

School of Computer Science Engineering and Technology
Assignment-08

Course- B.Tech
Code-23CS301PC206
Year- 2024-2025
Date- 09-09-2024

Type- Core
Course Name- Artificial Intelligence & Machine Learning
Semester- Even, **Instructor:** Prof. E.L.N. Kiran
Batch- AIML-A,B

Implement Decision Tree Classification using OpenML Car Dataset

In machine learning, decision trees are supervised learning models with associated learning algorithms that analyze data used for classification.

Let's say that you want to make a decision of the car to buy. In order to get the right car to buy, you could go on and evaluate the level of the safety, the number of seats and doors by asking series of if like questions.

1. Import the Libraries required for Decision Tree. [CO1]

Import all libraries required along with visualization to completed the task on DT

2. Import the OpenML Car Dataset from Sklearn Packages. [CO1]

- Once downloaded the car data, prepare the dictionary format to access the data using its keys.
- Describe all the features from the dictionary, feature names.
- Setup the dataframe , describe its details, check for missing values.
- Identify the target class and assign it to the dataframe.
- Perform exploratory analysis of dataframe using `seaborn - sns package`
- Draw the boxplot of first 10 columns to verify their role in decision to buy the car or not.

3. Train and Test Data. [CO2]

- Prepare the Train and Test data from the dataframe.
- Drop the `BinaryClass` column and define the dataframe with only `target` results
- Split the data into train, test using `train_test_split`

4. Train the DT using the Train Dataset. [CO3]

- `import sklearn.tree import DecisionClassifier`
- Apply the `model.fit` to dataset

5. Predict and Analysis the Performance of the DT Model. [CO4]

- Apply `model.predict(X_test)`
- Generate the `accuracy_score`, `classification_report`, `confusion_matrix` using `sklearn.metrics`
- Plot the decision tree and it branches using *from sklearn.tree import plot_tree, plot.figure*

6. Improve the Accuracy of Model using GridSearchCV Model. [CO4]

- Given the following parameters validate the GridSearchCV Model:
`param_grid = 'C': [0.1,1, 10, 100, 1000], 'gamma': [1,0.1,0.01,0.001,0.0001],
 'kernel': ['rbf']`
- Implement the `model.predict`
- Generate the `classification_report`, `confusion_matrix` using `sklearn.metrics`