Employee Attrition and Specialized Insights

P14 - Boscosylvester John Chittilapilly, Shlok Vivek Naik, Saksham Pandey
Department of Computer Science, North Carolina State University,
Raleigh, NC, 27695
bchitti, snaik2, spandey5@ncsu.edu

1 Data Set

https://www.kaggle.com/datasets/jpmiller/employee-attrition-for-healthcare
The dataset consists of over 1600 rows with 35 attributes for the data points.

2 Project Idea

Employee attrition is a serious problem faced by many companies, as the organization has to find an adequate replacement and make sure that there is no delay in the current deliverables. A solution to this problem is to identify employees who are likely to leave the workforce. The above mentioned dataset has several attributes which could be used to perform comparative analysis for detection of employees who might leave the company.
When an employee leaves the company we propose to find out at-risk employees, using clustering algorithms. The dataset has attributes such as employee experience, number of companies worked prior, education level etc. and could also be used to estimate the initial salary for a new employee. Thus, this tool would be useful for HR teams in handling current employees, rolling out new offers to prospective candidates and aiding organizations in analyzing different factors and the extent to which they impact employee attrition.

3 Software to Write

Python programming language, machine learning and data analysis/visualization libraries

4 References

[1] Shankar, R. S., Rajanikanth, J., Sivaramaraju, V. V., Murthy, K. V. S. S. R. (2018, July). Prediction of employee attrition using data mining. In 2018 ieee international conference on system, computation, automation and networking (icscan) (pp. 1-8). IEEE.

[2] Srivastava, D. K., Nair, P. (2017, March). Employee attrition analysis using predictive techniques. In International Conference on Information and Communication Technology for Intelligent Systems (pp. 293-300). Springer, Cham.

5 Division of Work

- Boscosylvester John Chittilapilly - Implementation and analysis of various machine learning algorithms for classification, data preparation for clustering algorithm.
- Shlok Vivek Naik - Data pre-processing, predicting estimated salaries for new employees using regression.
- Saksham Pandey - Model evaluation and parameter tuning, implementation of clustering model using existing database to find at risk employees.

6 Midterm Milestone

We plan to complete data preprocessing and preparation, and implement some of the planned machine learning algorithms for classification to benchmark results in terms of quantifiable metrics.