Decision Tree Classifier (Дерево решений)

Дорофеев Демид, Клюкин Павел
М8О-307Б-23

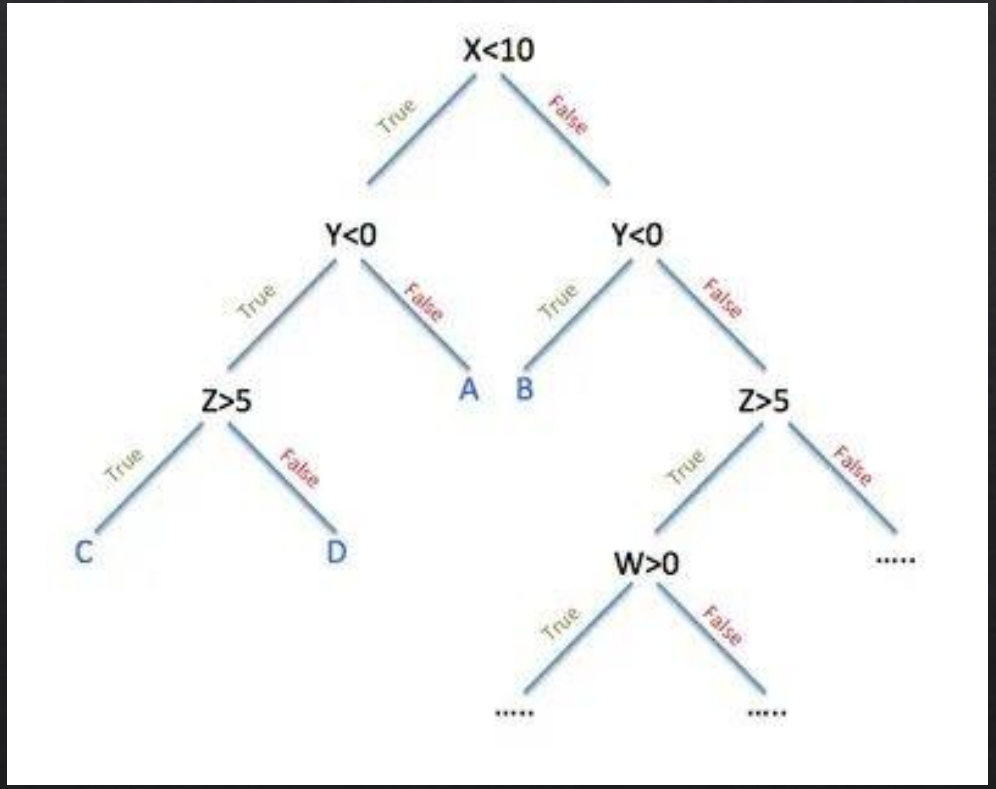
Дерево решений

Преимущества:

- Простая интерпретация (можно визуализировать дерево).
- Не требует масштабирования данных.

Недостатки:

- Склонно к **переобучению**, особенно при большой глубине дерева.
- Может быть нестабильным (небольшие изменения данных → другое дерево).



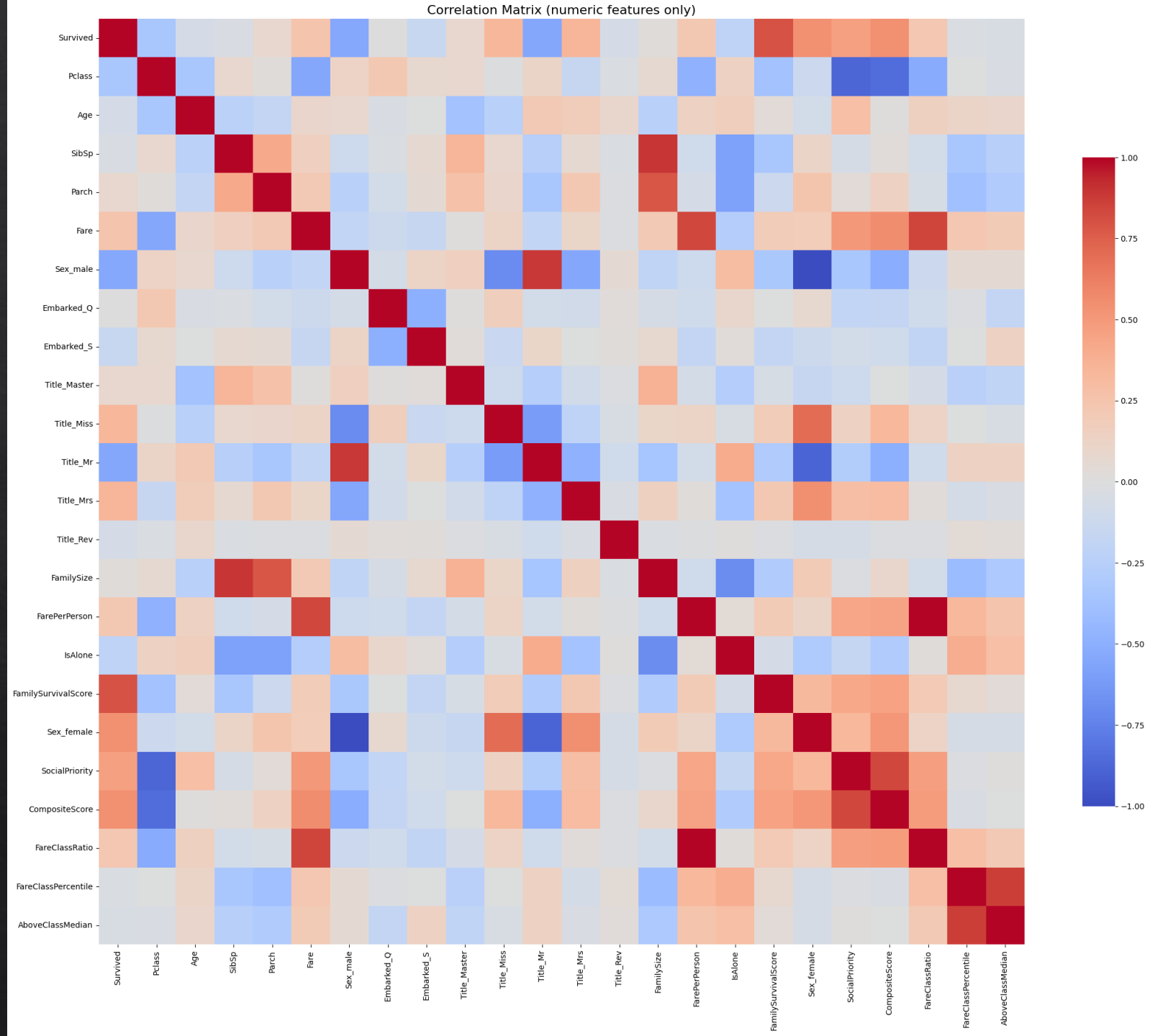
Подготовка

Добавление новых features

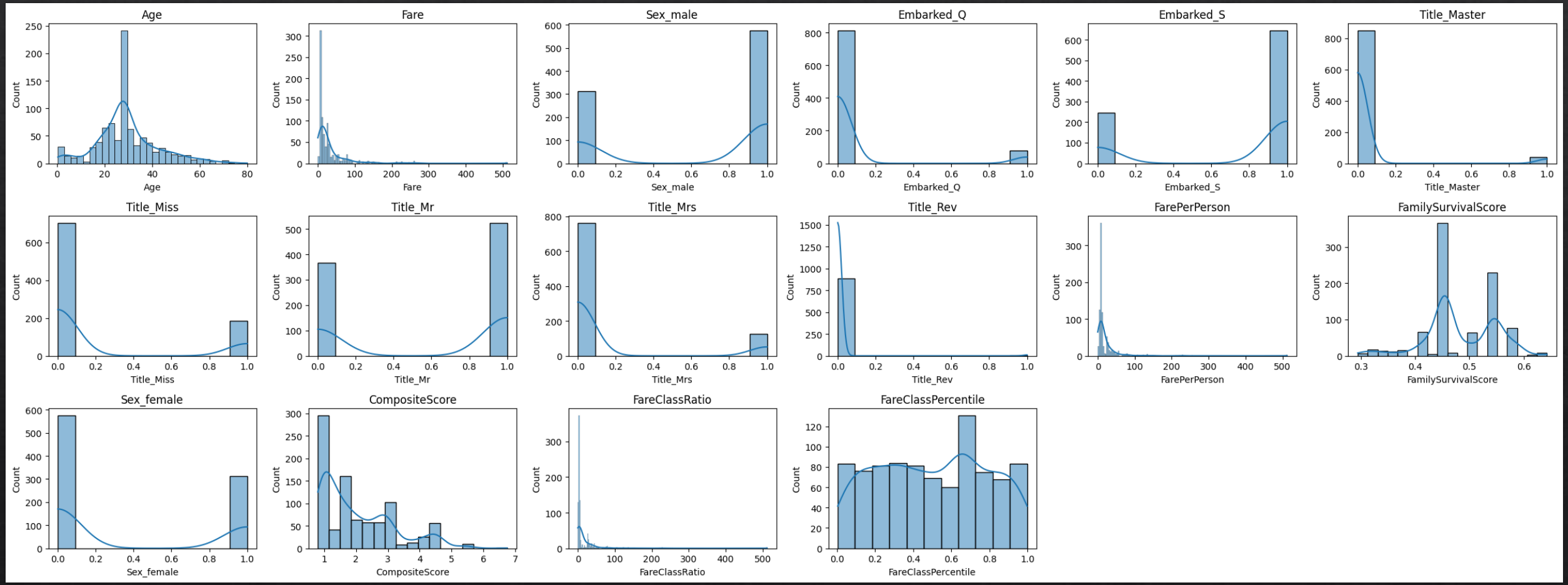
PassengerI,
Survived
Pclass,
Name, Sex,
Age, SibSp,
Parch,
Ticket,
Fare, Cabin,
Embarked

Survived, Pclass, Age, SibSp, Parch, Fare, Sex_male,
Embarked_Q, Embarked_S, Title_Master, Title_Miss, Title_Mr,
Title_Mrs, Title_Rev, FamilySize, FarePerPerson, IsAlone,
FamilySurvivalScore, Sex_female, SocialPriority,
CompositeScore, FareClassRatio, FareClassPercentile,
AboveClassMedian, Archetype_AveragePassenger,
Archetype_Businessman, Archetype_PoorChild,
Archetype_RichChild, Archetype_WealthyMother,
Archetype_WealthySingleWoman, Archetype_WorkingClassMan

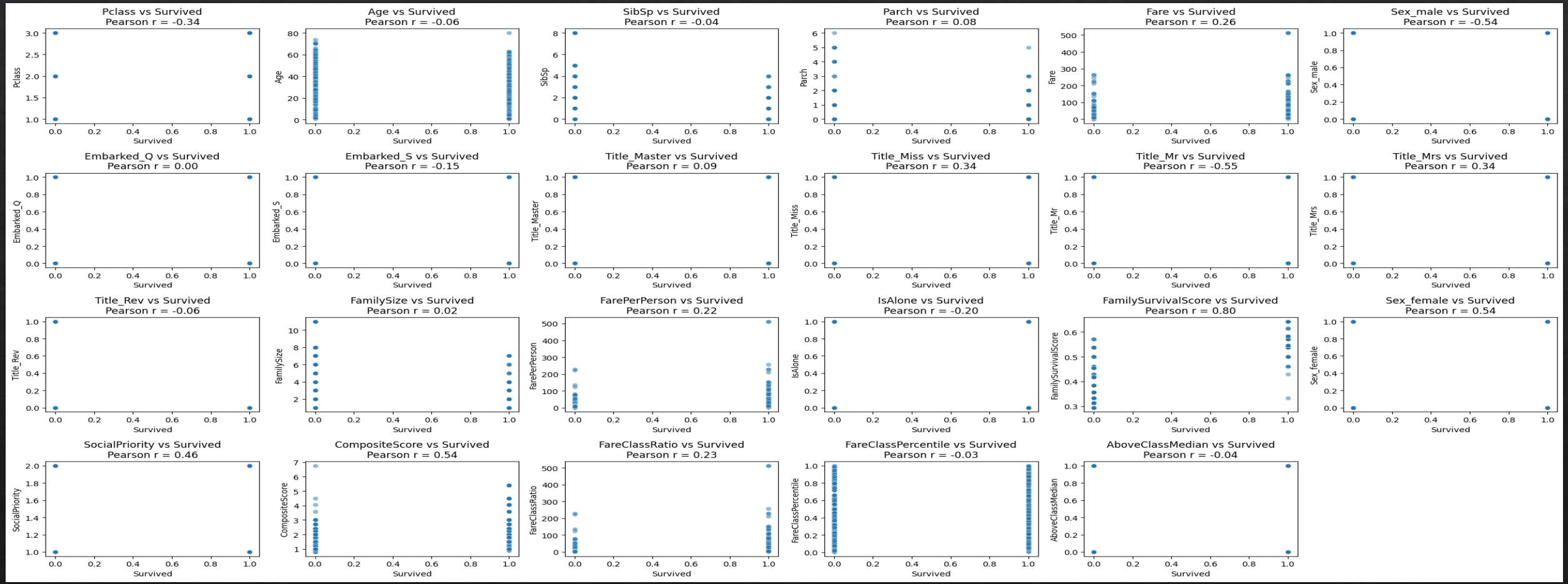
Корреляционная матрица признаков



Распределение данных



Соотношение с целевой переменной + коэфициент пирсона

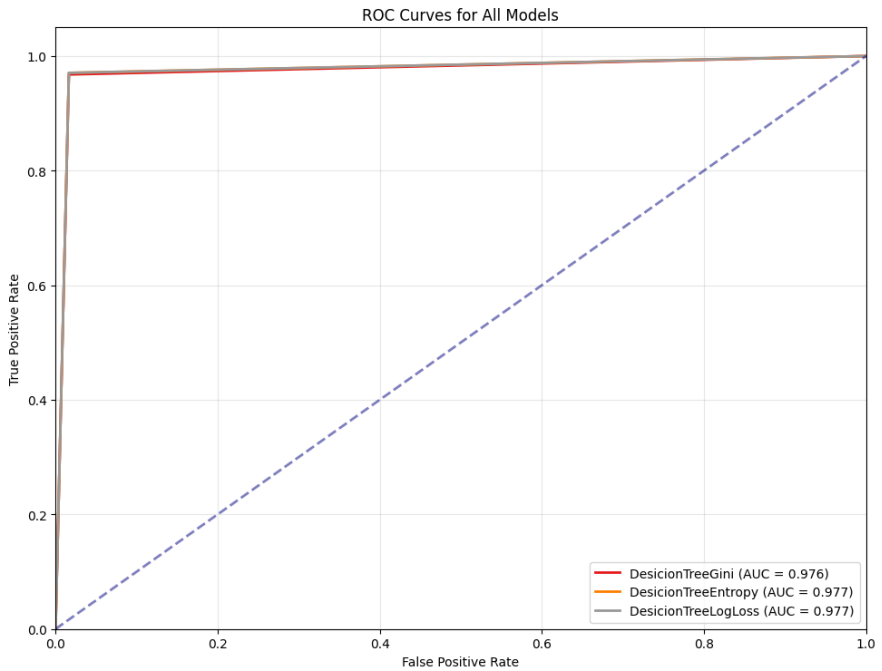


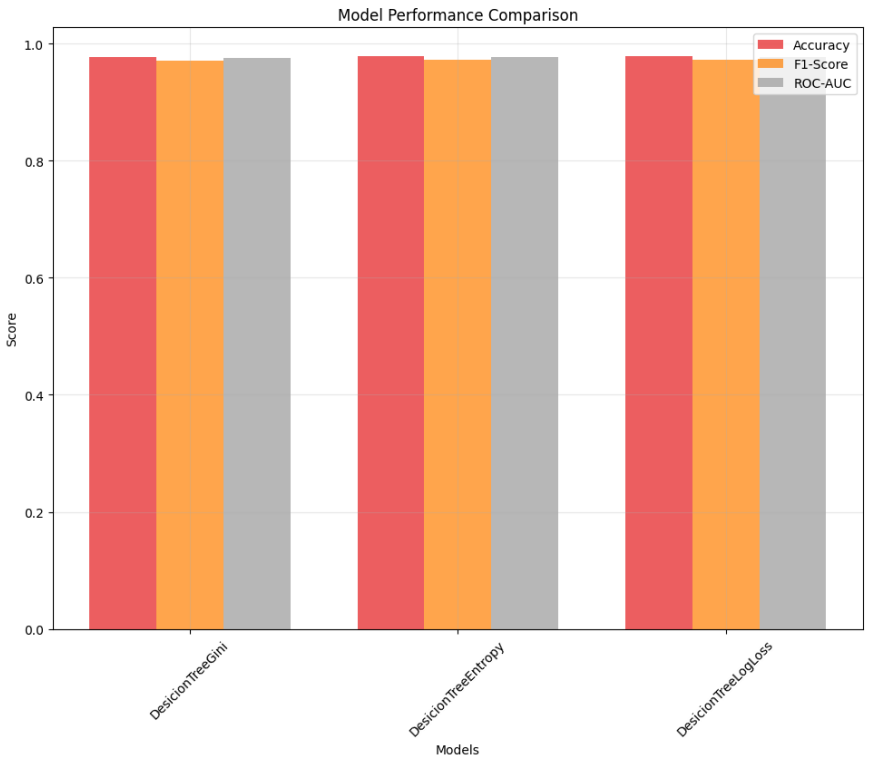
Критерий качества разбиения.

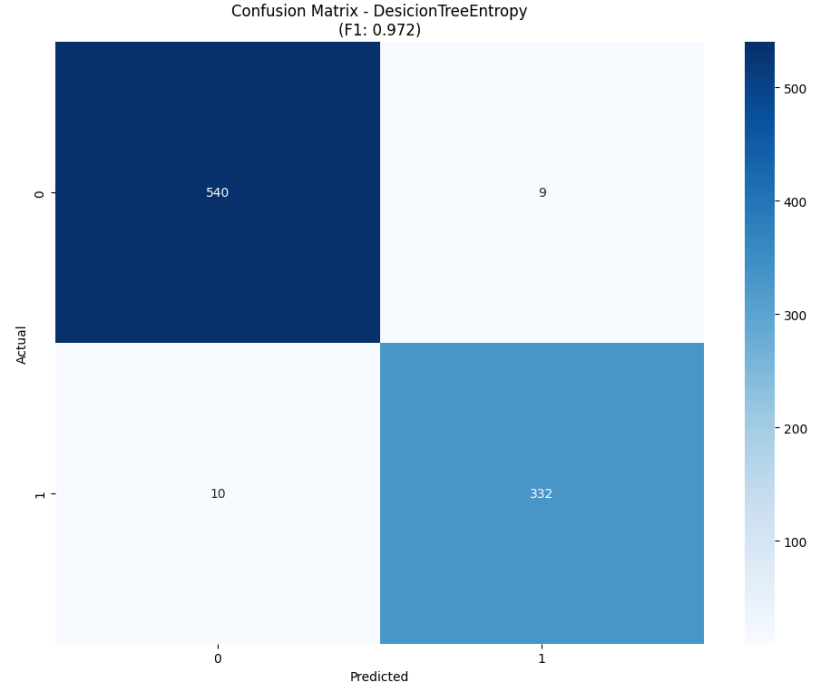
gini — индекс Джини

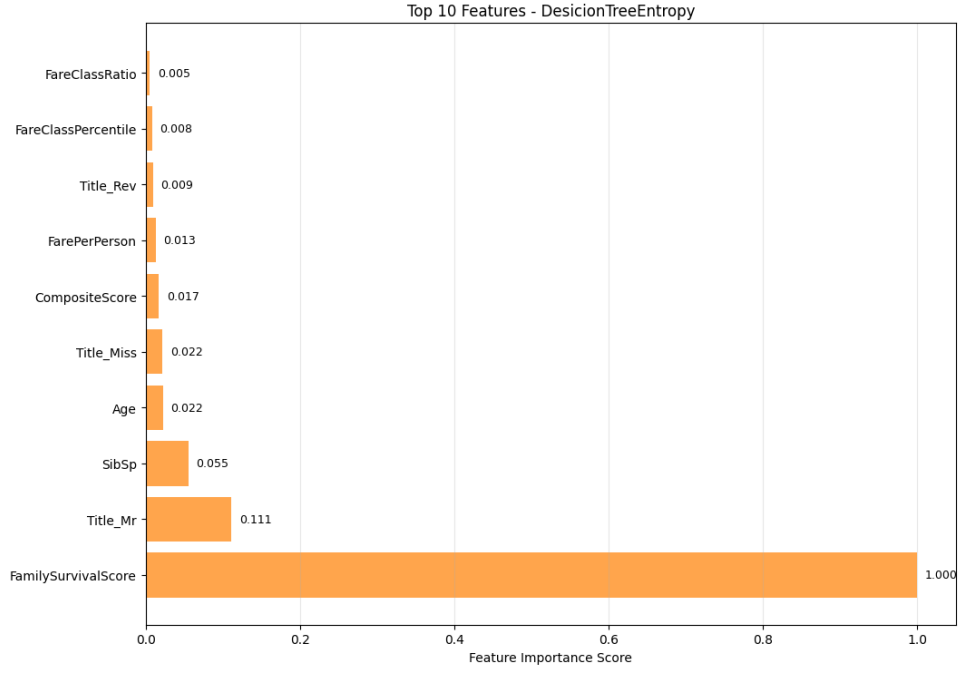
entropy —энтропия

log_loss — основан на логарифмической потере.





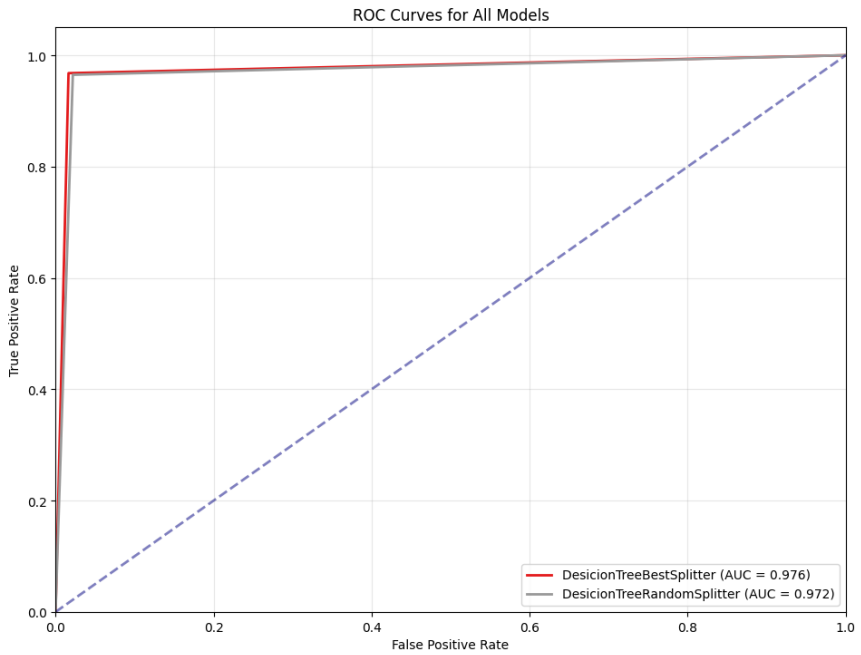


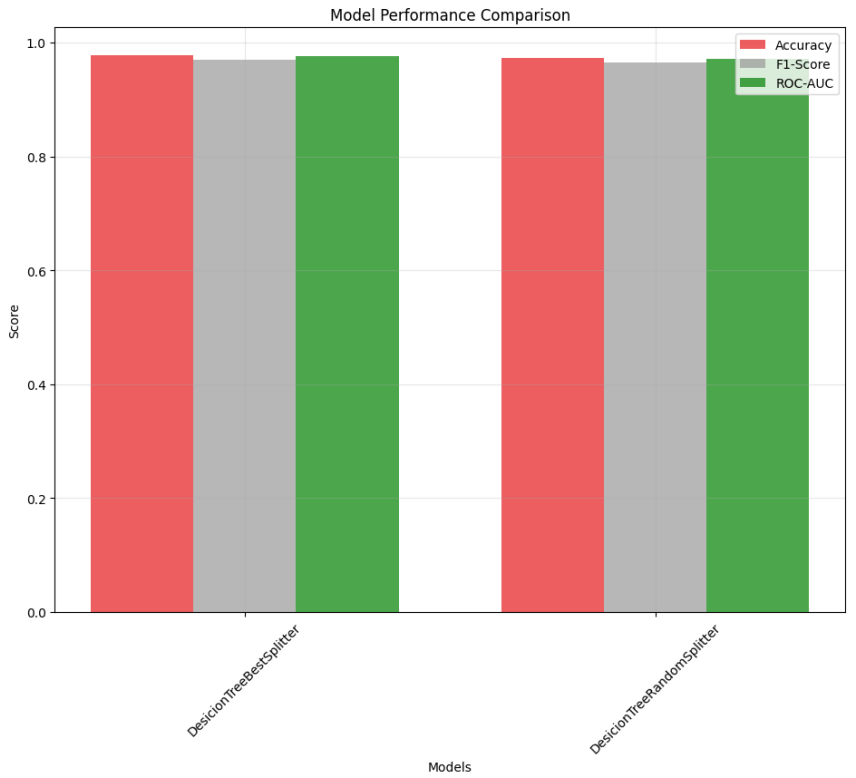


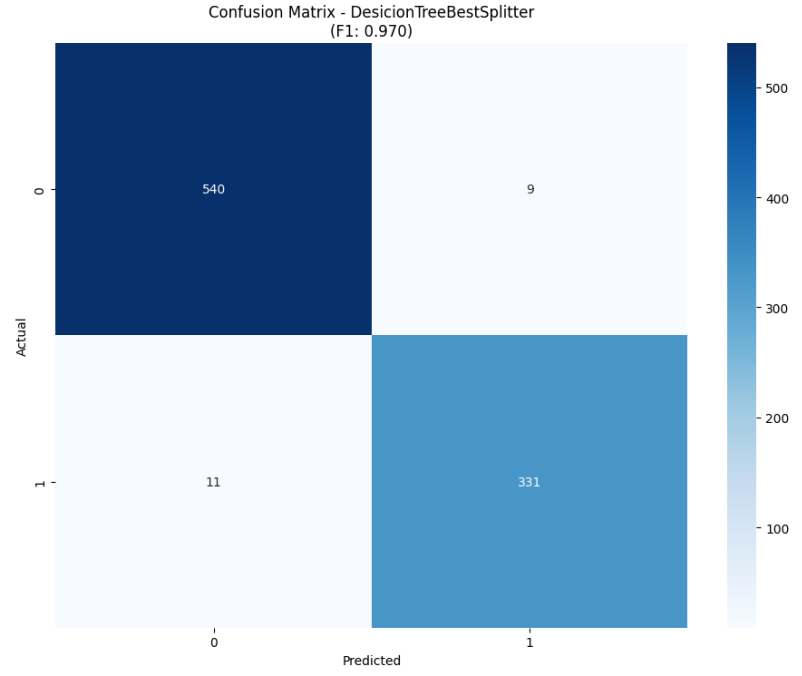
Стратегия выбора признака для разбиения.

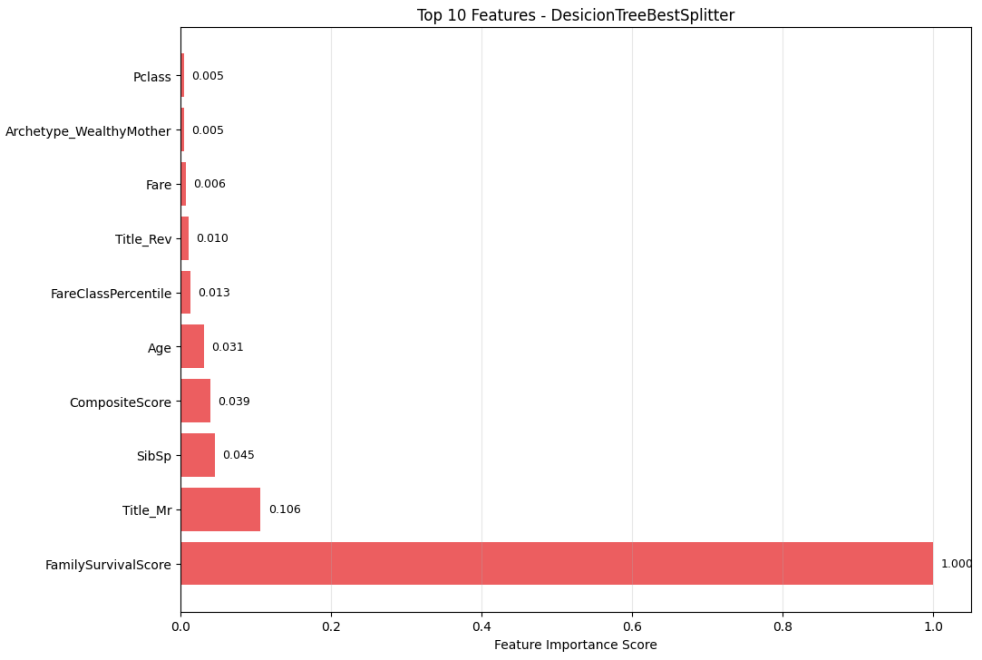
best — выбирает наилучшее разбиение

random' — выбирает случайный признак (для ускорения и разнообразия).



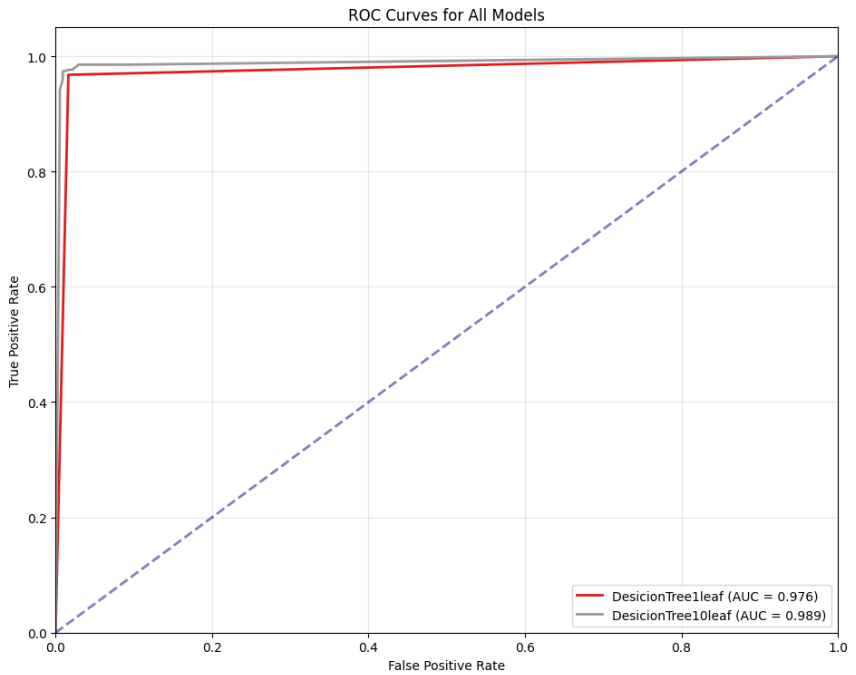


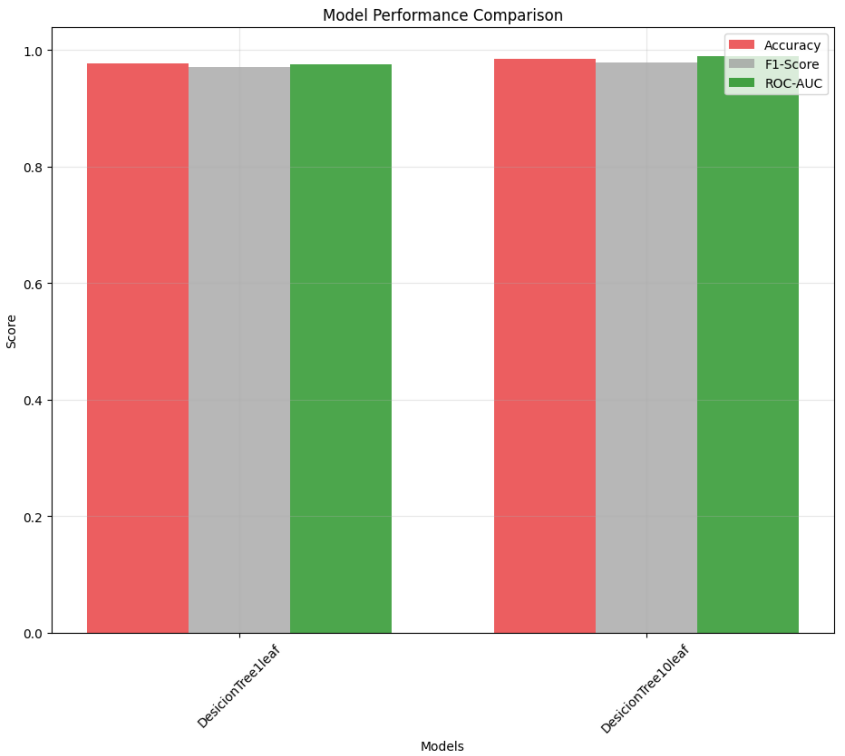


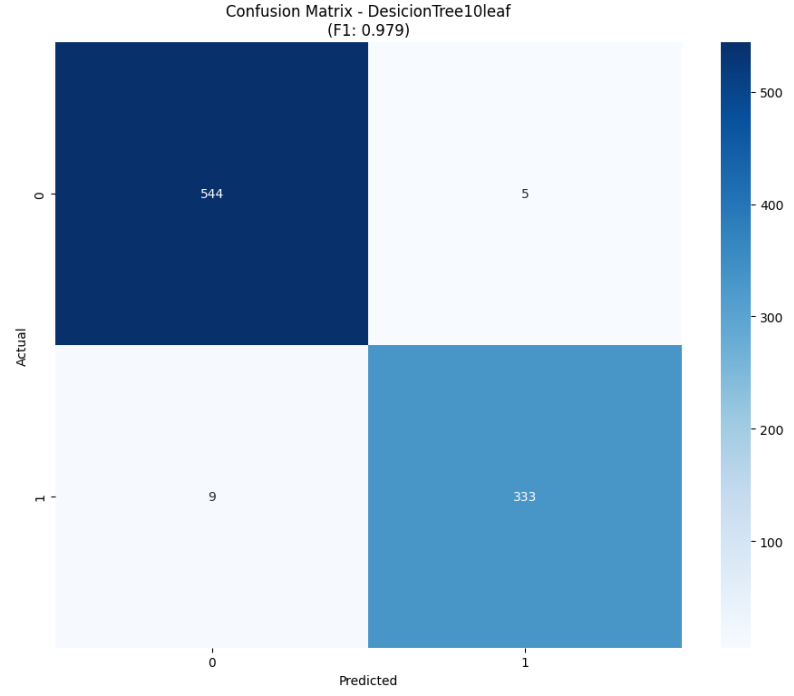


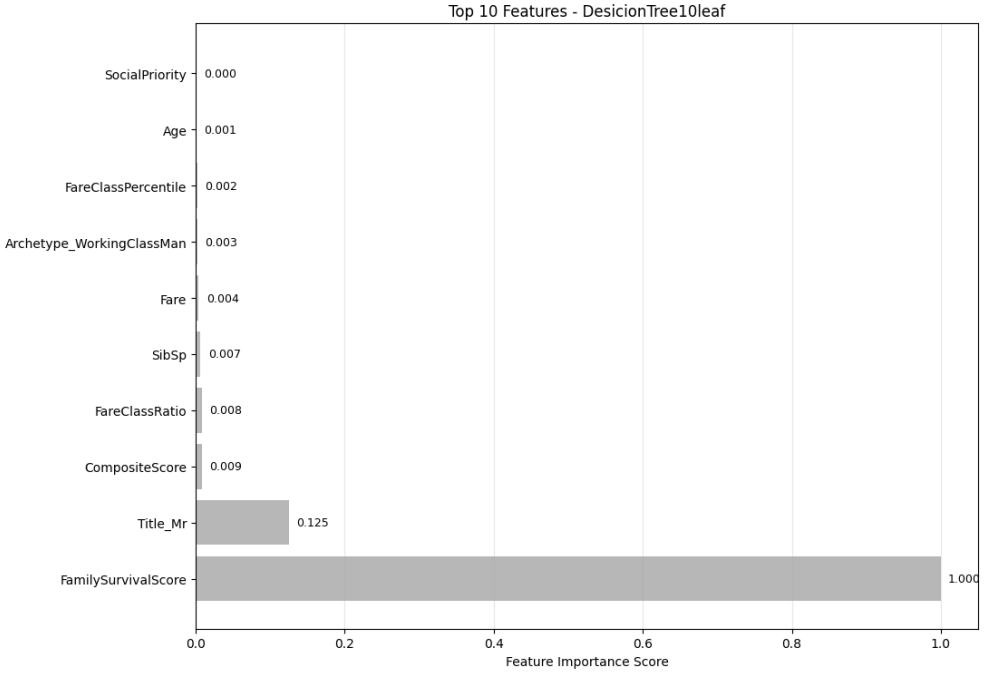
Минимальное количество образцов в листе.

Увеличение значения делает дерево более “гладким”.



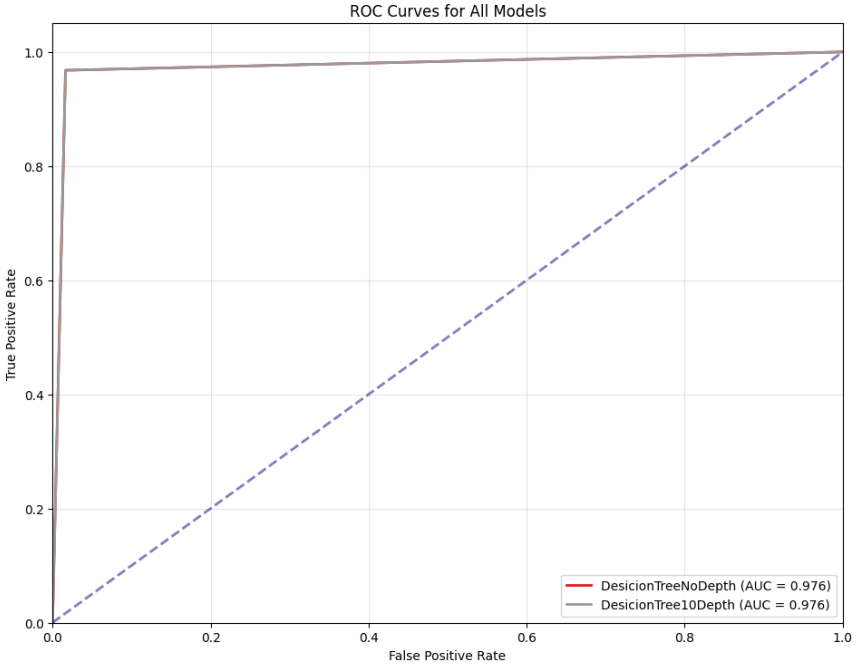


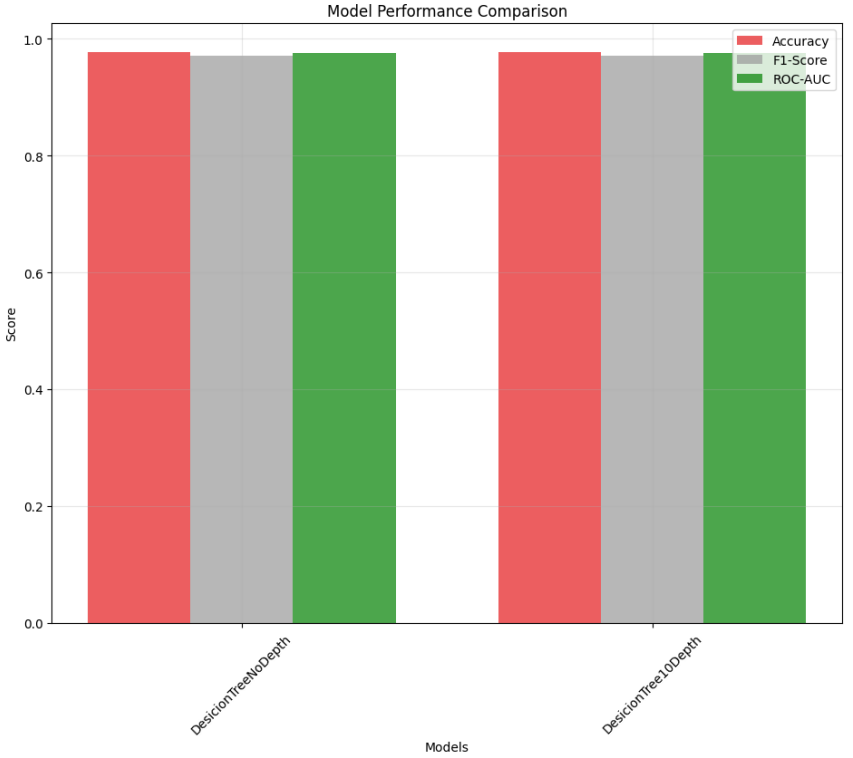


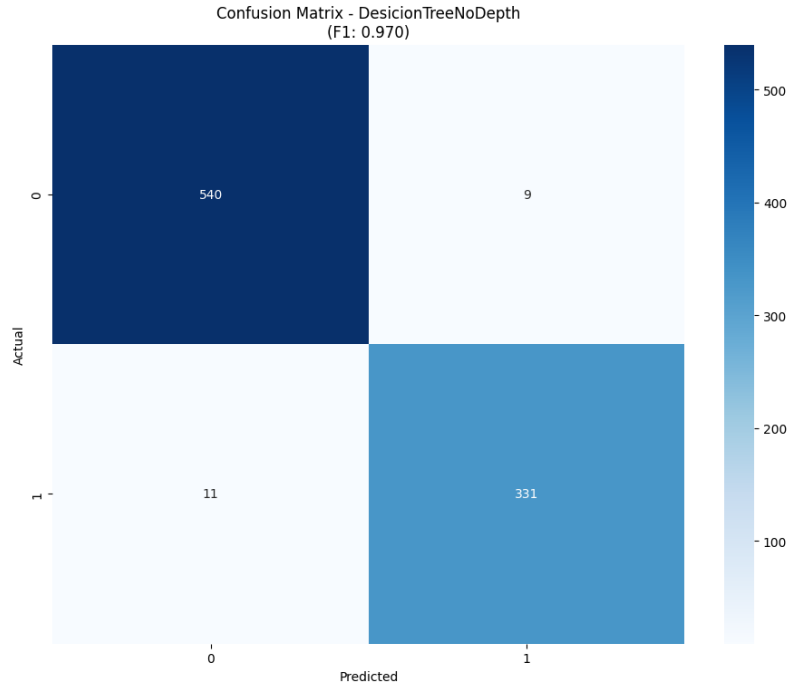


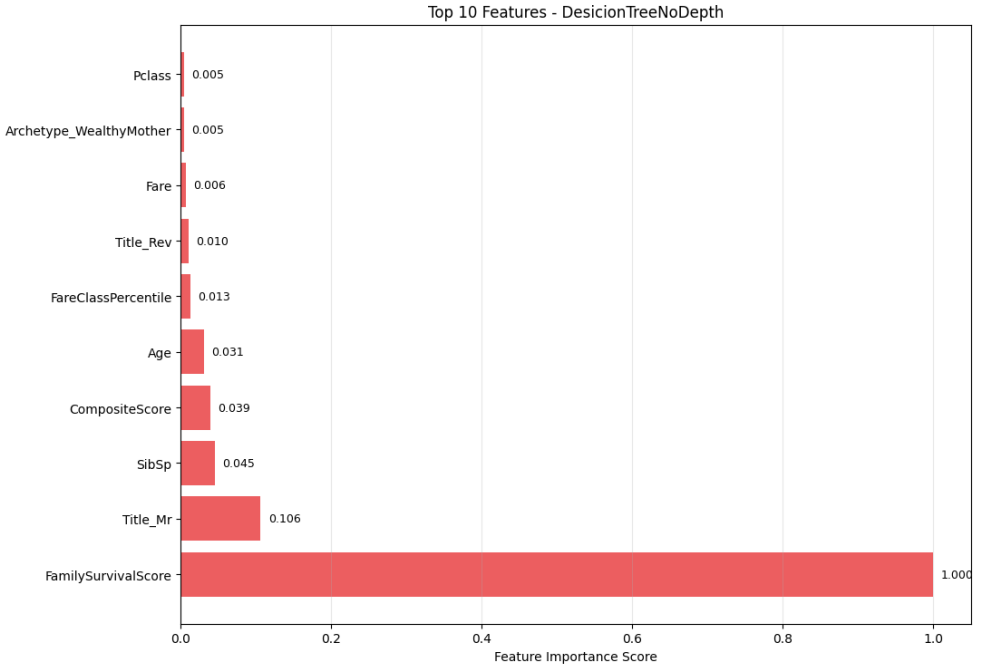
Максимальная глубина дерева.

None — без ограничения (может привести к переобучению).

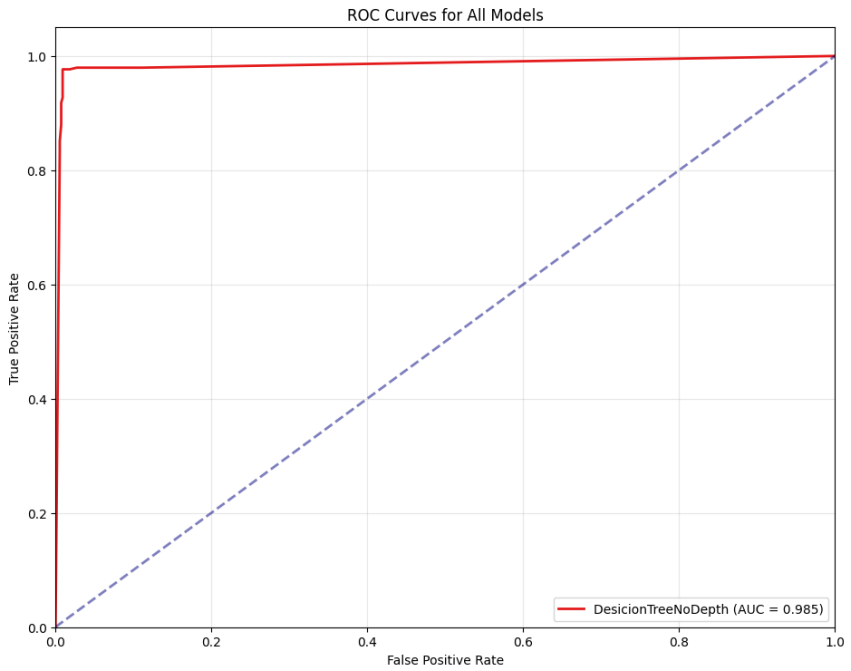


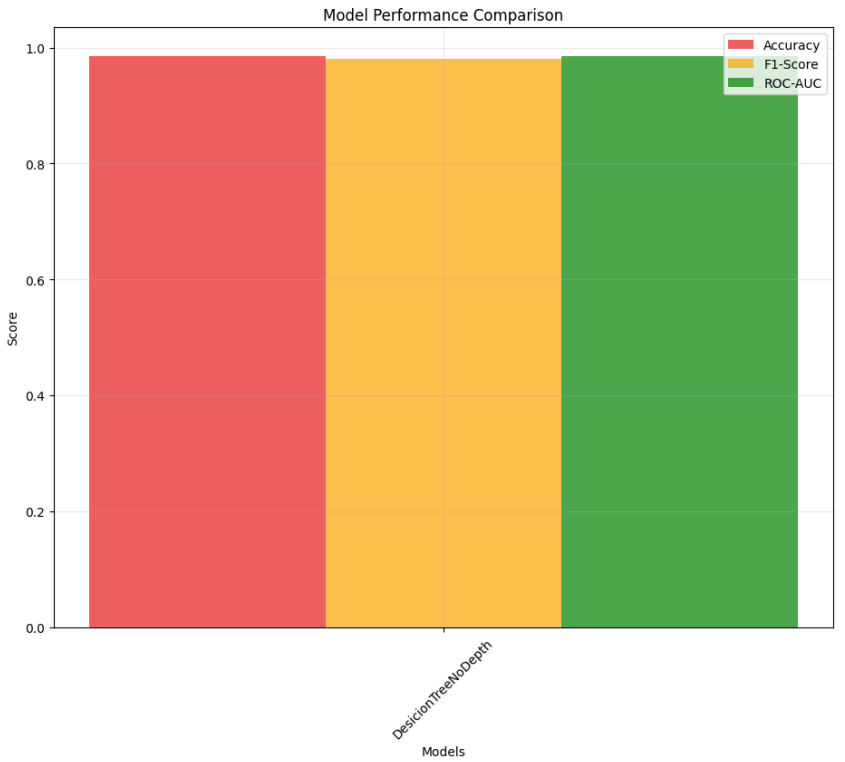


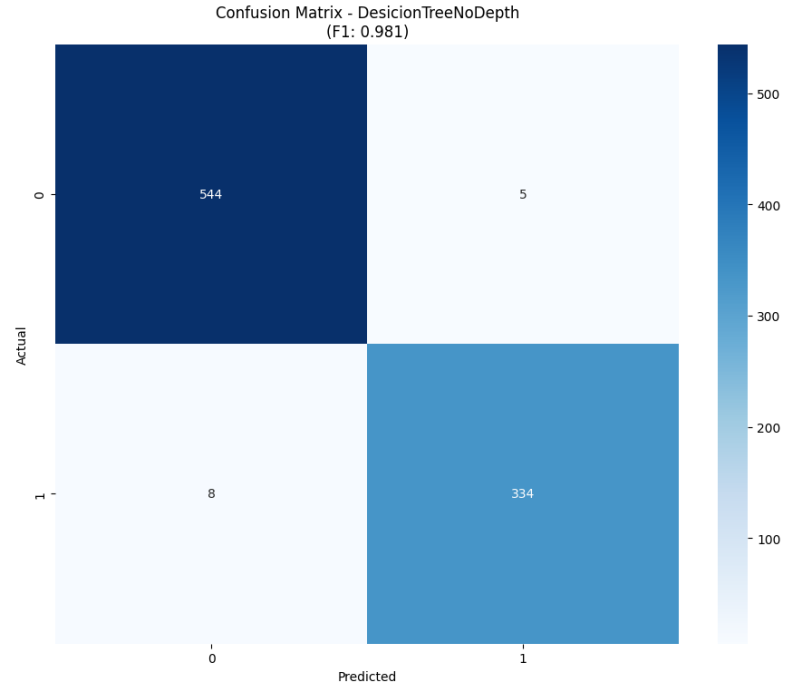


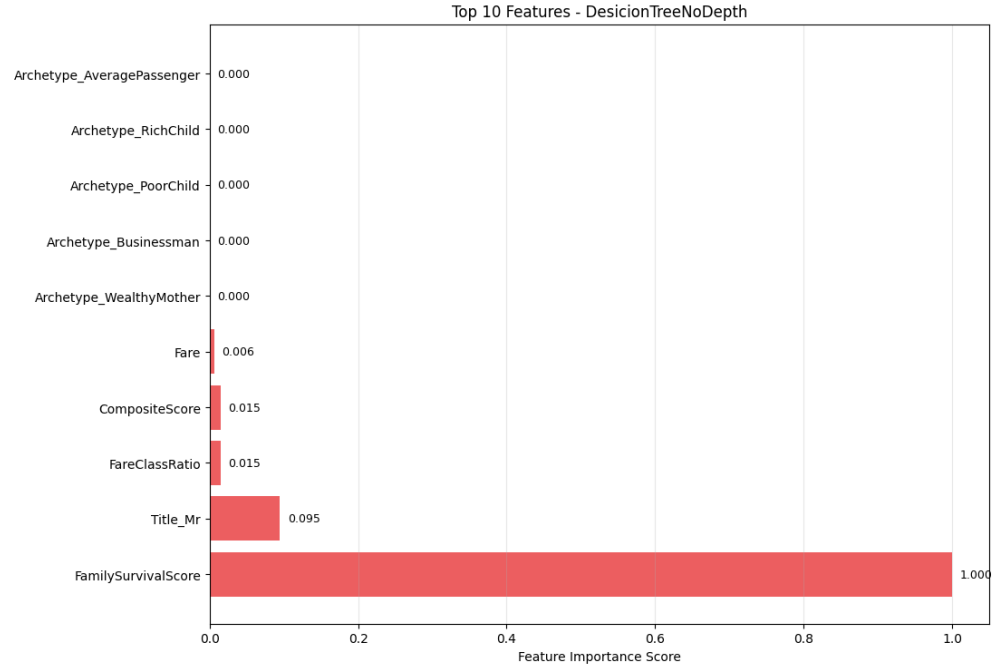


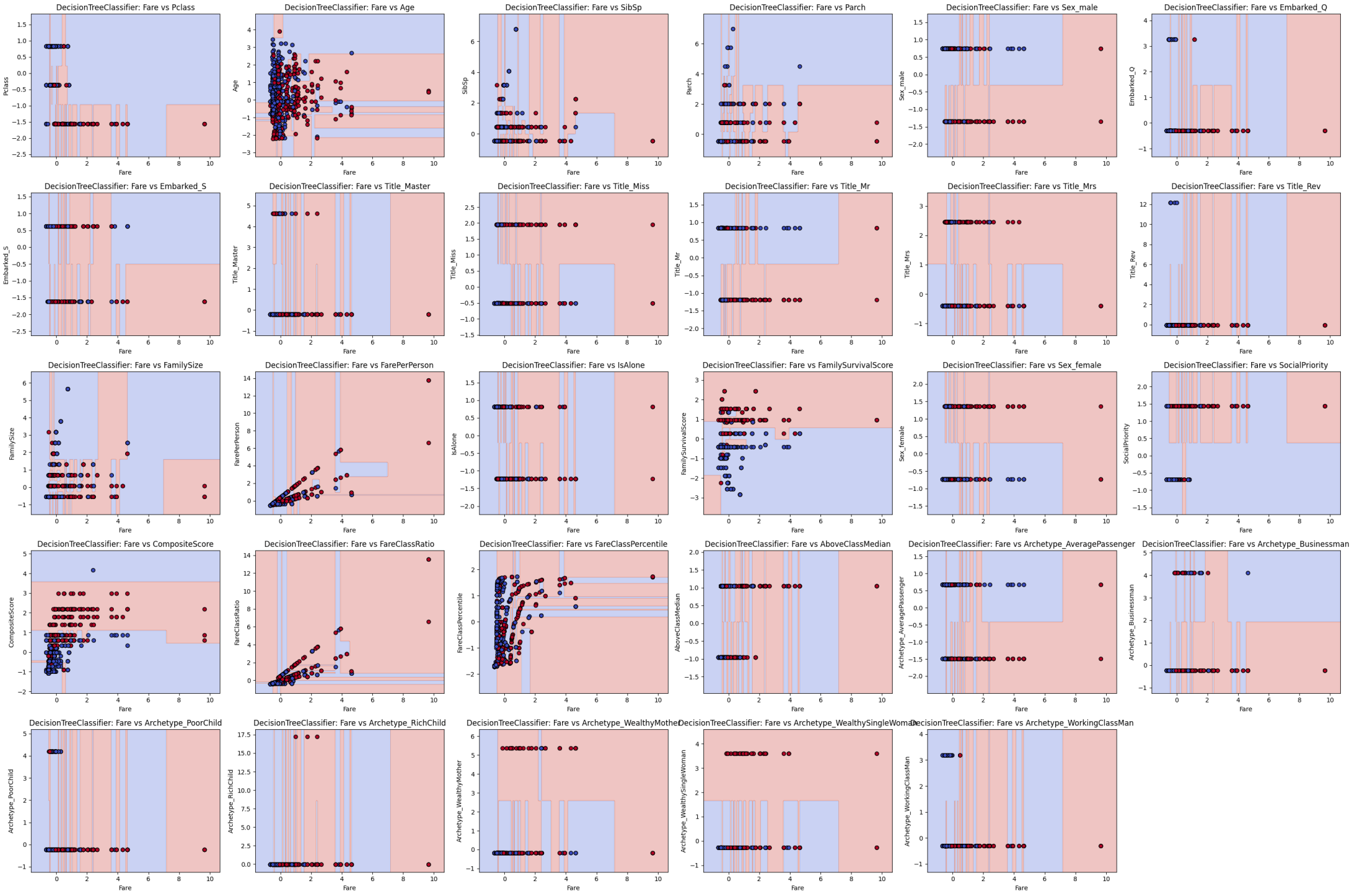
- criterion=gini
- splitter=best
- max_depth=None
- min_samples_split=2
- min_samples_leaf=15
- random_state=RANDOM_STATE
- max_features=None
- class_weight=None










DecisionTreeClassifier: Fare vs Pclass
DecisionTreeClassifier: Fare vs Age
DecisionTreeClassifier: Fare vs SibSp
DecisionTreeClassifier: Fare vs Parch
DecisionTreeClassifier: Fare vs Sex_male
DecisionTreeClassifier: Fare vs Embarked_Q
DecisionTreeClassifier: Fare vs Embarked_S
DecisionTreeClassifier: Fare vs Title_Master
DecisionTreeClassifier: Fare vs Title_Miss
DecisionTreeClassifier: Fare vs Title_Mr
DecisionTreeClassifier: Fare vs Title_Mrs
DecisionTreeClassifier: Fare vs Title_Rev
DecisionTreeClassifier: Fare vs FamilySize
DecisionTreeClassifier: Fare vs FarePerPerson
DecisionTreeClassifier: Fare vs IsAlone
DecisionTreeClassifier: Fare vs FamilySurvivalScore
DecisionTreeClassifier: Fare vs Sex_female
DecisionTreeClassifier: Fare vs SocialPriority
DecisionTreeClassifier: Fare vs CompositeScore
DecisionTreeClassifier: Fare vs FareClassRatio
DecisionTreeClassifier: Fare vs FareClassPercentile
DecisionTreeClassifier: Fare vs AboveClassMedian
DecisionTreeClassifier: Fare vs Archetype_AveragePassenger
DecisionTreeClassifier: Fare vs Archetype_Businessman
DecisionTreeClassifier: Fare vs Archetype_PoorChild
DecisionTreeClassifier: Fare vs Archetype_RichChild
DecisionTreeClassifier: Fare vs Archetype_WealthyMother
DecisionTreeClassifier: Fare vs Archetype_WealthySingleWoman
DecisionTreeClassifier: Fare vs Archetype_WorkingClassMan
Fare
Pclass
Age
SibSp
Parch
Sex_male
Embarked_Q
Embarked_S
Title_Master
Title_Miss
Title_Mr
Title_Mrs
Title_Rev
FamilySize
FarePerPerson
IsAlone
FamilySurvivalScore
Sex_female
SocialPriority
CompositeScore
FareClassRatio
FareClassPercentile
AboveClassMedian
Archetype_AveragePassenger
Archetype_Businessman
Archetype_PoorChild
Archetype_RichChild
Archetype_WealthyMother
Archetype_WealthySingleWoman
Archetype_WorkingClassMan