In [1]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestClassifier, ExtraTreesClassifier, A
from sklearn.metrics import accuracy_score, classification_report

# Загрузка данных
df = pd.read_csv("filtered_dataset.csv")

# Удалим колонку url
df = df.drop(columns=["url","status","status_encoded"])

# Кодирование категориальных признаков
cat_features = ["tld", "extension"]
df[cat_features] = df[cat_features].fillna("unknown")  # Заполним пропуски
df[cat_features] = df[cat_features].apply(LabelEncoder().fit_transform)

# Разделение на X и y
X = df.drop(columns=["is_malicious"])
y = df["is_malicious"]

# Train/test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, rar
```

In [2]:
```
# обучение моделей
```

In [3]:
```
# Модели бэггинга

model_rf = RandomForestClassifier(random_state=42)
model_et = ExtraTreesClassifier(random_state=42)

# Boosting

model_ada = AdaBoostClassifier(random_state=42)
model_gb = GradientBoostingClassifier(random_state=42)

# Обучение
print("Обучение...")
print("Бэггинг...")
model_rf.fit(X_train, y_train)
model_et.fit(X_train, y_train)
print("Boosting...")
model_ada.fit(X_train, y_train)
model_gb.fit(X_train, y_train)

print("OK")
```

```
Обучение...
Бэггинг...
Boosting...
```

```
/home/lexcf/.local/lib/python3.10/site-packages/sklearn/ensemble/_weight_boosting.py:527: FutureWarning: The SAMME.R algorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warning.
  warnings.warn(
OK
```

In [4]:
```python
# оценка качества
```

In [5]:
```python
models = {
    "Random Forest": model_rf,
    "Extra Trees": model_et,
    "AdaBoost": model_ada,
    "Gradient Boosting": model_gb
}

for name, model in models.items():
    y_pred = model.predict(X_test)
    acc = accuracy_score(y_test, y_pred)
    print(f"{name} Accuracy: {acc:.4f}")
    print(classification_report(y_test, y_pred))
```

```
Random Forest Accuracy: 0.8559
              precision    recall  f1-score   support

           0       0.85      0.87      0.86      1915
           1       0.87      0.85      0.86      1972

    accuracy                           0.86      3887
   macro avg       0.86      0.86      0.86      3887
weighted avg       0.86      0.86      0.86      3887

Extra Trees Accuracy: 0.8562
              precision    recall  f1-score   support

           0       0.84      0.88      0.86      1915
           1       0.88      0.83      0.85      1972

    accuracy                           0.86      3887
   macro avg       0.86      0.86      0.86      3887
weighted avg       0.86      0.86      0.86      3887

AdaBoost Accuracy: 0.7245
              precision    recall  f1-score   support

           0       0.70      0.76      0.73      1915
           1       0.75      0.69      0.72      1972

    accuracy                           0.72      3887
   macro avg       0.73      0.72      0.72      3887
weighted avg       0.73      0.72      0.72      3887

Gradient Boosting Accuracy: 0.7697
              precision    recall  f1-score   support

           0       0.73      0.86      0.79      1915
           1       0.83      0.69      0.75      1972

    accuracy                           0.77      3887
   macro avg       0.78      0.77      0.77      3887
weighted avg       0.78      0.77      0.77      3887
```

In [ ]:

In [ ]:

In [ ]: