Team 2

Ward Admission Classification

Connor Mignone, Amanda Konet, Yilin Ye,
Tonnar Castellano, Sarah Torrence

EXECUTIVE SUMMARY

Objectives

- Predict patient admission into hospital if test positive for COVID-19
- Avoid false negatives

Methodology

- Exploratory Data Analysis
- SMOTE and ROSE Oversampling
- KNN and Random Forests

Conclusions

- Blood Test results are most important for predicting admission
- “Accurately” predict true positives

Next Steps

- More Data
- Predict admission by ward
    - (ICU, semi-intensive, regular)
- Dashboard for medical professionals

OBJECTIVE

- Minimize the Externalities
- Save Hospital Resources
- Save Lives

DATA

- Only COVID-19 positive patients
- 558 observations and 63 variables
- Limitations: missing patient medical history and symptoms

ASSUMPTIONS

Covid Positives Only

Model can only be used to predict admission of covid positive patients.

Missing Tests

Imputed missing values based on assumption patients with no lab tests would have normal levels.

METHODS/APPROACH

Exploratory Data Analysis

Understanding lab tests, distributions of admitted patients

Data Cleaning

Handle missing data by removing columns or imputing values

Feature Engineering

Created features based on lab test variable groupings

Balancing Train Set

Oversample admitted patients, undersample discharged patients

Modeling

Random Forest 1
K Nearest Neighbors

Validation & Value

Estimate model results and value to the hospital

DATA CLEANING

Covid Positive

Filtered for only covid positive patients

Remove Same Values

Remove columns with no variation by patient

Combined Wards

Admitted vs. Discharged

Remove Missing Values

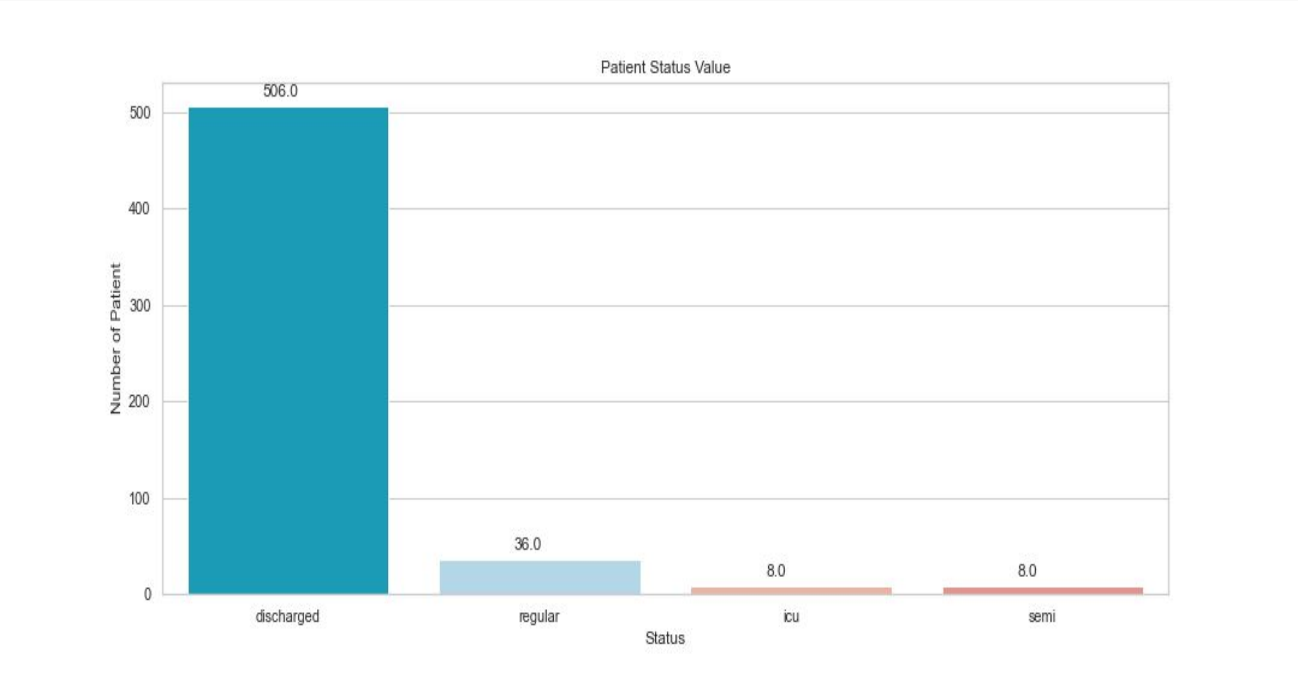
Remove columns with >= 98% missing values

Impute Missing Values

Randomly assign value from distribution N(0, 0.5)

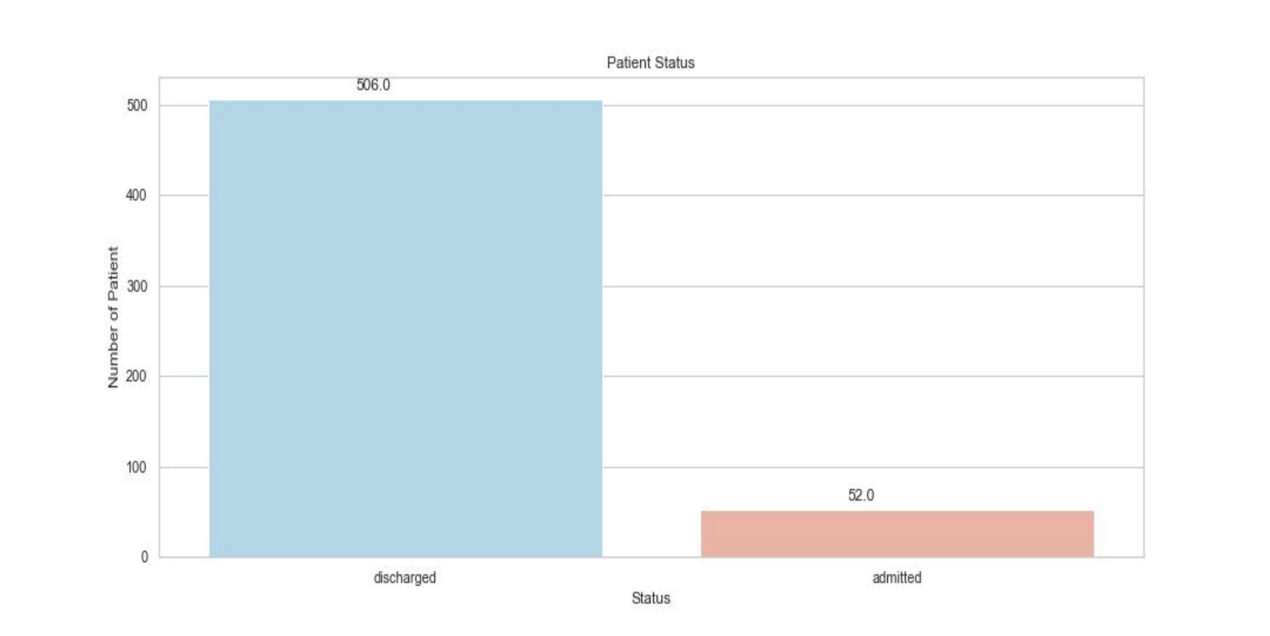
PATIENT STATUS

Before grouping



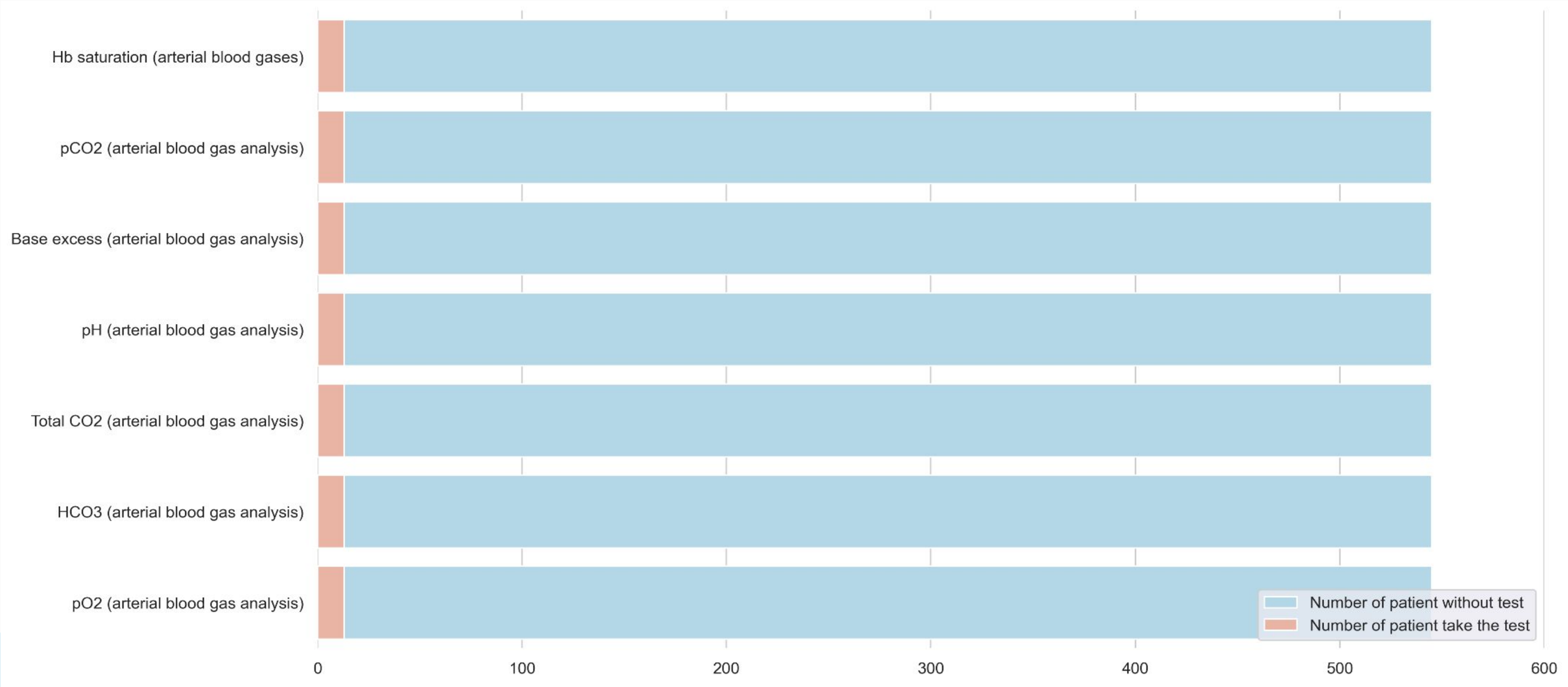
PATIENT STATUS

After grouping



GROUPED LAB TESTS

Ex. Blood gas grouping



FEATURE ENGINEERING

- Based on the ratio of missing value vs. non-missing value of each variable with respect to each patient, we grouped the variables with same ratio new feature groups
  - Binary: 1 indicates patient had test grouping done

- Lab groupings:
  - Urine test group
  - Virus test group: 1) whether virus tests given, 2) number of viruses detected
  - Blood test group
  - Blood gasses group: 1) arterial, 2) venous, 3) arterial & venous)
  - Bilirubin test group
  - Sodium & Potassium test group

MODELING

K-nearest Neighbors

Random Forest

Combined over-and under-sampling using SMOTE, ROSE

Cross Validation

Grid Tuning

TEST RESULTS

ROSE Sampling ($p_{admitted}$ = 0.3) with 5-fold Cross Validation

K-Nearest Neighbors

Admission	True Discharged	True Admitted
Predicted Discharge	147	**5**
Predicted Admitted (k = 6)	6	**10**

Random Forest 1

Admission	True Discharged	True Admitted
Predicted Discharge	145	**2**
Predicted Admitted	8	**13**

(mtry = 8, trees = 10, min n = 5)

N_{test} = 153

TEST RESULTS

K-Nearest Neighbors

Test Metric	Scores
Accuracy	93.5%
Sensitivity / Recall (TPR)	66.7%
Specificity (TNR)	96.0%
Precision	0.625
F1	0.645

(k = 6)

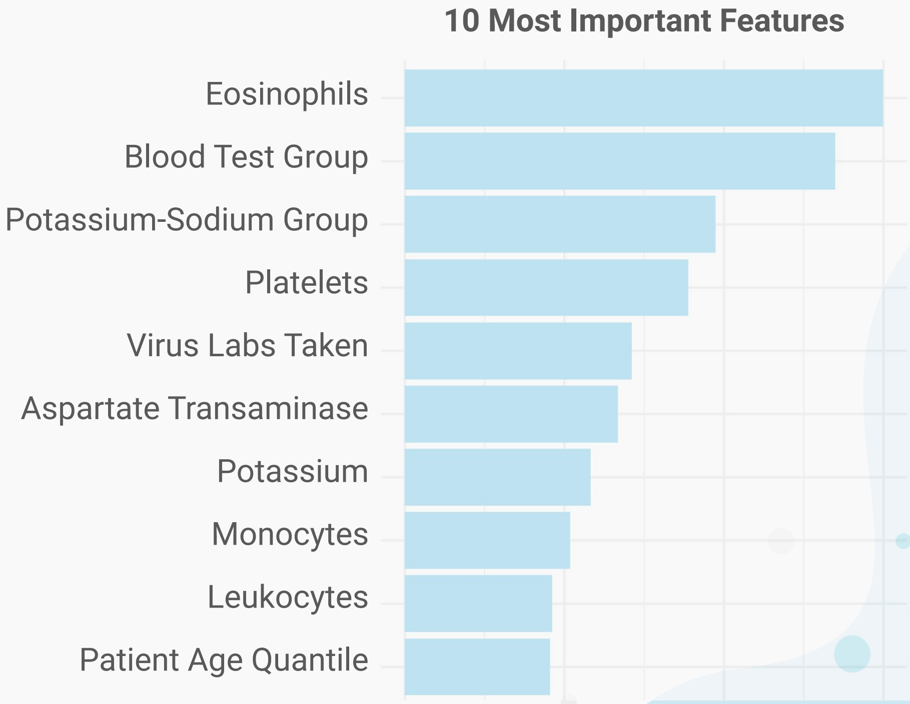
Random Forest

Test Metric	Scores
Accuracy	94.0%
Sensitivity / Recall (TPR)	**86.7%**
Specificity (TNR)	94.8%
Precision	0.612
F1	0.722

(mtry = 8, trees = 10, min n = 5)

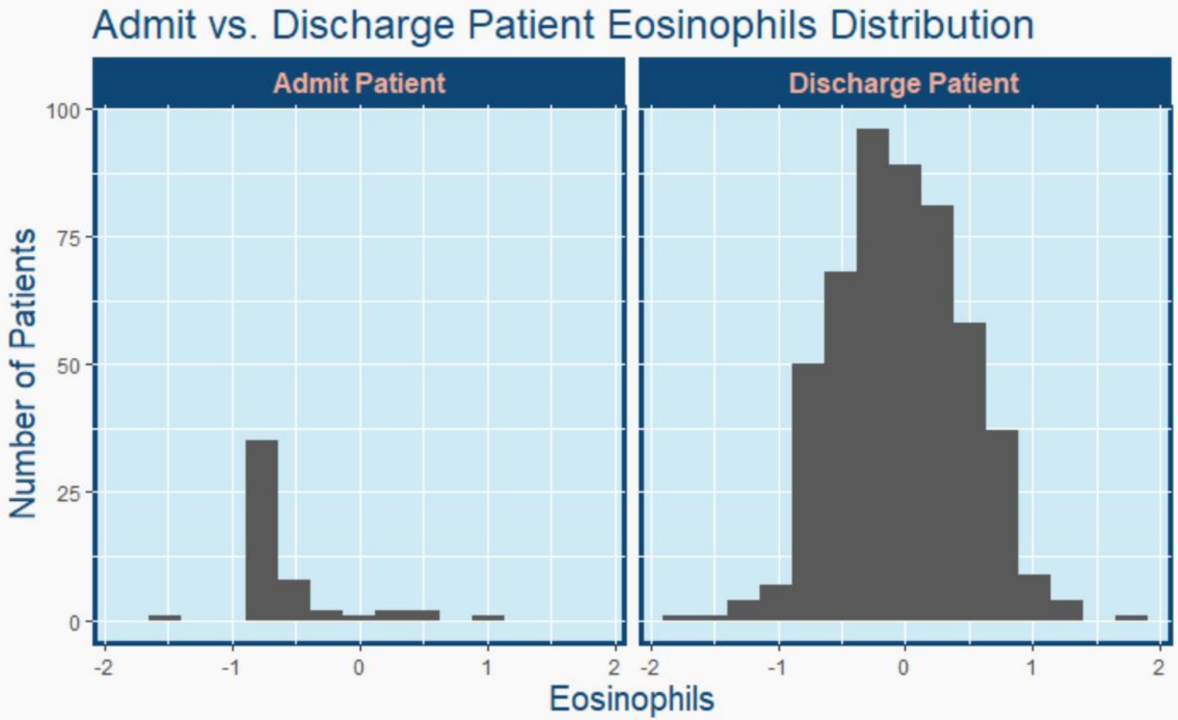
FEATURE IMPORTANCE

- Many important features related to blood tests:
    - Eosinophils
    - Whether blood tests were taken
    - Platelets
    - Monocytes
    - Leukocytes
- Other important features:
    - Whether sodium and potassium tests given
    - Whether other virus labs given
    - Patient age quantile



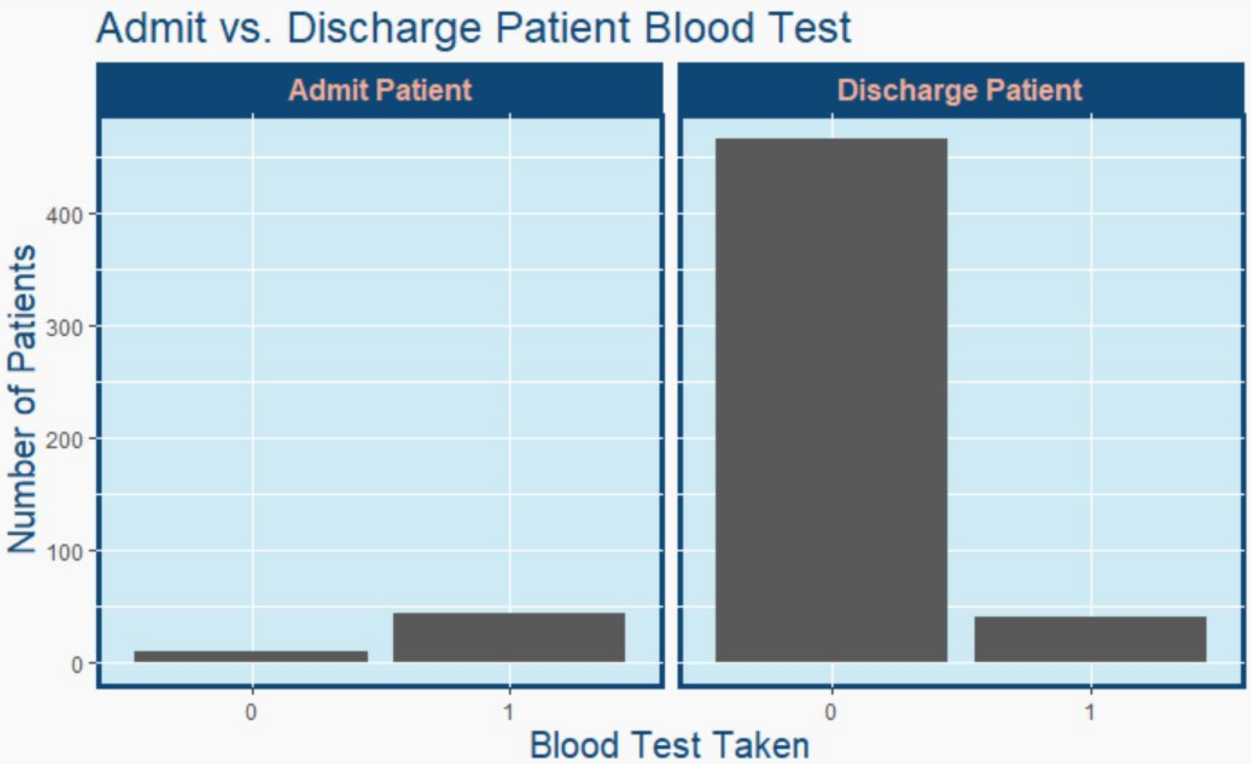
FEATURE IMPORTANCE

Low eosinophil (disease-fighting white blood cells) levels associated with mortality from COVID-19[1]



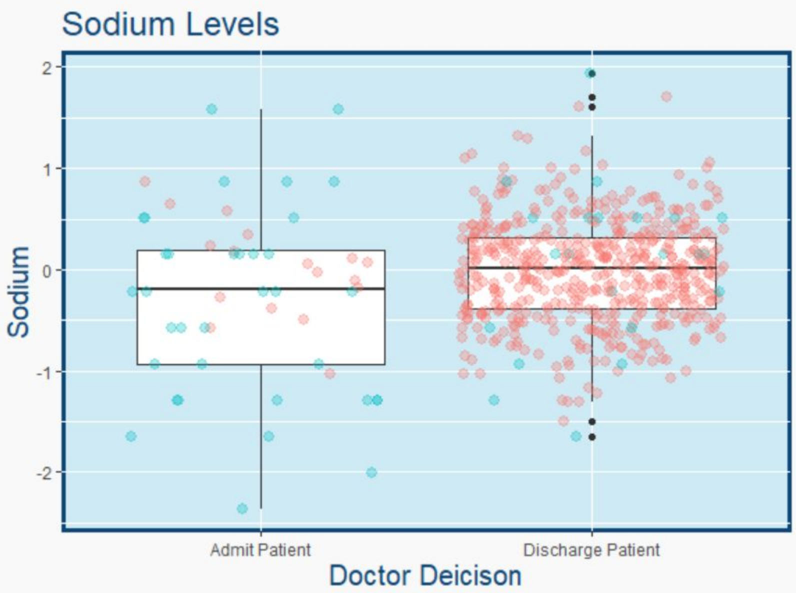
FEATURE IMPORTANCE

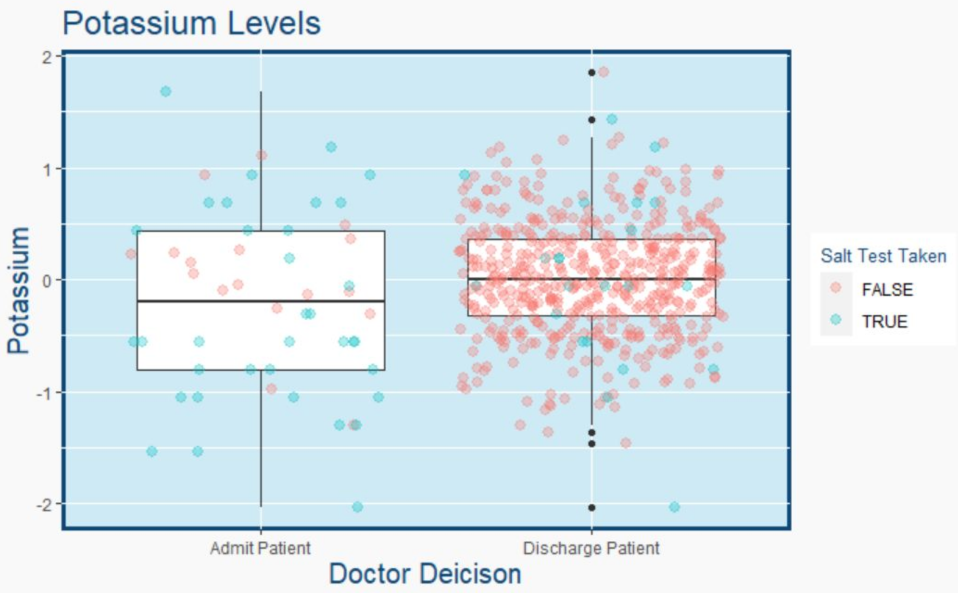
Most admitted patients had blood tests done



FEATURE IMPORTANCE

Low electrolyte levels associated with severe cases of COVID-19[2]





MODEL VALUE ADMITTANCE ASSUMPTIONS

- Our Model Admittance Rate: 13.69%
- Peak Hospitalization Rate: 15.2%
- Average Hospital Admittance Cost: $25,175.50
- Average Death Cost: $26,000
- Current Weekly Hospitalizations: 8,321
- Current Hospitalization Death Rate: 15%

VALUE - SURFACE LEVEL

Current Cost of Weekly Admissions:

8321 x 15.2% x $26,000 = $32,884,592.00

Current Weekly Cost of Model Admission:

8321x 13.69% x (($26000)+($26000) x 15%) = $29,618,797.62

Current Weekly Savings:

$32,884,592.00 - $29,618,797.62 = **$3,265,794.38**

VALUE SENSITIVITY ANALYSIS

Prediction Power	Model Admission Cost	Current Admission Cost	Savings
100%	$29,618,797.62	$32,884,592.00	$3,265,794.38
90%	$32,580,677.38	$32,884,592.00	$1,881,150.26
50%	$44,428,196.43	$32,884,592.00	$(11,543,604.43)

MODEL VALUE READMISSION ASSUMPTIONS

- The current COVID-19 readmission rate: 9%
- Our model false discharge rate: 1.1%
- Our model false admittance rate: 4.76%
- Our model true admittance rate: 7.7%
- Current Weekly Hospitalizations: 8321
- Current Weekly Readmissions: 8321 x .09 = 745

VALUE - MORE VARIABLES

Current Weekly Cost:

(8321 x 15.2% x $26,000)+(9% x 8321 x 26000) = $51,309,116.62

Current Weekly Cost of Model False Discharge Error:

((8321 x 1.1% x (($26000)+($26000) x 15%)) + (8321 x 9% x $26000)) x **150%** = $33,649,529.64

Current Weekly Cost of Model False Admit Error:

8321x 4.76% x $26000 = $10,302,190.48

Current Weekly Cost of Model True Admit:

(7.7% x 8321 x (26000 + (26000 x 15%)) + ($26000 x 9% x 8321) = $38,628,578.30

Current Weekly Savings:

$51,309,116.62 - ($38,628,578.30 + $33,649,529.64 + $10,302,190.48) = **-$31,271,181.80**

VALUE LIMITATIONS / SENSITIVITY ANALYSIS

A Change of Assumptions:

- If we assume our model lowers the admittance rate to around 3%, we break even in costs.

Model Readmittance Rate	Model Cost	Current Cost	Savings
3.218302%	$51,309,116.62	$51,309,116.62	$0
9% (Current)	$32,735,210.24	$51,309,116.62	$(31,271,181.80)
1%	$39,311,098.42	$51,309,116.62	$11,998,018.20

LIMITATIONS

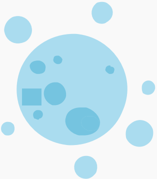
Data Size

Grouped Classes

Patient medical history and symptoms

NEXT STEPS

- Predict on each ward (ICU, semi-intensive, regular)
- Extend to all patients, not just covid positive
- Create dashboard to assist hospital staff in admitting patients
- Collect readmission data



Questions?

Appendix

MODELING DATASET

- Train/Test split
- Combined over-sampling majority and under-sampling minority
  - ROSE (balance such that minority proportion is one of: p = 0.3, 0.4, 0.5)
  - SMOTE (downsample majority 80%, upsample minority 50%)
- Cross-validation
  - 3-fold
  - 5-fold

MODELING

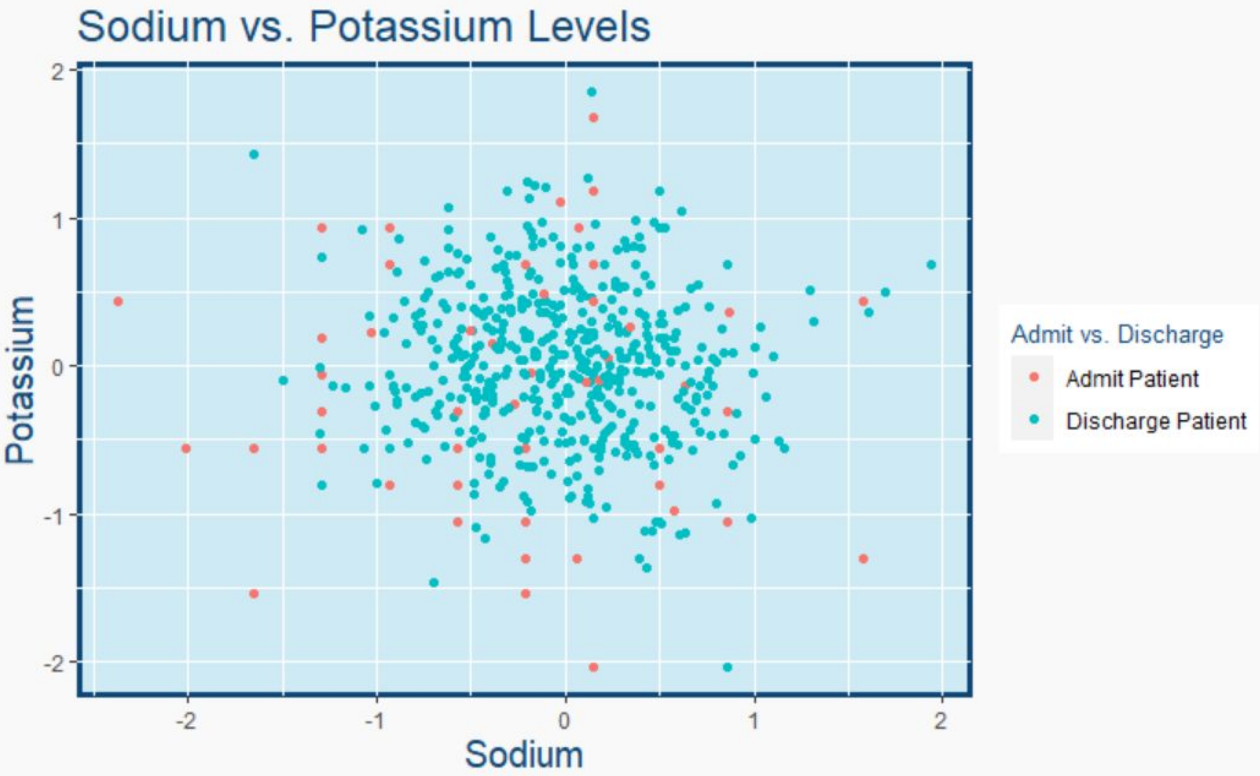
K-nearest Neighbors

Parameter	Values
k	[2, 10] and every 5th value [20, 80]

Random Forest

Parameter	Values
mtry	[2, 15]
Min n	[2, 15]
Trees	[10, 200]

FEATURE IMPORTANCE



CITATIONS

1. Yan et al. (2021) https://www.sciencedirect.com/science/article/pii/S1939455121000156
2. Lippi (2020) https://pubmed.ncbi.nlm.nih.gov/32266828/
3. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html
4. https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2777028
5. https://jamanetwork.com/journals/jama/fullarticle/2774420
6. https://www.healthleadersmedia.com/clinical-care/dying-hospital-costs-more-surviving-inpatient-stay

THANKS!

facebook.com/Freepik

@Freepik_Vectors

company/freepik-company

Please keep this slide for attribution