The Tree of Meowledge

Regression Tree

Prompts



- Create a Regression Tree using the preprocessed dataset 'data', with 'star_rating' as the target. Scale the features before training
- Plot the tree

Strategy



- Prepare the data
- Train the Regression Tree
- Check the Accuracy
- Visualise the tree

Code & Results

1. Prepare the data

This code 1) separates features and target, 2) scales the features, and 3) splits train and test sets.

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

# 1. Separate features and target
X = data.drop(columns=['star_rating'])
y = data['star_rating']

# 2. Scale features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# 3. Train-test split
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)
```

2. Train the Model

This code creates a regression tree model and trains it using the training set (Note max_depth=3).

```
from sklearn.tree import DecisionTreeRegressor, plot_tree

# 3. Train a Regression Tree
tree_model = DecisionTreeRegressor(max_depth=3, random_state=42) # you can tune max_depth
tree_model.fit(X_train, y_train)

# 4. Predictions
y_pred = tree_model.predict(X_test)
```

3. Check the Accuracy

This code evaluates model accuracy.

```
# 5. Evaluation

mse = mean_squared_error(y_test, y_pred)

r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse:.4f}")

print(f"R<sup>2</sup> Score: {r2:.4f}")

MSE, the closer the model predictions are to the actual values.

Mean Squared Error: 0.2777

R<sup>2</sup> Score: -0.3484

X Negative R<sup>2</sup> means that the tree predicts worse than the average!
```

value = 4.8

value = 5.0

4. Visualize the tree

This code plots the regression tree.

```
plt.figure(figsize=(16,8))
plot_tree(tree_model, feature_names=X.columns, filled=True, rounded=True, fontsize=10)
plt.show()
                                                                                                                    total comments <= -0.435
                                                                                                                      squared error = 0.07
                                                                                                                         samples = 96
                                                                                                                         value = 4.422
                                                                                  price <= 0.612
                                                                                                                                                               price <= 0.432
                                                                                quared error = 0.034
                                                                                                                                                              squared error = 0.05
                                                                                                                                                                samples = 86
                                                                                  samples = 10
                                                                                   value = 4.85
                                                                                                                                                                value = 4.372
                                                                                                                                       company whiskas <= 0.401
                                                              adult <= -0.062
                                                                                                total_comments <= -0.435
                                                                                                                                                                                   price <= 0.816
                                                                                                                                        squared error = 0.048
                                                            squared_error = 0.022
                                                                                                  squared error = 0.027
                                                                                                                                                                                squared_error = 0.018
                                                                                                                                             samples = 71
                                                                                                                                                                                   samples = 15
                                                                                                       samples = 4
                                                                samples = 6
                                                               value = 4.933
                                                                                                      value = 4.725
                                                                                                                                                                                    value = 4.553
                                                                                                                                             value = 4.334
                                                                                         squared_error = 0.0 squared_error = 0.02 squared_error = 0.05 squared_error = 0.01 squared_error = 0.01 squared_error = 0.01
                                                   squared_error = 0.04 squared_error = 0.0
                                                                                                                                   samples = 46
                                                                         samples = 4
                                                                                                                                                      samples = 25
                                                      samples = 2
                                                                                             samples = 2
                                                                                                                samples = 2
                                                                                                                                                                          samples = 5
                                                                                                                                                                                             samples = 10
```

value = 4.6

value = 4.85

value = 4.272

value = 4.448

value = 4.64

value = 4.51