**PROJECT PROPOSAL**

**LENGTH-OF-STAY PREDICTION IN HOSPITALS AT TIME OF ADMISSION**

**INTRODUCTION:**

Predictive analytics is growing day-by-day as an important tool in the healthcare industry because of its abilities to predict the individual outcomes of patients using machine learning (ML) techniques on vast volumes of data. Hospitals & Clinics can use Predictive analytics to better manage their time & resources and ultimately provide better services to their patients. For this project, we have chosen to focus on hospital length-of-stay (LOS), which is a very crucial aspect of determining the wait times and beds availability in the hospitals, especially in ICU (Intensive Care Unit). LOS is defined as the period of time, expressed in days, between hospital admission time and discharge time.

Through the project we initially plan to study the impact of factors, such as, patient age, gender, race, ethnicity, marital status, diagnosis category (heart disease, delivery, injury/poisoning, etc.) on a patient’s LOS. We want to observe the minimum, maximum and median LOS of patients across these factors and also determine the overall median LOS. Finally, we will be utilizing these factors to develop a predictive ML model (using a regressor) to predict the LOS of future patients, that the hospitals can utilize to optimize their medical resources and better manage bed availability in the emergency wards.

**MOTIVATION:**

According to a recent study from the Fraser Institute (*Comparing Performance of Universal Health Care Countries, 2019*), Canada has a relatively short supply of doctors and hospital beds—and the longest wait times, despite spending more on health care than most other developed countries with universal coverage. This problem is growing bigger, as Canada’s senior population grows in number with each passing day, seeking more medical care. As suggested by *Canadian Institute of Health Information* (CIHI), Over the next 20 years, Canada’s seniors’ population — those age 65 and older — is expected to grow by 68%.

Even from our personal experiences, we have witnessed our own friends and family members facing long wait times in the direst of situations, especially in emergency wards in Canada. This was one of our biggest concern and motivation to choose this topic for our research.

We want to tackle the problem of long wait time in hospitals at time of admission. One of the ways to deal with this issue was to estimate the Length-of-stay (LOS) of patients during the time of admission. This is due to the following reasons:

* Firstly, an accurate prediction of a patient's remaining LOS would be a useful guide for expectation management and would improve communication between doctors and their patients.
* Secondly, one of the primary goals of hospital administration is to allocate resources as efficiently as possible. They seek to improve the quality of healthcare services for patients (for example, on-time diagnosis and shorter wait times) while reducing costs.
* Lastly, Precise LOS prediction of patients is critical for efficient use of ICU resources such as staff, ventilators, and other medical devices.

**EVALUATION:**

The goal of this project is to predict the Length-of-stay (LOS) of patients at the time of admission. Hence, the measure of success of our project will be the accuracy of ML Model to predict the LOS of patients based on their demographic and health diagnosis features.

Furthermore, the project's success is also dependent on the methods and ways we can prescribe to hospital management to manage their resources better and minimize patients’ wait times, once the LOS of patients is predicted at the time of admission. This requires further research into the topic of hospital management, which would be done in the subsequent weeks.

**RESOURCES:**

For this project, we will be using the MIMIC-III Clinical dataset, which is a large, freely-available database comprising deidentified health-related data associated with over 40,000 patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012. This is a relational database consisting of 26 tables, that strongly simulates real life databases that are used in modern day hospitals.

Following methods and tools will be used to execute this project:

* Data Access for MIMIC-III (through completion of required trainings and becoming a credentialed PhysioNet user)
* Data Cleaning (MS Excel, SQL, Python)
* Exploratory Data Analysis (Python)
* Data Visualization (Python - Matplotlib & Seaborn)
* Build, Train and Test ML Models (Python)
* Model Refinement for better accuracy (Python)
* Prescribe a list of ways to optimize hospital resources and minimize wait times based on the LOS prediction.

**REFERENCES:**

* MIMIC-III Clinical Database (<https://physionet.org/content/mimiciii/1.4/>)
* Charles, D., King, J., Patel, V. & Furukawa, M. Adoption of Electronic Health record Systems among U.S. Non-federal Acute Care Hospitals. ONC Data Brief No. 9, 1–9 (2013).
* Collins, F. S. & Tabak, L. A. NIH plans to enhance reproducibility. Nature 505, 612–613 (2014).
* [Waiting Your Turn: Wait Times for Health Care in Canada, 2021](https://www.fraserinstitute.org/studies/waiting-your-turn-wait-times-for-health-care-in-canada-2021) (The Fraser Institute)
* [Comparing Performance of Universal Health Care Countries, 2019](https://www.fraserinstitute.org/studies/comparing-performance-of-universal-health-care-countries-2019?utm_source=Media-Releases&utm_campaign=Comparing-Performance-Universal-Health-Care-2019&utm_medium=Media&utm_content=Learn_More&utm_term=700) (The Fraser Institute)
* [Canada’s seniors population outlook](https://www.cihi.ca/en/infographic-canadas-seniors-population-outlook-uncharted-territory) (Canadian Institute of Health Information)