1. **Probabilistic Supervised Learning - Naive Bayes:**
   - **a)** Create a dataset from the sample given to you(e.g. “Play Tennis Probability”, “Shopper Buying Probability” etc.).
   - **b)** Perform the necessary pre-processing steps such as encoding.
   - **c)** Train the model using Naive Bayes Classifier.
   - **d)** Give new test data and predict the classification output.
   - **e)** Handcode the classification probability
   - **f)** Compare with the model output given by sklearn
   - **g)** Calculate the Accuracy of the model
   - **h)** Analyze and write the inference.

Expected Outcome:

Students will learn how to

- Convert a curated dataset into a usable form by Naive Bayes classifier and predict the outcome. **(Dataset #1 to #4)**
  Note: Naive Bayes expects numeric values for the features to predict the outcome.
- Use a given dataset (such as titanic.csv, iris.csv etc) on a Naive Bayes classifier and predict the outcome. **(Dataset #5 & #6)**

Dataset#1 : Play Tennis

Note: You should be able to create a workable Naive Bayes Classifier *out of any dataset given to you*, with proper data processing such as encoding.

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

OUTPUT ANALYSIS

```
Encoded value of weather outlook: [2 2 0 1 1 1 0 2 2 1 2 0 0 1]
Encoded value of Temp: [1 1 1 2 0 0 0 2 0 2 2 2 1 2]
Encoded value of humidity: [0 0 0 0 1 1 1 0 1 1 1 0 1 0]
Encoded value of windy: [1 0 1 1 1 0 0 1 1 1 0 0 1 0]
Encoded value of Play: [0 0 1 1 1 0 1 0 1 1 1 1 1 0] #This is outcome
```

```
Features and label variables are encoded in following way:
#weather - Overcast: 0, Rainy: 1, Sunny: 2
#temp - Hot: 1, Mild: 2, Cool: 0
#play - Yes: 1, No: 0
#humidity - Normal: 1, High: 0
#windy - weak: 1, strong: 0
This is by using sklearn's
```

- `preprocessing.LabelEncoder()`
- `labelEncoder.fit_transform(<<categorical feature>>)`

```
Encoded features for each row is
```

```
 ((2, 1, 0, 1), (2, 1, 0, 0), (0, 1, 0, 1), (1, 2, 0, 1), (1, 0, 1,
1), (1, 0, 1, 0), (0, 0, 1, 0), (2, 2, 0, 1), (2, 0, 1, 1), (1, 2, 1,
1), (2, 2, 1, 0), (0, 2, 0, 0), (0, 1, 1, 1), (1, 2, 0, 0))
```

```
This prints the encoded features of each row. There are 14 tuples as
output for the 14 rows in the dataset.
```

```
For new_data = [0,2,0,0] # 0:Overcast Weather, 2:Mild Temp, 0:Strong
Windy, 0:Humidity Strong
```

```
Predicted Value: [1] #meaning "Yes" to Play Tennis
Play Tennis: Yes
```

```
For new_data = [2,0,0,0] # 0:Sunny Weather, 2:Cool Temp, 0:Strong
Windy, 0:Humidity Strong
```

```
Predicted Value: [0] #meaning "No" to Play Tennis
Play Tennis: No
```

Apply Naive Bayes Classifier to the different datasets below and classify.

I) **Dataset #2:** Determine if a customer will buy a product on a *holiday*, *with discount and free delivery.*

	A	B	C	D
1	Day	Discount	Free Delivery	Purchase
2	Weekday	Yes	Yes	Yes
3	Weekday	Yes	Yes	Yes
4	Weekday	No	No	No
5	Holiday	Yes	Yes	Yes
6	Weekend	Yes	Yes	Yes
7	Holiday	No	No	No
8	Weekend	Yes	No	Yes
9	Weekday	Yes	Yes	Yes
10	Weekend	Yes	Yes	Yes
11	Holiday	Yes	Yes	Yes
12	Holiday	No	Yes	Yes
13	Holiday	No	No	No
14	Weekend	Yes	Yes	Yes
15	Holiday	Yes	Yes	Yes

Naive_Bayes_Dataset

II) **Dataset #3: - Classify as Mammals and Non-Mammals**

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Test the model to predict to which class (mammals or non - mammals)the following instance belongs to:

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

III) **Dataset #4:** Predict the outcome for *Test-XID (Refund ="No", "Married", Income = 120K)* for the following data set.

T-id	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No

7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

IV) Dataset #5: Titanic.csv

- Description of 887 passengers on the Titanic.
- Use Naive Bayes Classifier to Predict whether a given passenger survived the tragedy or not.

V) Dataset #6: iris.csv

- The data set contains 3 classes of 50 instances each
- Each class refers to a type of iris plant. (class: Iris Setosa, Iris Versicolour, Iris Virginica)
- Use Naive Bayes Classifier to predict the class a new dimension of iris flower belongs to.