In [8]:
```python
import pandas as pd
df=pd.read_csv('healthcare_dataset.csv')
df
```

Out[8]:

	Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date	Medication	Test_Results
0	Bobby JacksOn	30	Male	B-	Cancer	31-01-2024	Matthew Smith	Sons and Miller	Blue Cross	18856.281310	328	Urgent	02-02-2024	Paracetamol	Normal
1	LesLie TErRy	62	Male	A+	Obesity	20-08-2019	Samantha Davies	Kim Inc	Medicare	33643.327290	265	Emergency	26-08-2019	Ibuprofen	Inconclusive
2	DaNnY sMitH	76	Female	A-	Obesity	22-09-2022	Tiffany Mitchell	Cook PLC	Aetna	27955.096080	205	Emergency	07-10-2022	Aspirin	Normal
3	andrEw waTtS	28	Female	O+	Diabetes	18-11-2020	Kevin Wells	Hernandez Rogers and Vang,	Medicare	37909.782410	450	Elective	18-12-2020	Ibuprofen	Abnormal
4	adrIENNE bEll	43	Female	AB+	Cancer	19-09-2022	Kathleen Hanna	White-White	Aetna	14238.317810	458	Urgent	09-10-2022	Penicillin	Abnormal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
55495	eLIZABeTH jaCkSOn	42	Female	O+	Asthma	16-08-2020	Joshua Jarvis	Jones-Thompson	Blue Cross	2650.714952	417	Elective	15-09-2020	Penicillin	Abnormal
55496	KYle pEREz	61	Female	AB-	Obesity	23-01-2020	Taylor Sullivan	Tucker-Moyer	Cigna	31457.797310	316	Elective	01-02-2020	Aspirin	Normal
55497	HEATher WaNG	38	Female	B+	Hypertension	13-07-2020	Joe Jacobs DVM	and Mahoney Johnson Vasquez,	UnitedHealthcare	27620.764720	347	Urgent	10-08-2020	Ibuprofen	Abnormal
55498	JENniFER JOneS	43	Male	O-	Arthritis	25-05-2019	Kimberly Curry	Jackson Todd and Castro,	Medicare	32451.092360	321	Elective	31-05-2019	Ibuprofen	Abnormal
55499	jAMES GARCiA	53	Female	O+	Arthritis	02-04-2024	Dennis Warren	Henry Sons and	Aetna	4010.134172	448	Urgent	29-04-2024	Ibuprofen	Abnormal

55500 rows × 15 columns

In [9]:
```python
df.isnull().sum()
```

Out[9]:
```
Name                  0
Age                   0
Gender                0
Blood Type            0
Medical Condition     0
Date of Admission     0
Doctor                0
Hospital              0
Insurance Provider    0
Billing Amount        0
Room Number           0
Admission Type        0
Discharge Date        0
Medication            0
Test_Results          0
dtype: int64
```

Import libraries

In [10]:
```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report
from sklearn.feature_extraction.text import TfidfVectorizer
import numpy as np

data = pd.read_csv('healthcare_dataset.csv')  # Replace with your dataset path
data.columns = data.columns.str.strip()
print(data.head())
```

```
            Name  Age  Gender Blood Type Medical Condition Date of Admission  \
0  Bobby JacksOn   30    Male         B-            Cancer        31-01-2024   
1   LesLie TErRy   62    Male         A+           Obesity        20-08-2019   
2    DaNnY sMitH   76  Female         A-           Obesity        22-09-2022   
3   andrEw waTtS   28  Female         O+          Diabetes        18-11-2020   
4  adrIENNE bEll   43  Female        AB+            Cancer        19-09-2022   

             Doctor                    Hospital Insurance Provider  \
0     Matthew Smith             Sons and Miller         Blue Cross   
1   Samantha Davies                     Kim Inc           Medicare   
2  Tiffany Mitchell                    Cook PLC              Aetna   
3       Kevin Wells  Hernandez Rogers and Vang,           Medicare   
4    Kathleen Hanna                 White-White              Aetna   

   Billing Amount  Room Number Admission Type Discharge Date   Medication  \
0     18856.28131          328         Urgent     02-02-2024  Paracetamol   
1     33643.32729          265      Emergency     26-08-2019    Ibuprofen   
2     27955.09608          205      Emergency     07-10-2022      Aspirin   
3     37909.78241          450       Elective     18-12-2020    Ibuprofen   
4     14238.31781          458         Urgent     09-10-2022   Penicillin   

  Test_Results  
0       Normal  
1 Inconclusive  
2       Normal  
3     Abnormal  
4     Abnormal  
```

Encode Labels and Split Data

In [11]:
```python
# Encode medical conditions and medications
medical_condition_encoder = LabelEncoder()
medication_encoder = LabelEncoder()

data['Encoded_Condition'] = medical_condition_encoder.fit_transform(data['Medical Condition'])
data['Encoded_Medication'] = medication_encoder.fit_transform(data['Medication'])


X = data[['Encoded_Condition']]
y = data['Encoded_Medication']


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

Train the Model

In [12]:
```python
# Train a Random Forest Classifier
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

# Evaluate the model
y_pred = model.predict(X_test)
print("Classification Report:")
print(classification_report(y_test, y_pred, target_names=medication_encoder.classes_))
```

```
Classification Report:
              precision    recall  f1-score   support

     Aspirin       0.19      0.32      0.24      2211
   Ibuprofen       0.00      0.00      0.00      2271
     Lipitor       0.20      0.17      0.18      2224
 Paracetamol       0.20      0.33      0.25      2207
  Penicillin       0.18      0.16      0.17      2187

    accuracy                           0.19     11100
   macro avg       0.16      0.20      0.17     11100
weighted avg       0.15      0.19      0.17     11100
```

```
C:\Users\R.MUNIRANJANI\anaconda3\Lib\site-packages\sklearn\metrics\_classification.py:1531: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to con
trol this behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
C:\Users\R.MUNIRANJANI\anaconda3\Lib\site-packages\sklearn\metrics\_classification.py:1531: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to con
trol this behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
C:\Users\R.MUNIRANJANI\anaconda3\Lib\site-packages\sklearn\metrics\_classification.py:1531: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to con
trol this behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
```

Handle Undefined Conditions

In [13]:
```python
# TF-IDF Vectorizer for medical conditions
vectorizer = TfidfVectorizer()
condition_vectors = vectorizer.fit_transform(medical_condition_encoder.classes_)

def predict_medication(input_condition):
    try:

        encoded_condition = medical_condition_encoder.transform([input_condition])
        predicted_medication = model.predict([[encoded_condition[0]]])
    except ValueError:

        print(f"'{input_condition}' is not recognized. Searching for the closest match...")

        input_vector = vectorizer.transform([input_condition])
        similarities = np.dot(condition_vectors, input_vector.T).toarray().flatten()
        closest_idx = np.argmax(similarities)

        if similarities[closest_idx] > 0.1:  # Threshold to ensure meaningful similarity
            closest_condition = medical_condition_encoder.classes_[closest_idx]
            print(f"Closest known condition: {closest_condition}")
            encoded_condition = medical_condition_encoder.transform([closest_condition])
            predicted_medication = model.predict([[encoded_condition[0]]])
        else:
            print("No close match found. Assigning default medication.")
            predicted_medication = [medication_encoder.transform(['Default Medication'])[0]]  # Replace with actual default

    # Decode the predicted medication
    decoded_medication = medication_encoder.inverse_transform(predicted_medication)
    return decoded_medication[0]

# Input for prediction
input_condition = input("Enter a medical condition: ")
predicted_medication = predict_medication(input_condition)

print(f"Predicted Medication for '{input_condition}': {predicted_medication}")
```

```
'hypertension' is not recognized. Searching for the closest match...
Closest known condition: Hypertension
Predicted Medication for 'hypertension': Paracetamol
```

```
C:\Users\R.MUNIRANJANI\anaconda3\Lib\site-packages\sklearn\base.py:493: UserWarning: X does not have valid feature names, but RandomForestClassifier was fitted with feature names
  warnings.warn(
```

In [ ]: