Decision Support Tools for Health Emergency Care

PROBLEM: According to a John Hopkins study in 2016, Medical Error Is The Third Leading Cause Of Death In The U.S. , reasons go from poorly coordinated care to variation in doctor practice patterns. These issues are magnified in Emergencies departments where the diagnosis process and medical decisions need to be taken fast in overcrowded locations. This capstone project will focus on the first step in Emergency Department workflows, and maybe the second step as well depending on assigned resources and results from initial baselines. First two steps:

1. **Assigning an Emergency Severity Index (ESI).**
   When a patient arrives to a hospital emergency department (ED), clinical data is collected and interpreted to classify patients into the Emergency Severity Index (ESI), which goes from 1 (most urgent) to 5 (least urgent). The ESI sets the trajectory for further ED care, like location, queue position and timing.
2. **Deciding which Diagnostic Tests to perform.**
   The Nurse or Doctor orders diagnostic tests, results can change the planned treatment.

If the ESI is incorrect by underestimating urgency or the incorrect diagnostic test is ordered, the consequences would be a low quality health outcome, which could even mean death or other irreversible consequences

PROPOSAL: AI models, using latest techniques for better performance, to classify cases into the Emergency Severity Index (ESI) and a list of diagnostic tests. Build and API that let ED EHR or patient-centered applications make use of AI models for prediction.

POTENTIAL MARKET/CUSTOMERS:

- **Hospitals**' Electronic Health Record (EHR) applications and its Emergency Department Management modules. According to a **Gartner report in 2017** on how an Emergency Room Leverages the Real-Time Health System to Improve Efficiency: "Health Delivery Organizations transformation to a real-time health system (RTHS) exposes many opportunities to improve the quality of care and patient experience and to control costs", the report speaks how a RTHS can support the different phases in the workflow with analytic tools, it includes RTHS pre-assigning incoming patients to the appropriate level assessment room, based on availability and acuity level.

- **Patient Centered Healthcare** (Market in development). Patient-centered care has taken center stage in discussions of healthcare quality. The Patient-Centered Outcomes Research Institute (PCORI) was established in 2010 to fund research (~ $150 millions a year) that can help patients and those who care for them make better-informed decisions about the healthcare choices they face. PCORI and AHRQ had funded a 2016 conference regarding Patient centered Shared Decision Making in the Emergency Department, to create a Research Agenda which includes the development of applications for Patients. The market for mobile applications that let patients and their caregivers to access the patient's ED Information in real time is on the horizon, with support from the government and research organizations. This type of mobile applications can have a module that uses the machine learning models we create to confirm in real time the Emergency Severity Index assigned as well as diagnostic tests, information that patients (or family) can use for their own decisions (e.g: talk with the doctor, call a second opinion, etc). Handling disagreements between the doctor and patients is part of the whole Patient-centered movement being developed in the Health Care system. According to a **Gartner Report in 2017** regarding "Consumer Engagement With Healthcare and Wellness", the Market Penetration of Patient Decision Aids is less than 1% of target audience and its maturity is Embryonic.

DATASETS:

- NHAMCS by the CDC, ~26K by year, data since 2008. Open dataset. *https://www.cdc.gov/nchs/ahcd/datasets_documentation_related.htm*

Quote from a previous study using this dataset: ***"****NHAMCS is an annually collected, nationally representative probability sample survey of ED visits conducted by the Centers for Disease Control and Prevention's (CDC) National Center for Health Statistics. This study utilizes a pre-existing, publicly available, de-identified database" (de-identified data doesn't need IRB approval)*

- MIMIC III , ~60K patients and its EHR, de-identified information by MIT from a Hospital in Boston. It includes Emergency Admissions. Open dataset but requires taking the CITI "Data or Specimens Only Research" online course (takes about 6 hours and it has several short test quizzes that can be retaken as many times as necessary) https://physionet.org/physiobank/database/mimic3cdb

POSSIBLE DATA SCIENCE TECHNIQUES:

- Data Extraction, Pre-Processing, Aggregation
- Exploratory Data Analysis
- Deep Learning algorithms (How to explain results from deep learning models?)
- Natural Language Processing models
- Reinforcement Learning for Tuning models or even the model itself.
- Visualizations in D3

RELATED WORK:

(1) "Machine-Learning-Based Electronic Triage More Accurately Differentiates Patients With Respect to Clinical Outcomes Compared With the Emergency Severity Index". John Hopkins University, 2018, Science Direct. It is using a Random Forest Algorithm. Dataset: **~**173K records**,** Seems the dataset is not public yet

(2) "An Electronic Emergency Triage System to Improve Patient Distribution by Critical Outcomes". Department of Emergency Medicine, Johns Hopkins University, 2016, PubMed. It uses Logistic Regression for classifying patients. Dataset: NHAMCS by the CDC

(3) "Mortality prediction in the ICU: can we do better? Results from the Super ICU Learner Algorithm (SICULA) project, a population-based study". Division of Biostatistics, School of Public Health, University of Califonia, Berkeley, 2015, Pubmed, It is using an assemble of algorithms: logistic regression model. Stepwise regression, neural networks, etc. Dataset: MIMIC II dataset

RELATED PRODUCTS:

- **Hospitals Market**:
    a. Manchester Triage System (MTS), standard used in Europe. The system consists of around 50 flowcharts with standard definitions designed to categorize patients arriving to an emergency room based on their level of urgency.
    b. ClearTriage, supports Nurses receiving phone calls from patients. It is a decision tree with a set of questions for gathering information from the patient and returning a suggested treatment (e.g: actions to take at home). It is not used on emergency departments though.
    c. According to a paper in JMIR MEDICAL INFORMATICS: Lumiata had implemented A Web-Based Tool for Patient Triage in Emergency Department Settings, called cGPS. Its algorithm
    d. ER Mobile with TriageXpert, a DSHI proprietary traige expert system
- **Patient Centered Market**: No products for now.
    a. According to the 2016 Academic Emergency Medicine Consensus Conference's Keynote: Tools and Measurement for Shared Decision Making: "There are few high-quality studies on shared decision making in the ED. Future studies should adopt a user centered design approach when developing tools to facilitate shared decision making and use validated measures to determine their effect on decision quality."

COMPLEXITY LEVEL, EDA and MILESTONES:

Complexity and EDA Considerations:

- The CDC data is ready to be downloaded and it has already been used to predict the emergency score using a Logistic Regression model, the paper mentioned above (#2) lists the variables used for features and outcomes. Building a baseline to replicate this paper model should be straight forward.
- I had used the MIMIC III database before for classifying medical discharge notes into icd-9 codes (diagnoses codes), using deep learning models (w266 class: A+). The MIMIC III database has complete EHRs (not only discharge notes) and the related paper mentioned above (#3) explains which variables they used to predict mortality. We can use this database as an additional data source for the Emergency Index model, looking in the EHRs for the variables that were used in the model described by paper #2..

Tentative Milestones:

Week 4	Download CDC data (one year to begin with) and MIMIC data, place files in common repository, EDA
Week 4 and 5	Create baseline model with CDC data based on paper #2
Week 4 and 5	EDA on MIMIC III data to see if we have all variables to add it to the baseline model created by paper #2. If it is not the case, we just use the CDC data
Week 5	First set of visualizations of EDA in the CDC and MIMIC data
Week 6	Presentation #1: Propose improved models and Project Planning