M1 Presentation

BUSINESS SCHOOL
AALBORG UNIVERSITY

Kasper R. Haurum, Md. Raiyan Alam, Mike Christensen, Snorre K. Brouer.

BUSINESS SCHOOL
AALBORG UNIVERSITY

Problem Definition

Project Overview

Employee turnover is a costly problem for companies. In the HR dataset, we will find out the reason for employee turnover and the best possible reasons.
We will focus on the key factor of the dataset which can show us the best possible answers.
We will use Exploratory Data Analysis (EDA), Unsupervised Machine Learning (UML), & Supervised Machine Learning (SML).
These models will be used throughout the project for descriptive and predictive analysis.

Problem Statement.

In this study, we will attempt to solve the following problem statements:

- What is the likelihood of an active employee leaving the company?
- What are the key indicators of an employee leaving the company?
- What policies or strategies can be adopted based on the results to improve employee retention?

EDA

BUSINESS SCHOOL
AALBORG UNIVERSITY

Let's summarise EDA:

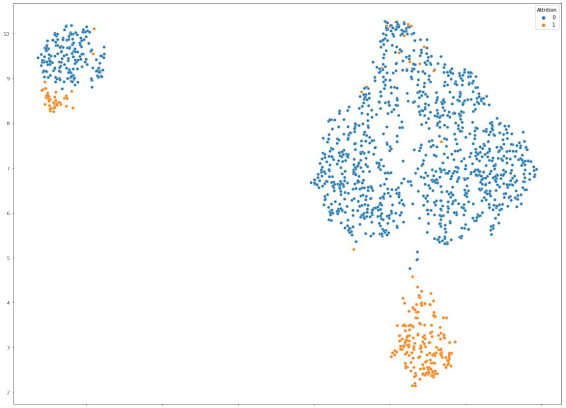
1. The dataset does not feature any missing or erroneous data values, and all features are of the correct data type.
2. There were some duplicate values. We removed it for our EDA, UML & SML. With duplicate value there were different sort of results (Every row had 2 exactly same rows. Befor removing duplicate values we measured the attrition value like #yes and #no percentage. After confirming the same weight of yes and no of duplicate values, we just removed that from our data set).
3. We removed the NAN values from some columns of the data set.
4. The strongest positive correlations with the target features are: Performance Rating, Monthly Rate, Num Companies Worked, Distance From Home.
5. The strongest negative correlations with the target features are: Total Working Years, Job Level, Years In Current Role, and Monthly Income.
6. The dataset is imbalanced with the majoriy of observations describing Currently Active Employees.
7. Several features (ie columns) are redundant for our analysis, namely: EmployeeCount, EmployeeNumber, StandardHours, and Over18.

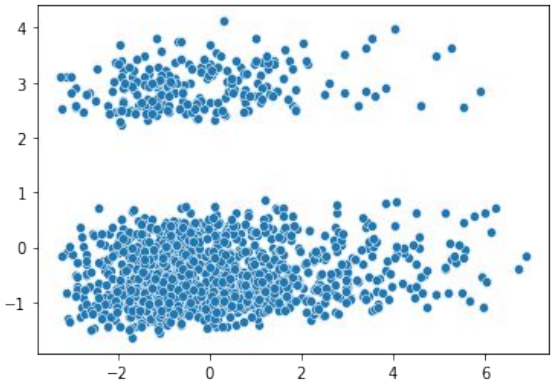
Unsupervised ML

BUSINESS SCHOOL
AALBORG UNIVERSITY

- In the UML part we used two different models:

  - PCA model
  - UMAP Model





Supervised ML

BUSINESS SCHOOL
AALBORG UNIVERSITY

- In the SML part we used 5 different models:

  - Baseline Algorithms
  - Logistic Regression & Confusion Matrics
  - Random Forest Classifier (RFC) & Confusion Matrics
  - XGBClassifier & Confusion Matrics
  - Elastic Net

Conclusion

BUSINESS SCHOOL
AALBORG UNIVERSITY

Strategic Retention Plan
The stronger indicators of people leaving include:

1. People on higher wages are less likely to leave the company.
2. A large number of leavers leave 1.5 year after being under their Current Managers. .
3. Age: Employees in relatively young age bracket 25-35 are more likely to leave.
4. DistanceFromHome: Employees who live further from home are more likely to leave the company.
5. TotalWorkingYears: The more experienced employees are less likely to leave.