CSE 6250 - Group 69

Mortality Prediction in Intensive Care Unit

Hyeon Woo Shim
Ishan Kaul
Nakul Patel

Understanding the problems

01 The intensive care unit is a challenging environment due to rapidly changing state of patients

02 Effective mortality prediction in such conditions can help us make informed, quicker decisions and save more lives in such critical states

Project objective

- Leverage big-data technologies for effective mortality prediction in ICU
- Implement the work done in the paper “Unfolding Physiological State: Mortality Modelling in Intensive Care Units” [1]
- Generate a better understanding of the severity of a patient's condition by using the clinical notes of doctors in order to create better context

[1] Ghassemi M, Naumann T, Doshi-Velez F, Brimmer N, Joshi R, Rumshisky A, Szolovits P. Unfolding physiological state. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 14. 2014

Pipeline

Data Preprocessing

Pull data from MIMIC database using pSQL queries

Feature Engineering

Create retrospective and time-varying baseline, derived and topic features

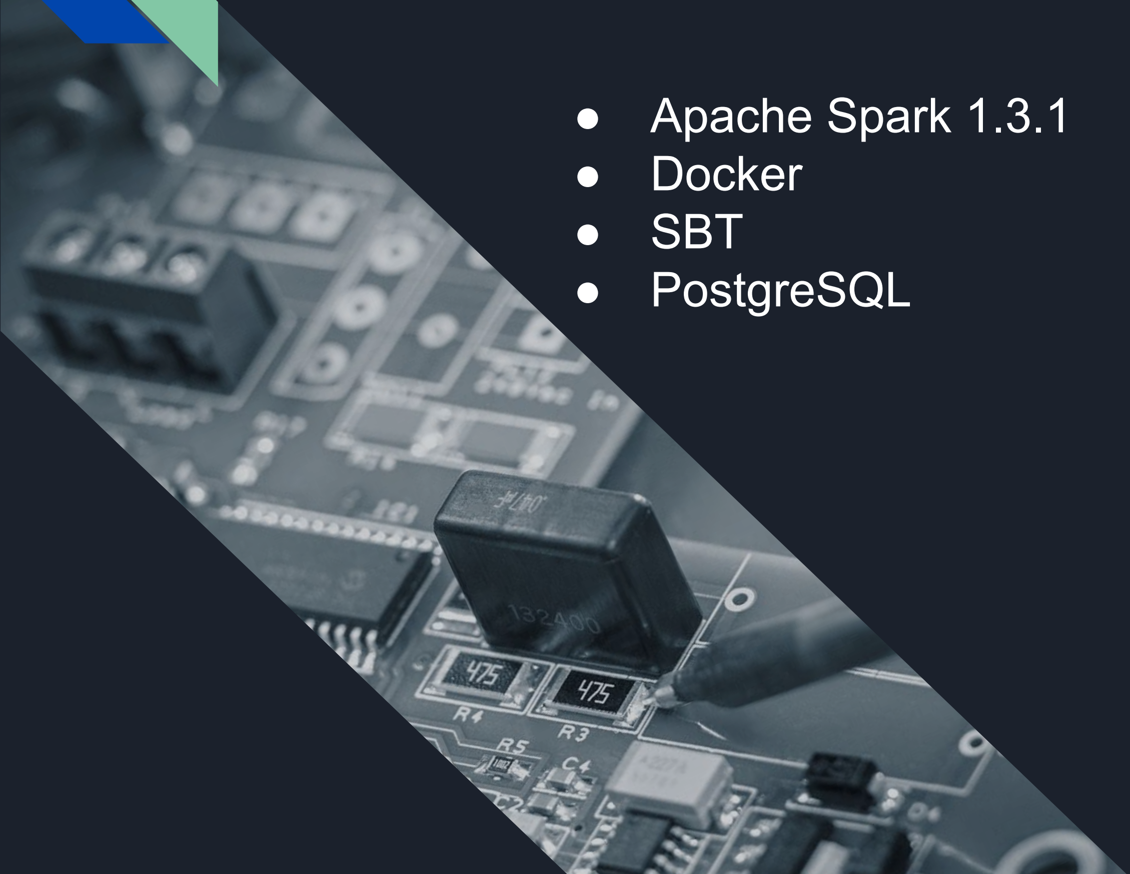
Prediction

Run **SVM** with parameter tuning to predict in-hospital, within 30 days and within 1 year mortality

Results

Aggregate best result for each model and compare with results from paper.

Software Stack & Dataset



- Apache Spark 1.3.1
- Docker
- SBT
- PostgreSQL

- MIMIC III
  - 46,520 patients
    - 26,121 male
    - 20,399 female
  - 2,078,705 clinical notes

Feature Engineering Models

We used a combination of different features for our prediction models

- **Admission Baseline Model** : benchmark model which are computed at the time of admission to ICU i.e. Age, Sex, SAPS score

- **Retrospective Derived Features Model** : derive additional features (30 EH comorbidities) from the admission data

- **Retrospective Notes Model** : derive contextual features with the help of clinical notes using LDA topic modelling

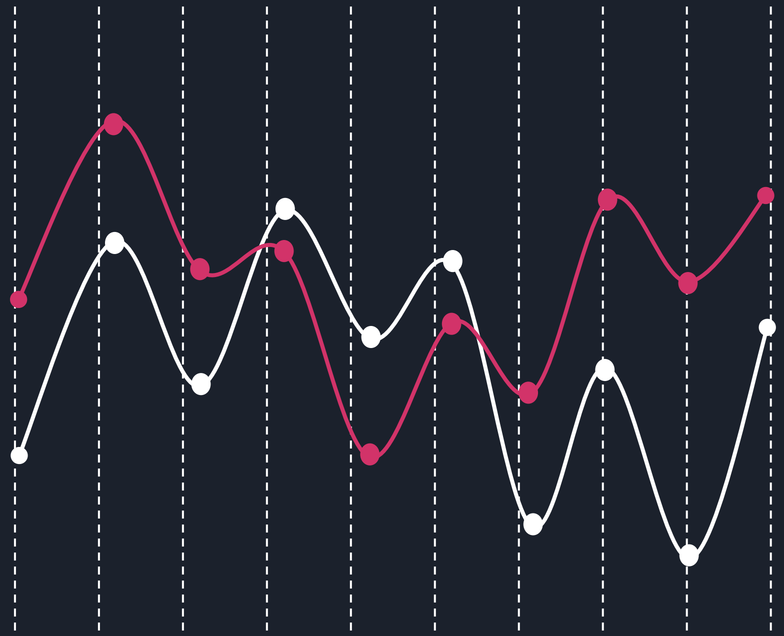
- **Time-varying model** : implemented real time dynamic features as a function of time by collecting all the notes available till a given time point

Predictive Models

Identified three types of prediction for our pipeline:

- Mortality in ICU
- Mortality within 30 days post-discharge
- Mortality within 1 year post-discharge

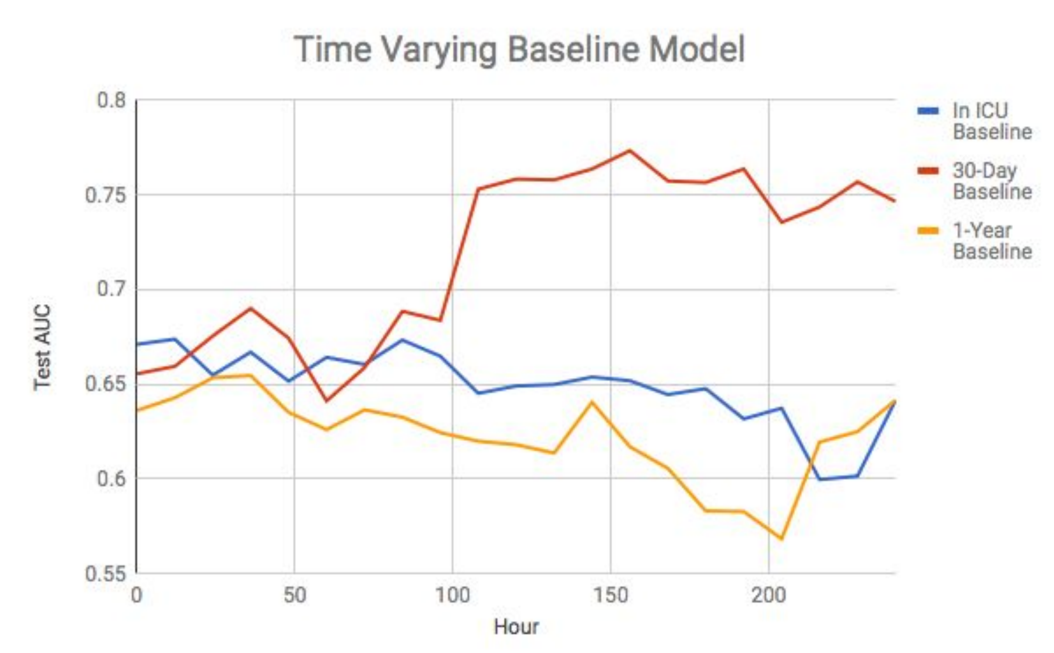
Randomly split data into train and test (70% - 30%) and ran SVM for each type of the prediction



Results

Model	In ICU	30 Day	1 Year
Retrospective Derived Features Model	0.6193971804	0.6233460113	0.6369264503
Retrospective Topic Model	0.6754885294	0.5726433431	0.5951190826

Results



Num Test Patients	Positive Test Patients
10365	1159
9883	982
8118	887
6745	808
5299	736
4518	673
3661	627
3197	593
2724	560
2457	521
2147	487
1943	430
1736	408
1602	379
1455	357
1357	329
1241	308
1146	277
1076	263
1004	250
940	241

Takeaways

- Convenience of big data tools
- Challenges of big data tools
- Parameter tuning
- Response of potential users

Thank you!