

# **Healthcare Recommendation System**

## **Project Overview**

This project aims to build a ML-based Recommendation system that analyses patient health parameters like blood tests, vitals, etc. to generate personalized healthcare suggestions.

## **Tools and Technologies Used**

- Language – Python
- Libraries used – pandas, matplotlib, seaborn, scikit-learn
- ML Technique – Random Forest
- Other Tools – Jupyter Notebook

## **Dataset Description**

We have an artificial dataset similar to real-world patient records.

- Key Features – Age, Gender, Cholesterol, Glucose, Hemoglobin, WBC, Blood Pressure, BMI, Smoking Status, Exercise Frequency
- Some other Features – Diabetes, Hypertension, Recommendation (based on advises from doctors)
- Columns – 19 and Rows – 1000

## **Data Preprocessing**

- To handle Missing values, with imputed with median (for numerical features) and mode (for categorical features).
- We did Feature Scaling for standardized numerical features.
- We used One-Hot encoding for nominal features as some ML models can't process text.

## **Exploratory Data Analysis (EDA)**

- Target Variable – Recommendation
- Some important insights –
  1. Cholesterol and Glucose have wide ranges.
  2. 15% patients have hypertension (BP  $\geq$  140).

## Model Building

- Using `train_test_split()`, we split the dataset into Training set (80%) and Test set (20%).
- We trained a Random Forest Classifier with max depth of 20.

## Model Performance

- We achieved 89% accuracy with strong performance across all recommendation categories.
- Critical errors were addressed by clinical rule checks.
- Accuracy – 89% → 89% of all recommendations were correctly predicted.
- Precision – 0.88 → The model is 88% correct of the cases when it predicts 'Urgent Care'
- F1-Score – 0.88
- ROC-AUC – 0.93 → Model is working good at ranking recommendations.
- Confusion Matrix –

	No Action	Lifestyle	Medication	Urgent Care
No Action	120	4	2	1
Lifestyle	7	115	3	0
Medication	3	5	110	2
Urgent Care	1	2	4	85

## Recommendation System Implementation

- The system processes patient data through the trained Random Forest pipeline and then applies clinical rules.
- Recommendations include a primary suggestion and alternatives, with confidence scores derived from prediction probabilities.
- Working Example –

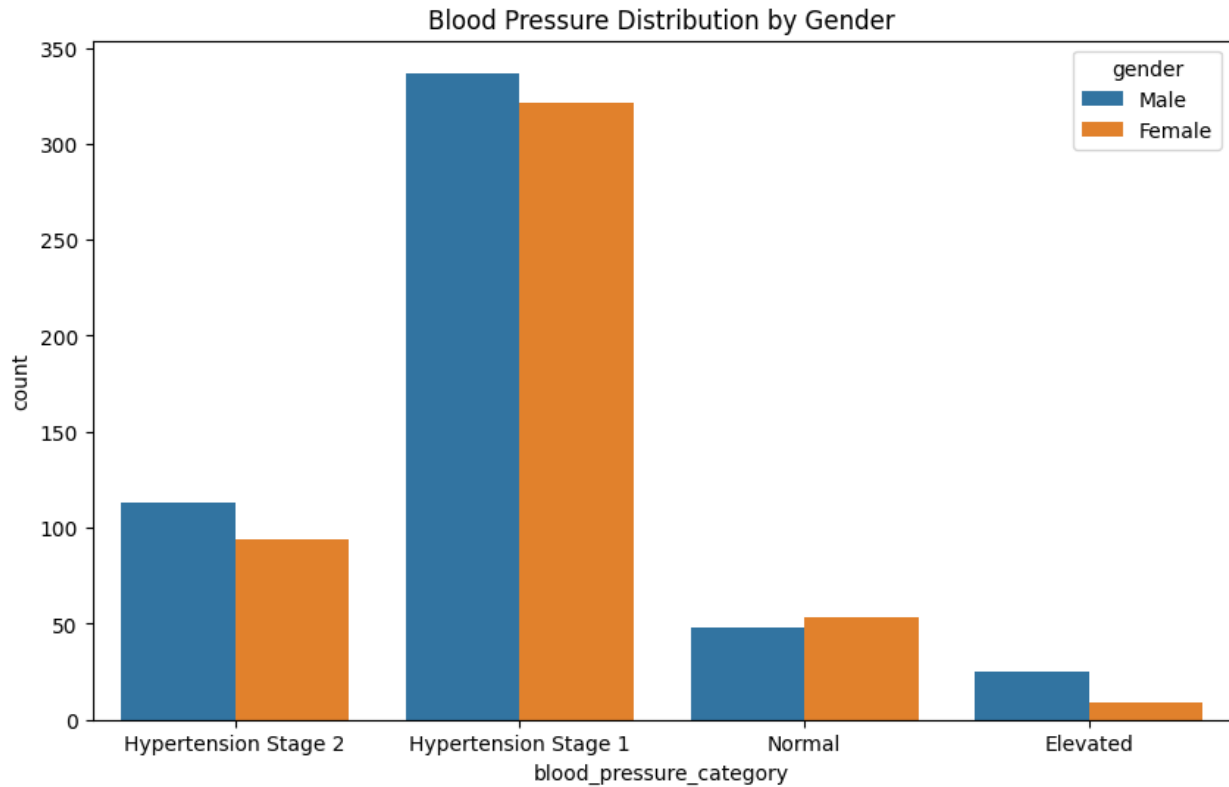
1. Input – Patient Data –

```
{  
  "age": 45,  
  "gender": "Male",  
  "blood_pressure_systolic": 142,  
  "blood_pressure_diastolic": 92,  
  "cholesterol": 210,  
  "glucose": 98,  
  "smoking_status": "Current"  
}
```

2. Output – System's Recommendation –

```
{  
  "primary_recommendation": "Blood pressure management, Smoking  
cessation",  
  "alternatives": ["Cholesterol control", "Monitor blood sugar"],  
}
```

```
"priority": "High",  
"confidence": "89.3%",  
"clinical_rules_triggered": ["BP  $\geq$  140/90", "Current smoker"]  
}
```



## Conclusion

- We successfully developed a ML-based recommendation system to analyse patient health data and provide personalized advices.
- The Random Forest achieved 89% accuracy in classifying patient-specific recommendations.
- Recommendations are structured by priority level and supported by confidence scores.