Tanmay Bhatkar TE INFT A Roll No 9

DSL Experiment 6

To implement Classification

Laboratory Exercise

A. Procedure: (Home _Loan Dataset)

Generated code may be subject to a license |

```python
import sklearn
import pandas as pd
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
loan_data = pd.read_csv("loan_train.csv")
loan_data.head()
```

	Unnamed: 0.1	Unnamed: 0	loan_status	Principal	terms	effective_date	due_date	age	education	Gender
0	0	0	PAIDOFF	1000	30	9/8/2016	10/7/2016	45	High School or Below	male
1	2	2	PAIDOFF	1000	30	9/8/2016	10/7/2016	33	Bechalor	female
2	3	3	PAIDOFF	1000	15	9/8/2016	9/22/2016	27	college	male
3	4	4	PAIDOFF	1000	30	9/9/2016	10/8/2016	28	college	female
4	6	6	PAIDOFF	1000	30	9/9/2016	10/8/2016	29	college	male

```python
print(loan_data.columns.values)
```

```
['Unnamed: 0.1' 'Unnamed: 0' 'loan_status' 'Principal' 'terms'
 'effective_date' 'due_date' 'age' 'education' 'Gender']
```

```python
loan_data["Gender"]
LabelEncoder().fit_transform(loan_data["Gender"].astype(str))
print(loan_data["Gender"])
```

```
0        male
1      female
2        male
3      female
4        male
        ...
341      male
342      male
343      male
344      male
345      male
Name: Gender, Length: 346, dtype: object
```

```python
loan_data["education"]
LabelEncoder().fit_transform(loan_data["education"].astype(str))
print(loan_data["education"])
```

```
0      High School or Below
1                  Bechalor
2                   college
3                   college
4                   college
               ...
341    High School or Below
342    High School or Below
343                 college
344                 college
345                 college
Name: education, Length: 346, dtype: object
```

```
X = loan_data.drop("loan_status",axis=1)
Y = loan_data["loan_status"]
Y = LabelEncoder().fit_transform(loan_data["loan_status"].astype(str))
X_train, X_test, y_train, y_test = train_test_split(X,Y,test_size=0.3,
random_state=1)
Naivebayes = GaussianNB()
b = Naivebayes.fit(X_train,y_train)
c = Naivebayes.predict(X_test)
```

```
print(c)
acc = sklearn.metrics.accuracy_score(y_test, c)
print(acc)
```

```
[0 1 0 1 0 0 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 0 1 1 1 1 1 1 1 1
 1 0 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 0 1 0 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1 0
 0 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1 1]
0.9807692307692307
```

98.07 percent accuracy achieved

8. Post-Experiments Exercise

A. Extended Theory: (Soft Copy)

● Types of classification.

Binary Classification

Involves two possible outcomes, such as True/False, Yes/No, or Fraud/Not Fraud. This type of classification uses models that output a probability score which is then thresholded to assign one of the two classes.

Multiclass Classification

Deals with scenarios where there are more than two classes, with each instance belonging to exactly one category. Examples include categorizing types of animals (Dog, Cat, Bird) or classifying news articles into different topics.

Multilabel Classification

Allows each instance to be associated with multiple labels simultaneously. For example, in text categorization, a single article might belong to multiple topics, such as both "Health" and "Technology." Models for multilabel classification are designed to predict a set of labels for each instance rather than a single label.

Ordinal Classification

Applies when the classes have a natural order or ranking, such as ratings (Low, Medium, High) or educational levels (Bachelor's, Master's, Doctorate). In ordinal classification, the model not only predicts the category but also respects the inherent order among the classes.

Hierarchical Classification

Involves classification tasks where categories are organized in a hierarchical structure. For example, in document categorization, a document might be first classified into a broad category such as "Science" and then further classified into subcategories like "Physics" or "Biology." This approach accounts for the relationships between higher-level and lower-level categories.