**Scenario 3:**

**A hospital has a large dataset of patient records, including information on demographics,**

**medical history, diagnoses, treatments, and outcomes. The hospital wants to use this data**

**to develop a machine learning model that can predict the risk of readmission for patients**

**after they are discharged from the hospital. Develop a ML solution for the aforesaid**

**prediction with an example Dataset.**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Project Title: Predicting Hospital Readmission Risk**

**Introduction:**

This documentation provides an overview of the Scenario 3 project, which aims to predict the risk of hospital readmission for patients after they are discharged from a hospital. The project utilizes a dataset containing patient records with information.

**Project Overview:**

Hospitals are interested in predicting hospital readmissions for discharged patients as it reflects the quality of care and can have financial implications. The project seeks to develop a machine learning solution to predict the likelihood of patients being readmitted within 30 days, after more than 30 days, or not at all.

**Dataset:**

<https://universe.roboflow.com/zakir-alam/breast-cancer-dataset/dataset/7>

The project uses a dataset named "diabetic\_data.csv" which contains a variety of independent variables and a dependent variable "readmitted" with three categories: No readmission, readmission in less than 30 days, and readmission in more than 30 days.

**Code Description:**

The project's code is written in Python and is organized into several sections. Here is a summary of the key components:

1. **Data Import and Preprocessing:** The project starts by importing necessary libraries, loading the dataset, and performing initial data analysis.
2. **Feature Engineering:** The code conducts feature engineering to prepare the data for machine learning. This includes:
   * Handling missing values and irrelevant columns.
   * Creating a new feature "service\_utilization" to measure medical care utilization.
   * Creating a feature "numchange" representing the number of medication changes for each patient.
   * Replacing and categorizing values in various columns like "admission\_type\_id," "discharge\_disposition\_id," "admission\_source\_id," and others.
   * Converting categorical variables to binary numeric values.
3. **Data Visualization:** The code includes data visualization using matplotlib and seaborn, although the specific visualizations are not provided in the code snippet.
4. **Handling Diagnosis Codes:** The code consolidates diagnosis codes into nine disease categories and assigns numeric values to primary diagnosis codes.
5. **Data Reduction:** The project also to handle cases where patients have multiple encounters. The code deduplicates records, consolidates multiple encounters into a single representation, and calculates average duration and medication changes.
6. **Target Variable Encoding:** The code encodes the dependent variable "readmitted" into three categories: No readmission (0), readmission in less than 30 days (1), and readmission in more than 30 days (0).

**Model Architecture:**

* A Sequential model is utilized, which is a linear sequence of neural network layers.
* The first layer is a dense layer with 64 units and a ReLU activation function. It accepts an input shape that matches the number of features in the training dataset.
* Subsequently, a second dense layer with 32 units and a ReLU activation function is added.
* Finally, a dense layer with a single unit and a sigmoid activation function is used to produce binary classification results.

**Model Compilation:**

* The model is compiled using the Adam optimizer, a common choice for gradient-based optimization.
* The loss function employed is binary\_crossentropy, which is well-suited for binary classification tasks.
* Accuracy is selected as a metric to monitor during training.

**Model Training:**

* The model is trained on the training data, specified as X\_train and y\_train, for 30 epochs.
* A batch size of 32 is used, and the model's performance is monitored on the validation data, represented by X\_test and y\_test.

**Model Evaluation:**

* After training, the model is evaluated using the test data, computing both the loss and accuracy.
* The code prints the accuracy of the model on the test data. The accuracy is obtained as 88%

**Conclusion:**

The Machine Learning project aims to predict hospital readmission risk using a dataset of patient records. The provided code preprocesses the data, engineers features, and encodes the target variable, builds a model and finds the accuracy.