Denise Dodd

Predicting Attrition

Proposal

TOPIC

This project will predict employee attrition and identify key factors that contribute to attrition.

BUSINESS PROBLEM

“One estimate places the cost to replace an employee at three to four times the position’s salary” (Wallace, 2023). When an employee leaves a company, the business must invest time and money to post a job, filter through resumes, conduct interviews, hire and onboard a new employee, train the replacement, and transfer duties to the new hire. Additionally, the historical knowledge that the original employee had is a resource that the company cannot regain. Therefore, it is to the company’s benefit to determine what contributes to attrition and predict if an employee will attrite so they deploy mitigating efforts, budget, and plan accordingly.

DATASET

For this study, I will be utilizing a dataset titled “Employee Attrition” (Kaggle, 2022). This is a dataset of about 4,500 employees of which roughly 16% are no longer with the company. There is a mix of 24 numerical and categorical variables. Numerical variables include data such as the employee’s age, commute distance, years with the company, and salary. Categorical variables include what department the employee works in, marital status, job role, and gender.

METHODS

After reviewing and cleaning the data, I will create dummy variables of the categorical columns. This will ensure that the classification models can accurately process the data. At this point I will determine if a scaler is necessary. Once the data has been prepared, I will assign the target variable to be “Attrition” and the features variables to be the rest of the dataframe. Training and testing variables will be created using an 80/20 split of the features and target data. The training data will be used to fit several classification models and predictions will be generated. I will analyze the effectiveness of these models by reviewing Precision, Recall, and Accuray scores. Once a final model has been selected, I will undergo any necessary hypertuning.

To visualize the model’s classification success, I will create a heatmap version of a confusion matrix. I will also generate a Precision-Recall Curve in lieu of an ROC curve due to the imbalance of attrition rates in the dataset (16% attrition vs 84% no attrition).

Finally, I will determine the 10 features that impact the model the most and turn these results into a bar graph. This will provide insight to the business of what factors they will want to focus on to improve attrition rates

ETHICAL CONSIDERATIONS

The company will want to be sure that they do not focus on the attrition rate to the point of discrimination. For example, if the company wants to hire someone who will be with the company for the long term but one gender attrites more than another or employees over or under a certain age tend to attrite more, discrimination laws prevent these factors from being considered when making a hiring decision.

CHALLENGES/ISSUES

A personal challenge for me, and the reason I chose this project, is that I do not have much experience predicting categorical or binary data. Most of my previous projects have involved predicting a continuous numerical variable such as price, distance, salary, etc. Therefore, dummy variables, classification models, and their associated metrics will require additional research on my end.

A challenge with the dataset is that I do not have any information about the company that this data comes from. For example, where is the company located? It is possible that outside factors not involved with the business could cause attrition rates for a New York company to be different from a company in Des Moines, IA. Also, what field of business does this company operate in? I have worked in both education and finance and the attrition rates and factors leading to attrition can vary between the two fields. The project I am undergoing applies directly to this company and this dataset. The findings in this study cannot be applied universally across all fields of business in every location.

REFERENCES

There are two articles that lend interesting perspectives to attrition which will likely be referenced in my presentation.

1) Five Hidden Costs Of Employee Attrition (Wallace, 2023) - This article details the expense of time and money that goes into processing a termination, searching for a replacement, and training the new employee. It also references additional factors such as loss of knowledge and the effect on the remaining employees when there is attrition.
2) 5 Key Factors Impacting Employee Attrition and Retention (Perceptyx, 2024) - This article offers insight into what causes attrition. It will be interesting to see if the factors referenced in this article are the same factors impacting attrition for the company in this study.

CITATION

Kaggle. (2022, November 3). *Employee attrition*. Kaggle. https://www.kaggle.com/datasets/ajayganga/employee-attrition

Perceptyx. (2024, February 6). *5 key factors impacting employee attrition and retention*. https://blog.perceptyx.com/5-key-factors-impacting-employee-attrition-and-retention

Wallace, L. (2023, March 21). *Forbes EQ Brand Voice: Five hidden costs of employee attrition*. Forbes. https://www.forbes.com/sites/forbeseq/2023/03/21/five-hidden-costs-of-employee-attrition/?sh=482ecd3562f4