```
import pandas as pd
import re
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, accuracy_score
import joblib

# Load dataset (example CSV structure: text,label)
# label: 1 = Depressed, 0 = Normal
data = pd.read_csv("bangla_dataset.csv")

# Bangla text cleaning function
def clean_text(text):
    text = str(text)
    text = re.sub(r"[^\u0980-\u09FF\s]", "", text)  # Keep Bangla Unicode only
    text = re.sub(r"\s+", " ", text)
    return text.strip()

# Preprocessing
data["cleaned"] = data["text"].apply(clean_text)

# Features and labels
X = data["cleaned"]
```

```
y = data["label"]


# Train-test split
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)


# Vectorization
vectorizer = TfidfVectorizer(max_features=3000)
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)


# Model
model = LogisticRegression()
model.fit(X_train_vec, y_train)


# Prediction
y_pred = model.predict(X_test_vec)


# Evaluation
print("Accuracy:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))


# Save model and vectorizer
```

```
joblib.dump(model, "depression_model.pkl")

joblib.dump(vectorizer, "vectorizer.pkl")


print("Model and Vectorizer saved successfully.")
```