Project Proposal – Employee Attrition Analysis

Xinlian Huang

Instructor: Prof. Mahmoud Daneshmand

I. Background Introduction

Attrition in human resources refers to the gradual loss of employees over time. Generally, relatively high employee attrition is problematic for companies. The cumulative cost of investing in employees can be relatively enormous because of training and time that it takes starts from rookie phases until employees become productive members. Obviously if HR does not pay attention in employee management to avoid the attrition, the profit of the company would definitely be affected in the long run. In order to find out which specific population is more likely to leave the company and why these people are leaving, this project focuses on those pressing issues and aims at providing HR some insights in managing a number of employees. There is a company with the issue of losing their most brilliant and experienced employees, so this project would make use of their employees' data to analyze and predict the trends of attrition.

II. Research Question

The company is facing an urgent phase: their valuable and experienced employees are leaving prematurely. They do not want

to lose more excellent employees and find out the solution source towards attrition. Therefore the real word research questions are:

- Why some of the best and experienced employees are leaving?
- How to predict which valuable employee will leave next?

III. Data Source

The data is simulated and comes from:
https://www.kaggle.com/colara/human-resource/data

Sample raw dataset:

	A	B	C	D	E	F	G	H	I	J
1	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	promotion_last_5years	department	salary	left
2	0.38	0.53	2	157	3	0	0	sales	low	1
3	0.8	0.86	5	262	6	0	0	sales	medium	1
4	0.11	0.88	7	272	4	0	0	sales	medium	1
5	0.72	0.87	5	223	5	0	0	sales	low	1
6	0.37	0.52	2	159	3	0	0	sales	low	1
7	0.41	0.5	2	153	3	0	0	sales	low	1
8	0.1	0.77	6	247	4	0	0	sales	low	1
9	0.92	0.85	5	259	5	0	0	sales	low	1
10	0.89	1	5	224	5	0	0	sales	low	1
11	0.42	0.53	2	142	3	0	0	sales	low	1
12	0.45	0.54	2	135	3	0	0	sales	low	1

Features explanation:

1. Satisfication_level:
Scaled in 0-1, the bigger the number is, the more satisfied the employee feel about the job.

2. Last_evaluation:
Time since last performance evaluation, scaled in 0-1, the bigger the number is, the time since last evaluation is longer

3. Number_project:
Number of project completed while at work

4. Average_montly_hours:
Average monthly hours at workplace

5. Time_spend_company:
Number of years spent at company

6. Work_accident:
Whether the employee had a workplace accident, 0 stands for no, 1 stands for yes

7. Left:
Whether the employee left the company or not, binary variable, 0 stands for still active at company, 1 stands for left the company.

8. Promotion_last_5years:
Whether the employee was promoted in the last five years, 0 stands for no, 1 stands for yes

9. Department:
Department in which they work for

10. Salary:
Relative level of salary (high/medium/low)

IV. Machine Learning Approach

Basically, the target variable is binary: whether the employee left the company or not, so Decision Tree and Logistic Regression fit the situation here. In this project, I will utilize both CART (Classification & Regression Tree) and Logistic Regression to build the model and compare the accuracy by examining the confusion matrix.

V. Real Word Impact & Application

First and foremost, model could be executed to get the insights on employee population with high attrition probability. HRs could take corresponding measures to manage this particular population.

Furthermore, by exploratory data analysis, we will know the patterns of employee attrition in advance so that company could adjust its strategy of employee recruitment and consider organization management.