

FistToFive

Final Project Proposal

Nov, 2018

Georgia Tech DataBootCamp

Team Members

Jeff Box, John Giglio, Bill Magill, John Commander, Luke Williams

Project Topic

Application of machine learning to healthcare data to draw insights and make shrewd predictions.

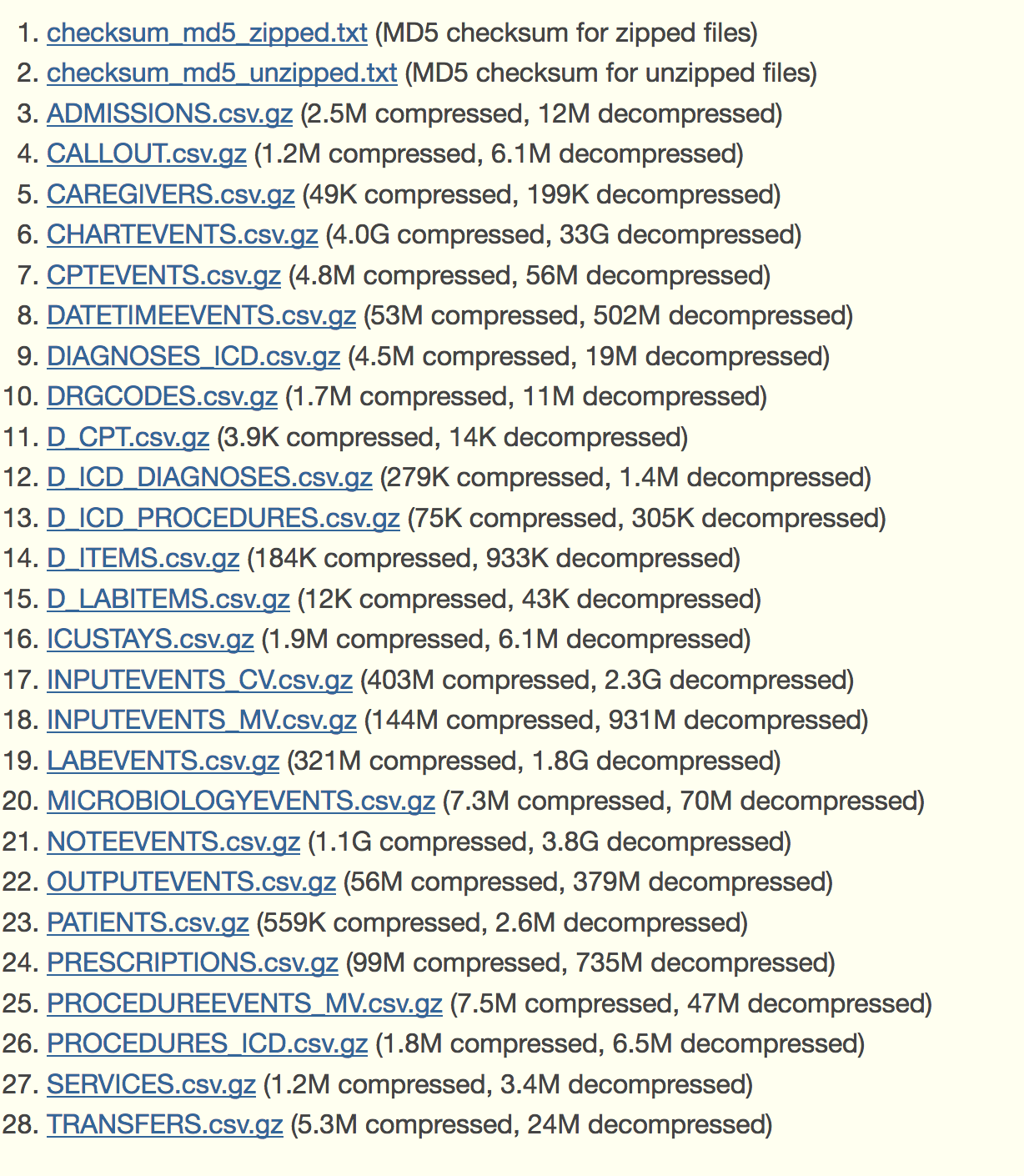
Dataset

Obtained from: <https://mimic.physionet.org/> through granted permission

The MIMIC-III (**M**edical **I**nformation **M**art for **I**ntensive **C**are) Clinical Database contains comprehensive clinical data from 40,000+ Intensive Care Unit (ICU) patients. Among the types of data included are:

* General - Patient demographics, hospital admissions & discharge dates, room tracking, death dates (in or out of the hospital), ICD-9 codes, unique code for health care provider and type (RN, MD, RT, etc). All dates are surrogate dates because of privacy issues, but time intervals (even those between multiple admissions of the same patient) are preserved.
* Physiological - Hourly vital sign metrics, SAPS, SOFA, ventilator settings, etc.
* Medications - IV meds, provider order entry data, etc.
* Lab Tests - Chemistry, hematology, ABGs, imaging, etc.
* Fluid Balance - Intake (solutions, blood, etc) and output (urine, estimated blood loss, etc).
* Notes & Reports - Discharge summary, nursing progress notes, etc; cardiac catheterization, ECG, radiology, and echo reports.

Contains 26 CSV files:



Starter code / scripts

<https://github.com/MIT-LCP/mimic-code>

Project Scope:

We aim to create an analysis of some of the data above by employing machine learning algorithms using open source libraries. Most likely the algorithms will utilize supervised learning with the goal of making accurate predictions or classifications, but we may also employ unsupervised models to extract features, detect underlying structure/patterns in the data, or to visualize it.

Examples of this are as follows:

1. Estimate (predict) day of mortality based on input of array of day 1 ICU admission data
2. Correctly classify patients as “critically ill” vs. “not critically ill” via input of day 1 ICU admission data
3. Predict sepsis (infection) based on an input array and compare this prediction with a validated sepsis scoring system such as SOFA