**Final Project Report**

**Machine Learning**

**Project Title:**

**Employee Attrition Prediction**

**Prepared by:**

**Darshankumar Patel (N01496781)**

**Vishesh Makwana(N01495694)**

**ITE 5310 – Winter 2022**

**Humber College**

**1. Problem Statement**

***“Employee Attrition Prediction”***

**2. Dataset Description**

The IBM HR Analytics Employee Attrition & Performance dataset is a fictional dataset consisting of employee records of a large corporation. The dataset includes 1,470 employee records and 35 attributes.

Graphical user interface, text, application

Description automatically generated

Text, table

Description automatically generated

**3. Dataset Analysis and Observations**

**Key Points:-**

1. JobLevel is highly related to Age as expected as aged employees will generally tend to occupy higher positions in the company.
2. MonthlyIncome is very strongly related to joblevel as expected as senior employees will definately earn more.
3. Also note that TotalWorkingYears is highly related to JobLevel which is expected as senior employees must have worked for a larger span of time.
4. YearsWithCurrManager has a strong relationship with YearsAtCompany.
5. YearsAtCompany is related to YearsInCurrentRole.

Key Note: - The number of observations in the 'No' category is far bigger than the number of observations in the 'Yes' category. As a result, we get skewed classes, which is a common example of the "Imbalanced Classification Problem." To deal with such difficulties, we must employ over-sampling or under-sampling strategies.

Chart

Description automatically generated Graphical user interface, application

Description automatically generated

**4. Proposed Analytical/Prediction Model**

* we have trained Decision Tree because it is the most basic model for predicting data and then we used Random Forest Regression Model because it gives most accurate prediction and good accuracy through cross validation. Moreover, we have also used artificial neural network i.e., ANN to compare it with Random Forest.

**5. Results and Discussions**

1. **Random-Forest 2.) FeedForward Neural Network (ANN)**

Graphical user interface, text, application, email

Description automatically generated Graphical user interface, application

Description automatically generated

Graphical user interface, chart

Description automatically generated with medium confidence

•ANN as more accuracy as compared to Random Forest

•In order to increase the model performance of the ANN we need to tune the hyperparameters. Some of the hyperparameters include Number of hidden layers, Number of neurons in a particular layer, activation function, etc... as the graph suggest no. of epochs increases accuracy.