Assignment No.5

```
import pandas as pd
```

```
df=pd.read_csv('car_evaluation.csv')
```

```
df.head(2)
```

	vhigh	vhigh.1	2	2.1	small	low	unacc
0	vhigh	vhigh	2	2	small	med	unacc
1	vhigh	vhigh	2	2	small	high	unacc

```
df.describe()
```

	vhigh	vhigh.1	2	2.1	small	low	unacc
count	1727	1727	1727	1727	1727	1727	1727
unique	4	4	4	3	3	3	4
top	high	high	3	4	med	med	unacc
freq	432	432	432	576	576	576	1209

```
col_names = ['buying', 'maint', 'doors', 'persons', 'lug_boot', 'safety', 'class']

df.columns=col_names
col_names
```

```
['buying', 'maint', 'doors', 'persons', 'lug_boot', 'safety', 'class']
```

```
for col in col_names:
    print(df[col].value_counts())
```

```
buying
high     432
med      432
low      432
vhigh    431
Name: count, dtype: int64
maint
high     432
med      432
low      432
vhigh    431
Name: count, dtype: int64
doors
3        432
4        432
5more    432
2        431
Name: count, dtype: int64
persons
4       576
more    576
2       575
Name: count, dtype: int64
lug_boot
med      576
big      576
small    575
Name: count, dtype: int64
safety
med     576
high    576
low     575
```

```
Name: count, dtype: int64
class
unacc    1209
acc       384
good       69
vgood      65
Name: count, dtype: int64
```

```
df.isnull().sum()
```

```
buying      0
maint       0
doors       0
persons     0
lug_boot    0
safety      0
class       0
dtype: int64
```

```
x=df.drop(['class'],axis=1)
y=df['class']
```

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=42)
```

```
x_train.shape,x_test.shape
```

```
((1208, 6), (519, 6))
```

```
import category_encoders as ce

encoder = ce.OrdinalEncoder(cols=['buying', 'maint', 'doors', 'persons', 'lug_boot', 'safety'])
x_train = encoder.fit_transform(x_train)
x_test = encoder.transform(x_test)
```

```
x_train.head()
```

	buying	maint	doors	persons	lug_boot	safety
1177	1	1	1	1	1	1
585	2	2	2	2	2	2
1551	3	1	2	1	3	2
727	2	1	3	2	1	1
707	2	1	3	3	1	3

```
from sklearn.ensemble import RandomForestClassifier

rfc=RandomForestClassifier(random_state=0)

rfc.fit(x_train,y_train)
```

```
RandomForestClassifier
RandomForestClassifier(random_state=0)
```

```
y_pred=rfc.predict(x_test)
```

```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

```
0.928709055876686
```

```
rfc_100 = RandomForestClassifier(n_estimators=100, random_state=0)
rfc_100.fit(x_train, y_train)
```

```
RandomForestClassifier
RandomForestClassifier(random_state=0)
```

```
y_pred_100=rfc_100.predict(x_test)
accuracy_score(y_test,y_pred_100)
```

```
0.928709055876686
```

```
from sklearn.ensemble import RandomForestClassifier
rfc_100=RandomForestClassifier(n_estimators=100,random_state=0)
rfc_100.fit(x_train,y_train)
y_pred_100 = rfc_100.predict(x_test)
print("model accuracy n_estimator=100: {0:0.4f}".format(accuracy_score(y_test,y_pred_100)))
```

```
model accuracy n_estimator=100: 0.9287
```

```
print(y_train)
```

```
1177    vgood
585     unacc
1551      acc
727       acc
707     unacc
        ...
1130    unacc
1294    vgood
860     unacc
1459    unacc
1126      acc
Name: class, Length: 1208, dtype: object
```