```
import os
import pickle
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.ensemble import IsolationForest

DATASET_DIR = "dataset_cleaned"

texts = []

for root, _, files in os.walk(DATASET_DIR):
  for file in files:
    if file.endswith(".txt"):
      with open(os.path.join(root, file), "r", encoding="utf-8") as f:
        texts.append(f.read())

print(f"Loaded {len(texts)} documents for anomaly detection training.")

# TF-IDF (classification ile aynı temsil uzayı)
vectorizer = TfidfVectorizer(stop_words="english", max_features=5000)
X = vectorizer.fit_transform(texts)

# Isolation Forest
anomaly_model = IsolationForest(
```

```
  n_estimators=100,
  contamination=0.02,  # %2 anomalous varsayımı
  random_state=42
)

anomaly_model.fit(X)

os.makedirs("model", exist_ok=True)
with open("model/anomaly_model.pkl", "wb") as f:
  pickle.dump((vectorizer, anomaly_model), f)

print("✅ Anomaly detection model saved: model/anomaly_model.pkl")
```