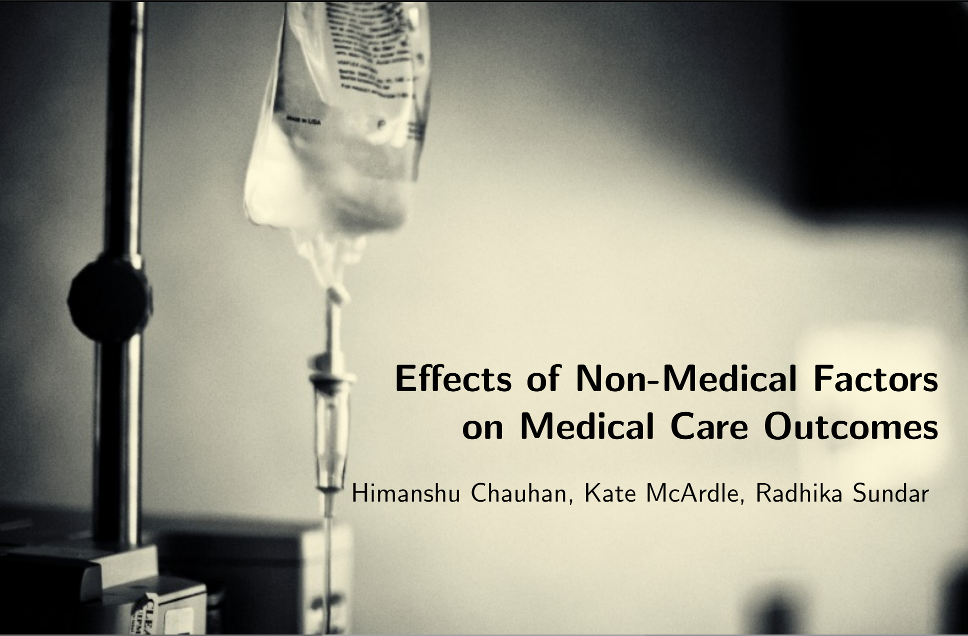


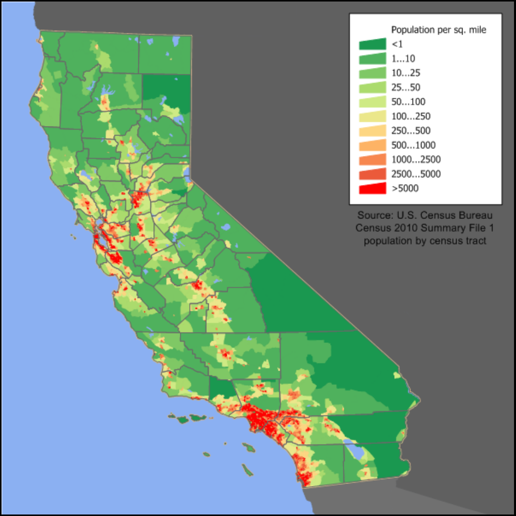
# Effects of Non-Medical Factors on Medical Care Outcomes

Himanshu Chauhan, Kate McArdle, Radhika Sundar

# Outline


1. Intro & Background
   - Motivation
   - Data
   - Other Studies
2. Problem Defn. & Approach
   - Objectives
   - Focus
   - Our Approach
3. Results
   - Insights/Inferences
   - Predictive Modeling Results - Heart Failure
   - Predictive Modeling Results - Cardiac Arrhythmia
4. Conclusion

# Motivation


Population per sq. mile
<1
1...10
10...25
25...50
50...100
100...250
250...500
500...1000
1000...2500
2500...5000
>5000
Source: U.S. Census Bureau
Census 2010 Summary File 1
population by census tract

# Data Source - CA Discharge Records (2009 – 2011)

| hospital | diagnosis | gender | race | county | age | source | pay type | los | charge |
|---|---|---|---|---|---|---|---|---|---|
| 130699 | 389 | 2 | 1 | 19 | | 231 | 3 | 0 | 18032 |
| 196404 | 250 | 2 | * | 38 | 42 | 231 | 1 | 44 | 67543 |
| 190017 | 428 | * | * | 43 | | 131 | 1 | 289 | 0 |

- Demographics
  - Age, Gender, Race, County, Insurance type, etc.

# Data Source - CA Discharge Records (2009 – 2011)

| hospital | diagnosis | gender | race | county | age | source | pay type | los | charge |
|---|---|---|---|---|---|---|---|---|---|
| 130699 | 389 | 2 | 1 | 19 | | 231 | 3 | 0 | 18032 |
| 196404 | 250 | 2 | * | 38 | 42 | 231 | 1 | 44 | 67543 |
| 190017 | 428 | * | * | 43 | | 131 | 1 | 289 | 0 |

- Demographics
  - Age, Gender, Race, County, Insurance type, etc.
- Diagnosis/Procedures performed
  - Primary diagnosis, Prinicipal procedure, Other diagnoses, etc.

# Data Source - CA Discharge Records (2009 – 2011)

| hospital | diagnosis | gender | race | county | age | source | pay type | los | charge |
|---|---|---|---|---|---|---|---|---|---|
| 130699 | 389 | 2 | 1 | 19 | | 231 | 3 | 0 | 18032 |
| 196404 | 250 | 2 | * | 38 | 42 | 231 | 1 | 44 | 67543 |
| 190017 | 428 | * | * | 43 | | 131 | 1 | 289 | 0 |

- Demographics
  - Age, Gender, Race, County, Insurance type, etc.
- Diagnosis/Procedures performed
  - Primary diagnosis, Prinicipal procedure, Other diagnoses, etc.
- Outcomes
  - Disposition, Length of Stay, Total Charges, etc.

# Data - Challenges

| hospital | diagnosis | gender | race | county | age | source | pay type | los | charge |
|---|---|---|---|---|---|---|---|---|---|
| 130699 | 389 | 2 | 1 | 19 | | 231 | 3 | 0 | 18032 |
| 196404 | 250 | 2 | * | 38 | 42 | 231 | 1 | 44 | 67543 |
| 190017 | 428 | * | * | 43 | | 131 | 1 | 289 | 0 |

- Masked Records

# Data - Challenges

| hospital | diagnosis | gender | race | county | age | source | pay type | los | charge |
|---|---|---|---|---|---|---|---|---|---|
| 130699 | 389 | 2 | 1 | 19 | | 231 | 3 | 0 | 18032 |
| 196404 | 250 | 2 | * | 38 | 42 | 231 | 1 | 44 | 67543 |
| 190017 | 428 | * | * | 43 | | 131 | 1 | 289 | 0 |

- Missing Entries

# Data - Challenges

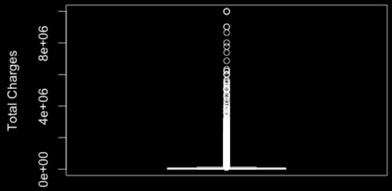
| hospital | diagnosis | gender | race | county | age | source | pay type | los | charge |
|---|---|---|---|---|---|---|---|---|---|
| 130699 | 389 | 2 | 1 | 19 | | 231 | 3 | 0 | 18032 |
| 196404 | 250 | 2 | $*$ | 38 | 42 | 231 | 1 | 44 | 67543 |
| 190017 | 428 | $*$ | $*$ | 43 | | 131 | 1 | 289 | 0 |

- Highly Categorical: $>$ 400 Hospitals, $\approx$ 13000 diagnosis codes

# Data - Challenges

| hospital | diagnosis | gender | race | county | age | source | pay type | los | charge |
|---|---|---|---|---|---|---|---|---|---|
| 130699 | 389 | 2 | 1 | 19 | | 231 | 3 | 0 | 18032 |
| 196404 | 250 | 2 | * | 38 | 42 | 231 | 1 | 44 | 67543 |
| 190017 | 428 | * | * | 43 | | 131 | 1 | 289 | 0 |

- Large variance in Charges/Length of Stay



# Related Work

- Acute Burn Patients [Carbonell et. al. (2005)]
  - Mortality and Length of Stay
- Bariatric Surgery Outcomes [Peters et. al. (1996)]
  - Surgery results/recovery times
- In-Hospital Mortality for Severe Sepsis [Banta et. al. (2012)]

# Outline

1 Intro & Background
- Motivation
- Data
- Other Studies

2 Problem Defn. & Approach
- Objectives
- Focus
- Our Approach

3 Results
- Insights/Inferences
- Predictive Modeling Results - Heart Failure
- Predictive Modeling Results - Cardiac Arrhythmia

4 Conclusion

# Records Dictated by Physicians

*Discharge status: Alive, but without permission.*

*The patient has no past history of suicides.*

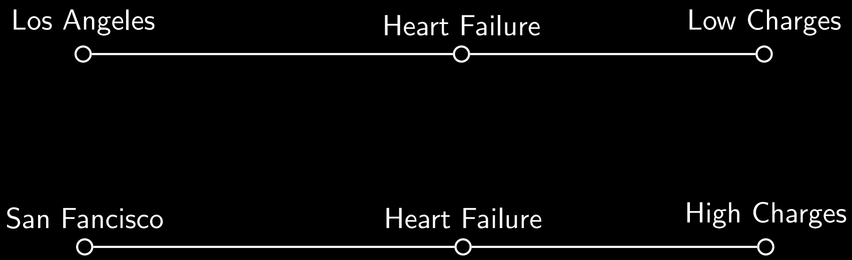
*By the time he was admitted, his rapid heart had stopped and he was feeling better.*

# Problem 1: Inference/Insight Mining

- Three Counties: Los Angeles, San Francisco, Santa Clara

# Problem 1: Inference/Insight Mining

- Three Counties: Los Angeles, San Francisco, Santa Clara

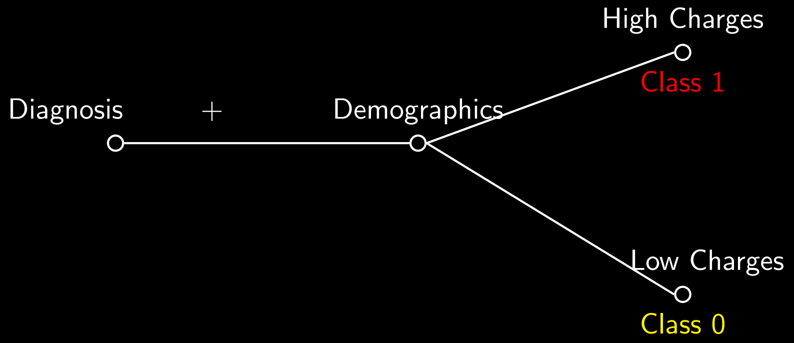


# Problem 2 - Predictions

- Predictions using demographic details
    - Charges
    - Length of Stay (LOS)

# Problem 2 - Predictions

- Predictions using demographic details
  - Charges
  - Length of Stay (LOS)



# Diseases Studied

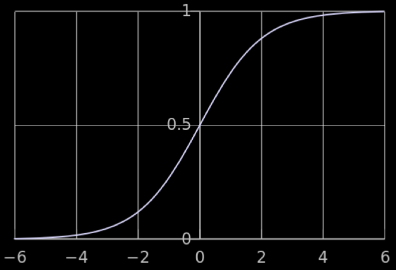
| Life Threatening | Non-Life Threatening |
|---|---|
| Heart Disease | Hypertension |
| Cancer | Diabetes |
| Heart Failure | Cardiac Arrhythmia |
| | Mental Illness |

# Diseases Studied

| Life Threatening | Non-Life Threatening |
|---|---|
| Heart Disease | Hypertension |
| Cancer | Diabetes |
| Heart Failure | Cardiac Arrhythmia |
| | Mental Illness |

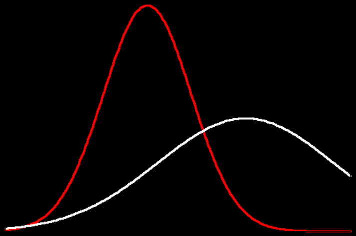
# Methods Used

- Logistic Regression



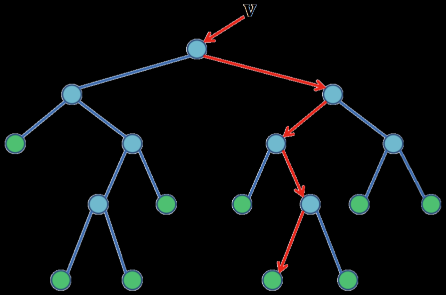
# Methods Used

- Naïve Bayes



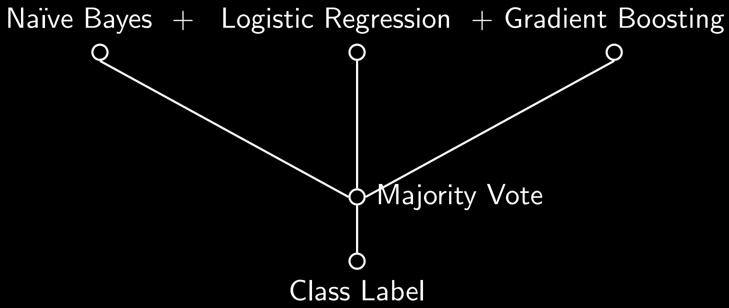
# Methods Used

- Generalized Boosting (gbm)



# Methods Used

- Ensemble



# Outline


1. Intro & Background
   - Motivation
   - Data
   - Other Studies
2. Problem Defn. & Approach
   - Objectives
   - Focus
   - Our Approach
3. Results
   - Insights/Inferences
   - Predictive Modeling Results - Heart Failure
   - Predictive Modeling Results - Cardiac Arrhythmia
4. Conclusion

# Results - Insights

- Heart Failure Charges across Hospitals:

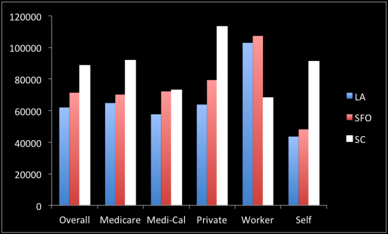
| Hospital | Mean value of Charges ($) |
|---|---:|
| Overall mean | 62,329 |
| Cedars Sinai - LA | 151,201 |
| Centinela Hospital - LA | 56,682 |
| St Agnes - Bay Area | 41,259 |

# Results - Insights

- Heart Failures Per Capita

| Los Angeles | Santa Clara | San Francisco |
|---|---|---|
| 0.0074 | 0.0048 | 0.0070 |

- Heart Failure - Insurace Payments



# Heart Failure Charge Predictions [Accuracy]

| Model | % Accuracy | AUROC |
|---|---|---|
| Baseline | 73.4 | - |
| Demographics only | 73 | 0.65 |
| Naïve Bayes | 79.5 | 0.79 |
| Logistic Regression | 79.7 | 0.79 |
| Gradient Boosting | 81.2 | 0.83 |
| Ensemble | 80.6 | 0.80 |

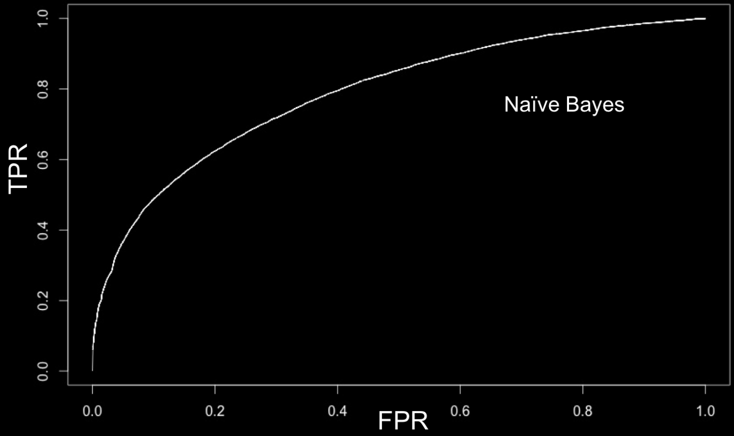
† AUROC : Area under ROC

# Heart Failure Charge Predictions [Accuracy]

| Model | % Accuracy | AUROC |
|---|---|---|
| Baseline | 73.4 | - |
| Demographics only | 73 | 0.65 |
| Naïve Bayes | 79.5 | 0.79 |
| Logistic Regression | 79.7 | 0.79 |
| Gradient Boosting | 81.2 | 0.83 |
| Ensemble | 80.6 | 0.80 |

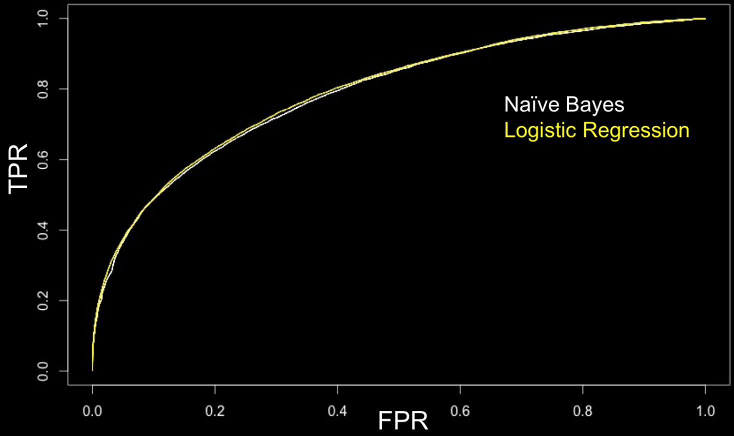
# Heart Failure Charge Predictions [ROC Plots]

Class 0: $\leq$ Mean Class 1: $>$ Mean



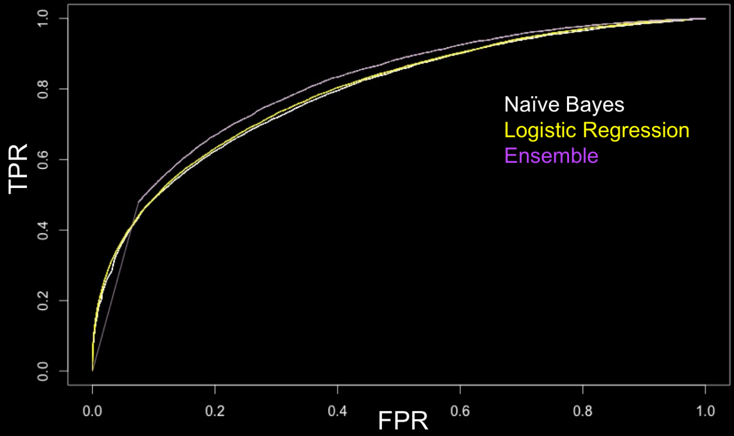
# Heart Failure Charge Predictions [ROC Plots]

Class 0: $\leq$ Mean Class 1: $>$ Mean



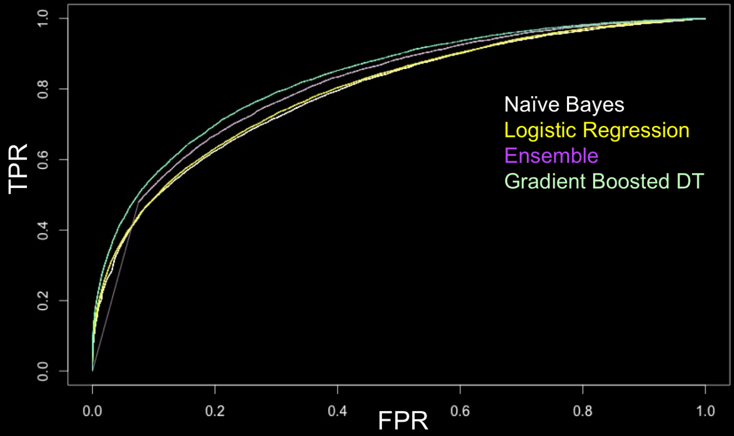
# Heart Failure Charge Predictions [ROC Plots]

Class 0: $\leq$ Mean      Class 1: $>$ Mean



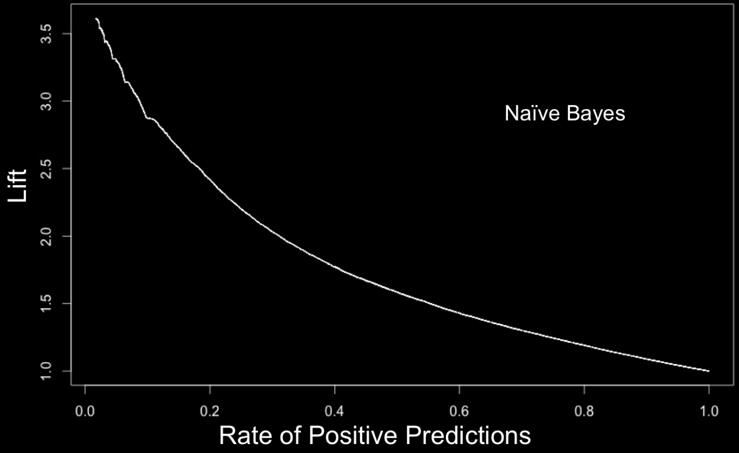
# Heart Failure Charge Predictions [ROC Plots]

Class 0: $\leq$ Mean Class 1: $>$ Mean



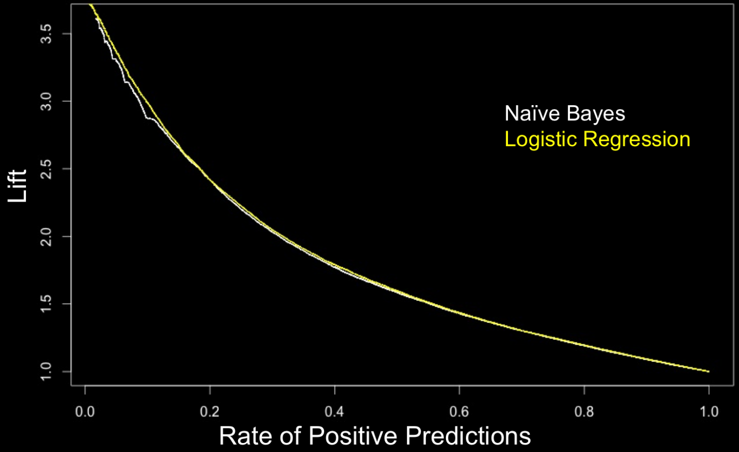
# Heart Failure Charge Predictions [Lift Charts]

Class 0: $\leq$ Mean Class 1: $>$ Mean



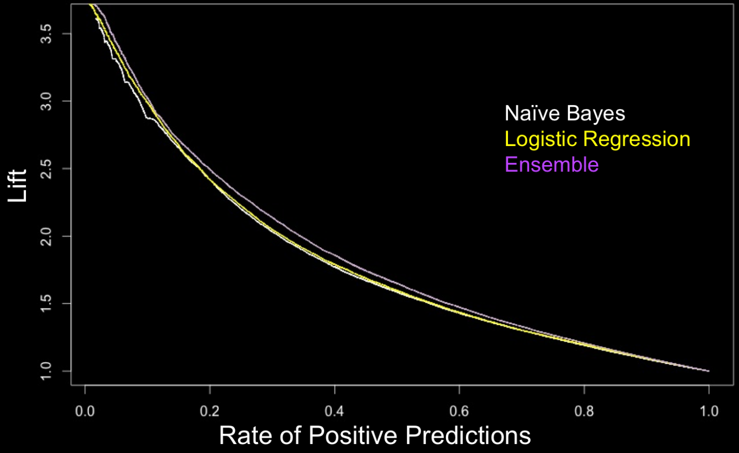
# Heart Failure Charge Predictions [Lift Charts]

Class 0: $\leq$ Mean Class 1: $>$ Mean



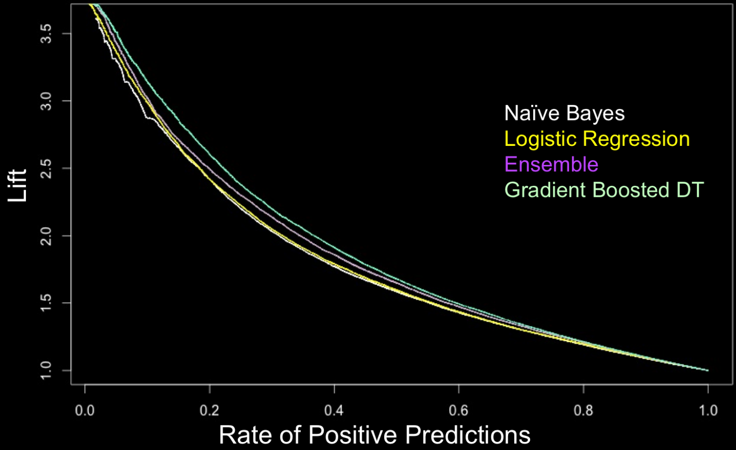
# Heart Failure Charge Predictions [Lift Charts]

Class 0: $\leq$ Mean Class 1: $>$ Mean



# Heart Failure Charge Predictions [Lift Charts]

Class 0: $\leq$ Mean Class 1: $>$ Mean



# Cardiac Arrhythmia Charge Predictions [Accuracy]

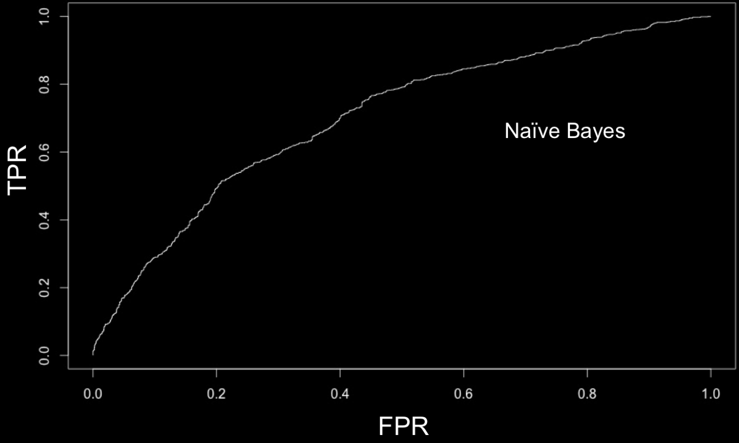
| Model | % Accuracy | AUROC |
|---|---|---|
| Baseline | 59.8 | - |
| Naïve Bayes | 65.5 | 0.70 |
| Logistic Regression | 67.9 | 0.69 |
| Gradient Boosting | 79.5 | 0.87 |
| Ensemble | 76 | 0.80 |

# Cardiac Arrhythmia Charge Predictions [Accuracy]

| Model | % Accuracy | AUROC |
|---|---|---|
| Baseline | 59.8 | - |
| Naïve Bayes | 65.5 | 0.70 |
| Logistic Regression | 67.9 | 0.69 |
| Gradient Boosting | 79.5 | 0.87 |
| Ensemble | 76 | 0.80 |

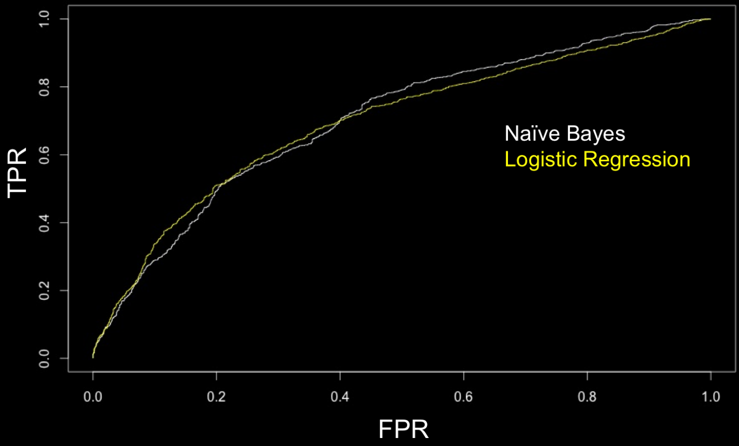
# Cardiac Arrhythmia Charge Predictions [ROC Plots]

Class 0: $\leq$ Mean Class 1: $>$ Mean



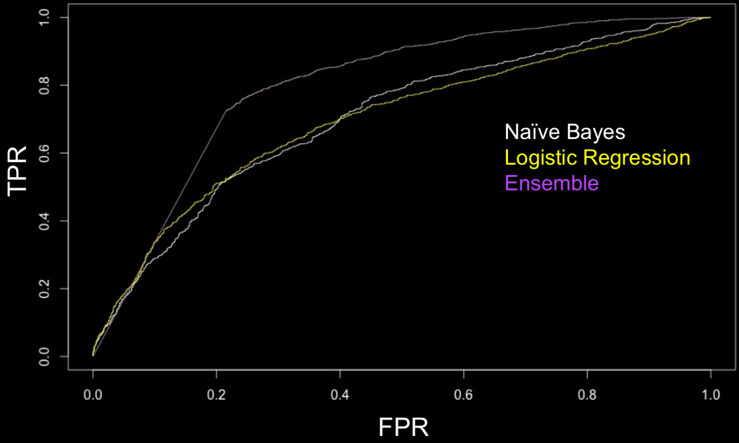
# Cardiac Arrhythmia Charge Predictions [ROC Plots]

Class 0: $\leq$ Mean    Class 1: $>$ Mean



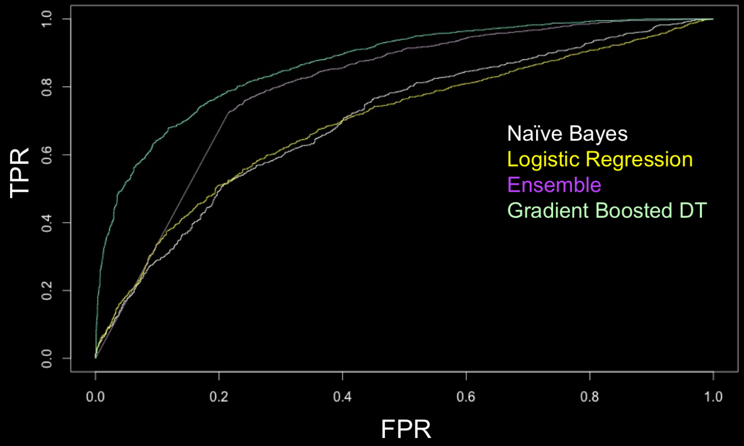
# Cardiac Arrhythmia Charge Predictions [ROC Plots]

Class 0: $\leq$ Mean    Class 1: $>$ Mean



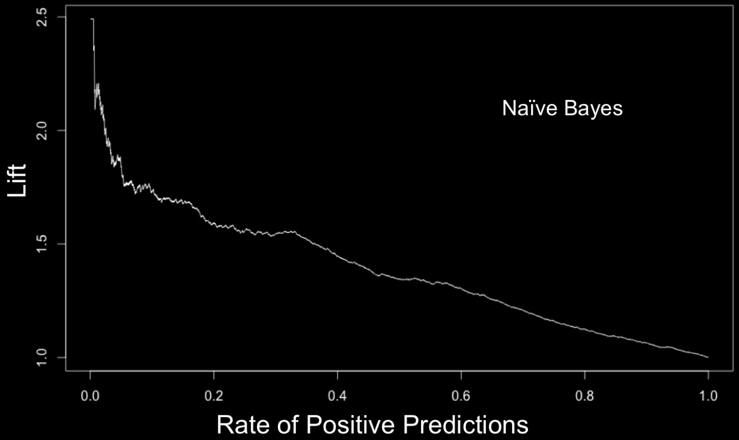
# Cardiac Arrhythmia Charge Predictions [ROC Plots]

Class 0: $\leq$ Mean Class 1: $>$ Mean



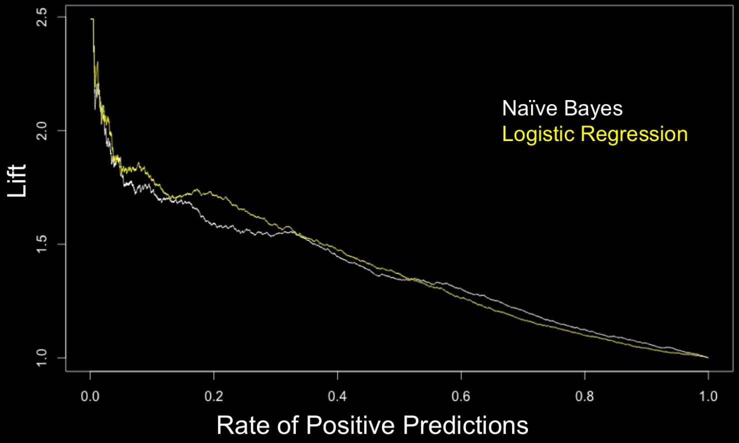
# Cardiac Arrhythmia Charge Predictions [Lift Charts]

Class 0: $\leq$ Mean    Class 1: $>$ Mean



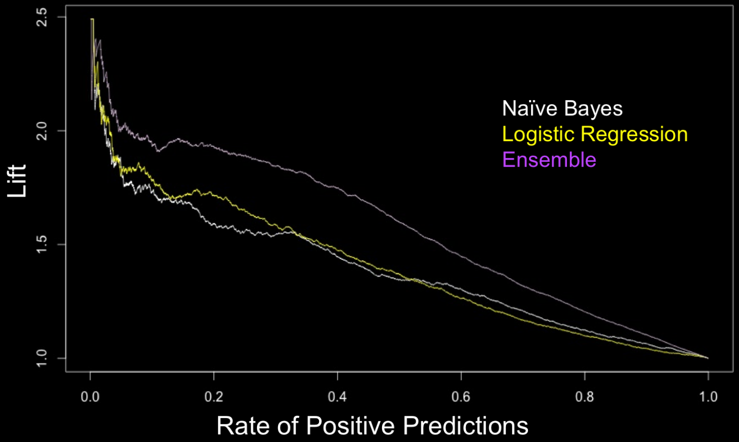
# Cardiac Arrhythmia Charge Predictions [Lift Charts]

Class 0: $\leq$ Mean Class 1: $>$ Mean



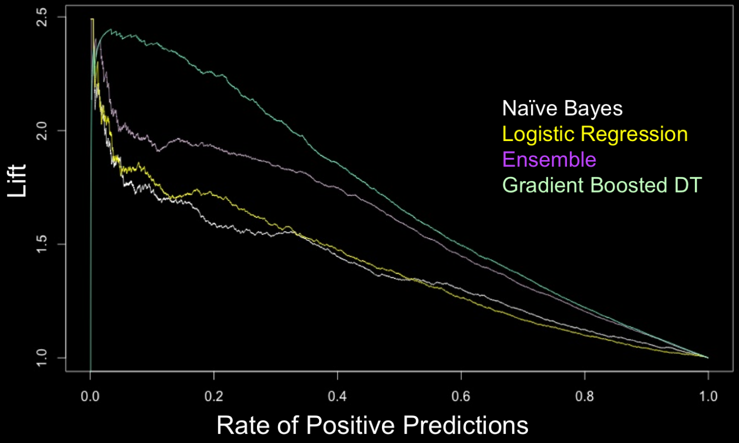
# Cardiac Arrhythmia Charge Predictions [Lift Charts]

Class 0: $\leq$ Mean Class 1: $>$ Mean



# Cardiac Arrhythmia Charge Predictions [Lift Charts]

Class 0: $\leq$ Mean Class 1: $>$ Mean



# Outline


1. Intro & Background
   - Motivation
   - Data
   - Other Studies
2. Problem Defn. & Approach
   - Objectives
   - Focus
   - Our Approach
3. Results
   - Insights/Inferences
   - Predictive Modeling Results - Heart Failure
   - Predictive Modeling Results - Cardiac Arrhythmia
4. Conclusion

## Conclusions

- Significant differences
  - Across locations
  - Among insurance types

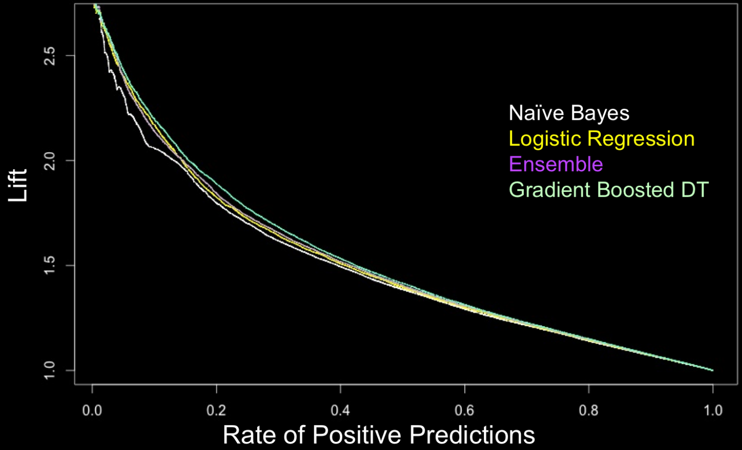
- Predictions on Medical Outcomes
  - Hard problem without considering all medical details

## Questions?

Thanks!

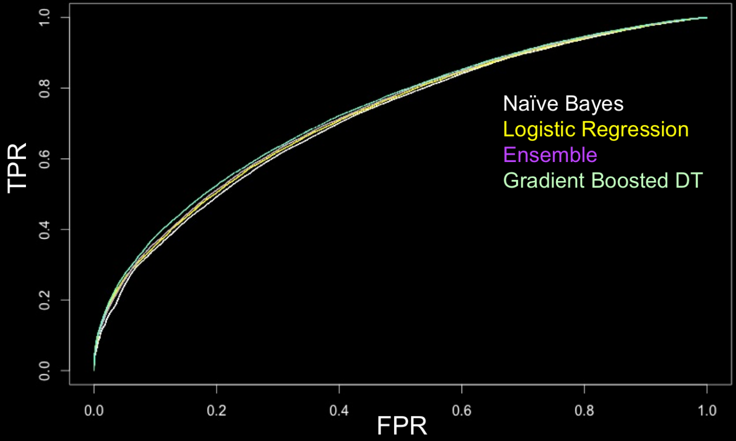
# Backup

## Heart Failure - LOS



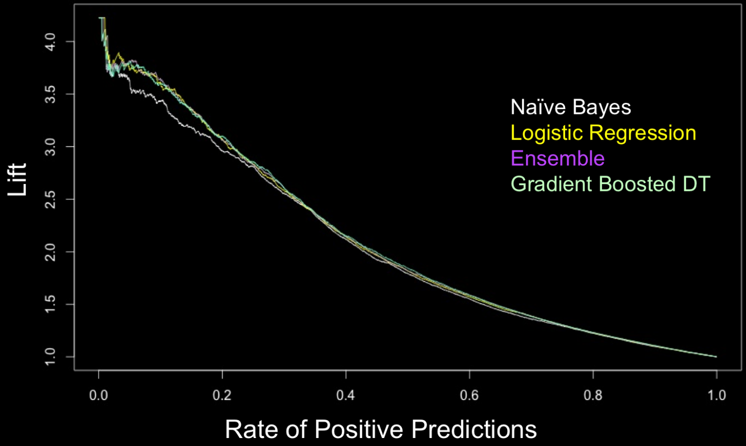
# Backup

## Heart Failure - LOS



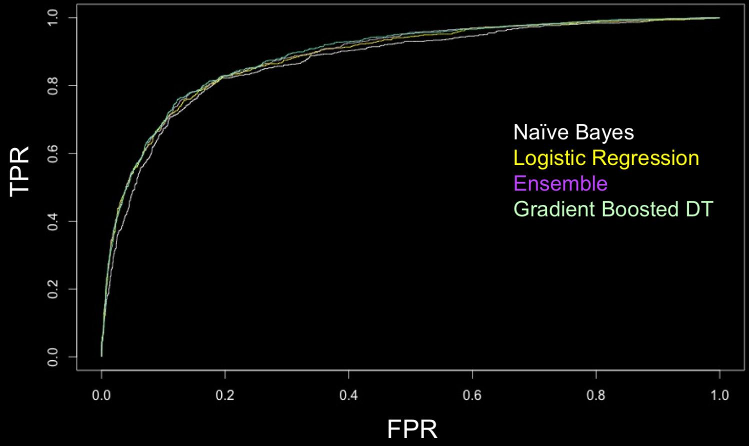
# Backup

## Cardiac Arrhythmia - LOS



# Backup

## Cardiac Arrhythmia - LOS



# Backup

Diabetes

| Model | Accuracy | AUROC |
|---|---|---|
| Baseline | 0.679 | - |
| Naïve Bayes | 0.770 | 0.792 |
| Logistic Regression | 0.779 | 0.809 |
| GBDT | 0.787 | 0.818 |

# Backup

Diabetes

| Model | Accuracy | AUROC |
|---|---|---|
| Baseline | 0.716 | - |
| Naïve Bayes | 0.809 | 0.819 |
| Logistic Regression | 0.812 | 0.836 |
| GBDT | 0.828 | 0.863 |

# Backup

Cardiac Arrhythmia LOS

| Model | Accuracy | AUROC |
|---|---|---|
| Baseline | 0.763 | - |
| Naïve Bayes | 0.846 | 0.871 |
| Logistic Regression | 0.853 | 0.884 |
| GBDT | 0.857 | 0.890 |
| Ensemble - Ave Prob | 0.853 | 0.888 |

# Backup

Hypertension

| Model | Accuracy | AUROC |
|---|---|---|
| Baseline | 0.771 | - |
| Naïve Bayes | 0.807 | 0.795 |
| Logistic Regression | 0.814 | 0.807 |
| GBDT | 0.820 | 0.816 |

# Backup

Hypertension

| Model | Accuracy | AUROC |
|---|---|---|
| Baseline | 0.731 | - |
| Naïve Bayes | 0.805 | 0.821 |
| Logistic Regression | 0.810 | 0.835 |
| GBDT | 0.832 | 0.865 |

# Backup

Prostate Cancer

| Model | Accuracy | AUROC |
|---|---|---|
| Baseline | 0.707 | - |
| Naïve Bayes | 0.765 | 0.749 |
| Logistic Regression | 0.775 | 0.773 |
| GBDT | 0.782 | 0.791 |

# Backup

Prostate Cancer

| Model | Accuracy | AUROC |
|---|---|---|
| Baseline | 0.659 | - |
| Naïve Bayes | 0.751 | 0.803 |
| Logistic Regression | 0.759 | 0.822 |
| GBDT | 0.792 | 0.855 |

# Backup

Schizophrenia

| Model | Accuracy | AUROC |
|---|---|---|
| Baseline | 0.707 | - |
| Naïve Bayes | 0.690 | 0.608 |
| Logistic Regression | 0.713 | 0.630 |
| GBDT | 0.721 | 0.696 |

# Backup

Schizophrenia

| Model | Accuracy | AUROC |
|---|---|---|
| Baseline | 0.712 | - |
| Naïve Bayes | 0.712 | 0.644 |
| Logistic Regression | 0.729 | 0.668 |
| GBDT | 0.758 | 0.776 |