```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pickle

from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, f1_score

df = pd.read_excel("twitter_data.xlsx”)
print("\nFirst 5 Rows:")
print(df.head())



df = df.dropna()

df["SENTIMENT_LABEL"] = df["SENTIMENT"].map({
   "NEGATIVE": 0,
   "NEUTRAL": 1,
   "POSITIVE": 2
})



sentiment_summary = df["SENTIMENT"].value_counts()
```

```
print("\nSentiment Summary (Pivot Table):")
print(sentiment_summary)



plt.figure(figsize=(15,6))

plt.subplot(1,2,1)
company_totals = df.groupby("COMPANY")["TOTAL"].sum().sort_values()
sns.barplot(x=company_totals.values,
            y=company_totals.index,
            palette="Blues_r")
plt.title("Total Mentions by Company")
plt.xlabel("Total")
plt.ylabel("Company")

plt.subplot(1,2,2)
colors = ["#1f77b4","#ff7f0e","#2ca02c"]

plt.pie(sentiment_summary.values,
        labels=sentiment_summary.index,
        autopct='%1.1f%%',
        startangle=140,
        colors=colors,
        shadow=True)

plt.title("Sentiment Distribution")

plt.tight_layout()
plt.show()
```

```
# Features
X = df[["COMPANY", "TOTAL"]]
y = df["SENTIMENT_LABEL"]

categorical_features = ["COMPANY"]
numeric_features = ["TOTAL"]

preprocessor = ColumnTransformer(
    transformers=[
        ("cat", OneHotEncoder(handle_unknown="ignore"), categorical_features),
        ("num", SimpleImputer(strategy="median"), numeric_features)
    ]
)

pipeline = Pipeline(steps=[
    ("preprocessor", preprocessor),
    ("classifier", DecisionTreeClassifier(random_state=42))
])

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

param_grid = {
    "classifier__max_depth": [3,5,7,10],
    "classifier__criterion": ["gini","entropy"]
}

grid = GridSearchCV(pipeline, param_grid, cv=3)
grid.fit(X_train, y_train)
```

```
best_model = grid.best_estimator_

print("\nBest Parameters:", grid.best_params_)


y_pred = best_model.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred, average="weighted")

print("\nModel Accuracy:", round(accuracy*100,2), "%")
print("F1 Score:", round(f1,4))

print("\nClassification Report:\n")
print(classification_report(y_test, y_pred))

cm = confusion_matrix(y_test, y_pred)

plt.figure(figsize=(5,4))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues")
plt.title("Confusion Matrix")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()


ohe = best_model.named_steps["preprocessor"]\
            .named_transformers_["cat"]

encoded_features = list(ohe.get_feature_names_out(categorical_features))
```

```
all_features = encoded_features + numeric_features

importances = best_model.named_steps["classifier"].feature_importances_

feature_importance_df = pd.DataFrame({
    "Feature": all_features,
    "Importance": importances
}).sort_values(by="Importance", ascending=False)

print("\nTop Feature Importance:")
print(feature_importance_df)

# Plot Feature Importance
plt.figure(figsize=(8,5))
sns.barplot(x="Importance", y="Feature", data=feature_importance_df)
plt.title("Feature Importance")
plt.show()

with open("sentiment_decision_tree_model.pkl", "wb") as f:
    pickle.dump(best_model, f)

print("\nModel Saved Successfully as sentiment_decision_tree_model.pkl")
```

SINO	COMPANY	SENTIMEN	TWEET TEXT	TOTA
1	AMAZON	NEUTRAL	BBC News Amazon boss Jeff Bezos rejects claims company acted like a 'drug dealer'	877
2	MICROSOFT	NEGATIVE	@Microsoft Why do I pay for WORD when it functions so poorly on my @SamsungUS Chromebook?	207
3	CS-GO	NEGATIVE	CSGO matchmaking is so full of closet hacking, it's a truly awful game.	947
4	GOOGLE	NEUTRAL	Now the President is slapping Americans in the face that he really did commit an unlawful act after	207
5	FIFA	NEGATIVE	Hi @EAHelp I've had Madeleine McCann in my cellar for the past 13 years and the little sneaky thing	967
6	MaddenNFL	POSITIVE	Thank you @EAMaddenNFL!! New TE Austin Hooper in the ORANGE & BROWN!! #Browns | @AustinHooper18	818
7	TomClancysRainbowSix	POSITIVE	Rocket League, Sea of Thieves or Rainbow Six: Siege ? I love playing all three on stream but which	47
8	Assassins Creed	POSITIVE	my ass still knee-deep in Assassins Creed Odyssey with no way out anytime soon lmao	504
9	CallofDuty	NEGATIVE	Please FIX IT! What In the world is going on here. @PlayStation @AskPlayStation @P1...	362
10	Dota2	POSITIVE	The professional dota 2 scene is fucking exploding and I completely welcome it. Get the garbage out...	645
11	AssassinsCreed	POSITIVE	Itching to assassinate #TCCGif #Assassins CreedBlack Flag #Assassins Creed #TheCapturedCollecti ve pic...	385
12	Verizon	NEGATIVE	@FredTJoseph hey fred, Comcast cut the cable and now Verizon stays calling me to shut that too pic.t...	369
13	CS-GO	NEGATIVE	CSGO WIngman (Im Silver dont bully)	276

Row Labels	Count of SENTIMENT
NEGATIVE	6
NEUTRAL	2
POSITIVE	5
Grand Total	13