KUMAR GAURAV

20122065

In [1]:

```
import pandas as pd
import numpy as np
```

In [51]:

```
df=pd.DataFrame({
    "Color":["Red","Red","Red","Yellow","Yellow","Yellow","Yellow","Yellow","Red","Red"],
    "Type":["Sports","Sports","Sports","Sports","Sports","SUV","SUV","SUV","SUV","Sports"],
"Origin":["Domestic","Domestic","Domestic","Domestic","Imported","Imported","Imported",
    "Stolen?":["Yes","No","Yes","No","Yes","No","Yes","No","No","Yes"]
})
```

In [52]:

```
df
```

Out[52]:

	Color	Type	Origin	Stolen?
0	Red	Sports	Domestic	Yes
1	Red	Sports	Domestic	No
2	Red	Sports	Domestic	Yes
3	Yellow	Sports	Domestic	No
4	Yellow	Sports	Imported	Yes
5	Yellow	SUV	Imported	No
6	Yellow	SUV	Imported	Yes
7	Yellow	SUV	Domestic	No
8	Red	SUV	Imported	No
9	Red	Sports	Imported	Yes

In [53]:

```
n_sample,n_features=df.shape
```

In [54]:

```
n_sample
```

Out[54]:

```
10
```

[55]:

In

```
target = df['Stolen?'].value_counts().to_dict()
target
```

Out[55]:

```
{'Yes': 5, 'No': 5}
```

In [56]:

```
p_yes=target['Yes']/n_sample
p_yes
```

Out[56]:

```
0.5
```

In [57]:

```
p_no=target['No']/n_sample
p_no
```

Out[57]:

```
0.5
```

In [58]:

```
y=df['Stolen?']
x=df.drop(columns='Stolen?')
y
```

Out[58]:

```
0     Yes
1      No
2     Yes
3      No
4     Yes
5      No
6     Yes
7      No
8      No
9     Yes
Name: Stolen?, dtype: object  [59]:
```

```
x
```

Out[59]:

	Color	Type	Origin
0	Red	Sports	Domestic
1	Red	Sports	Domestic

In

2	Red	Sports	Domestic
3	Yellow	Sports	Domestic
4	Yellow	Sports	Imported
5	Yellow	SUV	Imported
6	Yellow	SUV	Imported
7	Yellow	SUV	Domestic
8	Red	SUV	Imported
9	Red	Sports	Imported

In [60]:

```
temp={
    "Color":df["Color"].value_counts().to_dict(),
    "Type":df["Type"].value_counts().to_dict(),
    "Origin":df["Stolen?"].value_counts().to_dict()
}
temp
```

Out[60]:

```
{'Color': {'Red': 5, 'Yellow': 5},
 'Type': {'Sports': 6, 'SUV': 4},
 'Origin': {'Yes': 5, 'No': 5}}
```

In [61]:

```
y = "Stolen?"

dummy={}
for col in df.columns:
    dummy[col]={}
    for item in df[col].unique():
        dummy[col][item]={}
        for Stolen in df[y].unique():
            dummy[col][item][Stolen]=df[(df[col]==item)&(df[y]==Stolen)].shape[0]/target[St
            dummy
```

In [62]:

```
class NB:

    def __init__(self):
        self.dummy=dict()

    def fit(self, x, y):
        for i in x.columns:
            self.dummy[i]=dict()
            for c in x[i].unique():
                self.dummy[i][c]=dict()
                for ch in y.unique():
                    self.dummy[i][c][ch]=x[(x[i]==c) & (y==ch)].shape[0]/y[y==ch].shape[0]
        for i in y.unique():
            self.dummy[i]=y[y==i].shape[0]/y.shape[0]

    def predict(self, x):
        for i in x.columns:
            self.dummy[i]=dict()
            for c in x[i].unique():
                self.dummy[i][c]=dict()
```

In [63]:

```
p1=NB()
```

In [66]:

```
p1.dummy
```

Out[66]:

```
{'Color': {'Red': {}}}
```

In [67]:

```
df1 = pd.DataFrame({
    "color":["Red","Red","Red","Yellow"],
    "Type":["Sports","Sports","SUV","SUV"],
    "Origin":["Imported","Domestic","Domestic","Imported"]
})
```

In [68]:

```
df1
```

Out[68]:

	color	Type	Origin
0	Red	Sports	Imported
1	Red	Sports	Domestic
2	Red	SUV	Domestic

In

3	Yellow	SUV	Imported

[69]:

```
p1.predict(df1)
```

In [70]:

```
p1.dummy
```

Out[70]:

```
{'Color': {'Red': {}},
 'color': {'Red': {}, 'Yellow': {}},
 'Type': {'Sports': {}, 'SUV': {}},
 'Origin': {'Imported': {}, 'Domestic': {}}}
```

In [ ]: