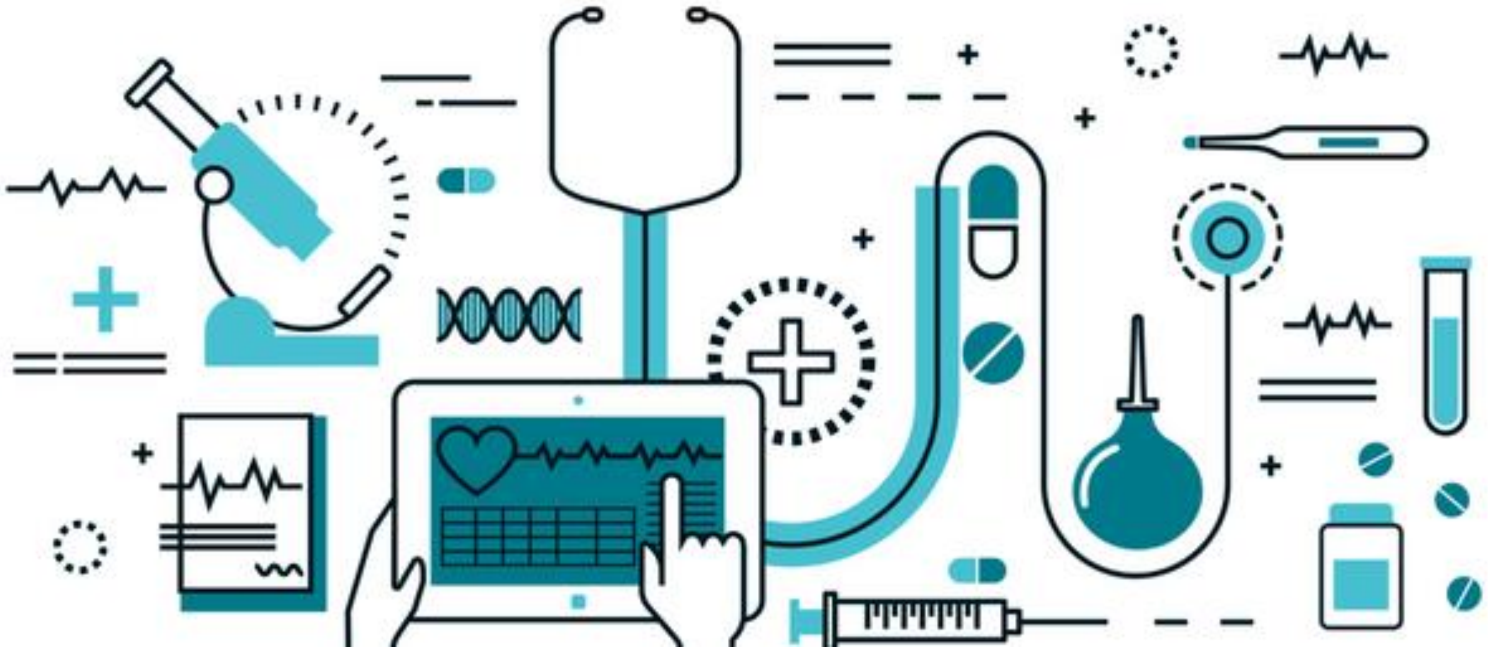


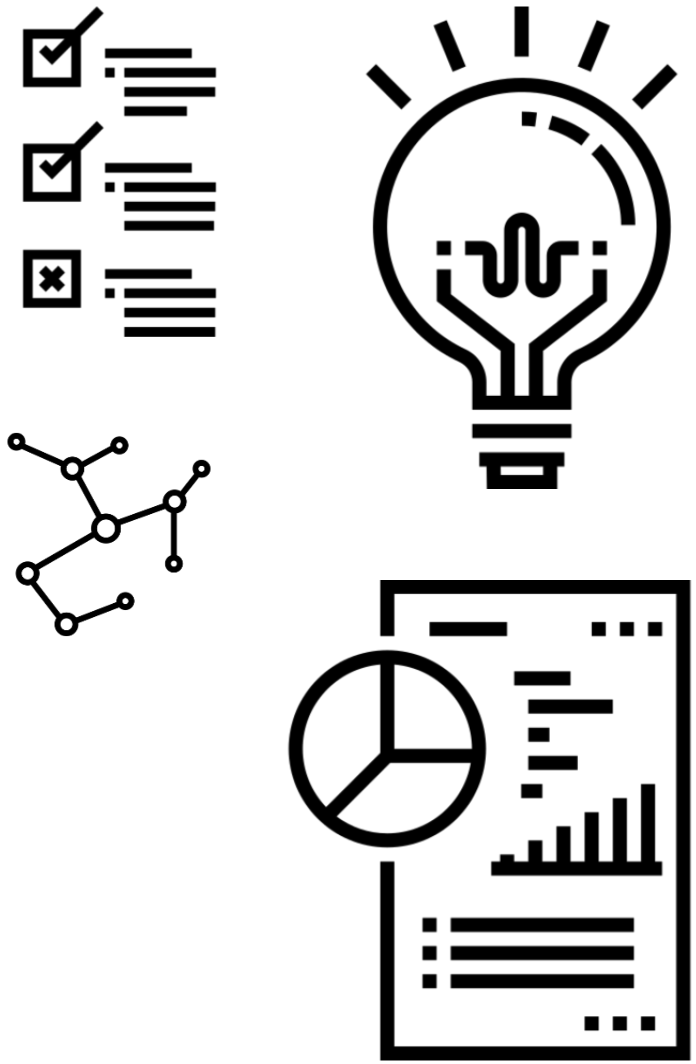
Predicting Clinical Outcomes

Faisal Maqbool
Supervisor: Dr. Saeed Ul Hassan

Information Technology University of the Punjab
Lahore, Pakistan

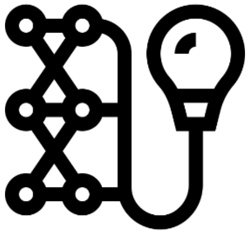
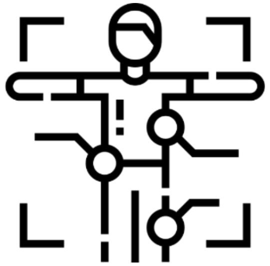
Outline

- Introduction
- Motivation
- Challenges
- Problem Statement
- Literature Review
- Dataset
- Proposed Technique



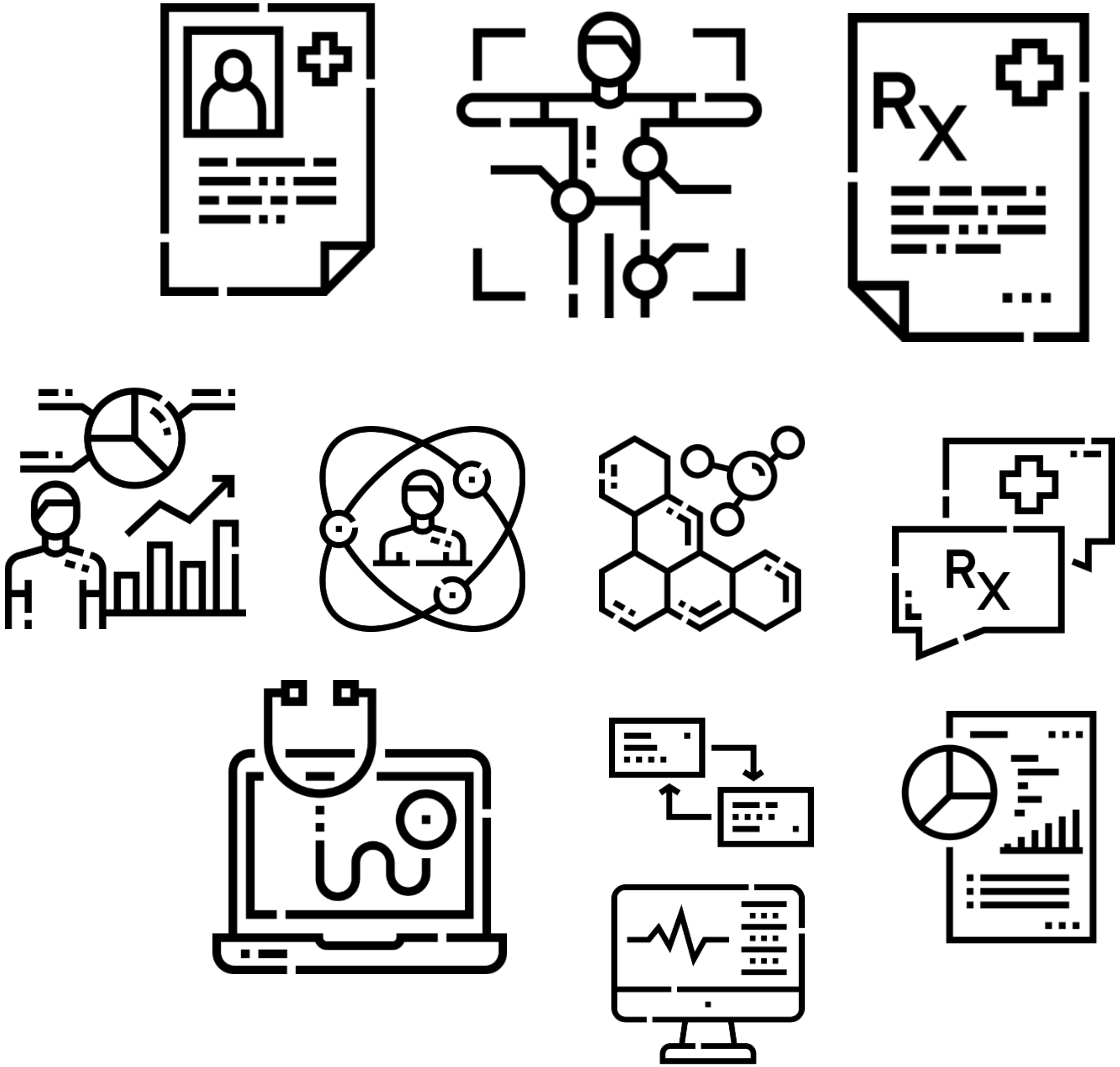
Introduction

- Deep learning models have revolutionized many fields including computer vision, natural language processing, speech recognition, and is being increasingly used in clinical healthcare applications
- Early hospital mortality, length of stay, obesity and ICD-9 code group prediction is critical as intensivists strive to make efficient decisions about ill patients admitted in hospitals

Motivation

- Extract sensitive information from admission and discharge summaries
- Extract symptoms
- Prescribe medicine & propose health matrices
- Aid doctors in giving accurate information
- Predict LOS, mortality, cause and other sensitive information using the clinical text



Literature Survey

Paper Title	Author(s)	Year	Journal	Data
Comparing deep learning and concept	Sebastian	2018	PLOS One	MIMIC (URL: https://physionet.org/works/
Clinical Text Classifcation with Rule-based	Liang Yao,	2018	IEEE International	https://www.i2b2.org/NLP/DataSets/
Bidirectional LSTM-CRF for Clinical Concept	Raghavendra	2016	Clinical Natural	https://www.i2b2.org/NLP/DataSets/
Identifying adverse drug event information in	Aron Henriksson	2015	Journal of Biomedical	Data used for this study was Stockholm
Predicting healthcare trajectories from	TrangPham	2017	Journal of Biomedical	Data for both cohorts were collected for 1
Recurrent neural networks for classifying	Yuan Luo	2017	Journal of Biomedical	https://www.i2b2.org/NLP/DataSets/
Multi-task learning for interpretable cause-of-d	Mireille Gomes,	2018	Proceedings of the	Proprietary data of Million Death Study
Predicting protein functions by applying	Kamal Taha,	2019	BMC Bioinformatics	Data for this study is Gene Ontology
Structured prediction models for RNN based s	Abhyuday N	2016	Proceedings of	
Identifying Risk Factors For Heart Disease in	Thanat Chokwijitkul	2018	Proceedings of the	Data for this study is from i2b2 risk factor
Extracting medication information from		2010	Journal of the	
2010 i2b2/VA challenge on concepts, assertions, and relations in c		2011		
Early hospital mortality prediction using vital	Sadeghi, R.,	2018	Smart Health	MIMIC (URL: https://physionet.org/works/
Benchmark of deep learning models on large	Purushotham, S.,	2017	Preprint	MIMIC (URL: https://physionet.org/works/
Mortality prediction in intensive care units (ICU	Davoodi, R., &	2018	Journal of Biomedical	MIMIC (URL: https://physionet.org/works/
Explainable prediction of medical codes from	Mullenbach, J.,	2018	Preprint	MIMIC (URL: https://physionet.org/works/

Problem Statement


Benchmarking results for several clinical prediction tasks such as **mortality prediction, length of stay prediction, and ICD-9 code group prediction** (International Statistical Classification of Diseases and Related Health Problems) and development of system predicting these valued information by combining NLP and Deep learning

Concrete Tasks

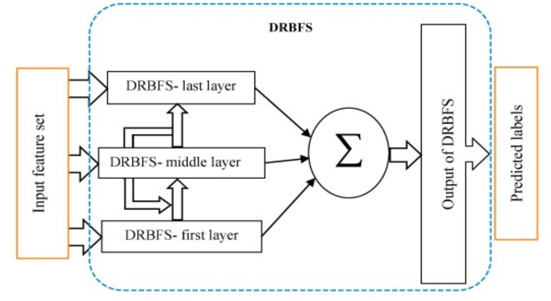
- Binary classifier for mortality or early mortality
- Regression for predicting the length of stay of patient at hospital
- ICD-9 group prediction, multi-class classification (n-groups depending on feature engineering)
- ICD-9 label prediction, multi-class classification (20 labels)
- Extracting above mentioned tasks and other clinical tasks using only discharge summaries (NLP specific)
- ****** For risk factors prediction, we are going to consult a person with domain knowledge after EID-Ul-AZHA

Objective

“Regardless of wide variety and complexity of features inherent in electronic health records, develop a model using NLP techniques and AI to predict several clinical tasks to aid doctors”

Challenges

- Terminological variations and irregularities in clinical information
- Morphological rules are required to cope with different variants in records
- Health records have specific formats
- Text retrieval, NLP and advanced DNN techniques are required to outperform already developed systems
- Approaches that enable faster and more collaborative research while protecting patient privacy and confidentiality are becoming more important

Paper	Dataset	Features	Model	Results
Early hospital mortality prediction using vital signals." ***Smart Health*9 (2018)**	**MIMIC-III**	Heart Rate Quantitative Features **(Max, Min, Skewness, Range etc.)**	**Random Forrest, Decision Trees**	**RF F1 Score : 0.97** **DT F1 Score : 0.91**
Benchmark of deep learning models on large healthcare mimic datasets. ***Journal of biomedical informatics* (2017) - *Preprint***	**MIMIC-III**	Three feature Sets for different prediction tasks	**FFN** **RNN** **MMDL : FFN + RNN**	**AUROC : 85 %** **MMDL**
Mortality prediction in intensive care units (ICUs) using a deep rule-based fuzzy classifier. ***Journal of biomedical informatics* 79 (2018)**	**MIMIC-III**	Demographics Vital Signs and Lab Events	 **DRBFS**	**AUROC : 73.60 %** **DRBFS**
Explainable prediction of medical codes from clinical text. ***arXiv preprint arXiv:1802.05695*(2018)**	**MIMIC-III**	**Discharge Summaries**	**Convolutional Attention for Multi-Label classification (CAML)**	**Precision@8 of 0.71 and a Micro-F1 of 0.54**

MIMIC-III
(Medical Information Mart for Intensive Care)

- A freely accessible critical care database
- Developed by the MIT Lab for Computational Physiology, comprising DE identified health data associated with ~40,000 critical care patients. It includes demographics, vital signs, laboratory tests, medications, and more

I2b2: Informatics for Integrating Biology and the Bedside

MIMIC-III Data Access

- Requires research ethics and compliance training courses (completed)
- Took “Data or Specimens Only Research under requirements set by “Massachusetts Institute of Technology Affiliates”
- 15 modules:
  - Research and Human Subjects
  - Privacy and Confidentiality
  - Assessing Risks
  - History and Ethical Principles
  - Regulations and Process
  - Genetics Research
  - International Research
  - HIPAA
  - Conflicts and Interest in Research

CITI PROGRAM

Completion Date 17-Apr-2019
Expiration Date 16-Apr-2022
Record ID 31301898

This is to certify that:

Faisal Maqbool

Has completed the following CITI Program course:

Human Research	(Curriculum Group)
Data or Specimens Only Research	(Course Learner Group)
2 - Refresher Course	(Stage)

Under requirements set by:

Massachusetts Institute of Technology Affiliates

CITI
Collaborative Institutional Training Initiative

Verify at www.citiprogram.org/verify/?wc1018ffc-fce8-45a9-8913-7a7bbfd3c313-31301898

Initial Exploration

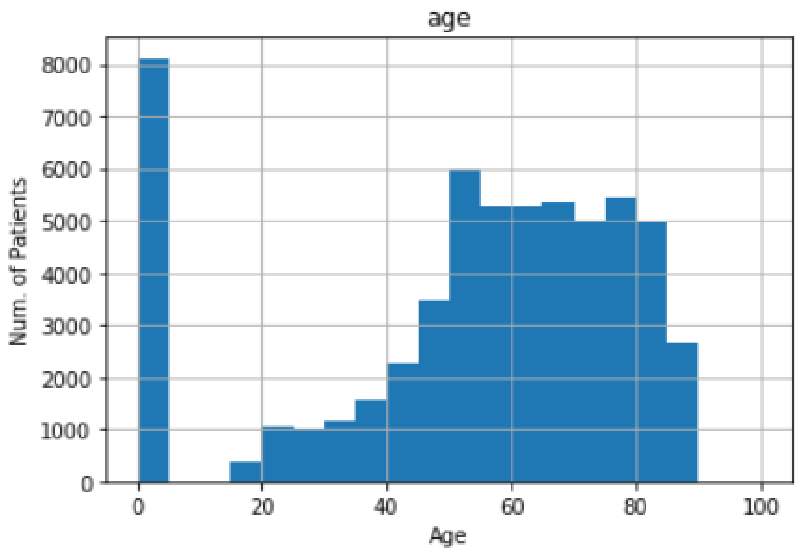


Figure 1: Age Distribution

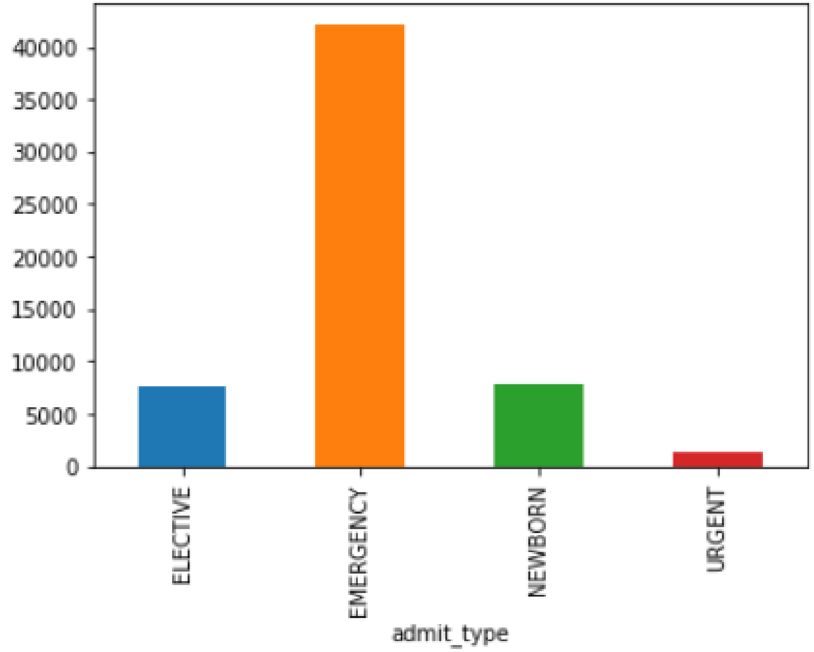


Figure 2: Admit Type Distribution

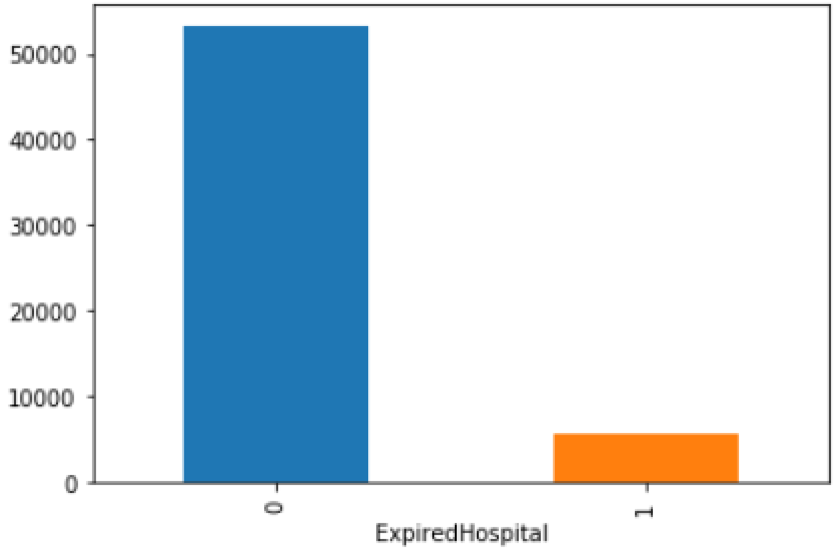
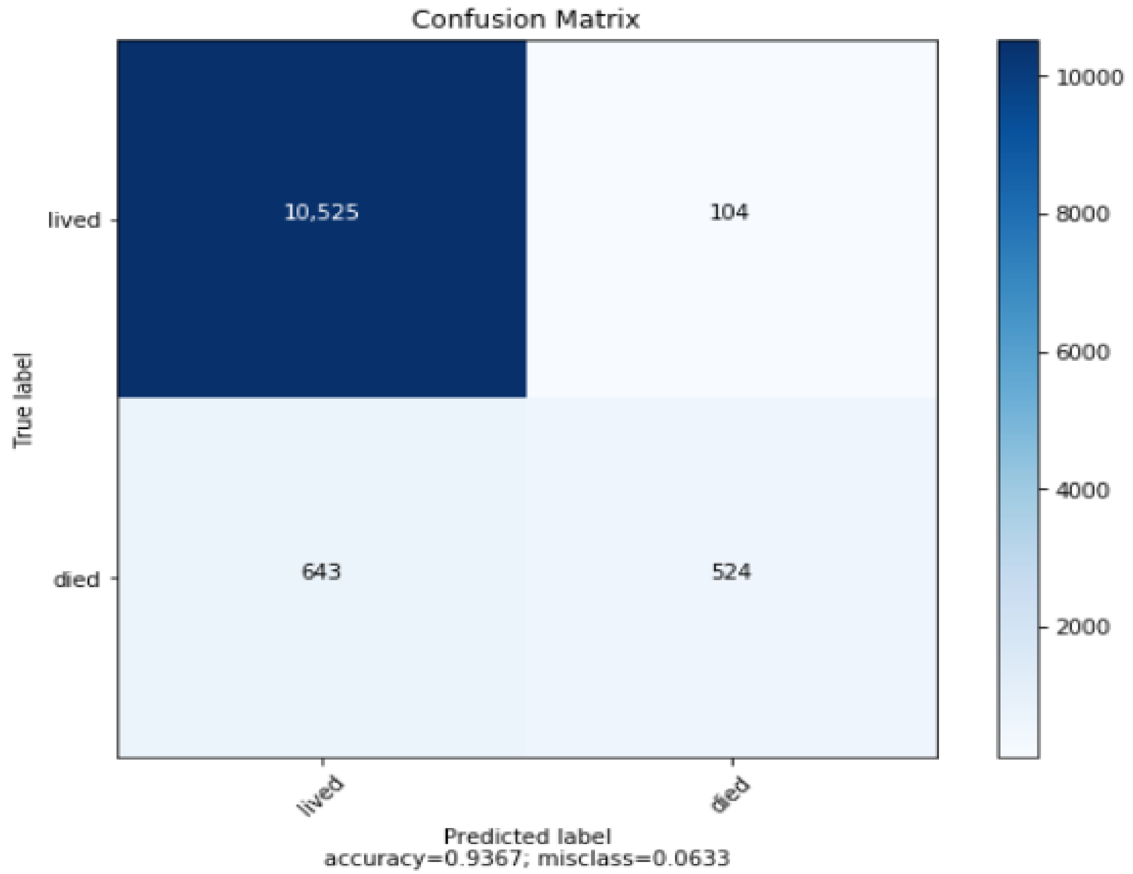


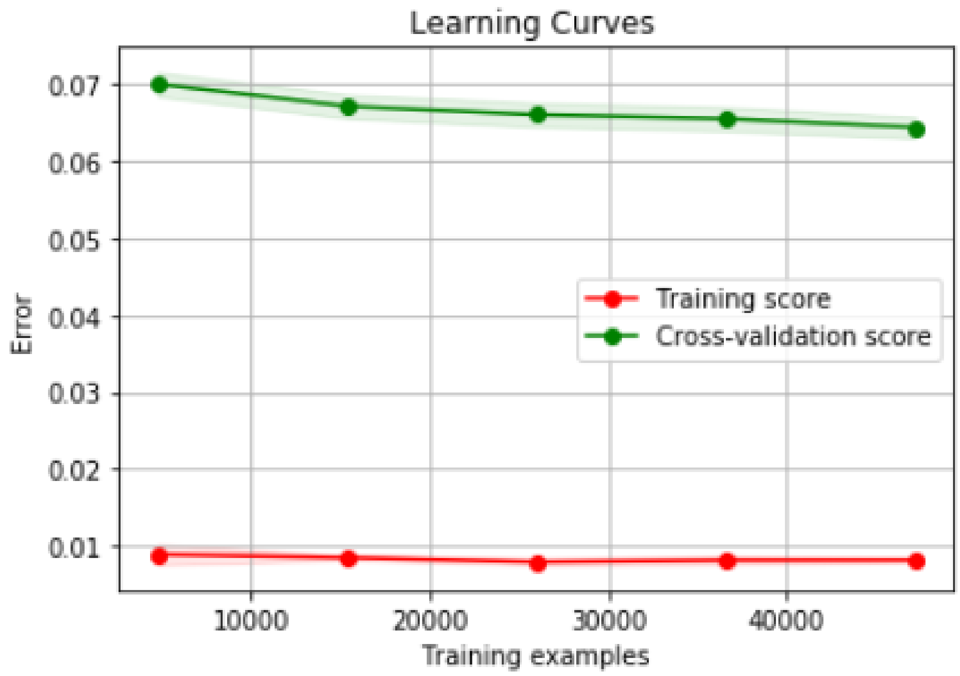
Figure 3: Expired At Hospital Dist.

Experiments Mortality Prediction

- **Random Forrest :**

```
precision  0.8344
recall  0.449
accuracy  0.9367
F1 score  0.5838
```

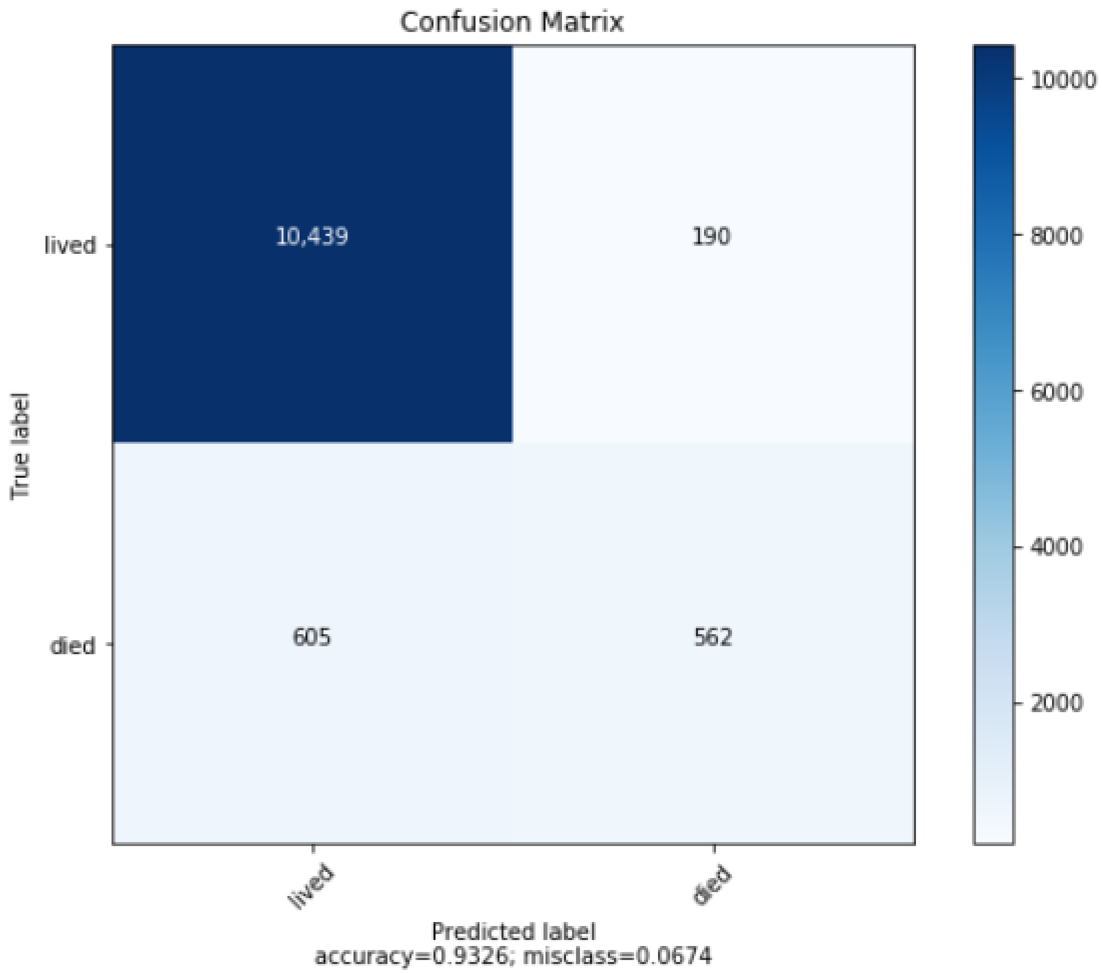


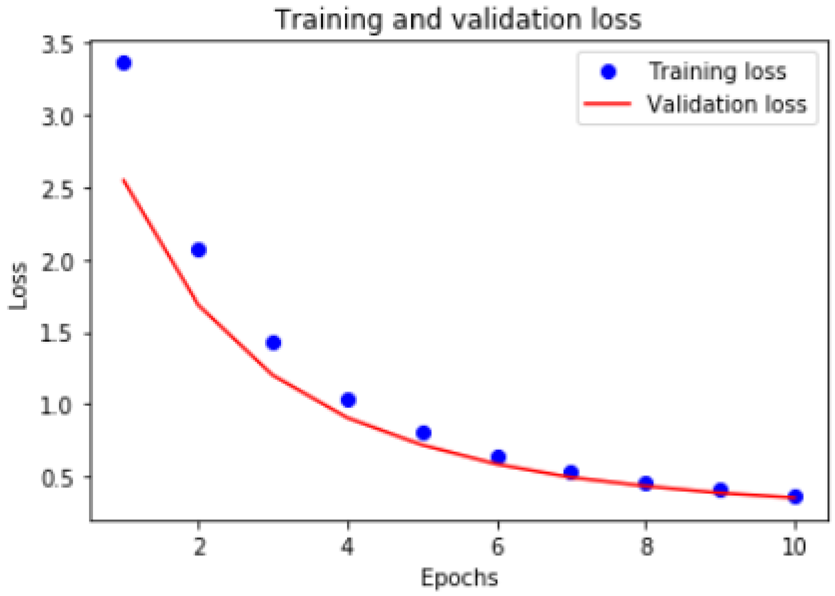


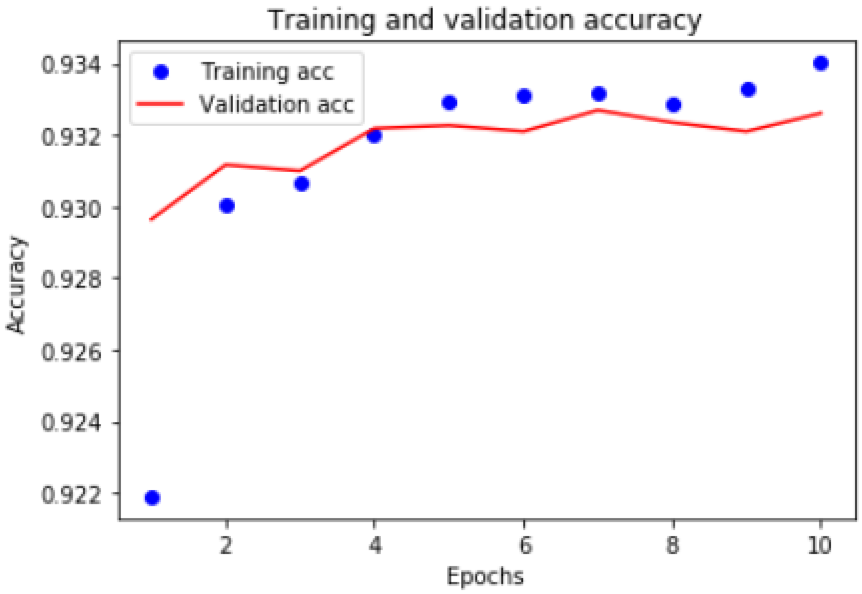
Cont..

NN:

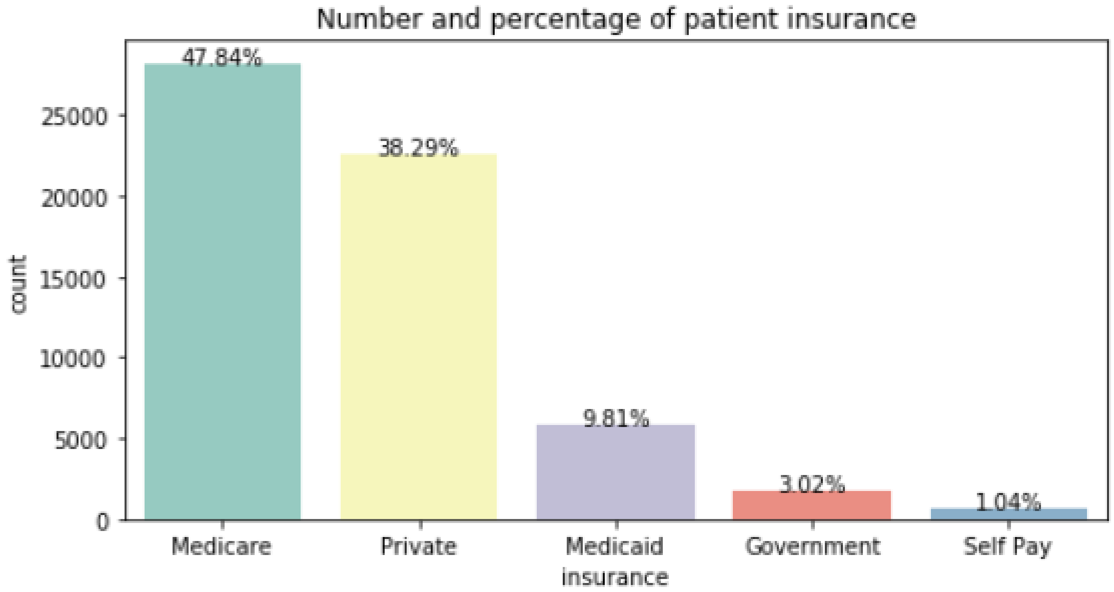
precision 0.7473404255319149
recall 0.48157669237360756
accuracy 0.9326042726347915
F1 score 0.5857217300677436

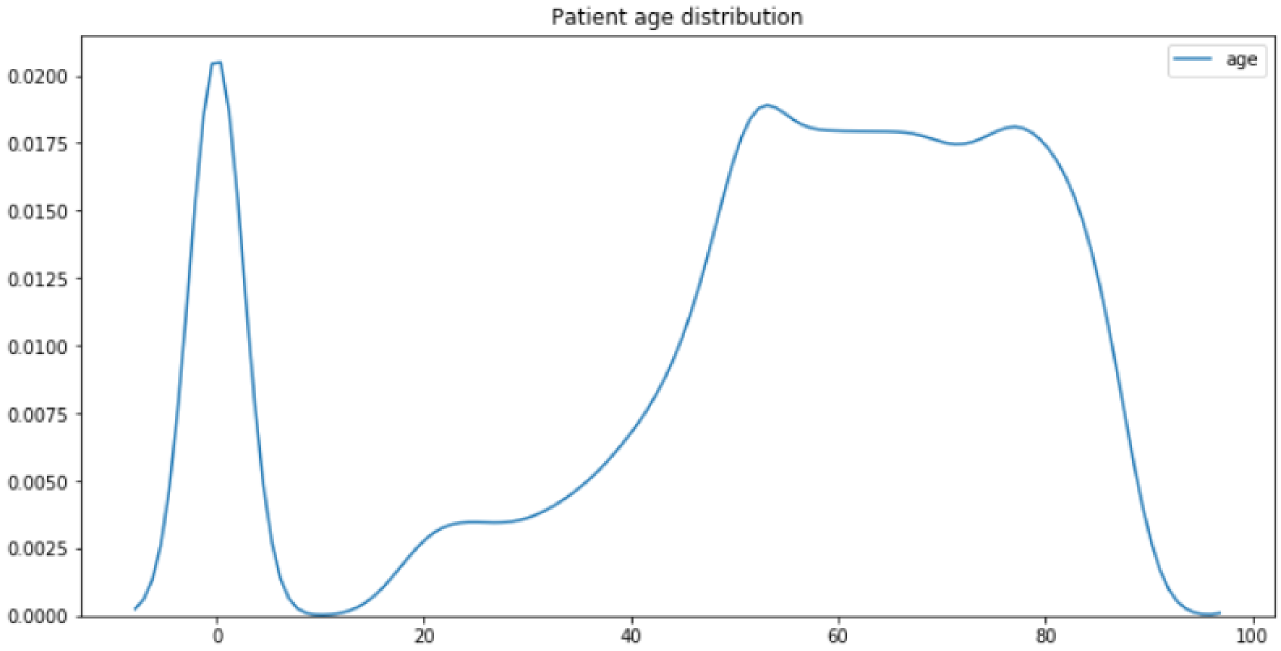


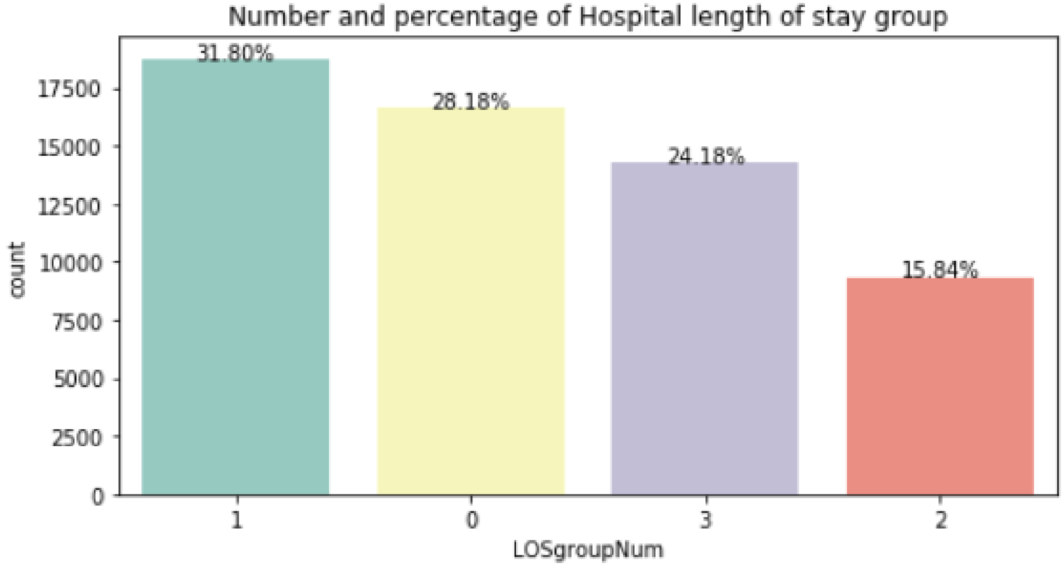


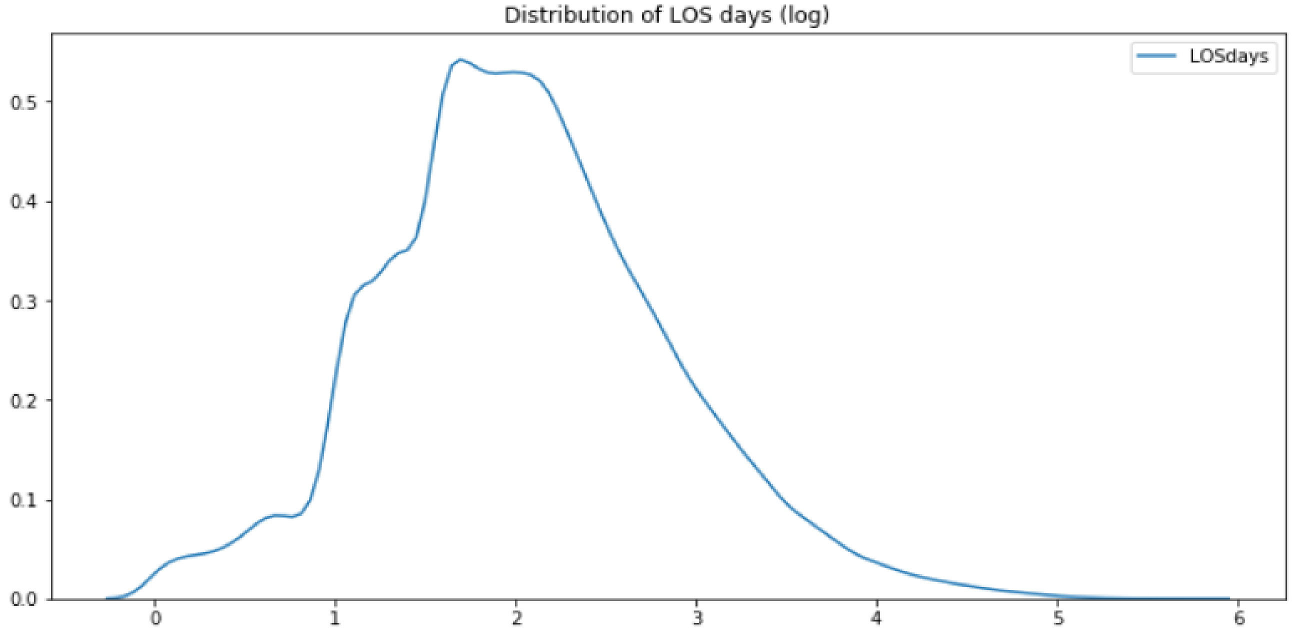


Initial Exploration Cont..









Experiments Length of Stay Prediction

Gradient Boosting Regression

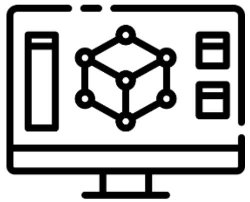
LOSgroupNum	predictions
1	1
3	3
0	0
0	0
1	1
2	2

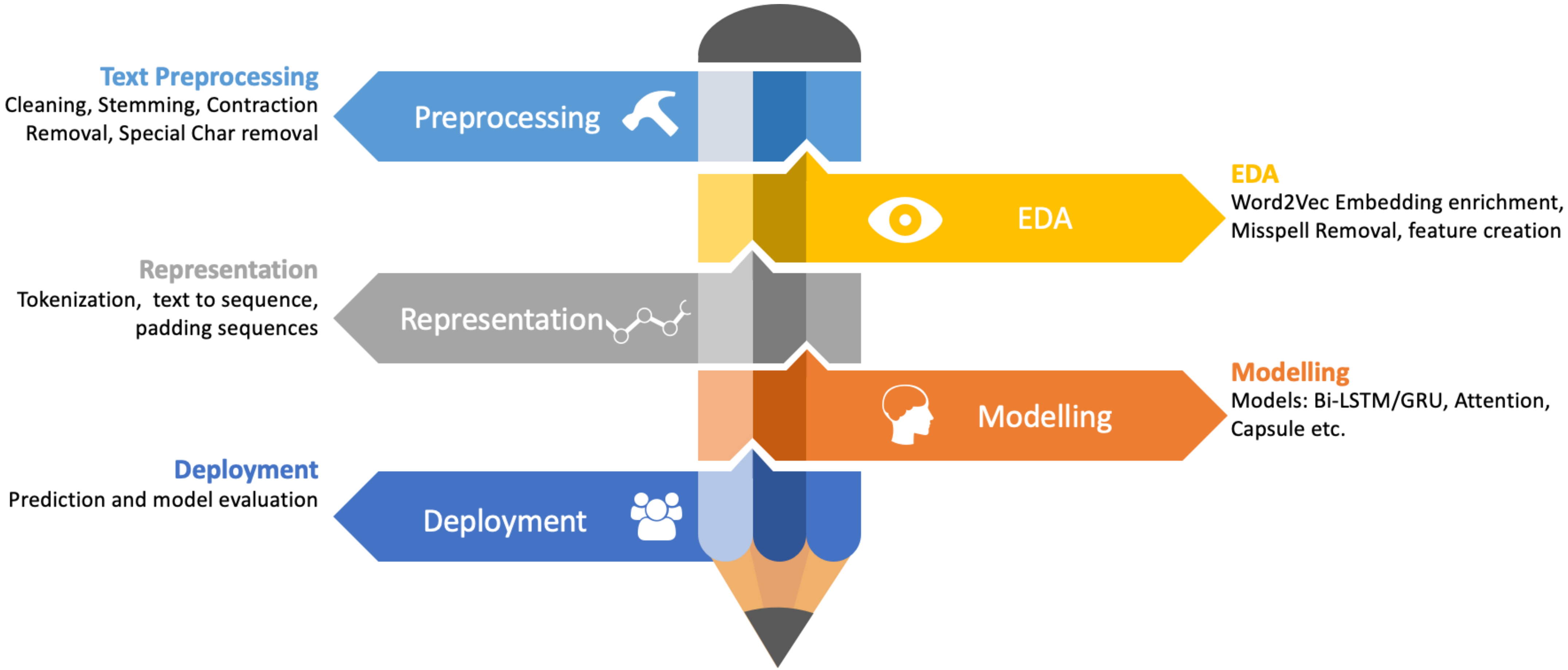
Figure 1. LOS Group Prediction (Labels vs Prediction)

LOSdays	predictions
6.79	6.974125
13.50	14.203448
3.38	7.075496
0.25	12.170375
7.08	10.801265
8.08	11.227309
16.04	18.490559

Figure 2. Length of Stay Prediction (Labels vs Prediction)

Proposed Technique





Milestones & Goals

1

July – 2019
Register Thesis

2

Aug – 2019
Feature Generation to use State of the art NLP + Deep Learning

3

Sep, Oct – 2019
Experiments and Tuning

4

Nov– 2019
Evaluation and checking dependencies

5

Dec 2019
Paper write-up, Submission

References

[1] Gentimis, Thanos, Alnaser Ala'J, Alex Durante, Kyle Cook, and Robert Steele. "**Predicting Hospital Length of Stay Using Neural Networks on MIMIC III Data**." In *2017 **IEEE** 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, pp. 1194-1201. IEEE, 2017.

[2] Sadeghi, Reza, Tanvi Banerjee, and William Romine. "**Early hospital mortality prediction using vital signals." Smart Health9 (2018)**: 265-274.

[3] Purushotham, Sanjay, Chuizheng Meng, Zhengping Che, and Yan Liu. "**Benchmark of deep learning models on large healthcare mimic datasets." *arXiv preprint arXiv:1710.08531*(2017)**.

[4] Purushotham, Sanjay, Chuizheng Meng, Zhengping Che, and Yan Liu. "**Benchmark of deep learning models on large healthcare mimic datasets." *arXiv preprint arXiv:1710.08531*(2017).**

[5] Mullenbach, James, Sarah Wiegreffe, Jon Duke, Jimeng Sun, and Jacob Eisenstein. "**Explainable prediction of medical codes from clinical text." *arXiv preprint arXiv:1802.05695*(2018).**

References

[6] http://kaggle.com/drscarlat/mimic3a/

Thank You