A Model to Flag Patients Likely to Have/Acquire a MRSA Infection

JoAnn Alvarez

July 26, 2018

Outline

Project Goal

Task: Flag patients who are likely to have/acquire MRSA

Project Goal

Task: Flag patients who are likely to have/acquire MRSA

Use case: Implement a realtime alert through the EHR.

Outline

What is MRSA?

Methicillin-resistant *Staphylococcus aureus*

Exists because of bacteria's evolution to survive different antibiotics

What is MRSA?

Methicillin-resistant *Staphylococcus aureus*

Exists because of bacteria's evolution to survive different antibiotics

80,000 life-threatening cases per year in US.

What is MRSA?

Many people have staph on skin or in nose.

What is MRSA?

Many people have staph on skin or in nose.
Hospital- or community- acquired
Spreads betweeen patients and providers.

Human Costs of MRSA

- If infection localized to skin or soft tissue, can be treated more easily
- Can spread to blood and internal organs
- Patients die

Financial Costs

Costs hospitals lots of money

Financial Costs

Costs hospitals lots of money

Hospital acquired:

- Liability
- Lose insurance reimbursements
- Impacts hospital quality ratings
- Hospitals are required to report to CMS and to the government in some states
- CMS reduces payment if the infection is preventable or if the hospital has high MRSA rates.

Financial Costs

Costs hospitals lots of money

Hospital acquired:

- Liability
- Lose insurance reimbursements
- Impacts hospital quality ratings
- Hospitals are required to report to CMS and to the government in some states
- CMS reduces payment if the infection is preventable or if the hospital has high MRSA rates.

Community acquired:

- Can save money with faster diagnosis and treatment

MRSA protocols

Hospitals develop protocols

- Screen to identify carriers.
- Decolonization

Prevent spread

- handwashing
- keep wounds clean and covered
- isolate patients

Treatment

Urgent
Oral or IV antibiotics
There are different tiers of abx that strains of MRSA are successively
resistant to.
vancomycin, teicoplanin, daptomycin, linezolid

Background
Data
Machine Learning Algorithm

Data Acquisition
Feature Engineering
Data Description

Outline

Background
Data
Machine Learning Algorithm

Data Acquisition
Feature Engineering
Data Description

Data Acquisition

Cerner EHR $\rightarrow$ integrated data mart

Background
Data
Machine Learning Algorithm
Data Acquisition
Feature Engineering
Data Description

Available Data

- patient encounter
    - admission date
    - discharge date
    - reason for visit
- patient data
    - birth date
    - race
    - gender
    - marital status

Background
Data
Machine Learning Algorithm

Data Acquisition
Feature Engineering
Data Description

Available Data

- diagnoses
- labs
- vital signs
- medications administered
- location
    - timestamps
    - location
- billing
    - insurance type

Background
Data
Machine Learning Algorithm

Data Acquisition
Feature Engineering
Data Description

Data Acquisition

Identifed training cases

Used ICD 10 diagnosis codes.

- A41.02 Sepsis due to MRSA
- J15.212 Pneumonia due to MRSA
- A49.02 MRSA infection, unspecified site
- B95.62 MRSA infection causing disease classified elsewhere

Background
Data
Machine Learning Algorithm

Data Acquisition
Feature Engineering
Data Description

Features

Started with those recommended by infection control person

Background
Data
Machine Learning Algorithm

Data Acquisition
Feature Engineering
Data Description

Features

Started with those recommended by infection control person

- Nursing home patients
- Immunocompromised
- Any access to the body such as central line or catheter
- On dialysis

Background
Data
Machine Learning Algorithm
Data Acquisition
Feature Engineering
Data Description

Features

Location within hospital

Background
Data
Machine Learning Algorithm

Data Acquisition
Feature Engineering
Data Description

Features

Location within hospital

- Surgery
- Trauma
- Obstetrics
- Rehab

Features

Prior hospital/ED visits

Background
Data
Machine Learning Algorithm
Data Acquisition
Feature Engineering
Data Description

Features

Prior hospital/ED visits

- Number of visits in last year
- Time since last visit (inpatient, ED, ICU)
- Length of stay at last visit

Features

Labs and vitals

Background
Data
Machine Learning Algorithm

Data Acquisition
Feature Engineering
Data Description

Features

Labs and vitals

- White blood cell count
- Height, weight, BMI
- Lactic acid
- Age
- Shock index by age: (HR/SBP) * Age

Background
Data
Machine Learning Algorithm

Data Acquisition
Feature Engineering
Data Description

Features

Words in 'Reason for visit'

Background
Data
Machine Learning Algorithm

Data Acquisition
Feature Engineering
Data Description

Features

Words in 'Reason for visit'

- Cellulitis
- Abscess
- Sepsis

Background
Data
Machine Learning Algorithm
Data Acquisition
Feature Engineering
Data Description

Features

More features:

Background
Data
Machine Learning Algorithm

Data Acquisition
Feature Engineering
Data Description

Features

More features:

- Hospital's bed count
- Hospital region
- Patient's ED arrival method
- Time since admission

Preprocessing

Used `recipes` R package

Preprocessing

Used `recipes` R package

- Impute missing numeric with extreme value
- Impute missing factors with mode
- One-hot encode
- Create extra categories for new, unseen values.

Challenges

Very low prevalence

Background
Data
Machine Learning Algorithm
Data Acquisition
Feature Engineering
Data Description

Challenges

Very low prevalence

Dependence on past visits which are often not observable

Background
Data
Machine Learning Algorithm

Data Acquisition
Feature Engineering
Data Description

Challenges

Very low prevalence

Dependence on past visits which are often not observable

Patients cannot be identified if they went to a different hospital (within or outside Tenet)

Background
Data
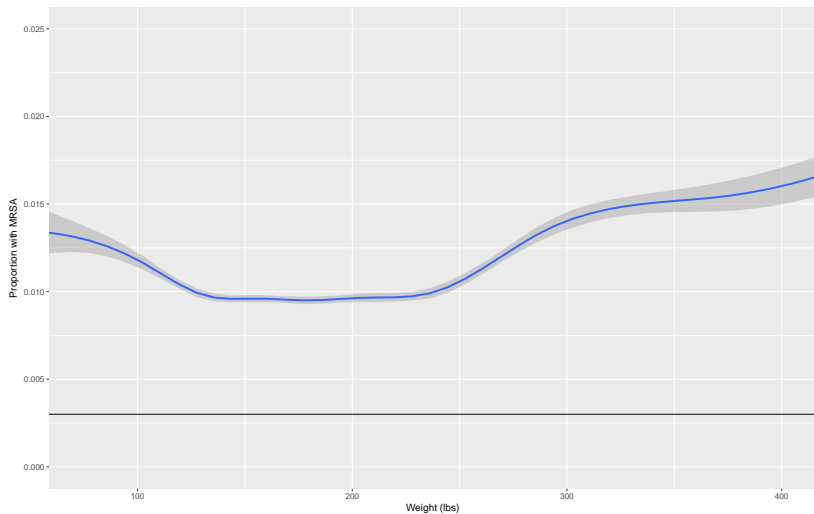Machine Learning Algorithm

Data Acquisition
Feature Engineering
Data Description

Challenges

Very low prevalence

Dependence on past visits which are often not observable

Patients cannot be identified if they went to a different hospital (within or outside Tenet)

Complexities with time

Background
Data
Machine Learning Algorithm

Data Acquisition
Feature Engineering
Data Description

Data Description

- Identified 3,000,000 inpatient/ED encounters
- 10,000 MRSA cases over about 2 years
- Prevalence: 0.3%
- Mostly community acquired

Relationship with Age

Background
Data
Machine Learning Algorithm
Data Acquisition
Feature Engineering
Data Description
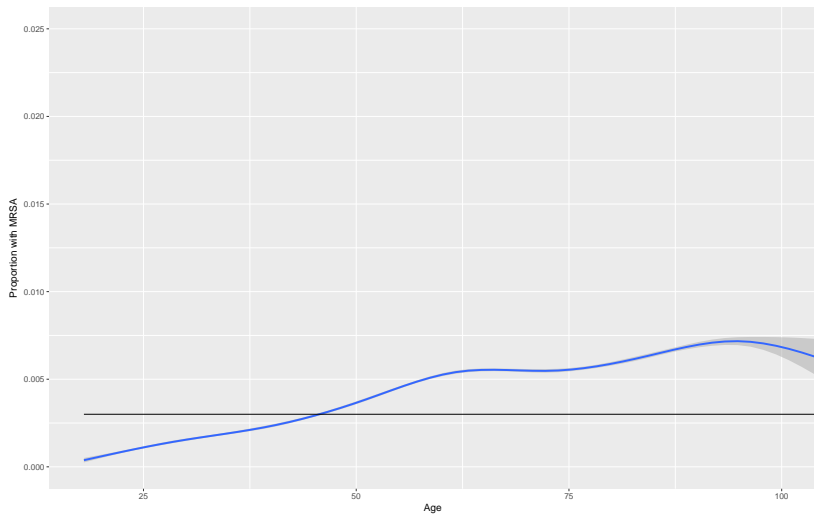
Relationship with Weight

Outline

xgboost!

xgboost!

R packages:

xgboost!

R packages:

- xgboost

- caret

- recipes

- mlr

Variable Importance
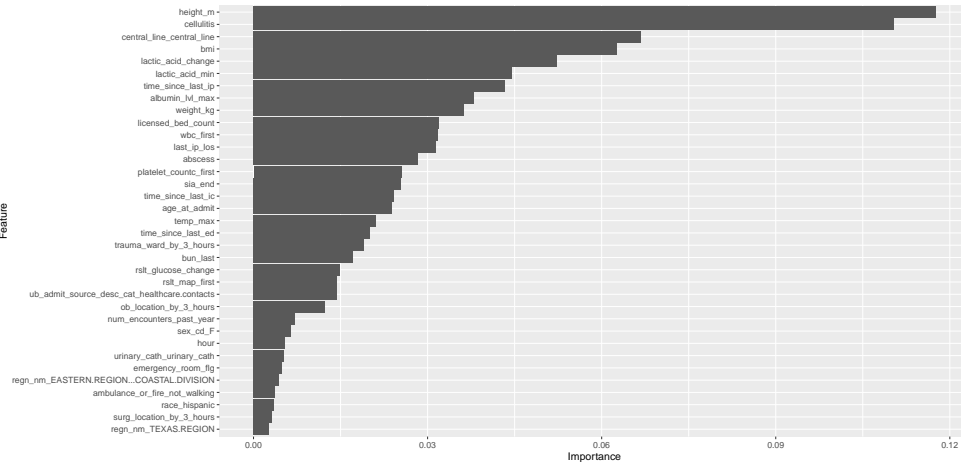
Tuning

xgboost tuning parameters:

- Learning rate/shrinkage

Tuning

xgboost tuning parameters:

- Learning rate/shrinkage
- Number of boosting rounds

Tuning

xgboost tuning parameters:

- Learning rate/shrinkage
- Number of boosting rounds
- Maximum tree depth
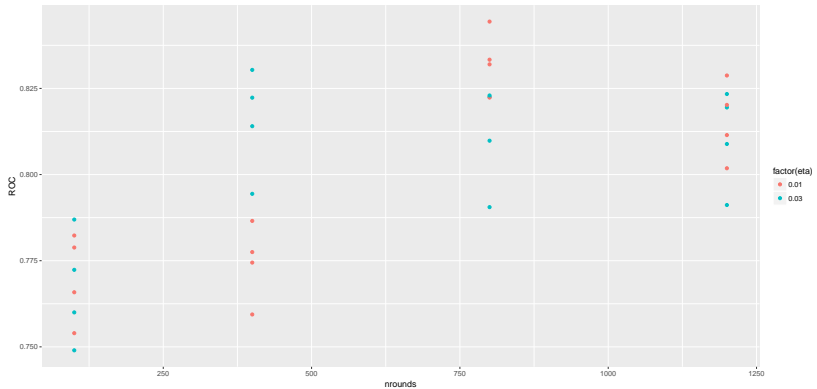
Tuning

xgboost tuning parameters:

- Learning rate/shrinkage
- Number of boosting rounds
- Maximum tree depth
- Fraction of observations to subsample

Tuning

xgboost tuning parameters:

- Learning rate/shrinkage
- Number of boosting rounds
- Maximum tree depth
- Fraction of observations to subsample
- Fraction of features to subsample

Effects of tuning parameters

Effects of tuning parameters

See figure

Alert rate

"Alert fatigue" Want a very low alert rate because event very rare

Performance at different alert rates and times

Table : Performance at selected alert rates

Alert Rate	Cutoff	PPV	Sensitivity	Specificity
1%	0.049	0.09	0.26	0.99
2%	0.026	0.07	0.42	0.98
3%	0.017	0.05	0.50	0.97

Next Steps

- Tuning!
- Previous antibiotic exposure
- Find evidence of wound
- Try some interactions with whether came through ED
- Expand data