Presentation Notes

SEAGATE EMPLOYEE CHURN PROJECT

Submitted by:

Charles Biddle Porter
charles.biddleporter@colorado.edu
Harshit Gole
harshit.gole@colorado.edu
Manan Dhanteja
Manan.dhanteja@colorado.edu

Duc Hoang
duc.hoang@colorado.edu
Kirti Vatsh
kirti.vatsh@colorado.edu

TABLE OF CONTENTS

INTRODUCTION

The Society for Human Resource Management states that for corporations the replacement cost is 50% - 60% with overall costs lying anywhere from 90% - 200%. Such staggering numbers is one of the main reasons for Seagate HR's drive to leverage predictive analytics and machine learning to enhance their strategic role and provide insights on hiring and employee churn to business leaders.

Our team here will now present how it can be executed.

MISSION STATEMENT

Our mission is to develop a specialized data analysis framework for Seagate that meticulously identifies and analyzes the factors leading to voluntary employee churn. This framework is engineered to assess the financial repercussions of turnover on the organization and forecast a 2-year hiring pipeline, strategically optimizing Seagate's workforce for future challenges and growth. By integrating diverse HR data streams—from performance evaluations to exit interviews—our goal is to equip Seagate with actionable insights that enable precise workforce planning and enhance overall organizational resilience.

DATA ANALYSIS

As our main objective was to develop a predictive model, we began with 6 different machine learning techniques including Logistic regression, KNN-5, KNN-10, Ada Boost, Bagging Classifier and Random Forest. After developing and testing the models we found Random Forest provided us the best accuracy (82%) and was most suited to the task when drawing in other considerations such as F1 score and AUC. A Random Forest model is a type of ensemble learning that creates multiple decision trees during training and selects the most frequent outcome among these trees to make predictions. Its design effectively prevents overfitting, a common problem with regular decision trees, enhancing its generalization capability across various datasets. This makes it particularly useful for addressing employee turnover issues, especially when overfitting was a prevalent problem in preliminary models using other methods. Our performance metrics are represented in a confusion matrix, which indicated the model correctly identified employees who did not churn 2917 times and those who did churn 922 times, with 310 false positives and 543 false negatives. This level of accuracy is essential for formulating effective strategies to retain employees. Using our model, we can predict the upcoming voluntary churn and develop a hiring pipeline for HR to use to reduce time without roles filled.

MODEL

As our main objective was to develop a predictive model, we began with 6 different machine learning techniques including Logistic regression, KNN-5, KNN-10, Ada Boost, Bagging Classifier and Random Forest. After developing and testing the models we found Random Forest provided us the best accuracy (82%) and was most suited to the task when drawing in other considerations such as F1 score and AUC. A Random Forest model is a type of ensemble learning that creates multiple decision trees during training and selects the most frequent outcome among these trees to make predictions. Its design effectively prevents overfitting, a common problem with regular decision trees, enhancing its generalization capability across various datasets. This makes it particularly useful for addressing employee turnover issues, especially when overfitting was a prevalent problem in preliminary models using other methods. Our performance metrics are represented in a confusion matrix, which indicated the model correctly identified employees who did not churn 2917 times and those who did churn 922 times, with 310 false positives and 543 false negatives. This level of accuracy is essential for formulating effective strategies to retain employees. Using our model, we can predict the upcoming voluntary churn and develop a hiring pipeline for HR to use to reduce time without roles filled.

HR INSIGHTS

Significant insights in the final analysis were the expected two-year employee churn by Country, Region, Department. We found that Employees are 1.8x more likely to churn if they work in the Asia Pacific Region than the next closest.

When looking more closely we found that Thailand had the highest expected churn, beating the USA in second place by 31%.

On an even finer review, we find that Engineering teams are the most likely to experience churn considerably beating the next closest department by 114%. This means that a developed hiring pipeline will likely draw more resources in these areas. A next step could be to examine closer why people are leaving those teams and how you could reduce their churn.

Another key insight from our work was the total cost of our expectations forecasted by the model. We found voluntary churn would cost Seagate $43.6 million over the next two years in hiring associated expenses. We achieved this by multiplying the base pay by the cost-to-replace multiplier for those individuals with 42% or higher probability of leaving in the next 2 years. This number is relevant as it allows Seagate to budget for the expected churn more effectively, in addition to quantifying the economic impact of churn.