

Capstone 2 Project Proposal

Isabela Caetano

June 2021

1 Context

IBM is a technology company that is based out of 170 countries around the world. Their current work force is comprised of 345,900 employees. With such a massive number of employees, it is critical to them that they retain their talent and ensure their productivity.

IBM data scientists have developed a data set that captures current employee attrition along with job function, department, last pay raise, etc. The goal is model employee attrition, based on the data's other available features, in order to highlight which types of employees are the susceptible to feeling worn down.

2 Criteria For Success

Potential criteria for success in this project could be:

- Employee retention
- Higher productivity amongst employees
- Fewer personal/sick/vacation days taken off per year per employee

3 Scope of Solution Space

The only data to be considered will come from the given IBM dataset. No other data sets or APIs will be used.

4 Constraints

One constraint to this problem is that person attrition is subjective. What may feel like being worn-down to one employee may not feel so to another employee.

Another constraint is data imbalance. Currently, there are significantly more employees who aren't feeling as depleted as ones that are. This will make the modelling a bit tricky.

5 Stakeholders

Some potential key stakeholders in this project are:

- The IBM data scientists who put together the data set
- HR leads
- People analytics department

6 Data Sources

The main source of data for this project will come from a [Kaggle](#) data set that was created by IBM data scientists.