

Lab 13 – Assignment 04: Naïve Bayes Classifier

Submission: In groups

Naïve Bayes Classifier:

Naïve Bayes is a classification algorithm that works based on the Bayes theorem. Bayes theorem is used to find the probability of a hypothesis with given evidence.

In this, using Bayes theorem we can find the probability of A, given that B occurred. A is the hypothesis and B is the evidence.

$P(B|A)$ is the probability of B given that A is True.
 $P(A)$ and $P(B)$ is the independent probabilities of A and B.

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

Provide answers to ONE of the following:

1. Diabetes Prediction:

You are required to do diabetes prediction for the given dataset. The dataset contains the following features:

Pregnancies	Glucose	BloodPressure	SkinThickness
Insulin	BMI	DiabetesPedigreeFunction	Age

The label is provided in column ‘Outcome’.

You need to:

1. Read the csv file and add data to an array.
2. From sklearn use naïve_bayes classifier for actual classification
`from sklearn.naive_bayes import GaussianNB`
3. Print the accuracy of the prediction

Note:

X,Y can be split as:

```
file = 'iris.csv'
iris = pd.read_csv(file)
print(iris.head())
data = iris.values;
X = data[:,0:4]
Y = data[:, -1]
```

2. Churn Analysis:

You are required to do Churn analysis on the given dataset. The dataset contains the following features:

customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService
MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	
StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges

The label is provided in column 'Churn'

You need to:

1. Read the csv file and add data to an array.
2. Convert all non-numeric data into numeric data using python's sklearn LabelEncoder. (You can also do it manually in the excel sheet provided). The features that need changing are: Gender (M=0,F=1), Partner (Y=1, N=0), Dependent (Y=1, N=0), PhoneService (Y=1, N=0), and so on.
3. Replace empty values with zeros or defaults.
4. From sklearn use naïve_bayes classifier for actual classification
`from sklearn.naive_bayes import GaussianNB`
5. Print the accuracy of the prediction