Date ___ / ___ / _____

Assignment- C2

* **Title:-** Naive Bayes Algorithm.

* **Problem statement:-**

Download Prima Indians diabetes dataset. Use Naive Bayes algorithm for classification.

1) Load the data into CSV file and split it into training and test datasets.

2) Summarize properties in training dataset so that we can calculate probability and make predictions.

3) Classify samples from the test dataset and a summarized training dataset.

* **Objectives:-**

- To learn classification algorithm like naive-bayes.
- To implement such algorithm to predict data.

* **Outcomes:-**

Students will be able to:-

- learn classification algorithms.
- make predictions using training dataset.

* **S/W & H/W Requirements:-**

- OS : Windows 10 / Ubuntu (64-bit).
- Python scipy libraries / R studio.
- Gedit Editor.
- 4GB RAM, 500GB HDD.

* <u>Theory</u> :-

⇒ <u>Bayes Theorem</u> :-
It is a way of finding a probability when we know certain other probabilities.

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$$

where,

$P(A|B)$ = how often A happens given that B happens.
$P(B|A)$ = how often B happens given A happens.
$P(A)$ = how likely A is on it's own.
$P(B)$ = how likely B is on its own.

⇒ <u>Naive-Bayes Classification</u> :-
It is a simple yet effective and commonly used machine learning classification. It is a probabilistic classifier that makes classification using the maximum "Aposteriori" decision rule in a Bayesian setting. It can be represented using a very Bayesian network. It is especially popular for text classification and is a traditional solution for problems such as spam detection.

* <u>Application</u> :-

1) <u>Real time Prediction</u> :- Naive Bayes is also an eager learning classifier and it is very fast. Thus, it could be used to make predictions in real time.

2) <u>Multi-Class Prediction</u> : This algorithm is also known for multi-class prediction feature. Here, we can predict the probability of multiple classes of target variable.

2) Text Classification:-

It is used to have higher success rate as compared to other algorithms. As a result, it is widely-used in spam filtering. & sentiment analysis.

from sklearn.naive-bayes import Gaussion NB.
```
gnb = Gaussion NB ().
gnb.fit (x-train, y-train).
```

& Test-Case:

Input:- Diabetes Dataset.

Output:- Confusion Matrix.   Accuracy = 0.74479.

	0	1
0	109	16
1	33	34

Test set was 25% of the data-set & 75% of predicted values were obtained correctly.

* Conclusion:- we have successfully learnt & implemented Naive-Bayes classification algorithm.