# Topic: Naive Bayes

**Instructions**

Please share your answers filled inline in the word document. Submit Python code and R code files wherever applicable.

Please ensure you update all the details:

**Name: Shahina Athar**

**Batch Id:**  **10122020**

**Topic: Naïve Bayes**

1. **Business Problem**
   1. **Objective**
   2. **Constraints (if any)**
2. **Work on each feature of the dataset to create a data dictionary as displayed in the below image:**



**2.1 Make a table as shown above and provide information about the features such as its Data type and its relevance to the model building, if not relevant provide reasons and provide description of the feature.**

**Using R and Python codes perform:**

1. **Data Pre-processing**

**2.1 Data Cleaning, Feature Engineering, etc.**

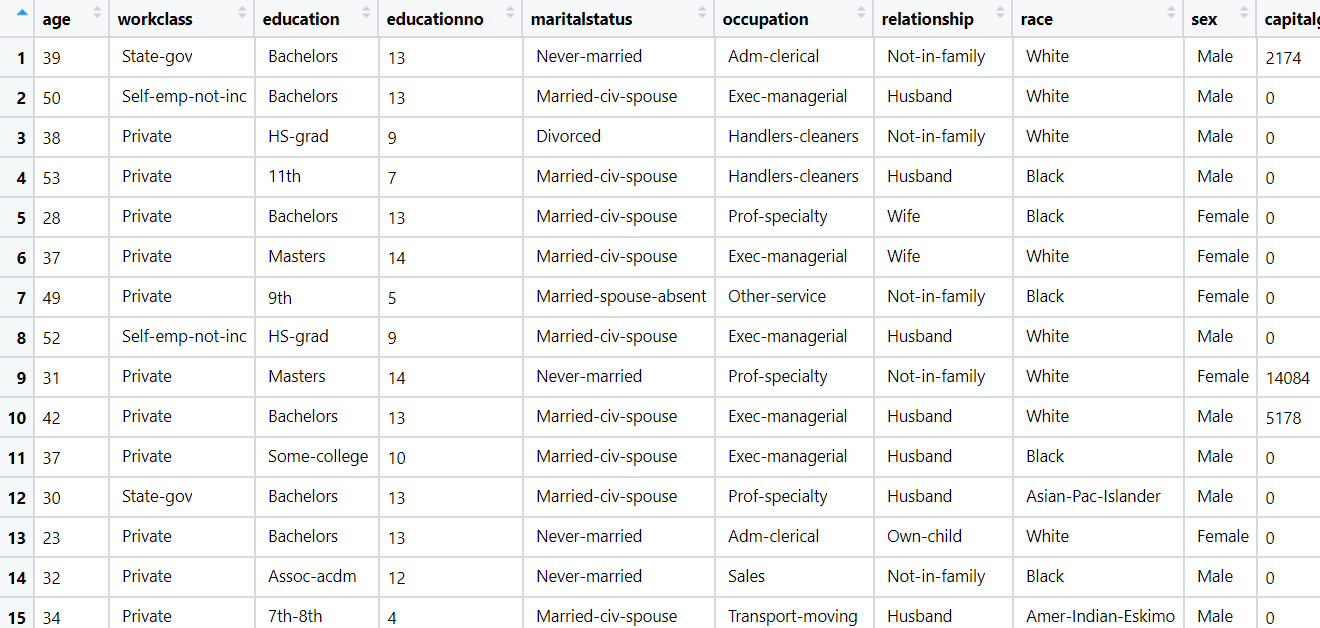
1. **Exploratory Data Analysis (EDA):**
   1. **Summary**
   2. **Univariate analysis**
   3. **Bivariate analysis**
2. **Model Building**
   1. **Build the model on the scaled data (try multiple options)**
   2. **Perform the Naïve Bayes Machine Learning Algorithm**
   3. **Validate the train and test data and perform confusion matrix, get precision, recall and accuracy from it.**
3. **Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.**

**Note:**

The assignment should be submitted in the following format:

* R code
* Python code
* Code Modularization should be maintained
* Documentation of the modules (elaborating on steps mentioned above)

1.) Prepare a classification model using Naive Bayes for Salary dataset, train and test datasets are given separately use both datasets for model building.



Pro

* 1. **Validate the train and test data and perform confusion matrix, get precision, recall and accuracy from it.**

Ans**: For Gaussian Naive Bayes:**

Precision : (True Positive/True Positive + False Positive) = 81%

Recall: TruePositives / (TruePositives + FalseNegatives) = 95%

Accuracy: 79%

**For Multinomial Naive Bayes:**

Precision : (True Positive/True Positive + False Positive) = 79%

Recall: TruePositives / (TruePositives + FalseNegatives) = 95%

Accuracy: 77%

1. **Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.**

Ans: After Prepared a classification model using Naive Bayes for Salary dataset, If we give any new data to this model they can easily identify if the person salary is <=50K Or >50K With the 79% of accuracy.

Problem Statement: -

This dataset contains information of users in social network. This social network has several business clients which can put their ads on social network and one of the Client has a car company who has just launched a luxury SUV for ridiculous price. Build the Bernoulli Naïve Bayes model using this dataset and classify which of the users of the social network are going to purchase this luxury SUV.

Purchased: - 1 and Not Purchased: - 0

A screenshot of a cell phone

Description automatically generated

* 1. **Validate the train and test data and perform confusion matrix, get precision, recall and accuracy from it.**

Ans**: For Bernoulli Naïve Bayes model:**

Precision : (True Positive/True Positive + False Positive) = 80%

Recall: TruePositives / (TruePositives + FalseNegatives) = 93%

Accuracy: 79%

1. **Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.**

Ans: After Prepared a classification model using Bernoulli Naive Bayes for social network dataset, We can easily identify which of the users of the social network are going to purchase this luxury SUV or Not With the 79% of accuracy

Problem Statement: -

In this case study you have been given with tweeter data collected from an anonymous twitter handle, with the help of Naïve Bayes algorithm predict a given tweet is Fake or Real about real disaster occurring.

Real tweet: - 1 and Fake tweet: - 0

A screenshot of a cell phone

Description automatically generated

* 1. **Validate the train and test data and perform confusion matrix, get precision, recall and accuracy from it.**

Ans**: For Naïve Bayes model:**

Precision : (True Positive/True Positive + False Positive) = 74%

Recall: TruePositives / (TruePositives + FalseNegatives) = 94%

Accuracy: 86%

1. **Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.**

Ans: After Prepared a classification model using Naive Bayes for tweeter data dataset, we can easily identify given tweet is Fake or Real about real disaster occurring with the 86% of accuracy.