**Block Diagram for Patient Readmission Prediction**

**(For Specific Disease)**

**Step1: Patient Data Collection**

* Collect all the necessary data requires for prediction.
* We have Electronic Health records in the form of text files.
* Age, Diagnosis, Complaints on Admission, Medical History, Course in Hospital, Treatment Given, Condition at discharge, Medicine given. (Primary Findings)
* Some of the other but important entities like Temp, Bp, Pulse, Weight, CNS, RS, PA, CVS. (Secondary Findings)
* Primary Findings shows the most recent disease.
* Secondary Findings Suggest the overall health of patient whether the patient is recovering well or not or any side effects are troubling the patient during treatment.

**Step2: Data Pre-Processing**

* In this step, we need to clean the data, remove the duplicates and outlier.
* Now, Convert the data into more suitable format.
* In our data, most of the data is in medical naming convention. For ease to understand the data, we need to convert the data into general naming convention. By this we can easily identify the patient with same disease and symptoms.
* To identify the medical naming convention and making the data in structured format we can use Name Entity Recognition Tool or either we can train the model manually.
* For Manually training of data, we need quality of data.

**Step3: Feature Engineering**

* In this step, we have to consider only features which have high impact on readmission prediction.
* If the dataset is imbalanced, try to balance the dataset using the oversampling methods (SMOTE)
* We can use Statistical analysis or correlation analysis for finding out the features.

**Step4: Identify the features which are necessary for 30 days, 60 days and 90 days classification.**

**Features that can be used for 30 days readmission prediction and why?**

If we could convert the treatment given into these three main types –

* **Medical Oncologists**
* **Radiation Oncologists**
* **Surgical Oncologists**

**Why these three categories** – Because we can easily categorize the risk associated with these types. Because Radiation and Surgical Oncologists treatment are more critical and severe and these types of cancer treatment have severe adverse effect as well on patient health.

* **Medical History** – Using medical history, we can identify the type of disease and affected body part.
* **Complaint on Admission** – In most of the case, this is a patient follow-up treatment part. So, we can get central idea that after getting nth number of treatments what’s the patient health? Is it normal or need to worry.
* **Condition at Discharge** – We can consider this step as dynamic. Even after getting treatment patient might feel sick after a day.
* **Medicine at discharge** – Most of the readmission are not due to original diagnosed disease it’s because of other disease associated with it. If the patient whose treatment falls under surgical oncologist and have any respiratory disease along with it. So, it becomes difficult to recover from both the disease because of its severity. So, patient might get readmitted within 30 days.
* So, its very important to check the secondary findings of the patient. This finding indicated that patient is recovering all good.
* Mostly surgical treatment patient falls under 30 days readmission because due to surgery there might be a chance that patient will get any kind of infection.

**Features for 60 days patient readmission –**

For, 60 days prediction. The treatment type should fall under medical oncologist.

The side effect for medical oncologist is very common such as fatigue, hair loss, vomiting etc.

If the side effects stay more than 60 days. Then we can get idea that patient is suffering from additional disease or the current treatment is not suitable for the patient.

Age will always remain a constant factor whether we have to predict it for 30 ,60 or 90 days.

In 60 days, prediction, we have to check whether patient is taking regular follow-up treatment or not and how well he is responding to the treatment.

Medicine at Discharge – Even after regular follow- up, if the medicine power is not decreased, then patient might be readmitted. In this scenario, patient will suffer from one additional problem of high-power medicine side effects.

Phycological factors – (this entity is not present in our dataset) But physiological factors can play important role in readmission. This includes stress, family support and financial condition.

**Features for 90 days patient readmission –**

In 90 days, readmission, we have to classify the patient with long term recovery and side effects associated with me.

For 90 days prediction, if the patient falls under radiology oncologist treatment. In this treatment radiation are used to kill the cancer blood cells. If the patient has excess cancer blood cells, and it cannot be treated with the help of chemotherapy then doctor prescribes the radiological therapy. The main advantage is that it kills cancer blood cells in very short duration. But the disadvantage associated with this therapy is that because of radiological treatment, new cancer cells can be produced and another disadvantage is due to radiation it can affects the nearby organs as well.

Another reason can be the organ failure due to radiation therapy and skin disease may arise due to radiation.

Disease progression check-up and whether the patient has changed his daily life style or not.

**Factors that can contribute for patient readmission can be –**

* **Age**
* **Diagnosis**
* **Treatment**
* **Vital Signs**
* **Medicine prescribed**
* **Condition at Discharge**
* **Side effects**
* **Cancer stage**

**Step5: Model Training**

**What is LSTM, Topic Modelling and how it works? By combining is there any advantage or not?**

**LSTM -**

In Patient re-admission condition, we mostly have historical data for making the prediction. So, LSTM can be used for patient readmission because LSTM have three main gates.

1st Remember Gate

2nd Forgot Gate

3rd Output Gate

Let’s take an example, suppose we are predicting something and we have features like age, disease, days requires for recover.

So, In the Stage- 1, We are analyzing disease with days requires for recover, in this step, we have not included the age but the age column is also the factor in our prediction. So, LSTM have that capability to remember the age column while comparing it in later stage. As well as it can forgot if the column work is already done and output gate make sure that it is taking that column into consideration on that particular stage to return the output.

**Topic Modelling** –

Topic Modelling is statistical technique used to identify some words in a corpus based on the data which is falling under one category.

**Comparison based on Performance -**

**LSTM -**

May perform well on tasks involving short to medium-length sequences but might struggle with longer documents where maintaining context is challenging.

**Topic LSTM -**

Generally, performs better on tasks involving long documents or complex themes, as it leverages both local sequential information and global thematic context.

**Step6: Continuous Improvement.**

Setup automated alerts if the performance metrics falls under certain threshold.

This can be done using AWS or by writing manual script.

**Step 7: Performance Evaluation Metrics**

We will consider F1-Score and AUC-ROC curve.

**Step 8: Deployment of Model**

We will deploy the model using Stream lit for checking purpose.