Employee Attrition - Prediction

# **Vinay Nagaraj DSC680 - T302 Applied Data Science (2217-1) https://vinaynagaraj88.github.io/**

# Which Domain?

For an organization to perform successfully, it is important that the employer and the employee have a good relationship and understanding. It is known that if a strong relationship is in place employees will be more productive, more efficient, create less conflict and will be more loyal.

When an employee decides to quit, there will be a lot of challenges for the employer. It will impact their productivity, revenue, experience and also time invested in training the employee. So, it is important for employers to understand why employees are leaving the company.

I choose to work on this topic as it will provide a good overview about the reasons which cause Employee attrition.

Below are several references I am planning to refer as part this project:

1. <https://medium.com/analytics-vidhya/predict-employee-attrition-a34e2c5a972d>
2. <https://www.toolbox.com/hr/engagement-retention/articles/what-is-attrition-complete-guide/>
3. <https://www.abacademies.org/articles/predictive-analytics-in-employee-churn-a-systematic-literature-review.pdf>
4. <https://esource.dbs.ie/bitstream/handle/10788/3497/msc_salunkhe_t_p_2018.pdf?sequence=1&isAllowed=y>
5. <https://www.knime.com/blog/predicting-employee-attrition-with-machine-learning>
6. <https://towardsdatascience.com/people-analytics-with-attrition-predictions-12adcce9573f>
7. <https://blog.perceptyx.com/employee-attrition-analytics>
8. <https://medium.com/swlh/exploratory-data-analysis-employee-attrition-rate-591ce8e7518f>
9. <https://www.linkedin.com/pulse/analyzing-employee-attrition-mike-west>
10. <https://analyticsindiamag.com/predictive-attrition-model/>

# Which Data?

I am planning to use the below dataset from Kaggle which is a fictional data set created by IBM data scientists. I am planning to use this data to train the model as part of this project.

Dataset Link - <https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>

This dataset contains 1,470 rows of data which describes details about the Employee’s age, Attrition, Monthly income of the employee, department of work, distance from home to work for the employee among other columns.

Column Attrition looks like the feature I will have to predict which will tell us whether the Employee will leave the company or not. So, feature “Attrition” will mostly be my target variable.

# Research Questions? Benefits? Why analyze these data?

Below are some of the research questions I would like to find answers for as attrition has become a big challenge for the managers and answering these research questions will help them understand the reason behind the attrition rate.

* Which are the important factors which the contributing to employees leaving the current job?
* Is Monthly income a critical factor to keep employees happy?
* Is Travel time between home to work a factor and if so, can the option of WFH be considered?
* Are highly educated employees more risker of leaving the company?
* How does satisfaction and work-life balance relate to employee retention?
* How many new hires leave in less than a year and why?

# What Method?

* I am planning on using Python language and work on Jupyter notebook for this project.
* I will initially clean the dataset for any missing values or outliers.
* Perform some EDA on the dataset to understand the trend of the data.
* I will perform Feature reduction/selection and then build or use some of the classification models to make predictions on employee attrition.

# Potential Issues?

My initial review of the data and the columns do provide some good insights and I don’t see any major issues or roadblocks at this point.

# Concluding Remarks

Attrition can make a big dent in your organization’s bottom line as well as its culture. Some forms of attrition are unavoidable, like if an employee is retiring or is moving to another city. Organization's performance is heavily based on the quality of the employees. So as an organization it is important to understand the causes of employee attrition and see how the rate can be kept below a certain acceptable threshold.

Through this project we will identify the factors that lead to employee attrition and build a classifier model that would help an organization in predicting the employees that can leave the organization so that they can work with employees and take corrective measures to reduce the attrition problem related to employees leaving the organization.