

Healthcare Analytics Project

Vineet Kumar

Rohit Bajpai

April 19, 2020

Abstract

We are working on three projects

1. Estimating Load of TB patients in India
2. Sentiment Mining from Nursing notes for Mortality Prediction
3. Age Stratified Modeling for Micro level Hospital Bed Forecast

Contents

1	Introduction
---	--------------

2

Chapter 1

Introduction

Data were extracted from the MIMIC databases using **Google Big Query** with the help of dialect of Structured Query Language (SQL)

Inclusion Criterion

1. Notes identified by physicians as errors (indicated by the `iserror` column in the MIMIC-III nursing notes table) were excluded.
2. We considered only patients whose age $\in [15, 85]$.

```
select * from mimic.ICU_Admitted where ICUSTAY_AGE_GROUP = "adult"
and hospital_expire_flag="N" and ICUSTAY_EXPIRE_FLAG="N"
```

Total Nursing Notes = 191836

```
select a.HADM_ID,p.Text from mimic.ICU_Admitted as a
inner join 'physionet-data.mimiciii_notes.noteevents' as p on a.HADM_ID = p.HADM_ID
```

134615 Nursing Progress Notes extracted unique entries : 2121379

```
select *
from 'physionet-data.mimiciii_notes.noteevents'
where hadm_id in (select hadm_id from mimic.tab3) and Description in ('Nursing Progress Note')
order by subject_id
```

----- SAPSII Scores

```
select *
from 'physionet-data.mimiciii_derived.sapsii'
where icustay_id in (select icustay_id from mimic.tab4)
order by subject_id
```

Only 134615 notes are considered

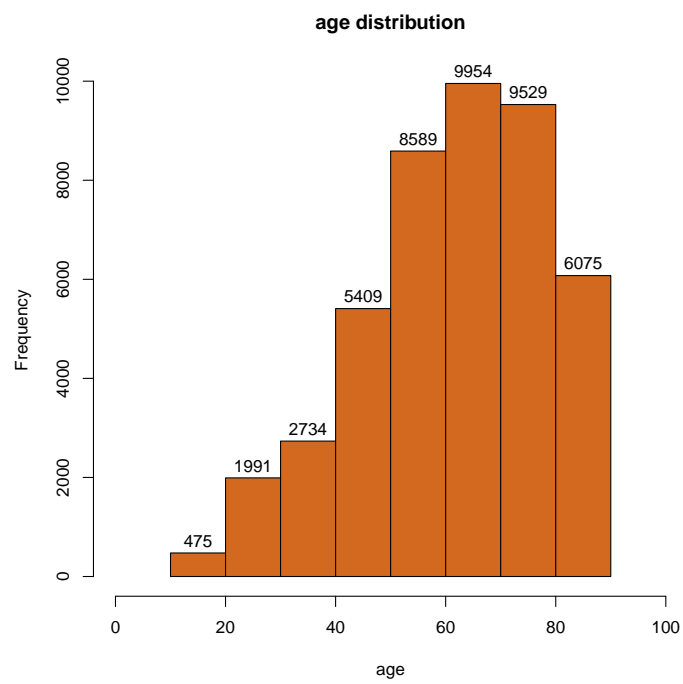


Figure 1.1: Age distribution of patients selected