem:

1. Logistic regression

2. Random Forest

3. Gradient Boosting

1. **If needed, please formulate other questions that may be answered using the data you have.**

What type of measures should the company take in order to retain their employees?

1. **What other data sets could you use to answer your questions?**

I could benefit from more HR data since my dataset have a relatively smaller size.