



Glass Identification Using Decision Tree and K-Nearest Neighbors

Presented By :

Name : Akshay Yashwant Pharande
Roll No. : 37
Division : CI

Under the guidance of :

Asst. Prof. V. L. Badadare
MCA, SET (Computer Application)



Objectives Of The Study

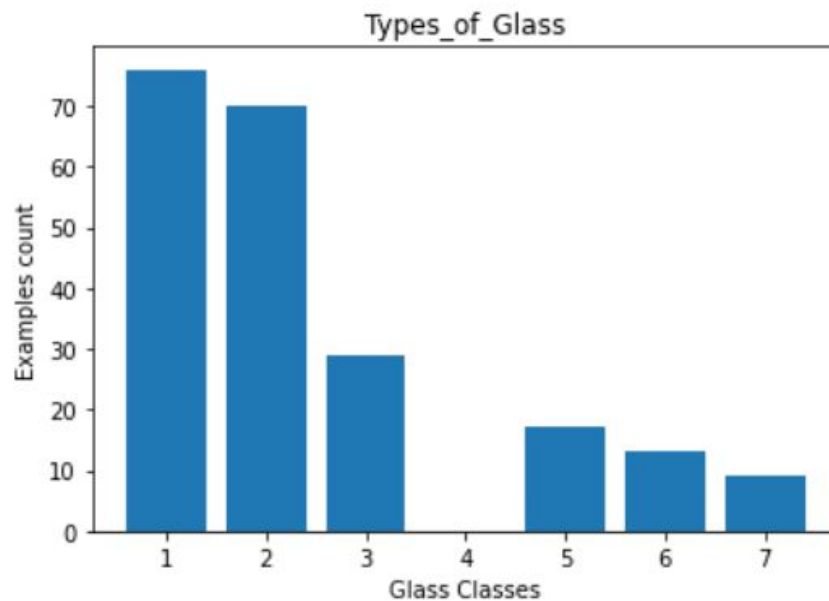
- ❑ To Predict type of glass.
- ❑ Compare Decision Tree and K-Nearest Neighbors Classifiers.



Research methodology

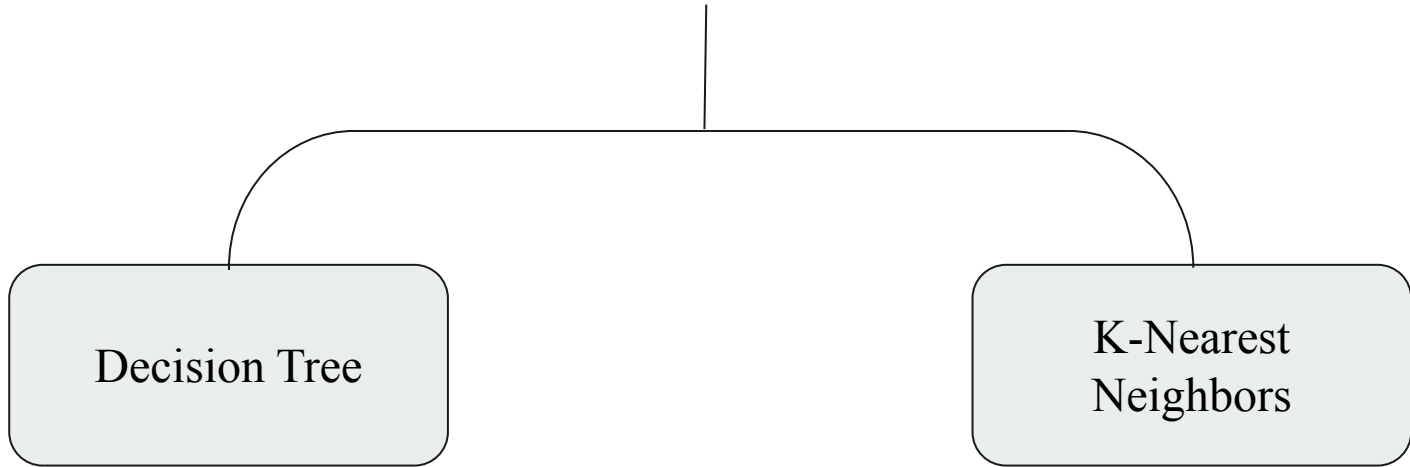
- ❑ For the research Secondary data is collected.
- ❑ The data is collected from the website named “UCI Repository”.
- ❑ Analytical Tool : Python.

Data Visualization





Model Building





Decision Tree Classifier

Training Set Contains 80% Of The Data And The Testing Set Contains 20% Of Data.

Accuracy score of validation set :

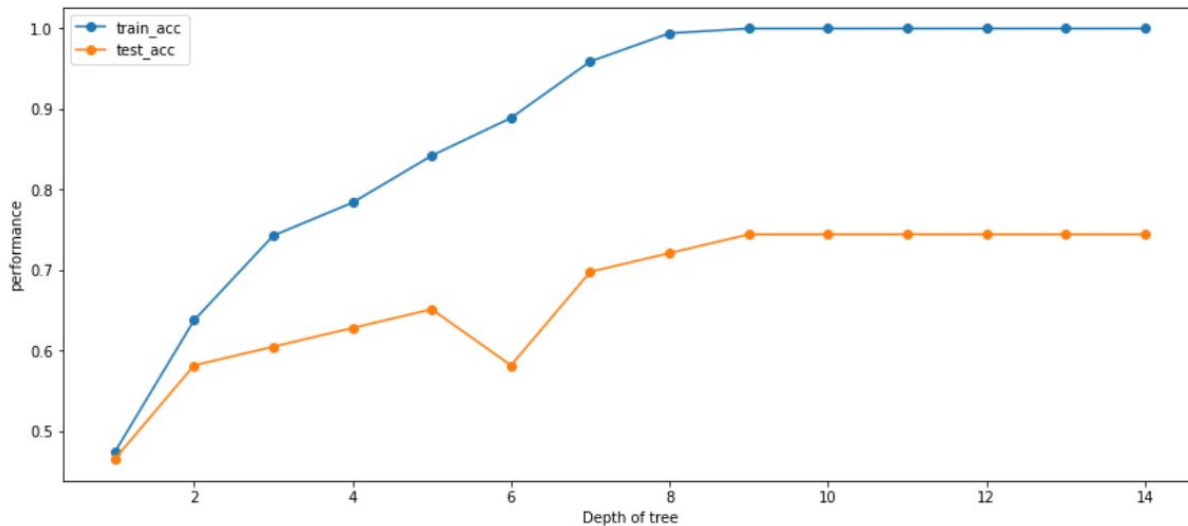
```
from sklearn.metrics import accuracy_score  
accuracy_score(y_test, dt_predict)
```

```
0.6511627906976745
```

Optimizing performance : Max. depth of tree method.

Visualisation of training and testing accuracy

: <matplotlib.legend.Legend at 0x1e3ee532a30>





Predictions on validation set:

Input:

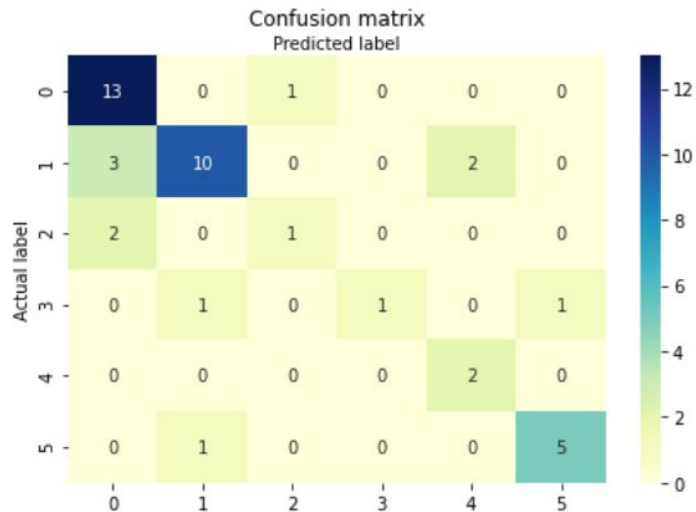
```
: #predictions on validation set  
dt_predict=dt_model.predict(X_test)
```

Output:

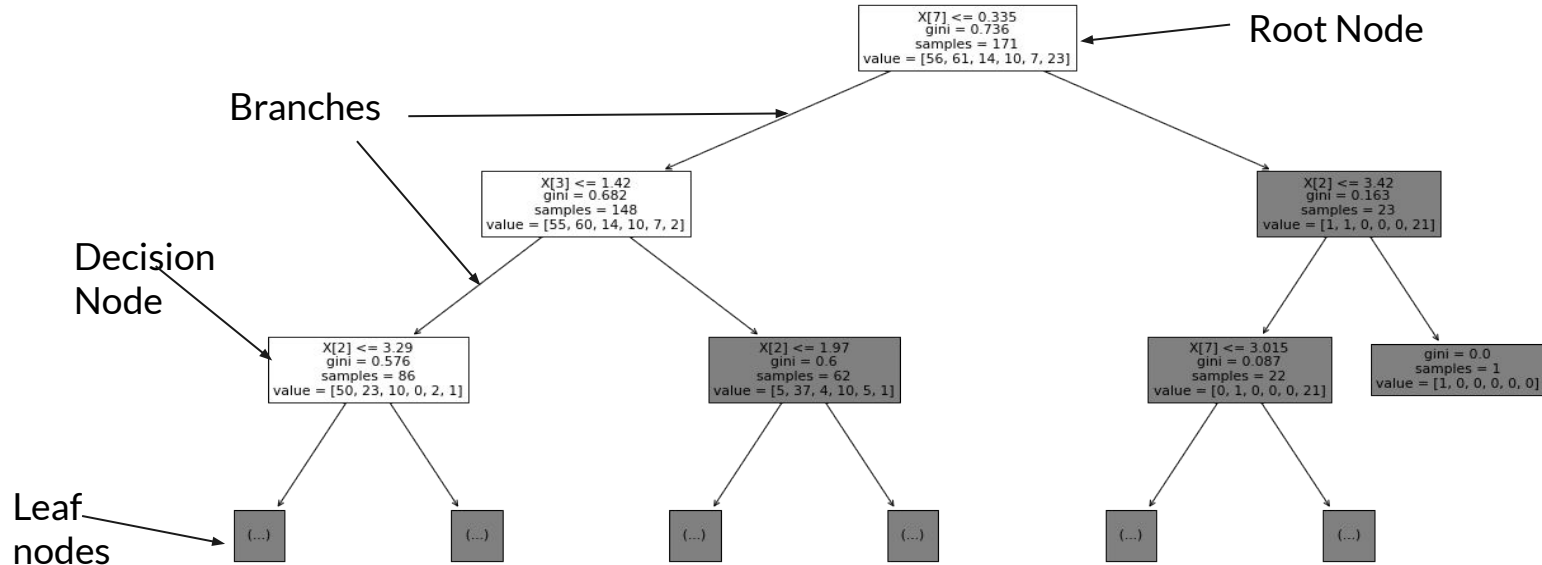
```
[1]: array([1, 2, 7, 1, 2, 7, 2, 2, 1, 3, 7, 7, 1, 1, 6, 1, 1, 2, 6, 1, 1, 1,  
          7, 2, 2, 1, 5, 2, 1, 1, 1, 2, 1, 1, 2, 6, 6, 2, 1, 2, 2, 1, 1],  
        dtype=int64)
```


- **Accuracy after optimizing the performance : 0.74**
- **Confusion Matrix:**

: Text(0.5, 257.44, 'Predicted label')



Decision Tree





K-Nearest Neighbors

- **Accuracy score of validation set**

```
: from sklearn.metrics import accuracy_score  
accuracy_score(y_test, predict_type_n)
```

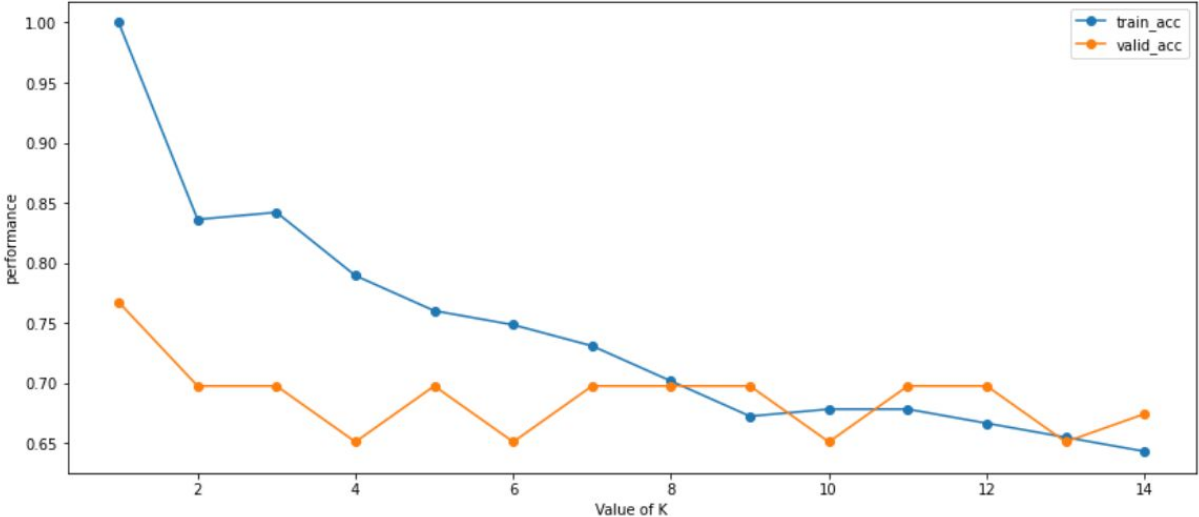
```
: 0.6976744186046512
```

- **To choose optimal value of K**



Visualisation of training and testing accuracy

: <matplotlib.legend.Legend at 0x1e3ef0fb670>





Predictions on validation set:

Input:

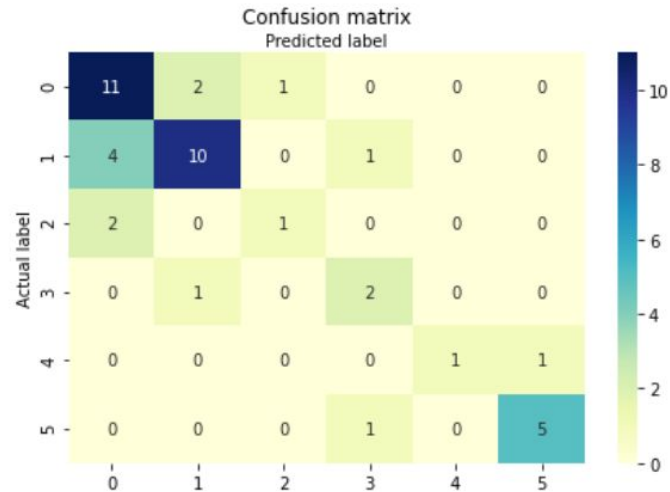
```
In]: predict_type_n_2 = clf.predict(X_test)
      predict_type_n_2
```

Output:

```
In]: array([2, 2, 7, 1, 1, 7, 2, 2, 1, 1, 7, 7, 2, 1, 1, 1, 2, 2, 5, 1, 2, 1,
            7, 5, 2, 3, 5, 1, 1, 2, 3, 2, 1, 1, 5, 6, 7, 1, 1, 2, 2, 1, 1],
          dtype=int64)
```

- Accuracy after optimizing the performance: 0.69
- Confusion Matrix:

Text(0.5, 257.44, 'Predicted label')





Findings & Conclusion

Report of :

Decision Tree Model

	precision	recall	f1-score	support
1	0.92	0.97	0.94	70
2	0.96	0.93	0.95	76
3	0.93	0.82	0.87	17
5	1.00	0.85	0.92	13
6	0.82	1.00	0.90	9
7	0.97	0.97	0.97	29
accuracy			0.94	214
macro avg	0.93	0.92	0.92	214
weighted avg	0.94	0.94	0.94	214

KNN(K-nearest neighbors) model

	precision	recall	f1-score	support
1	0.63	0.86	0.73	70
2	0.75	0.71	0.73	76
3	0.50	0.12	0.19	17
5	0.62	0.62	0.62	13
6	0.80	0.44	0.57	9
7	0.88	0.76	0.81	29
accuracy			0.70	214
macro avg	0.70	0.58	0.61	214
weighted avg	0.70	0.70	0.68	214

Suggests to use the decision tree model over K-nearest neighbors model.