

EAS 509: Statistical Learning II

Group Project Report

GROUP 12:

Kishan Patel (kpatel39)
Geethika Bedadhala (gbedadha)
Ish Kumar (ishkumar)

TITLE: *Employee Turnover Prediction*

Goal: *Using Employee churn dataset to predict an employee's risk of quitting with Survival Analysis Model.*

Report

This report presents a survival analysis model to predict an employee's risk of quitting based on various factors. The dataset contains information on employee experience, gender, age, industry, profession, traffic source, coach, wage type, commuting mode, and personality traits.

Exploratory data analysis reveals that there is no strong correlation present between all the attributes, and that employees with experience less than 50 months tend to dominate the dataset, and employees in the age range of 27-30 are more likely to quit.

Principal component analysis (PCA) is performed to identify underlying patterns and relationships between variables. However, the elbow method shows that almost all principal components are required to capture a significant amount of variance.

Cluster analysis is performed to group employees based on their characteristics. Three clusters are identified based on gender, age, wage, personality traits, supervisor gender, and quitting probability. We observed that cluster with higher female employees with lower wage and high innovation and extraversion scores tend to quit more compared to other clusters.

Kaplan-Meier curves and log-rank tests are performed to determine the survival probability and the impact of gender, profession, industry, way of transportation on employee churn. The results suggest that the survival rate decreases with time, the median survival time is around 50 months, and gender does not significantly affect survival rates. However, profession does have a significant impact on survival rates, with employees in IT, Law having a steeper survival rate curve compared to others. Similarly, industry we can say that employees from Real Estate, Telecom and Retail industry do not have a higher risk of quitting compared to other industries. Way of transportation and wages also had a significant impact on the survival rate of the employees.

Finally, Cox-proportional hazard models are fitted to identify the variables significantly associated with the survival outcome. The coefficients and p-values of different industries are also analyzed,

and the results show that agriculture industry has the highest impact on survival rates. Profession does have a significant impact on survival outcome, with employees in management, marketing, consulting, engineering, and teaching having a higher risk of quitting compared to others.

In conclusion, this study shows that survival analysis can be an effective method to predict employee churn. The factors that impact employee churn can be identified using various statistical techniques, including Kaplan-Meier curves, log-rank test, and Cox-proportional hazard models. These insights can help organizations take proactive steps to retain valuable employees and improve their overall performance.