

Jungle Rumble

Random vs CatBoost

Prompts



- Compare RandomForest vs CatBoost using the preprocessed dataset 'data', with 'star_rating' as the target

Strategy



- Prepare the data
- Train RandomForest and CatBoost Regressors
- Compare the Accuracy

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 Models

This code trains RandomForest Regressor.

```
from sklearn.ensemble import RandomForestRegressor

# 3. Random Forest
rf_model = RandomForestRegressor(
    n_estimators=200,
    max_depth=None,
    random_state=42,
    n_jobs=-1
)
rf_model.fit(X_train, y_train)
rf_pred = rf_model.predict(X_test)
```

This code trains CatBoost Regressor.

```
from catboost import CatBoostRegressor

# 4. CatBoost (silent mode)
cat_model = CatBoostRegressor(
    iterations=500,
    depth=8,
    learning_rate=0.1,
    random_state=42,
    verbose=0
)
cat_model.fit(X_train, y_train)
cat_pred = cat_model.predict(X_test)
```

3. Compare the Accuracy

This code evaluates models accuracy.

```
# 5. Evaluation
results = {
    "RandomForest": {
        "MSE": mean_squared_error(y_test, rf_pred),
        "R²": r2_score(y_test, rf_pred)
    },
    "CatBoost": {
        "MSE": mean_squared_error(y_test, cat_pred),
        "R²": r2_score(y_test, cat_pred)
    }
}
```

```
# Display results in a nice table
results_df = pd.DataFrame(results).T
print(results_df)
```

	MSE	R²
RandomForest	0.234249	-0.137221
CatBoost	0.215931	-0.048288

✅ The MSE helps to see if your model is good enough. The smaller the MSE, the closer the model predictions are to the actual values.

❌ Negative R² means that models predict worse than the average!