# Desh Iyer | 500081889 | Experiment - 6

April 2, 2023

## 1 Experiment 6 - Implement a Naive-Bayes Classifier

### 1.1 Import Libraries

[ ]:
```
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score

import pickle

import pandas as pd
```

### 1.2 Load Data

[ ]:
```
data = pd.read_csv('naive-bayes-classification-data.csv')
print(data)
```

```
     glucose  bloodpressure  diabetes
0         40             85         0
1         40             92         0
2         45             63         1
3         45             80         0
4         40             73         1
..       …              …        …
990       45             87         0
991       40             83         0
992       40             83         0
993       40             60         1
994       45             82         0

[995 rows x 3 columns]
```

### 1.3 Train-test Split

```
[ ]: X_train, X_test, y_train, y_test = train_test_split(data[['glucose',␣
      ↪'bloodpressure']], data['diabetes'], test_size=0.3, random_state=42)

     print(f'X_Train shape: {X_train.shape}\ny_train shape: {y_train.shape}')
```

```
X_Train shape: (696, 2)
y_train shape: (696,)
```

### 1.4 Define Gaussian Naive-Bayes Classifier

```
[ ]: classifier = GaussianNB()
     y_predicted = classifier.fit(X_train, y_train).predict(X_test)
```

### 1.5 Print Accuracy

```
[ ]: print("Number of mislabeled points out of a total %d points : %d" % (X_test.
      ↪shape[0], (y_test != y_predicted).sum()))
     print("The resultant accuracy of the Gaussian Naive Bayes classifier is: %f" %␣
      ↪accuracy_score(y_test, y_predicted))
```

```
Number of mislabeled points out of a total 299 points : 20
The resultant accuracy of the Gaussian Naive Bayes classifier is: 0.933110
```

### 1.6 Save Model using Pickle

```
[ ]: with open('model.pickle', 'wb') as f:
         pickle.dump(classifier, f)
```