

Topic Submission Form – Submitted by 14th Dec 2020

This form should be submitted by the mentioned deadline.

Name: Anish Mahapatra

Student Number: 9880383565

Course: MS in Data Science

Fill your topic/s below

Project Title/Area 1:

Prediction of Employee Attrition using Machine Learning Techniques

Dataset:

[IBM HR Analytics Employee Attrition & Performance | Kaggle](#)

Description:

The IBM HR Analytics Attrition & Performance Dataset is used to identify the factors that lead to employee attrition and understand the role of various components that can lead to employee attrition. The dataset consists of 1470 rows and 35 columns relevant to information regarding the employee. With 26 numerical and 8 categorical columns, we can do follow the procedure to extract the most out of the dataset.

It has categorical columns such as Education, Environment Satisfaction, Job Involvement, Job Satisfaction, Performance Rating, Relationship Satisfaction, Work-Life Balance etc. The target variable Attrition classifies if the employee has left the company or not. Using classification machine learning techniques, we will need to classify the probability of the employee leaving the company voluntarily.

Post Business understanding and Data understanding, we shall proceed to Data preparation. We will iterate on the modelling and then proceed to incorporate appropriate evaluation post which we can focus on the final step of deployment of the model.

Project Title/Area 2:

Prediction of Customer Churn using Machine Learning Techniques

Dataset:

[Telco Customer Churn | Kaggle](#)

Description:

The Telecom customer churn dataset gives us the customer details to be able to identify if a telecom customer has churned in the last month from the company or not. The data contains 3 numerical values and 18 categorical columns. This dataset is to be used to predict behaviour to retain customers. We will need to analyse the relevant customer data and develop focused customer retention programs. Each row represents a customer, and each column contains customer attributes.

The target variable Churn helps us understand the customers that have left the company. The data also contains information on the services that each customer has signed up for such as phone, multiple lines, internet, online security, online backup, device protection, tech support, streaming TV and movies. It also consists of customer account information such as how long they have been a customer, payment method, paperless billing, monthly charges and total charges. Demographic information such as gender, age range, and if they have partners and dependents.

Project Title/Area 3:

Credit Card Fraud Detection using Machine Learning Techniques

Dataset: [Credit Card Fraud Detection | Kaggle](#)

Description:

When fraudulent transactions are made, credit card companies must be able to recognize the fraudulent transactions so that customers are not charged for items they did not purchase.

The dataset contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days where we have 492 frauds out of 284,807 transactions. This dataset is highly imbalanced, where the positive class (frauds) account for 0.172% of all transactions.

It contains only numerical input variables which are the result of a PCA transformation, with variables ranging from V1 to V28. The Feature Class represents whether the transaction is fraudulent or not, with 1 indicating fraud, and 0 indicated not a fraudulent transaction. Given the class-imbalance ratio, various imbalance techniques can be applied to the dataset to be able to get data that is less biased. A lot of information about the transactions cannot be derived from this dataset as Principal Component Analysis has already been performed on this dataset. So, the focus of this problem is to be able to identify as many fraudulent transactions as possible.

Fill in this section if a member of staff has agreed to be your supervisor:

Member of Staff: _____

If you have found a supervisor, then you and the member of staff who agreed to supervise your project should sign below.

Student Signature

Supervisor Signature

Date

Date