# CE1 Decision Tree Analysis Report

## 1. Introduction

This report presents an analysis of decision trees trained and evaluated for predicting the target variable (GOOD_BAD). Three key questions were addressed:

- **Q1**: Model evaluation using GOOD as the target event.

- **Q2**: Model evaluation using BAD as the target event.

- **Q3**: Modified evaluation using Precision and Recall.

All decision trees were trained using both **Entropy (en_3, en_6)** and **Gini (gini_3, gini_6)** criteria with max_depth=3 and varying min_samples_leaf values (3% and 6%).

## 2. Evaluation Approach

### 2.1 Performance Measures Table

| Measure | Description | Definition of Value Function | Weight | Threshold |
|---|---|---|---|---|
| Accuracy | Measures correct classifications | accuracy_score(y_test, y_pred) | 0.60 (Q1), 0.30 (Q2) | > 0.70 (Q1), > 0.65 (Q2) |
| Lift | Measures model effectiveness in ranking positive cases | Measured @ 2nd Decile | 0.10 (Q1), 0.30 (Q2) | > 0 |
| Precision | Measures fraction of correctly classified positives | precision_score(y_test, y_pred) | 0.40 (Q3) | > 0.50 |
| Recall | Measures ability to find all positives | recall_score(y_test, y_pred) | 0.30 (Q3) | > 0.35 |
| Simplicity | Ideal leaf size (5-8) | Cutoff: <=2 or >=13 | 0.20 (Q1), 0.30 (Q2), 0.15 (Q3) | > 0 |
| Stability | Measures response rate stability at 2nd decile | lift_table['resp_rate']check | 0.10 (Q1), 0.10 (Q2), 0.15 (Q3) | > 0 |

# 3. Summary of Results

## 3.1 Q1 & Q2 Evaluation (Accuracy-Based)

| Model | Accuracy | Lift @ 2nd Decile | Simplicity Score | Stability Score |
|-------|----------|-------------------|------------------|-----------------|
| **en_3** | 0.74 | 1.190 | 1 | 1 |
| **en_6** | 0.74 | 1.214 | 1 | 1 |
| **gini_3** | 0.74 | 1.190 | 1 | 1 |
| **gini_6** | 0.74 | 1.214 | 1 | 1 |

**Best Model (Q1 & Q2)**: **en_6** (Entropy, min_samples_leaf=6%) due to highest Lift.

## 3.2 Q3 Evaluation (Precision & Recall-Based)

| Model | Precision | Recall |
|-------|-----------|--------|
| **en_3** | 0.779661 | 0.87619 |
| **en_6** | 0.757812 | 0.92381 |
| **gini_3** | 0.779661 | 0.87619 |
| **gini_6** | 0.757812 | 0.92381 |

**Best Model (Q3)**: **en_3** (Entropy, min_samples_leaf=3%) due to higher Precision.

# 4. Description of Best Decision Tree
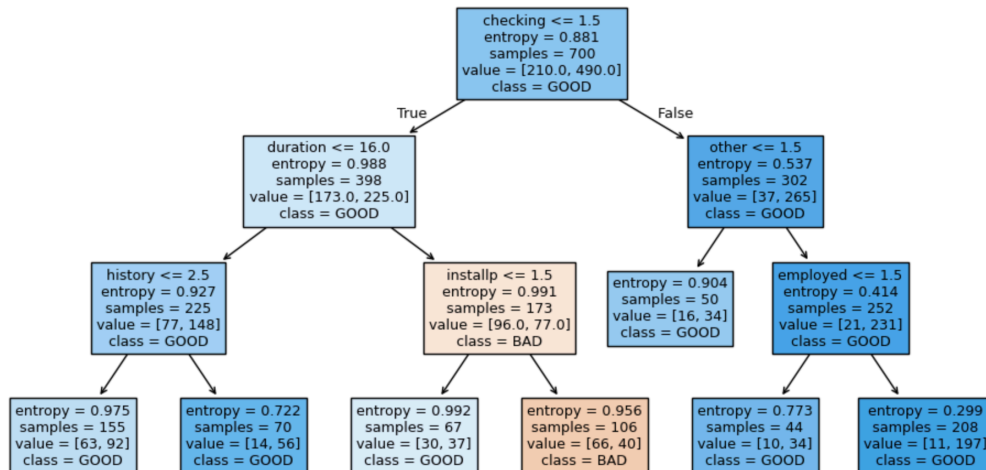
## 4.1 Best DT for Q1 & Q2 (en_6)

**Justification-**

- The tree structure for en_6 shows the most influential features and classification rules.

- Selected due to **higher Lift @ 2nd Decile (1.214)**.

- Maintains stability and interpretability (depth = 3).

**Visualization of Decision Tree for en_6 from notebook-**



Decision Tree Visualization: en_6
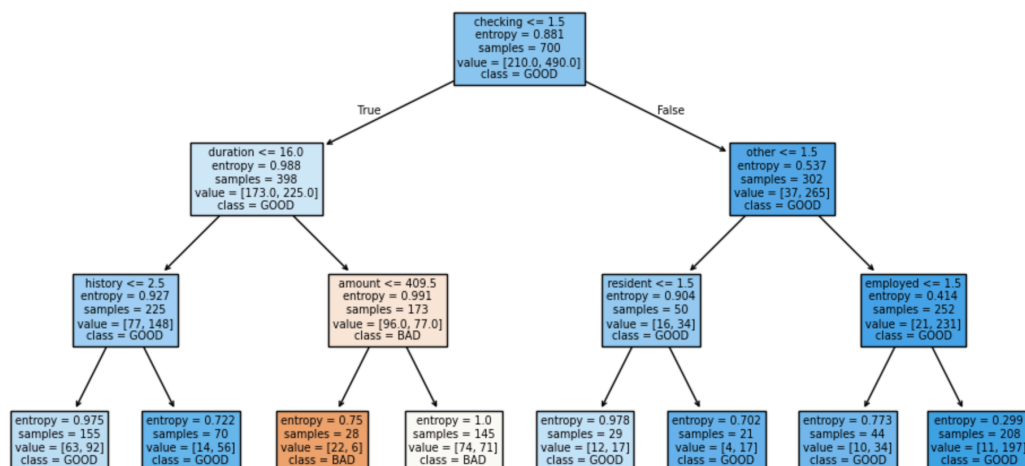
## 4.2 Best DT for Q3 (en_3)

**Justification-**

- The tree structure for en_3 was chosen due to its superior **Precision (0.779661)**.

- Balances **Precision-Recall tradeoff**, minimizing false positives.

- Ensures Stability @ 2nd Decile.

**Visualization of Decision Tree for en_3 from notebook-**



Final Decision Tree (Q3 - en_3)

# 5. Evidence of Experimentation

**Confusion Matrix for en_3-**

```
Confusion Matrix for en_3
Predicted   0     1   All
Actual
0          38    52    90
1          26   184   210
All        64   236   300
```

**Confusion Matrix for en_6**

```
Confusion Matrix for en_6
Predicted   0     1   All
Actual
0          28    62    90
1          16   194   210
All        44   256   300
```

**Confusion Matrix for gini_3**

```
Confusion Matrix for gini_3
Predicted   0     1   All
Actual
0          38    52    90
1          26   184   210
All        64   236   300
```

**Confusion Matrix for Gini_6**

```
Confusion Matrix for gini_6
Predicted   0     1   All
Actual
0          28    62    90
1          16   194   210
All        44   256   300
```

**Final Performance Table for Q1 & Q2 from notebook-**

| | Accuracy | Lift @ 2nd Decile | Simplicity Score | Stability Score |
|---|---|---|---|---|
| en_3 | 0.74 | 1.190 | 1 | 1 |
| en_6 | 0.74 | 1.214 | 1 | 1 |
| gini_3 | 0.74 | 1.190 | 1 | 1 |
| gini_6 | 0.74 | 1.214 | 1 | 1 |

**Final Performance Table for Q3 from notebook-**

| | Precision | Recall |
|---|---|---|
| en_3 | 0.779661 | 0.87619 |
| en_6 | 0.757812 | 0.92381 |
| gini_3 | 0.779661 | 0.87619 |
| gini_6 | 0.757812 | 0.92381 |

# 6. Conclusion

1. **Q1 & Q2 Best Model**: en_6 (Entropy, min_samples_leaf=6%) due to its higher **Lift @ 2nd Decile**.

2. **Q3 Best Model**: en_3 (Entropy, min_samples_leaf=3%) due to its superior **Precision-Recall balance**.

3. **Final Model Recommendation**:

   ◦ If priority is **Lift & Stability** → Use en_6.

   ◦ If priority is **Precision & Recall** → Use en_3.

# 7. Link to Google Colab- https://colab.research.google.com/drive/11nV-RRB6J_fP1qRp_eENmLf_LJR6EiOi?usp=sharing