Predicting Patient Survival in the Intensive Care Unit

The University of Chicago

Meet the Team

Vamika Venkatesan

Jenny Zhen

Jessica Wang

Viviana Hernandez

Meghna Diwan

Agenda

01

ABOUT THE PROJECT

The motivation of our project

02

EXPLORING THE DATA

How does the raw data look?

03

PREPARING THE DATA

The techniques we used to extract relevant information

04

MODELS USED

The models we implemented to predict patient outcome

05

EVALUATING THE MODELS

The evaluation of our models performances

06

REFLECTION & CONCLUSION

Reflection on methodology and conclusion

01

ABOUT THE PROJECT

Why did we choose this?

The Intensive Care Unit

1

The ICU ranks number 1 in highest mortality unit in hospitals

4,000,000

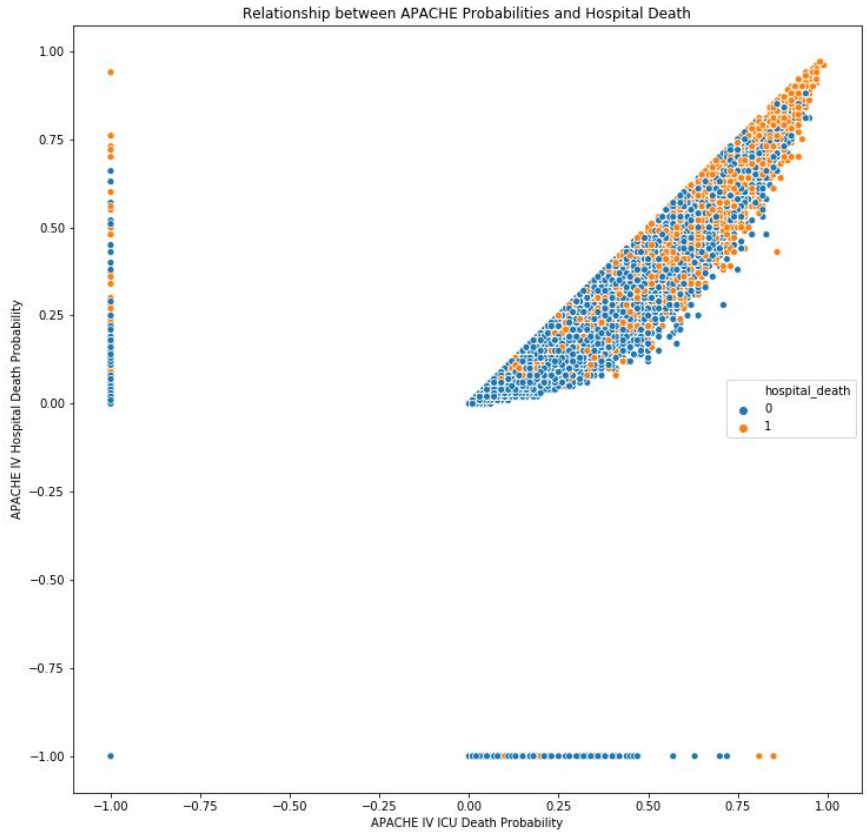
The number of annual ICU admissions in the US

500,000

The number of annual deaths in the ICU in the US

Challenges faced by the ICU

- Prone to error because of complexity of care
- Large variation in mortality among ICU patients
- Resource allocation is critical to patient welfare
- Current assessments do not accurately capture patient outcomes



Acute Physiology and Chronic Health Evaluation (APACHE) IV scores currently used -- there is no clear grouping of hospital death

The Goal

Provide hospitals with an additional measurement to better assess patient's mortality based on the first 24 hours at the ICU

02:

EXPLORING THE DATA

What does the dataset contain?

Dataset

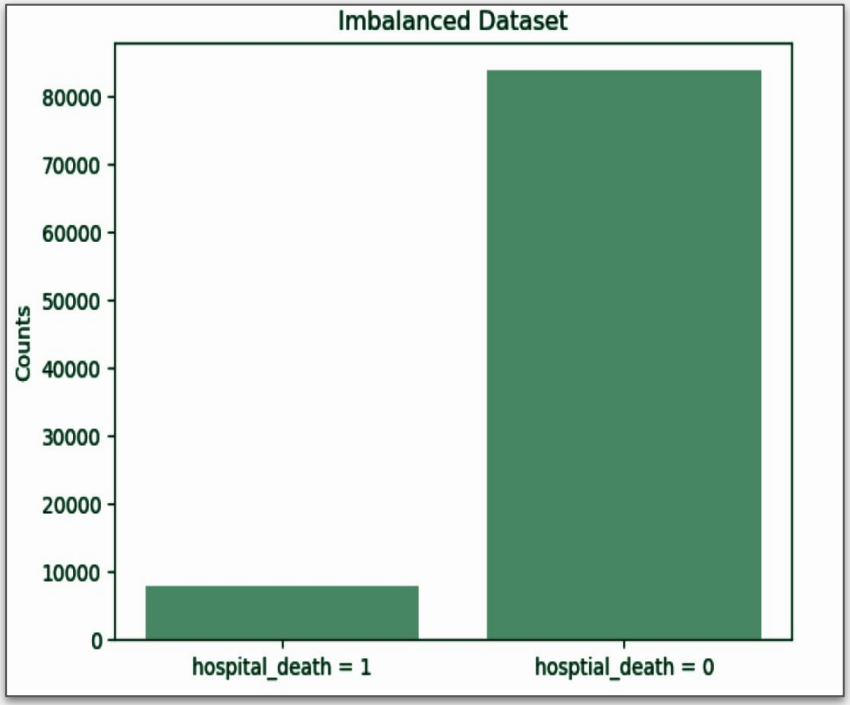
- Patient health data from MIT's GOSSIS (Global Open Source Severity of Illness Score) initiative
- More than 130,000 hospital Intensive Care Unit (ICU) visits from patients, spanning a one-year timeframe
- Each row corresponds to an encounter for a unique patient
- Comprising Argentina, Australia, New Zealand, Sri Lanka, Brazil, and more than 200 hospitals in the United States

Exploratory Data Analysis

- 91,713 encounters
- 186 columns (185 features)
  - Identifier
  - Demographic
  - APACHE covariate
  - Vitals
  - Labs
  - Lab blood gas
  - APACHE prediction
  - APACHE comorbidity
  - APACHE grouping
- Target variable: hospital_death (1/0)
- Missing values
  - 175 columns have missing values
  - 56 columns have missing values more than 70%

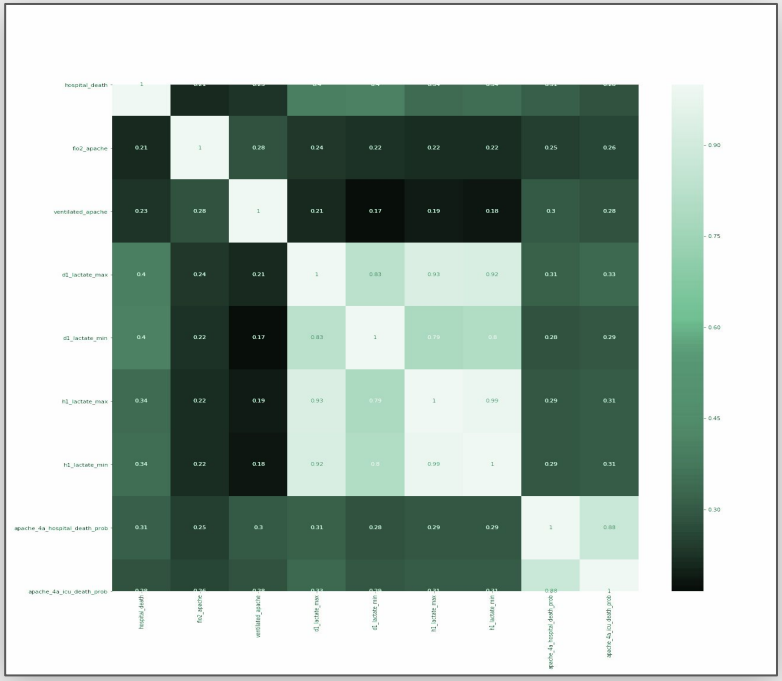
Exploratory Data Analysis

- Extremely unbalanced dataset
  - **1 : 10.6**
  - **8.6% death**

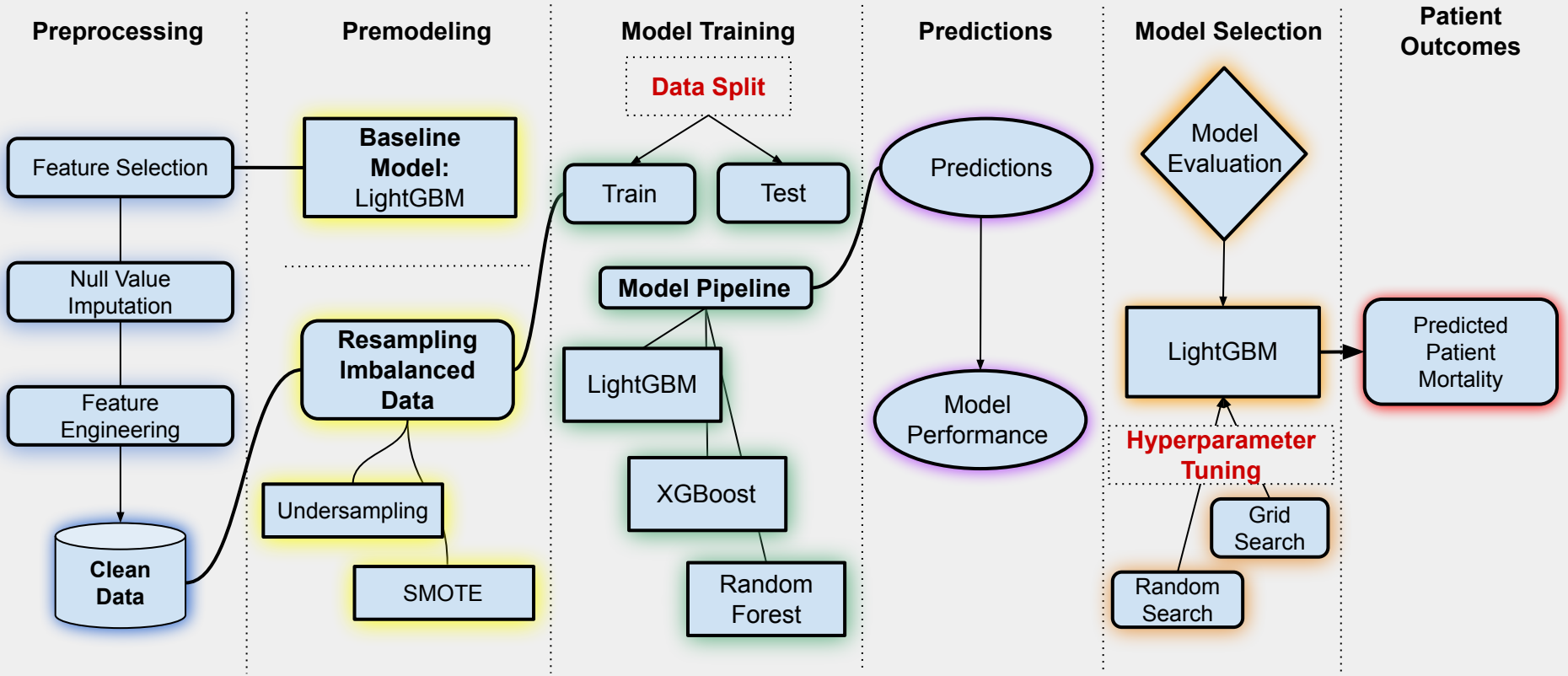


Exploratory Data Analysis

- Strong correlation between *Apache_4a_hosital_death_prob* and *Apache_4a_icu_death_prob*
- Using *Apache_4a_icu_death_prob for* imputation
- Since we have 185 features in the data, there are not many significant correlations between any of the features and target variable. However, in the hospital setting, multiple features together will create a greater effect. So we need feature engineering in the next step.



Pipeline Development



03:

PREPARING THE DATA

How did we process the data?

Missing Values

Missing < 10%

General imputation using Median as their distributions were not skewed with our target

70% > Missing > 10%

Imputed by group median of APACHE ICU Death Probability

Missing > 70%

Dropped so as to not introduce bias

Feature Engineering

- Glasgow Coma Scale Score
- High Bilirubin Levels
- High Creatine Levels
- # of tests not administered in 1 hr / 24 hrs
- Pulse Pressure Variation in 1 hr / 24 hrs
- High Lactate Levels
- Level of White Blood Cells
- High Urine Output

Feature Engineering

- Glasgow Coma Scale Score
- High Bilirubin Levels
- High Creatine Levels
- # of Test not administered in 1 hr / 24 hrs
- Pulse Pressure Variation in 1 hr / 24 hrs
- High Lactate Levels
- Level of White Blood Cells
- High Urine Output

The level of consciousness used to gauge deterioration of a patient's condition

Feature Engineering

- Glasgow Coma Scale Score
- High Bilirubin Levels
- High Creatine Levels
- # of Test not administered in 1 hr / 24 hrs
- Pulse Pressure Variation in 1 hr / 24 hrs
- High Lactate Levels
- Level of White Blood Cells
- High Urine Output

High creatine levels can indicate kidney injury in patients

Feature Engineering

- Glasgow Coma Scale Score
- High Bilirubin Levels
- High Creatine Levels
- # of Test not administered in 1 hr / 24 hrs
- Pulse Pressure Variation in 1 hr / 24 hrs
- High Lactate Levels
- Level of White Blood Cells
- High Urine Output

The number of tests missing could be indicative of a patient's death

Feature Engineering

- Glasgow Coma Scale Score
- High Bilirubin Levels
- High Creatine Levels
- # of Test not administered in 1 hr / 24 hrs
- Pulse Pressure Variation in 1 hr / 24 hrs
- High Lactate Levels
- Level of White Blood Cells
- High Urine Output

Ratio between systolic and diastolic blood pressure could indicate heart risk

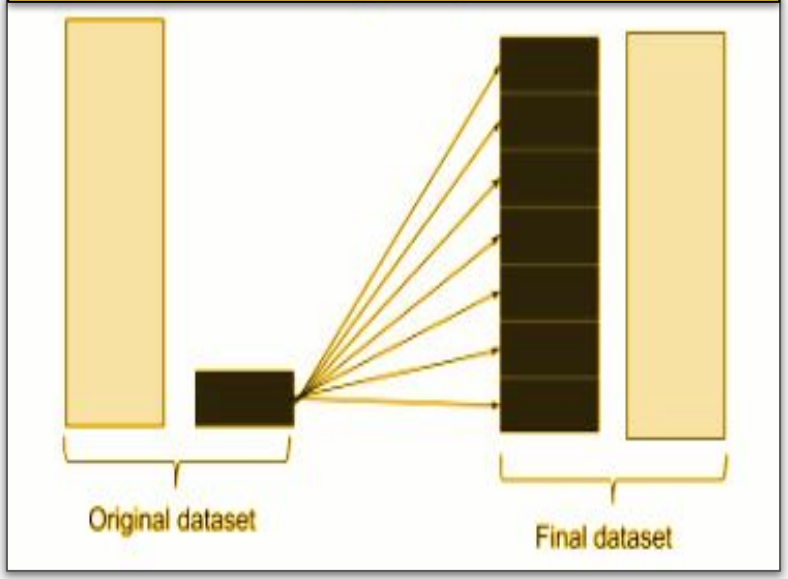
Feature Engineering

- Glasgow Coma Scale Score
- High Bilirubin Levels
- High Creatine Levels
- # of Test not administered in 1 hr / 24 hrs
- Pulse Pressure Variation in 1 hr / 24 hrs
- High Lactate Levels
- Level of White Blood Cells
- High Urine Output

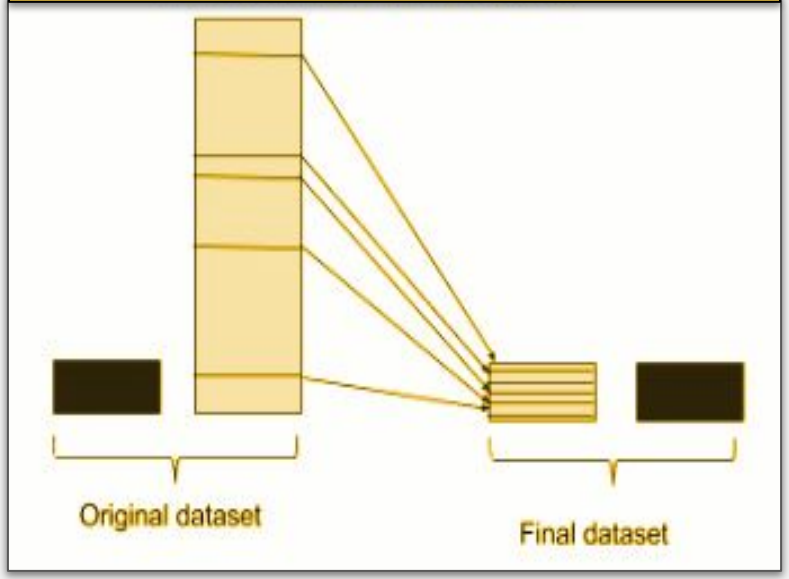
High Level of White Blood Cells can indicate inflection

Resampling Imbalanced Class

Synthetic Minority Oversampling

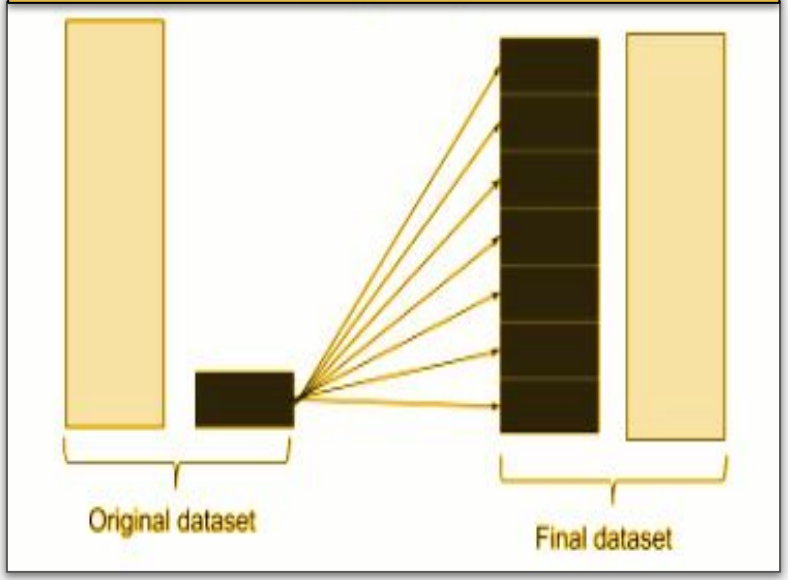


Random Undersampling



Resampling Imbalanced Class

Synthetic Minority Oversampling



What?
SMOTE is used to create fake patients who died from the 5-nn of the patients that had died in our original dataset

Why?
To create an equal number of patients who survived and who died in our training dataset

But...
Creating synthetic data can potentially create more variance in our data

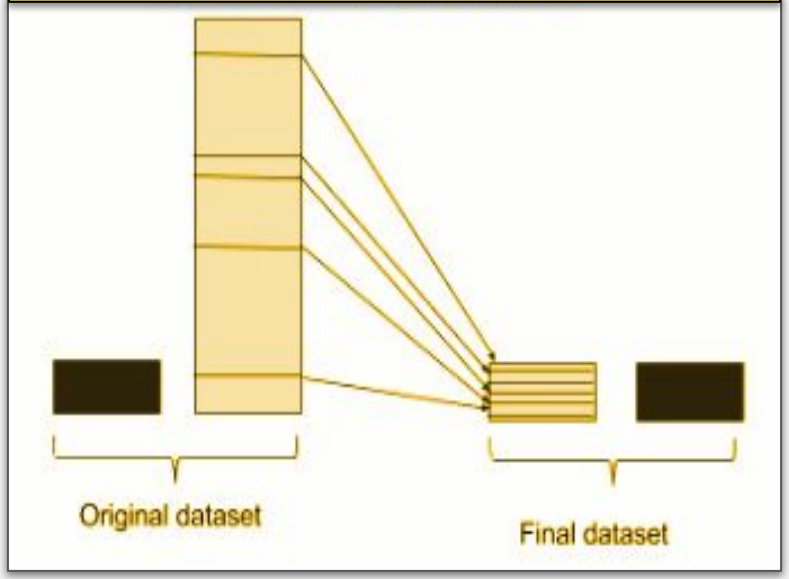
Resampling Imbalanced Class

What?
Random Undersampling is used to select patients who survived in our original dataset

Why?
To create an equal number of patients who survived and who died in our training dataset

But...
Useful information could be lost and the sample could be biased

Random Undersampling



04
CREATING MODELS

What models did we use?

Our Machine Learning Problem

Given a patient's vitals and lab results in the first 24 hours, how likely is it for them to **die** ?

Where Do Our Labels Come From ?

- Our labels are derived from patient outcomes
- Represented as “hospital death” in our dataset

Survived - 0	Died - 1
83,798	7,915

Baseline Model

What :

- Dropped variables with > 70% missing values
- **Did not** Impute any missing values may bring bias to our model
- We used LightGBM to predict our target since the model takes missing values.

Why :

- Missing values in our data may be indicative to hospital death (our target)
- To see improvement through imputing and feature engineering

model	precision_test	recall_test	FPR_test	AUROC_test	f1_test	f1_train	f1_CV
LGBMClassifier(boosting_type='gbdt', class_wei...	0.707812	0.286166	0.011158	0.889366	0.407557	0.999763	0.431752

Model Selection

1. The three models are strong learners because of its ensemble methods
2. Provide feature importance scores, which gives us good interpretation of mortality in ICU

LightGBM

A gradient boosting framework using tree-based learning algorithm

- Very fast
- Grows tree leaf-wise instead of level-wise so reduce more loss
- Not good for small data due to sensitivity to overfitting

XGBOOST

A more regularized gradient boosting machine

- Controls overfitting
- High efficiency of compute time and memory resources
- Meant for structure datasets on classification
- Can be slow on large data

Random Forest

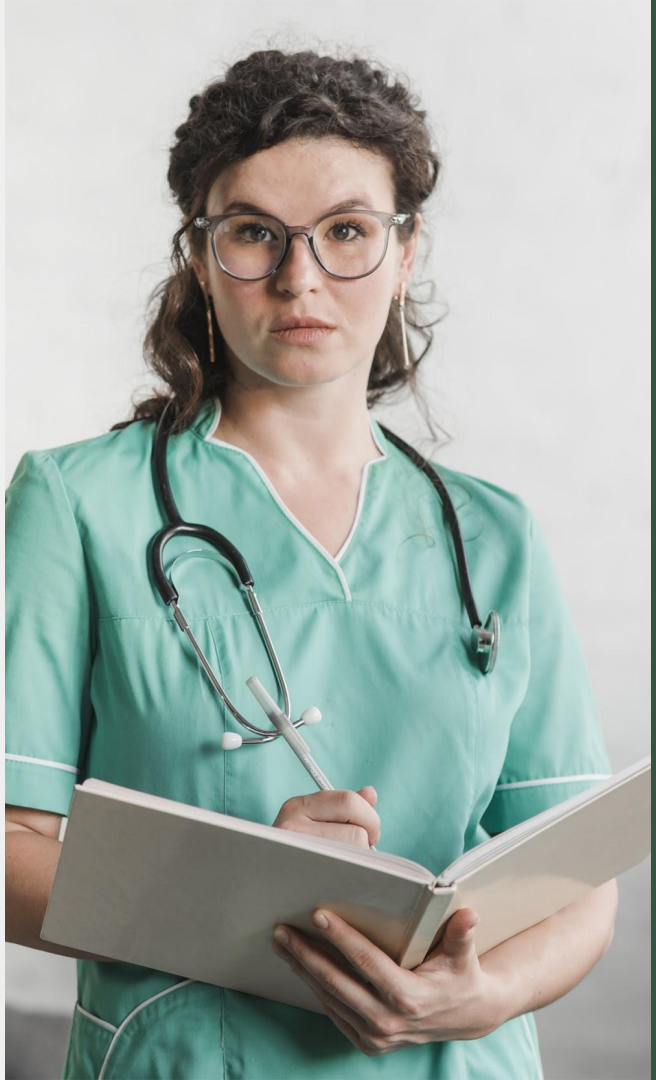
An ensemble of decision trees

- More stable using bootstrapping aggregation
- Good interpretability
- Solves overfitting
- Capture feature interactions
- Can be slow on large data

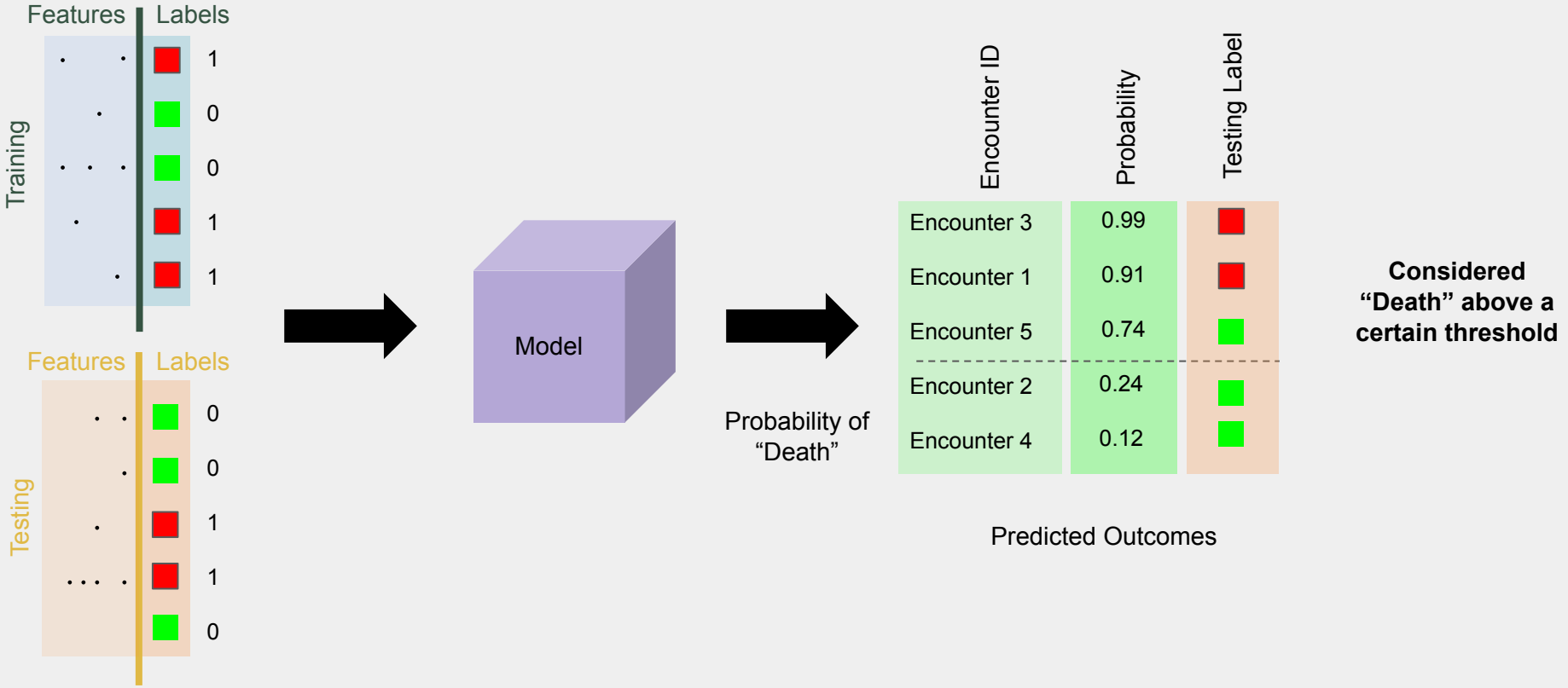
05:

EVALUATING THE MODELS

How we decided which model performed best?



How will we evaluate our models?

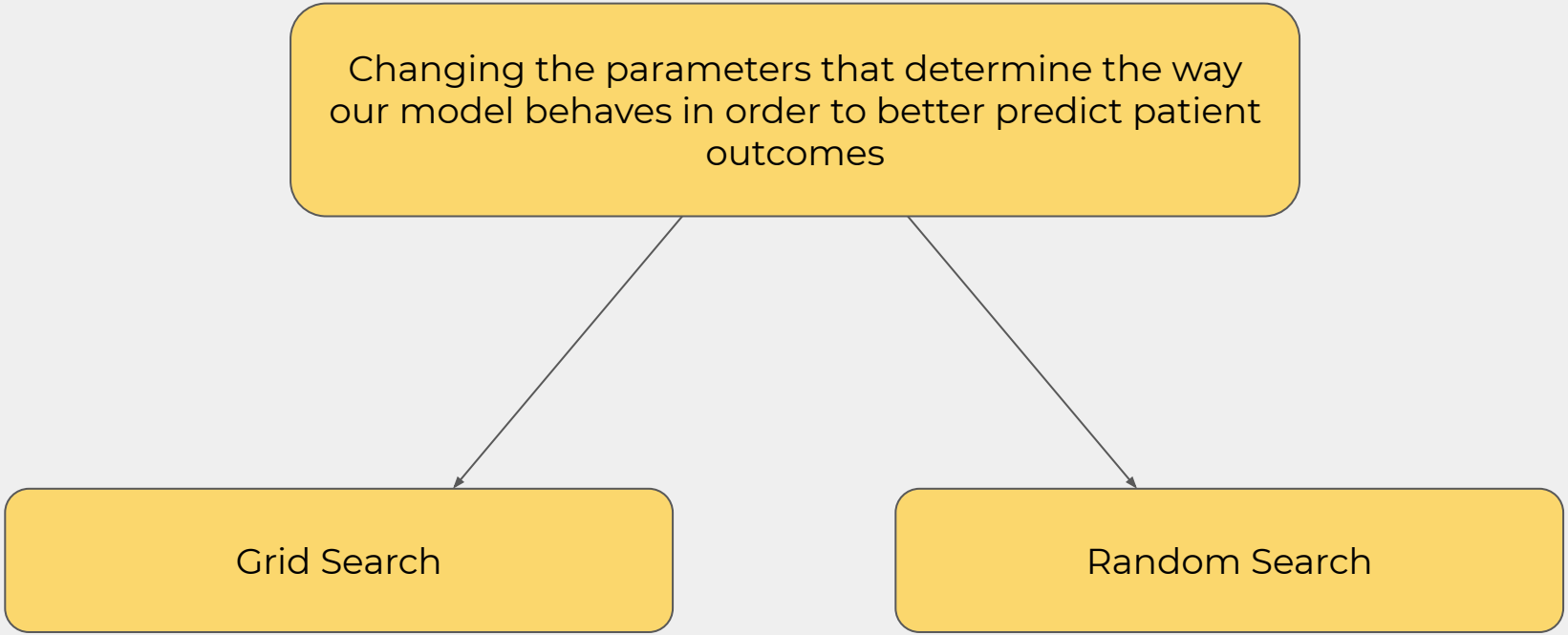


LightGBM and Undersample

1. LightGBM has a **higher AUROC** (Area under the Receiver Operating Characteristic) score
    a. AUROC score tells if a classifier is good or not (Best is 1, worst is 0.5)
2. Undersampling gives a **higher recall** score
    a. Recall (True Positive Rate) shows how much deaths we correctly predicted out of all deaths
    b. We want to a high recall to not miss any patient with mortality risks
3. LightGBM is **faster** in terms of efficiency and has a better overfitting control

	Recall	AUROC	Overfit
LGBM	0.82	0.90	No
XGB	0.79	0.86	No
RF	0.72	0.80	Slightly

Hyperparameter Tuning

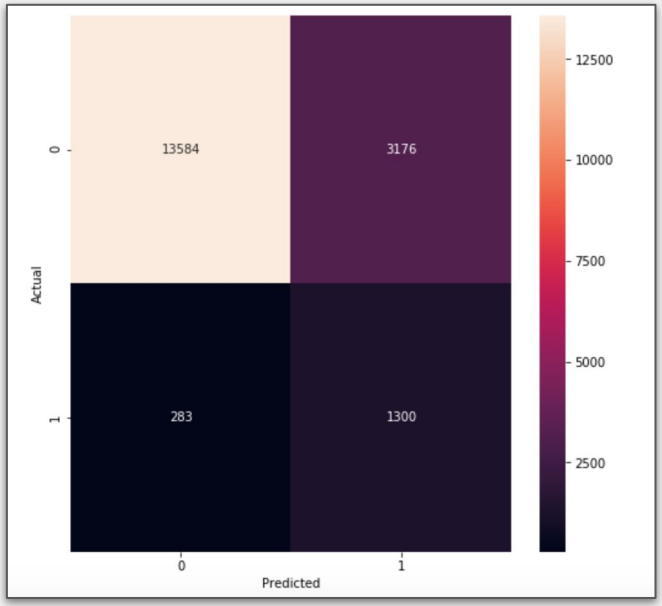


06
CONCLUSION & REFLECTIONS

What is the impact of this project?

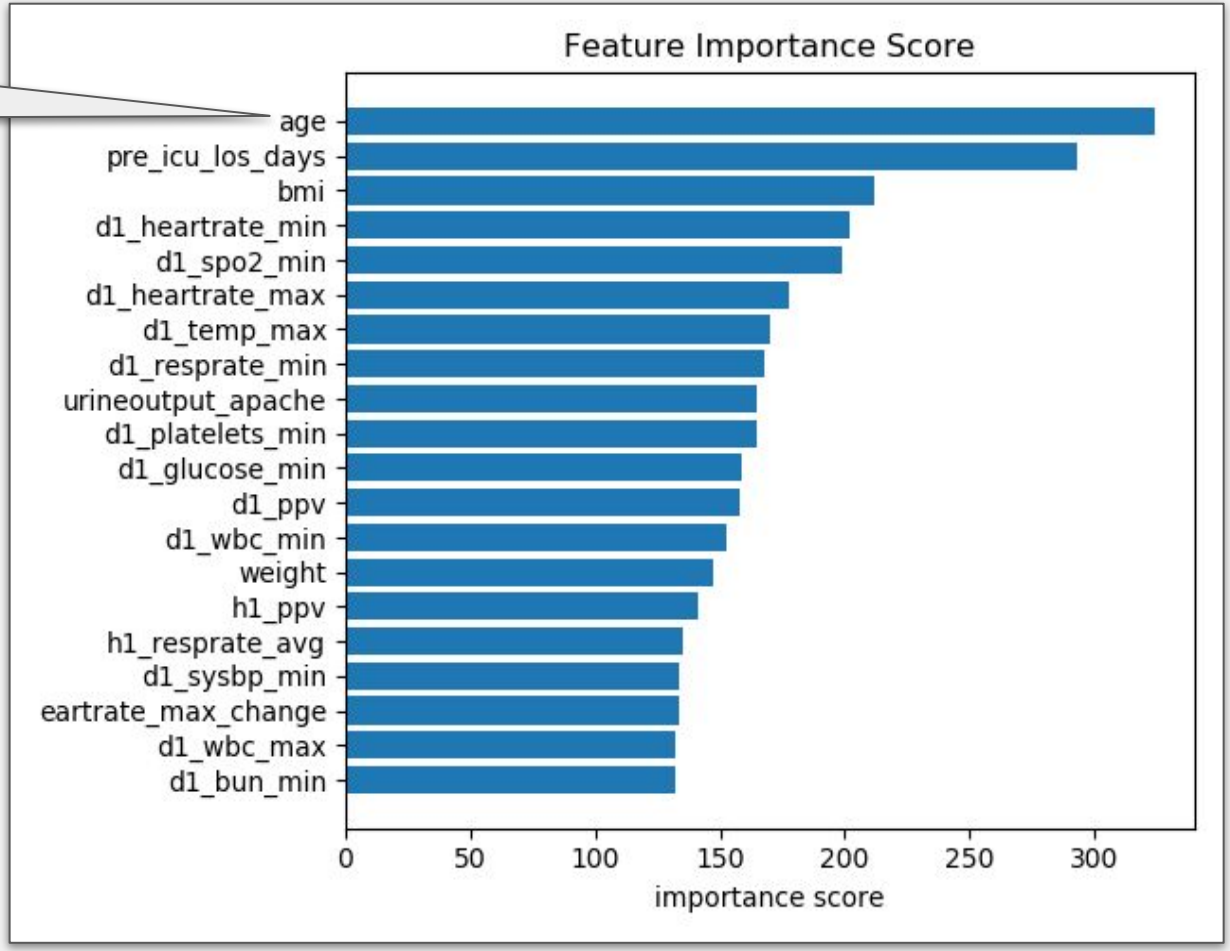
Final Results

	Recall	AUROC	F1	Precision
LGBM	0.82	0.90	0.43	0.3

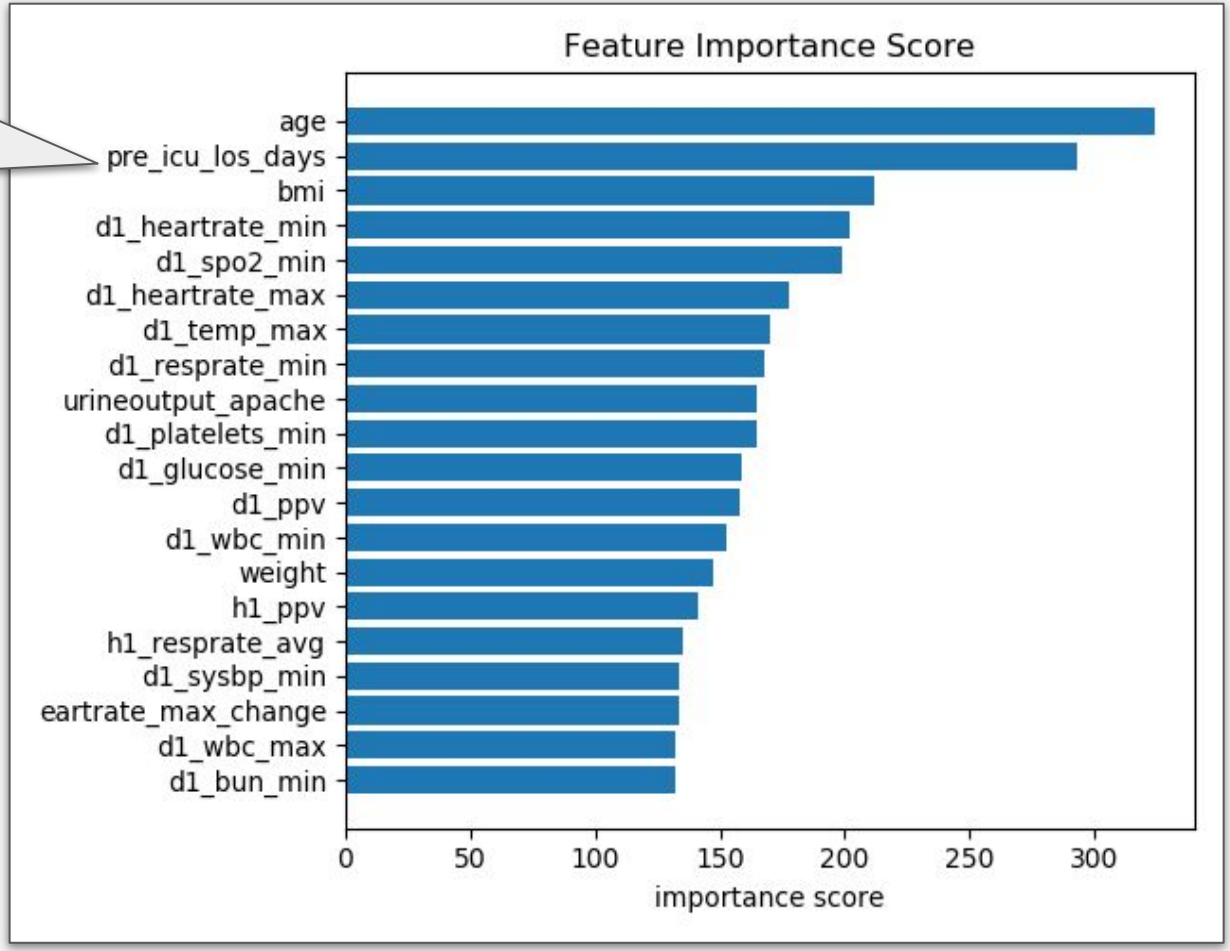


<u>True Negative</u> Predicted Survival Actual Survival	<u>False Positive</u> Predicted Death Actual Survival
<u>False Negative</u> Predicted Survival Actual Death	<u>True Positive</u> Predicted Death Actual Death

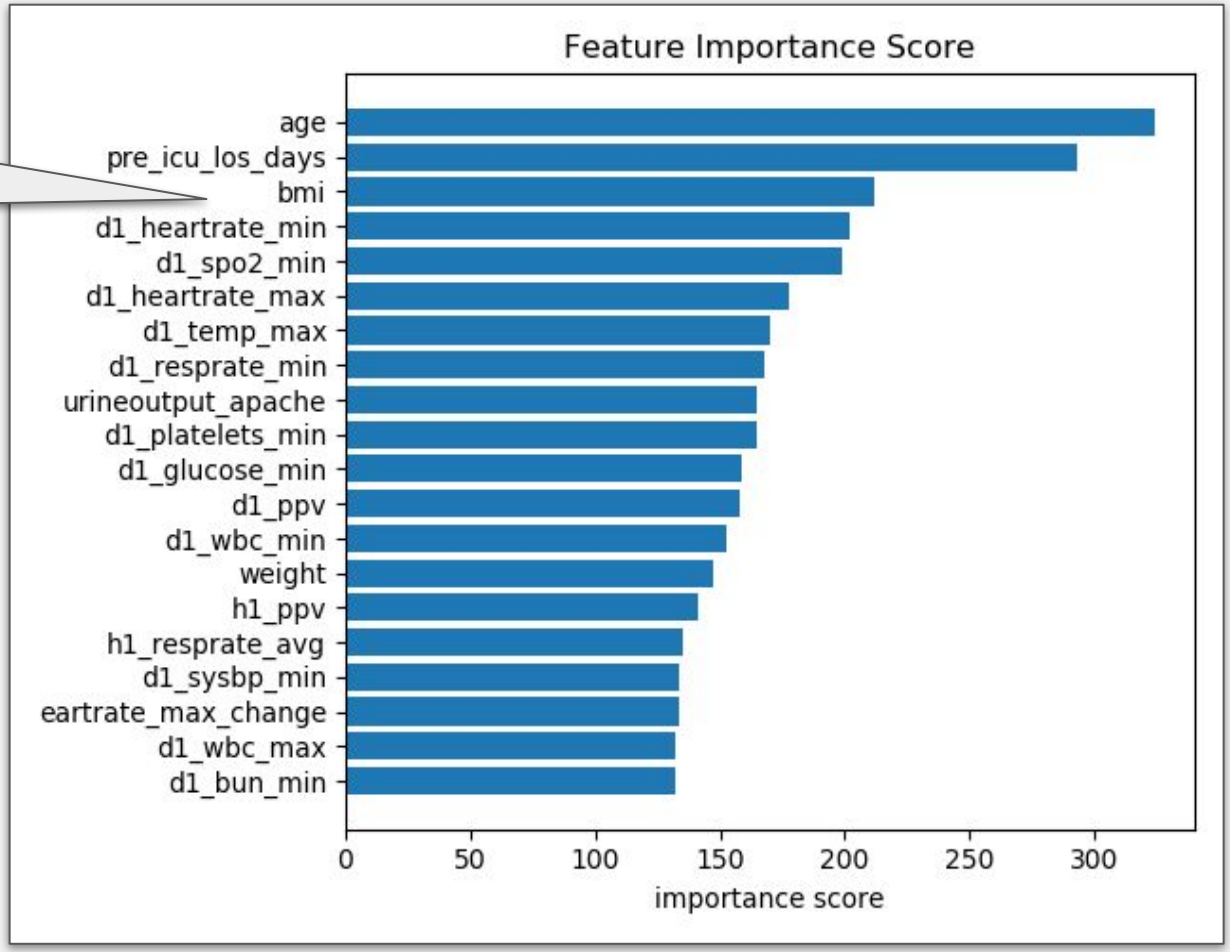
Age of people seem help predict death



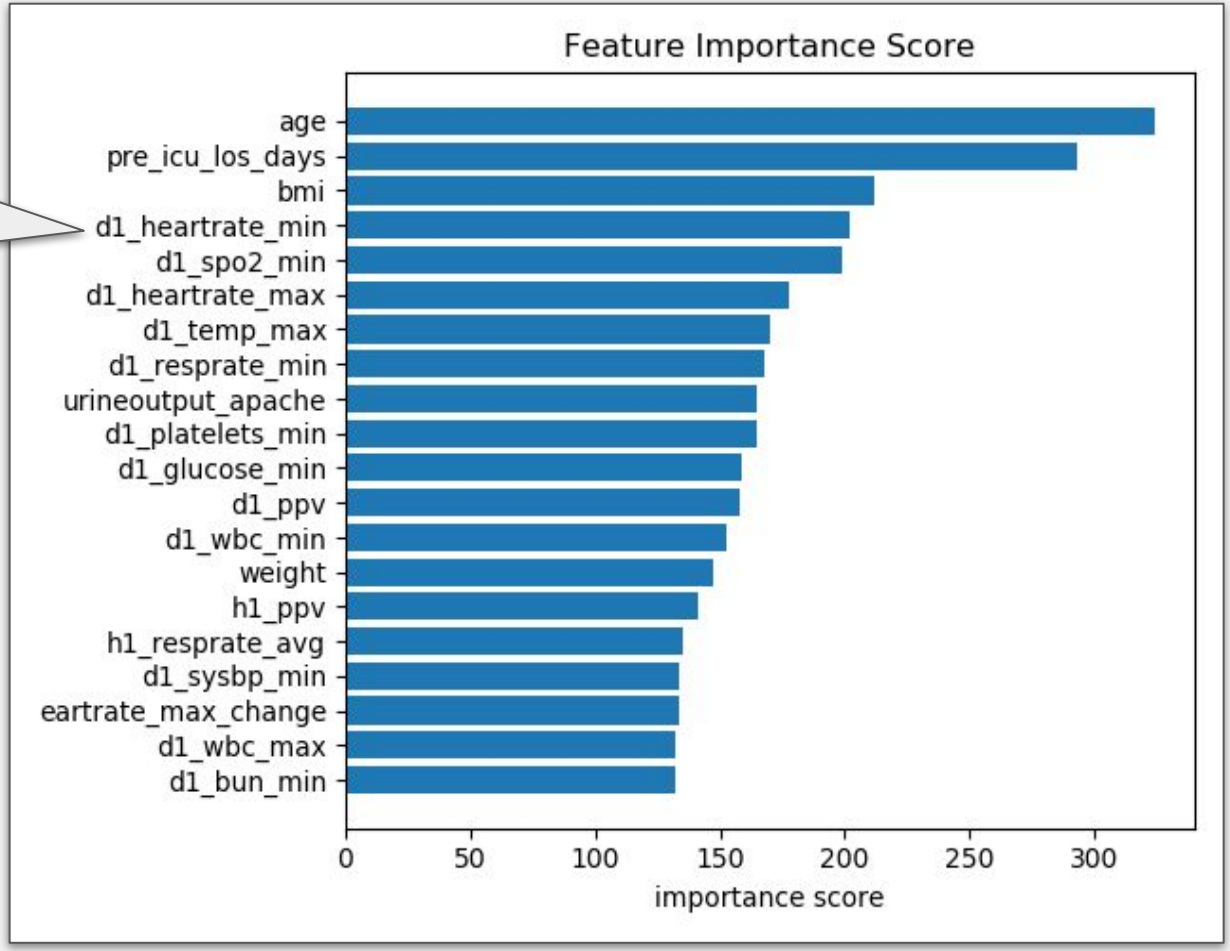
Length of stay for patient at hospital before ICU admit seems to impact death



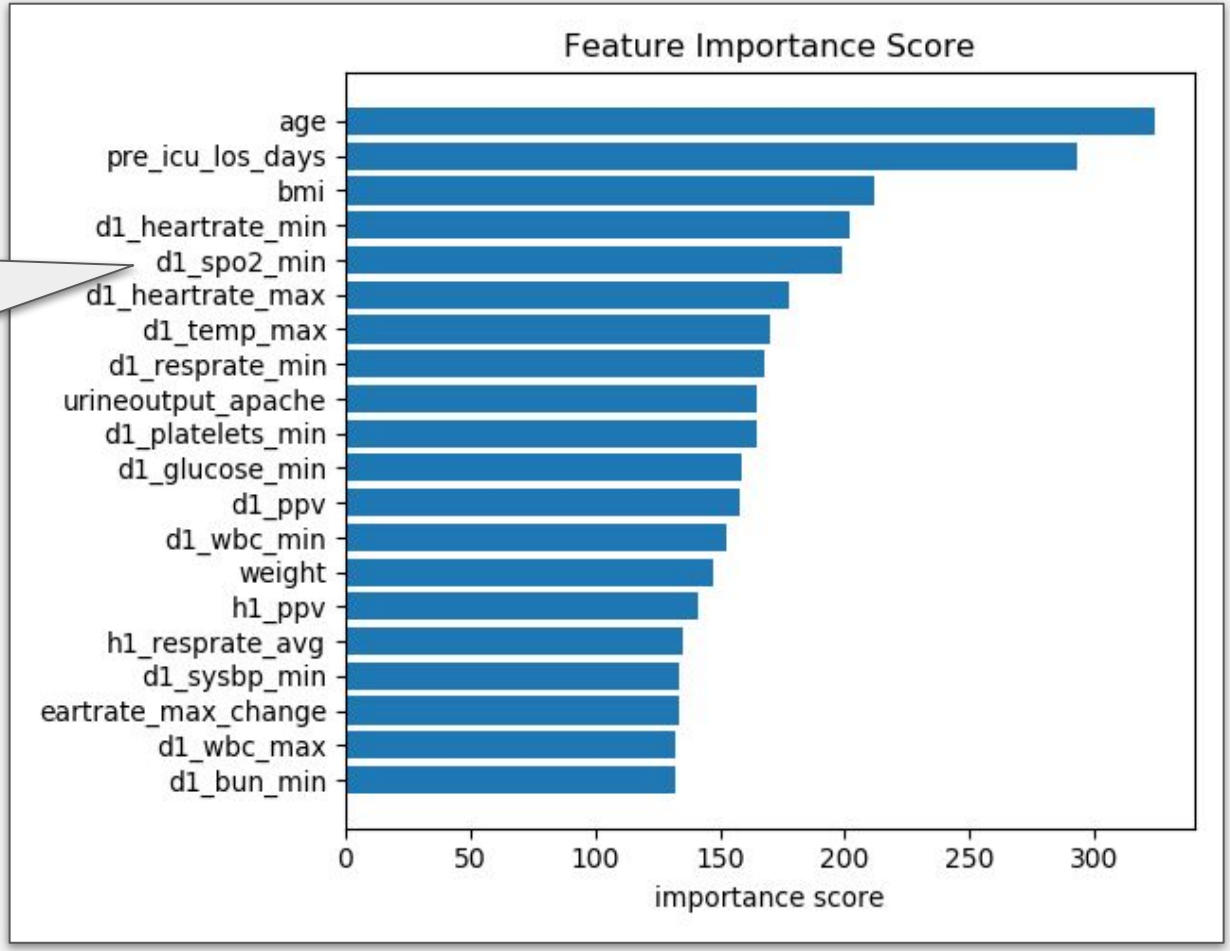
BMI standards seems to impact death



Minimum heart rate in the first 24 hrs has a large effect on patient outcome



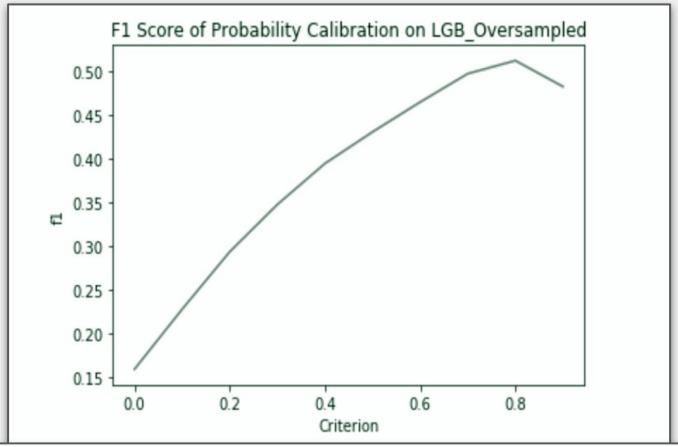
Amount of oxygen in the blood during the first 24 hours indicate mortality



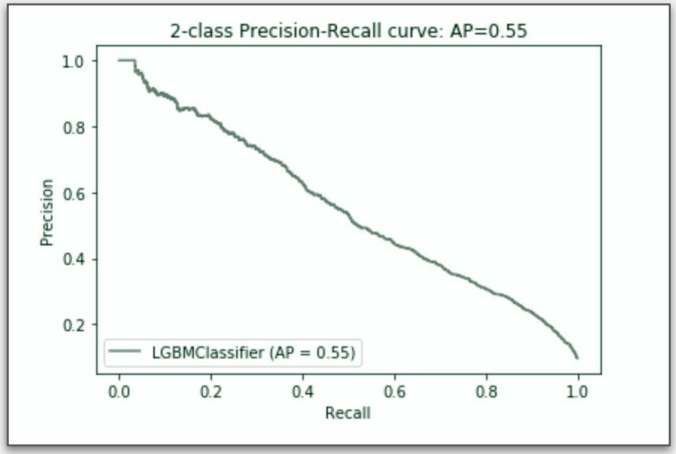
Final Results

	Recall	AUROC	F1	Precision
LGBM	0.82	0.90	0.43	0.3

Probability need not be calibrated at this point

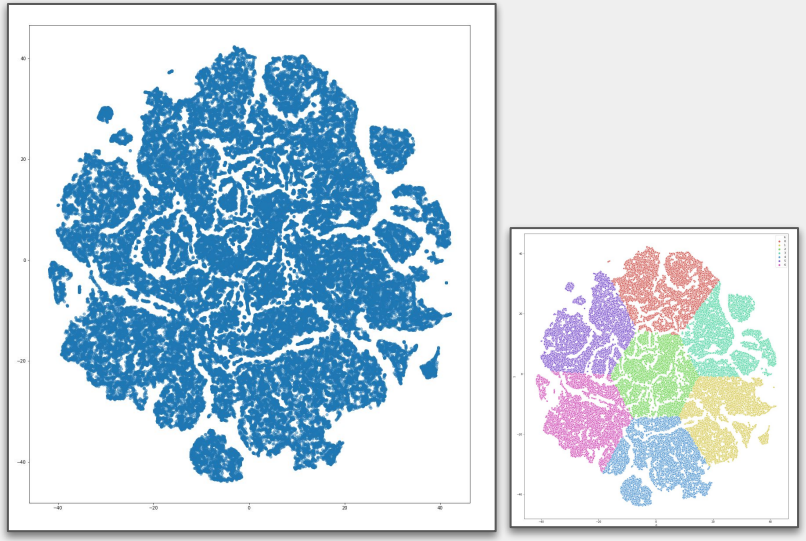


Maximum f1-score is 0.5120375807398708 when criterion is 0.8

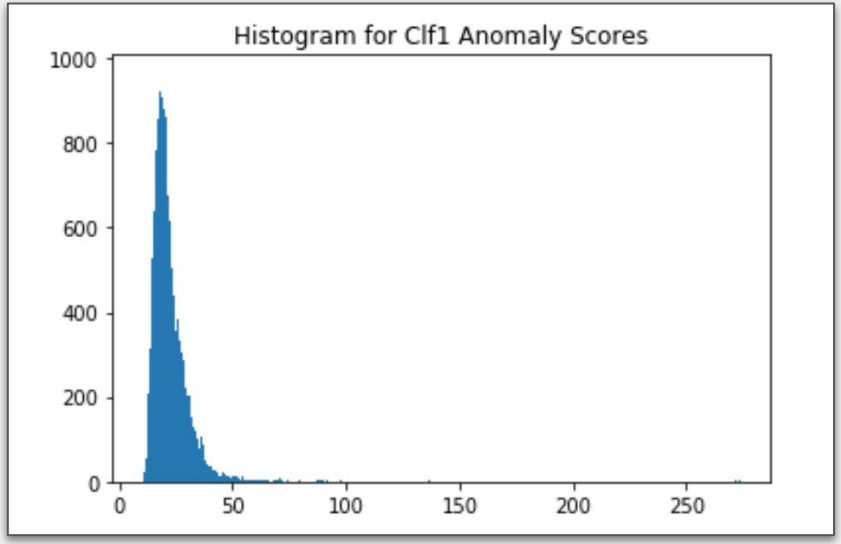


Potential Improvements

Imputing Missing Value using Clustering (t-SNE)



Use AutoEncoder to better detect death



(Neural network reducing signal “noises”)

Impact of Results

We estimate a total savings of $5.1B - $12.2B for the overall critical care medicine industry with the implementation of this new metric

An additional resource for doctors to reduce medical error

Hospitals could better allocate resources in the ICU to patients at most risk

Could potentially save lives

Lessons Learned

Test are not very indicative of death: we only have min/max for 1 hr and 24 hr.

Time series data is a huge factor in ICU

We could create better features

Domain expertise is vital to predict better outcomes

Get yourself on Git and drop out of Google Collab

Collaboration is key

Document everything so that your code is readable

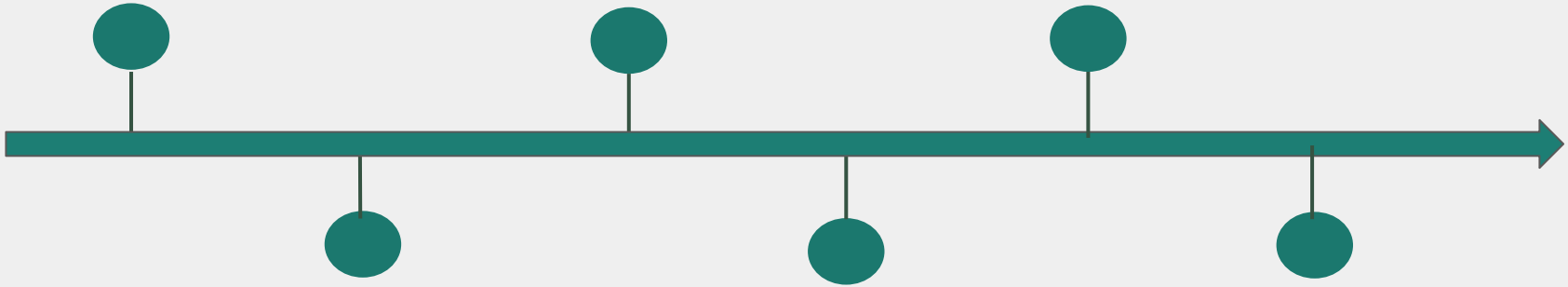
Comments are essential

Don’t underestimate how long data cleaning takes

80% of the work is EDA

Although they process faster, Windows can’t be trusted

Macbooks are forever



Questions?

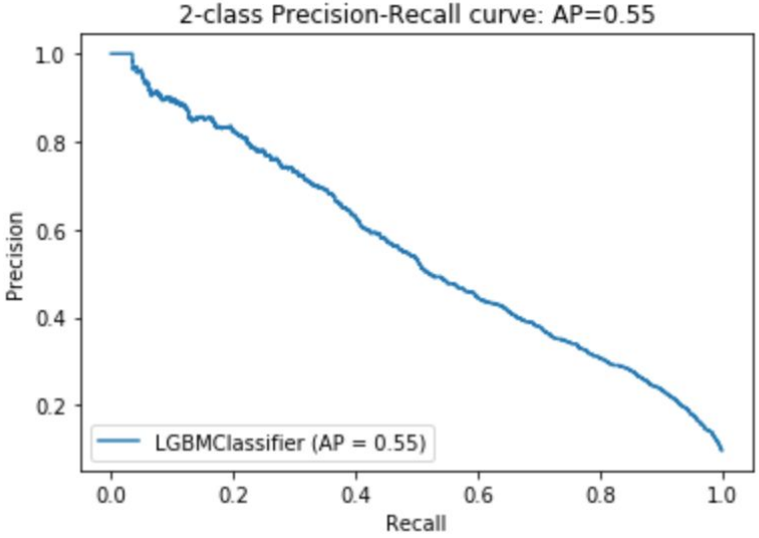
Appendix -- PIC

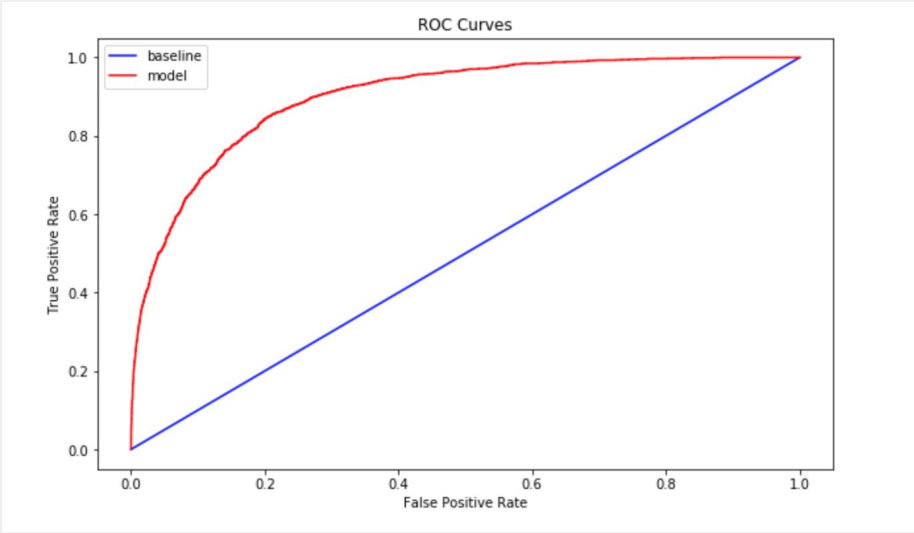
	model	precision_test	recall_test	FPR_test	AUROC_test	f1_test	f1_train		f1_CV	recall_CV
0	LGB_U	0.291413	0.816804	0.187589	0.897324	0.429568	0.481614	[0.41290111329404067, 0.43291351805205713, 0.4...		0.796589
0	LGB_U	0.291413	0.816804	0.187589	0.897324	0.429568	0.481614	[0.41290111329404067, 0.43291351805205713, 0.4...		0.814277
0	LGB_U	0.291413	0.816804	0.187589	0.897324	0.429568	0.481614	[0.41290111329404067, 0.43291351805205713, 0.4...		0.814277
0	LGB_U	0.291413	0.816804	0.187589	0.897324	0.429568	0.481614	[0.41290111329404067, 0.43291351805205713, 0.4...		0.810486
0	RF_U	0.296658	0.728996	0.163246	0.868660	0.421707	0.433578	[0.41074306177260517, 0.42295081967213105, 0.4...		0.724574
0	RF_U	0.296658	0.728996	0.163246	0.868660	0.421707	0.433578	[0.41074306177260517, 0.42295081967213105, 0.4...		0.73784
0	RF_U	0.296658	0.728996	0.163246	0.868660	0.421707	0.433578	[0.41074306177260517, 0.42295081967213105, 0.4...		0.728996
0	RF_U	0.296658	0.728996	0.163246	0.868660	0.421707	0.433578	[0.41074306177260517, 0.42295081967213105, 0.4...		0.731522
0	XG_U	0.273380	0.794062	0.199344	0.797359	0.406730	0.482665	[0.39461301314294983, 0.4004452926208651, 0.40...		0.768162
0	XG_U	0.273380	0.794062	0.199344	0.797359	0.406730	0.482665	[0.39461301314294983, 0.4004452926208651, 0.40...		0.795325
0	XG_U	0.273380	0.794062	0.199344	0.797359	0.406730	0.482665	[0.39461301314294983, 0.4004452926208651, 0.40...		0.79343
0	XG_U	0.273380	0.794062	0.199344	0.797359	0.406730	0.482665	[0.39461301314294983, 0.4004452926208651, 0.40...		0.782691

	model	precision_test	recall_test	FPR_test	AUROC_test	f1_test	f1_train		f1_CV	recall_cv
0	LGB_O	0.677223	0.351232	0.015811	0.895908	0.462562	0.537690	[0.4504950495049505, 0.4691565253881662, 0.466...		0.344915
0	LGB_O	0.677223	0.351232	0.015811	0.895908	0.462562	0.537690	[0.4504950495049505, 0.4691565253881662, 0.466...		0.353127
0	LGB_O	0.677223	0.351232	0.015811	0.895908	0.462562	0.537690	[0.4504950495049505, 0.4691565253881662, 0.466...		0.355022
0	LGB_O	0.677223	0.351232	0.015811	0.895908	0.462562	0.537690	[0.4504950495049505, 0.4691565253881662, 0.466...		0.360708
0	RF_O	0.339616	0.614656	0.112888	0.858008	0.437500	0.448524	[0.42686499887311247, 0.44404168554599, 0.4413...		0.598231
0	RF_O	0.339616	0.614656	0.112888	0.858008	0.437500	0.448524	[0.42686499887311247, 0.44404168554599, 0.4413...		0.616551
0	RF_O	0.339616	0.614656	0.112888	0.858008	0.437500	0.448524	[0.42686499887311247, 0.44404168554599, 0.4413...		0.619078
0	RF_O	0.339616	0.614656	0.112888	0.858008	0.437500	0.448524	[0.42686499887311247, 0.44404168554599, 0.4413...		0.621605
0	XG_O	0.732824	0.303222	0.010442	0.646390	0.428954	0.606538	[0.40703296703296704, 0.414721723518851, 0.417...		0.292483
0	XG_O	0.732824	0.303222	0.010442	0.646390	0.428954	0.606538	[0.40703296703296704, 0.414721723518851, 0.417...		0.291851
0	XG_O	0.732824	0.303222	0.010442	0.646390	0.428954	0.606538	[0.40703296703296704, 0.414721723518851, 0.417...		0.295009
0	XG_O	0.732824	0.303222	0.010442	0.646390	0.428954	0.606538	[0.40703296703296704, 0.414721723518851, 0.417...		0.319015

	model	precision_test	recall_test	FPR_test	AUROC_test	f1_test	f1_train	f1_CV
0	LGB	0.703896	0.342388	0.013604	0.901638	0.460688	0.594870	0.440461
0	LGB	0.703896	0.342388	0.013604	0.901638	0.460688	0.594870	0.462649
0	LGB	0.703896	0.342388	0.013604	0.901638	0.460688	0.594870	0.451915
0	LGB	0.703896	0.342388	0.013604	0.901638	0.460688	0.594870	0.464225
0	RF	0.757143	0.167404	0.005072	0.865350	0.274185	0.327635	0.281137
0	RF	0.757143	0.167404	0.005072	0.865350	0.274185	0.327635	0.273859
0	RF	0.757143	0.167404	0.005072	0.865350	0.274185	0.327635	0.261506
0	RF	0.757143	0.167404	0.005072	0.865350	0.274185	0.327635	0.294661
0	XG	0.714719	0.297536	0.011217	0.643160	0.420161	0.646855	0.402294
0	XG	0.714719	0.297536	0.011217	0.643160	0.420161	0.646855	0.418522
0	XG	0.714719	0.297536	0.011217	0.643160	0.420161	0.646855	0.398927
0	XG	0.714719	0.297536	0.011217	0.643160	0.420161	0.646855	0.40949

	model	precision_test	recall_test	FPR_test	AUROC_test	f1_test	f1_train	f1_CV	recall_CV
0	LGB_U_tuned	0.290438	0.821226	0.189499	0.898039	0.429114	0.512608	0.415793	[0.8016424510423247, 0.8294377763739734, 0.814...
0	LGB_U_tuned	0.290438	0.821226	0.189499	0.898039	0.429114	0.512608	0.43891	[0.8016424510423247, 0.8294377763739734, 0.814...
0	LGB_U_tuned	0.290438	0.821226	0.189499	0.898039	0.429114	0.512608	0.421586	[0.8016424510423247, 0.8294377763739734, 0.814...
0	LGB_U_tuned	0.290438	0.821226	0.189499	0.898039	0.429114	0.512608	0.418756	[0.8016424510423247, 0.8294377763739734, 0.814...

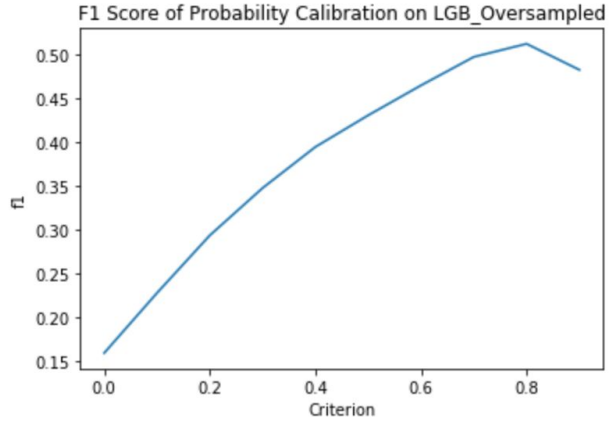




After lifting our threshold to 80%

```
Precision = 0.47833
Recall (TPR) = 0.55085
Fallout (FPR) = 0.05674
Roc_auc_score = 0.89970
f1 = 0.5120375807398708
```

```
Maximum f1-score is 0.5120375807398708 when criterion is
```



Hyperparameter Tuning

Parameters	What they do
num_leaves	Number of tree leaves
learning _rate	Learning rate for boosting
n_estimators	Number of trees or boosting rounds
max_depth	Max level of depth of trees

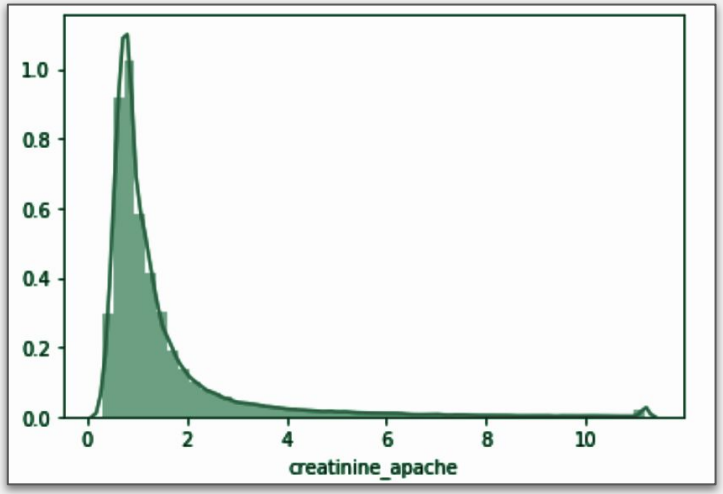
Results Recall

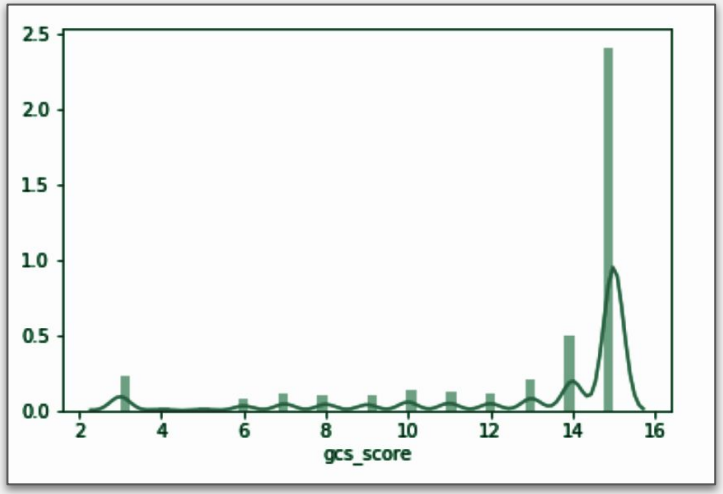
```
our model
0.8212255211623499
icu probability
0.17435249526216046
hospital probability
0.2874289324068225
```

Results F1 Scores

```
our model
0.4291137151345107
icu probability
0.27353815659068387
hospital probability
0.37775010377750096
```

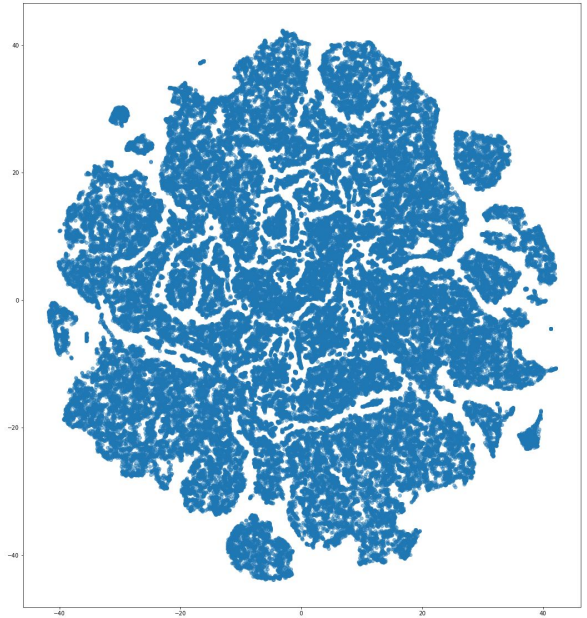
Exploratory Data Analysis





- high creatinine values result in death and low gsc scores result in death
- from the skewed distribution plots, we can also identify the imbalance in the dataset: the majority of patients are survived

Clustering Analysis



]:

DBSCAN_label	hospital_death	num_patients	median_age	median_bmi	top_elective	top_gender	top_ethnicity	top_apache_3j_diagnosis_desc
-1	0	52701	58.0	29.014391	0	M	Caucasian	Sepsis__other_than_urinary
	1	5610	69.0	27.270568	0	M	Caucasian	Sepsis__other_than_urinary
0	0	30921	70.0	26.880273	0	M	Caucasian	Sepsis__other_than_urinary
	1	2294	74.0	25.995912	0	M	Caucasian	Sepsis__other_than_urinary
1	0	50	55.0	29.743746	0	M	Caucasian	Acute_myocardial_infarction
2	0	43	87.0	23.215788	0	M	Caucasian	Congestive_heart_failure
	1	7	87.0	22.881253	0	F	Caucasian	Sepsis__other_than_urinary
3	0	37	51.0	25.234375	0	M	Caucasian	Other_neurologic_disease
	1	2	50.5	24.854656	0	F	African_American	Drug_overdose
4	0	46	74.0	31.712448	0	M	Caucasian	Rhythm_disturbance
	1	2	74.0	31.980525	0	M	Other_Unknown	Sepsis__other_than_urinary