



Deep dive in data

The 6th DeepDive Challenge

Early Prediction of Patient Hospitalization

Mehr 1402



Innovation Center
for Data Science

Emergency departments (EDs) can get overcrowded and get near to a breaking point. Alongside ever-increasing demand, one of the leading causes of ED overcrowding is the boarding of hospitalized patients in the ED as they await bed placement.

Early identification of emergency department (ED) patients who need hospitalization can help to improve the overcrowding problem of Eds and is essential for quality of care and patient safety.

The goal of this challenge is to build effective and reliable machine learning models for early prediction of hospitalization in emergency departments. The model will use a few simple demographic, and medical features that are present when the patient is being treated in the ED.

We have provided you with a real-world but anonymized dataset that includes ground-truth indicating whether that each patient has been hospitalized after receiving treatment in the emergency department or not. You will find a data dictionary describing the dataset features at the end of this document.

For evaluating the models, you must pay attention to the imbalanced nature of this problem. Despite being a simple binary classification problem, the class distribution is very imbalanced, since only a small percentage of the patients undergoing treatment in the EDs are hospitalized.

Field Name	Description
triage	Triage is the process of sorting people based on their need for immediate medical treatment as compared to their chance of benefiting from such care. Triage is done in emergency rooms, disasters, and wars, when limited medical resources must be allocated to maximize the number of survivors. This field is an integer variable, and represents the triage value associated to the patient. <u>Smaller values indicate more severe and higher priority.</u>
accident	This is a binary variable indicating if the patient has come to the emergency because of an accident.
successful_cpr	Count of successful CPR's performed on the patient.
failed_cpr	Count of unsuccessful CPR's performed on the patient.
origin	Indicating the origin of the patient. Has 3 values for 1. Coming from the same city 2. Coming from another city 3. International tourist.
reason_code	This field holds codes from a categorization of medical conditions that represents the main complaint of the patient.
hospitalized	Ground-truth. Binary variable indicating whether the patient has been hospitalized or not.
gender	Patient gender.
age	Patient age.
triage_hour	The hour in which triage has been performed on the patient.