# **Hospital Admissions Prediction Using Random Forest**

## **Project Documentation**

### **Table of Contents**

1. **Introduction**
2. **Project Overview**
3. **Installation and Setup**
4. **Dataset Description**
5. **Preprocessing Steps**
6. **Model Training**
7. **Results and Evaluation**
8. **Prediction Example**
9. **Conclusion**
10. **License**

## **1. Introduction**

This document provides a detailed explanation of the **Hospital Admissions Prediction** project. The goal of this project is to predict the number of hospital admissions based on features such as arrival month, age, and gender. The project uses a **Random Forest Regressor** model, and the data is preprocessed using techniques like One-Hot Encoding and Standard Scaling.

## **2. Project Overview**

The project aims to predict the number of hospital admissions using the following features:

* **Arrival Month**: The month of admission (e.g., January, February).
* **Age**: The age of the patient.
* **Gender**: The gender of the patient.

The project involves:

* Preprocessing the data (encoding categorical variables and scaling numerical features).
* Training a Random Forest Regressor model.
* Evaluating the model using Mean Squared Error (MSE).
* Making predictions for specific inputs.

## **3. Installation and Setup**

To run this project, you need the following Python libraries:

* Pandas
* NumPy
* Scikit-learn
* Matplotlib (optional, for visualization)

You can install the required libraries using the pip package manager.

## **4. Dataset Description**

The dataset should contain the following columns:

* **Arrival Month**: The month of admission (e.g., "January", "February").
* **Age**: The age of the patient.
* **Gender**: The gender of the patient (e.g., "Male", "Female").
* **Admissions**: The number of admissions (used for grouping and summing).

## **5. Preprocessing Steps**

The data is preprocessed as follows:

1. **Group Admissions by Month**: The dataset is grouped by the arrival month, and the total admissions per month are calculated.
2. **Encode Categorical Variables**: The arrival month and gender are encoded using One-Hot Encoding.
3. **Scale Numerical Features**: The age feature is scaled using Standard Scaling.

## **6. Model Training**

A **Random Forest Regressor** is used to train the model. The dataset is split into training and testing sets, with 80% of the data used for training and 20% for testing. The model is trained on the training set and evaluated on the testing set.

## **7. Results and Evaluation**

The model is evaluated using **Mean Squared Error (MSE)**, which measures the average squared difference between the predicted and actual values. A lower MSE indicates better model performance.

## **8. Prediction Example**

The trained model can be used to predict the number of admissions for a specific month, age, and gender. For example, you can predict the number of admissions for a 28-year-old male in January.

## **9. Conclusion**

This project demonstrates how to predict hospital admissions using a Random Forest Regressor. The model is trained on preprocessed data and evaluated using MSE. The project can be extended by incorporating additional features or experimenting with other machine learning algorithms.

## **10. License**

This project is licensed under the MIT License. See the LICENSE file for details.