

# Predictive Modeling Using Decision Trees

SELVARAJ X(2577127)

01-NOV-2023

# Introduction

- ❖ Decision trees are a type of machine learning algorithm that can be used for both classification and regression tasks.
- ❖ They are relatively easy to understand and interpret, making them a popular choice for predictive modeling in many different industries.
- ❖ In this case study, we will use decision trees to build predictive models for customer satisfaction and sales prediction in a retail business.
- ❖ We will explore the dataset, implement the decision tree models, evaluate their performance, visualize the decision trees, and discuss the interpretability of decision trees in predictive modeling.

# Classification Task - Predicting Customer Satisfaction

- ❖ Customer satisfaction is essential for the success of any retail business. Satisfied customers are more likely to make repeat purchases and recommend the business to others.
- ❖ Sales prediction is also important, as it allows retailers to plan their inventory and marketing campaigns more effectively.
- ❖ Implement a decision tree classifier using Python libraries like scikit-learn to predict customer satisfaction.

# Classification Task

```
from sklearn.tree import DecisionTreeClassifier
dt_class = DecisionTreeClassifier()
dt_class.fit(xtrain,ytrain)

print('Hyperparameter:\n',dt_class.get_params())
print()

ypred = dt_class.predict(xtest)

print('Predicted class labels for the test samples:\n',ypred)
print()
```

# Regression Task

- ❑ Implement a decision tree regression model using Python libraries to predict sales based on customer attributes and behavior.

```
from sklearn.tree import DecisionTreeRegressor
dt_reg = DecisionTreeRegressor()
dt_reg.fit(xtrain,ytrain)

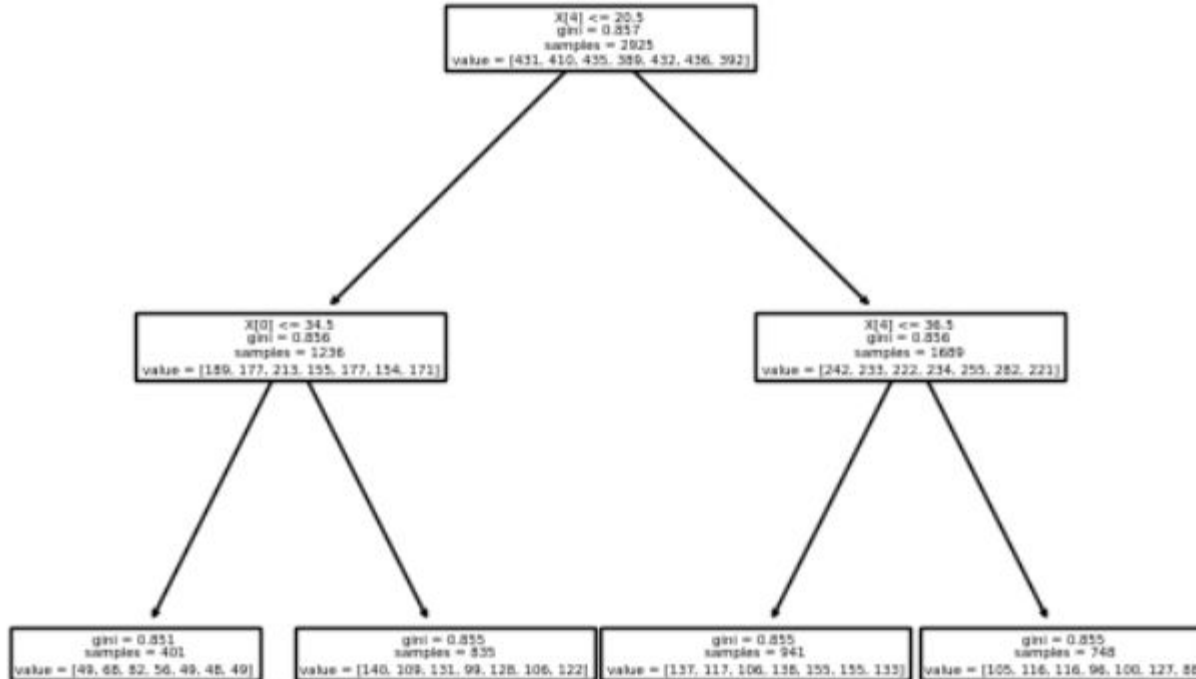
print('Hyperparameter:\n',dt_reg.get_params())
print()

ypred = dt_reg.predict(xtest)

print('Predicted class labels for the test samples:\n',ypred)
```

# Decision Tree Visualization

- Decision trees can be visualized using Python libraries such as graphviz.



# Overfitting

- ❖ Overfitting is a phenomenon where a machine learning model learns the training data too well and is unable to generalize to new data. This can happen with decision trees if they are allowed to grow too large.
- ❖ Decision trees grow by recursively splitting the data into smaller and smaller subsets, until each subset contains only one class or a certain number of data points. This process is called tree construction.
- ❖ If the tree is allowed to grow too large, it will learn all of the details of the training data, including the noise and outliers. This means that the tree will be able to make accurate predictions on the training data, but it will not be able to generalize to new data.

# Methods for reducing overfitting

- ❖ Methods for reducing overfitting, such as pruning, minimum samples per leaf, and maximum depth.
- ❖ There are a few different methods that can be used to reduce overfitting in decision trees:
  - Pruning: Pruning is a technique where nodes are removed from the decision tree to reduce its size. This can be done using a variety of algorithms, such as cost complexity pruning and reduced error pruning.
  - Minimum samples per leaf: This parameter specifies the minimum number of data points that must be in a leaf node before it can be split. Increasing the minimum samples per leaf can help to reduce overfitting.



## Methods for reducing overfitting- cont...

- Maximum depth: This parameter specifies the maximum depth of the decision tree. Increasing the maximum depth can lead to overfitting, while decreasing the maximum depth can lead to underfitting.

```
# Pruning the Decision Tree  
# Pruning can be implemented by setting parameters like max_depth,  
# min_samples_split, min_samples_leaf, or max_leaf_nodes.  
  
clf_pruned = DecisionTreeClassifier(max_depth=2)  
clf_pruned.fit(xtrain, ytrain)  
  
# Evaluate the pruned classification model  
ypred = clf_pruned.predict(xtest)
```

# Model Comparison

Accuracy score : 0.13435897435897437

Confusion Martix :

```
[[17 21 23 15 17 21 27]
 [27 16 20 16 20 18 20]
 [28 16 19 16 28 25 17]
 [27 18 25 21 16 28 18]
 [14 18 14 22 17 21 15]
 [17 19 19 16 14 24 18]
 [34 24 22 16 21 13 17]]
```

Accuracy score : 0.15487179487179487

Confusion Martix :

```
[[44 0 26 0 42 29 0]
 [47 0 24 0 42 24 0]
 [39 0 21 0 41 48 0]
 [43 0 22 0 50 38 0]
 [32 0 18 0 46 25 0]
 [35 0 18 0 34 40 0]
 [34 0 18 0 49 46 0]]
```

★ Accuracy improves after pruning the trees

# Real-World Application

- ❖ Customer satisfaction prediction and sales forecasting can be used in a variety of ways in the retail industry.
- ❖ For example, retailers can use these models to:
  - Identify customers who are at risk of churning, so that they can take steps to retain them.
  - Develop targeted marketing campaigns for different customer segments.
  - Optimize inventory levels and pricing.
  - Plan for future growth.

# Benefits of using predictive models

- ❖ potential benefits of using predictive models in retail business operations and decision making.
- ❖ There are many potential benefits to using predictive models in retail business operations and decision-making. For example, predictive models can help retailers to:
  - Improve customer satisfaction
  - Increase sales
  - Reduce costs
  - Make better business decisions

# Conclusion

- ❖ Decision trees are a powerful machine learning algorithm that can be used for both classification and regression tasks. They are relatively easy to understand and interpret, making them a popular choice for predictive modeling in many different industries.
- ❖ In this case study, we have shown how to use decision trees to build predictive models for customer satisfaction and sales prediction in a retail business. We have also discussed how to reduce overfitting in decision trees and the potential benefits of using predictive models in retail business operations and decision-making.