Hospice Utilization by Minorities

J. Patrick Bernard

Regis University

MSDS696 Data Science Practicum II

J Patrick Analytics

Agenda

- Purpose
- Assumptions
- Data Engineering
- Process
- Results
- Next Steps
- Lessons Learn
- Questions

Purpose

- Every community deserves the high-quality health and end of life care provided by hospice.
- The objective of this project is to identify hospice utilization gaps by race.

J Patrick Analytics

Assumptions/Concerns

- This study will not determine a cause for any disparities
  - This is to inform those that are better qualified
- Personal compute environment would be insufficient
- Personal bias could influence results

Data Engineering

- Data source: Medicare Data from 2018
  - ~ 65 million record
  - Data was limited to beneficiaries that reside in the 50 states and the District of Columbia
- Data was partition by state units so it could be processed
  - Used pickle files to store data frames for each state
- Data was summarized and saved into pickle files and csv files for further analysis
- Focus on Louisiana and Mississippi

J Patrick Analytics

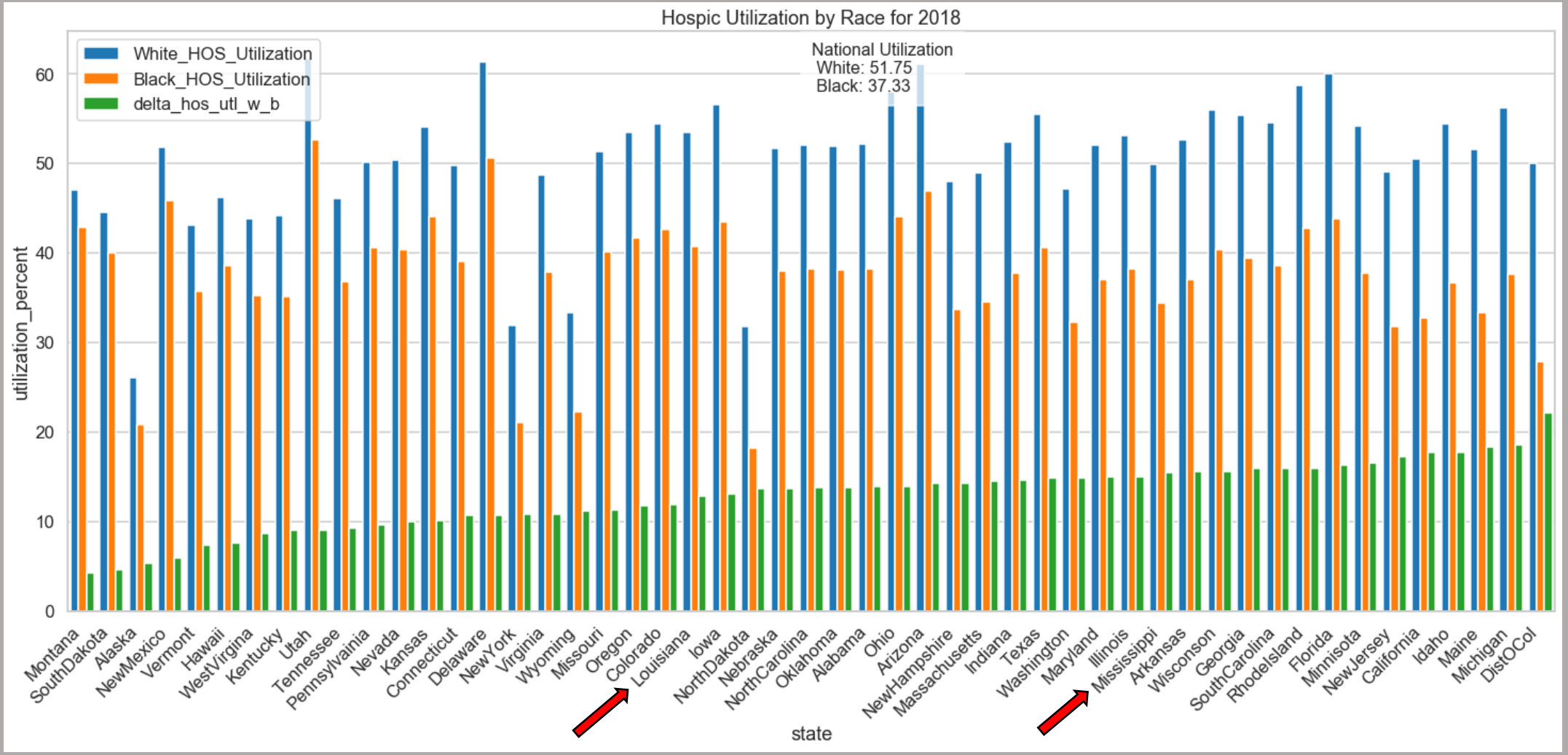
Determine hospice utilization by race

- How is Utilization Calculated

$$\frac{number\ of beneficiaries\ who\ died\ in\ hospice\ in\ a\ racial\ group}{number\ of \text{beneficiar}ies\ who\ died\ in\ that\ racial\ group}$$

J Patrick Analytics

Hospice Utilization



J Patrick Analytics

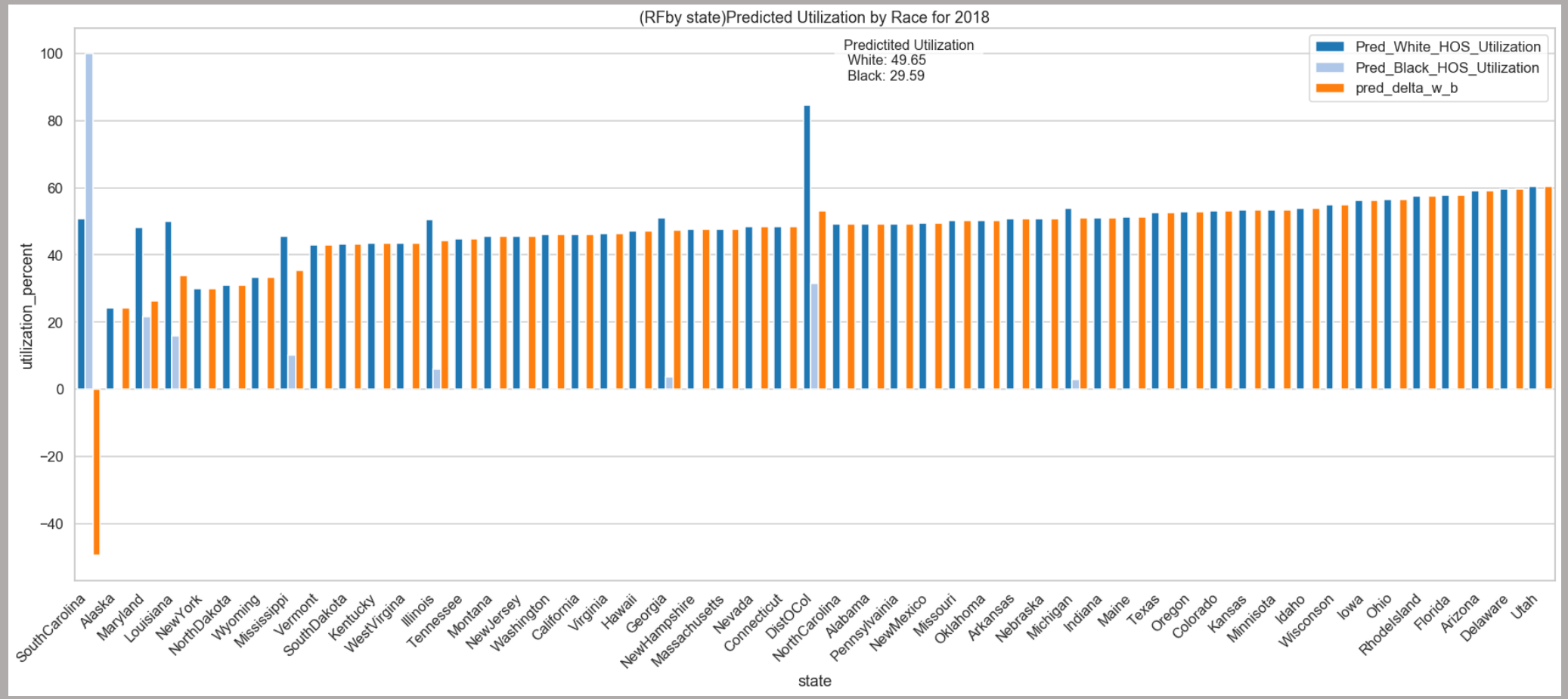
Can the records be classified by race?

- The Model
  - Random forest
  - Built a training set for the classifier
  - Train the classifier
  - Use the trained model on the data set

Can the records be classified by race?

- How model 1 was built
  - For each state
    - split the data into a training and a test set
    - train the model
    - Classify the entire state data set on the model
    - Record the feature importance for that model
    - Save the predicted race for each record
  - Open the summary file and plot the Hospice Utilization by predicted race

Predicted Hospice Utilization (RF by state)

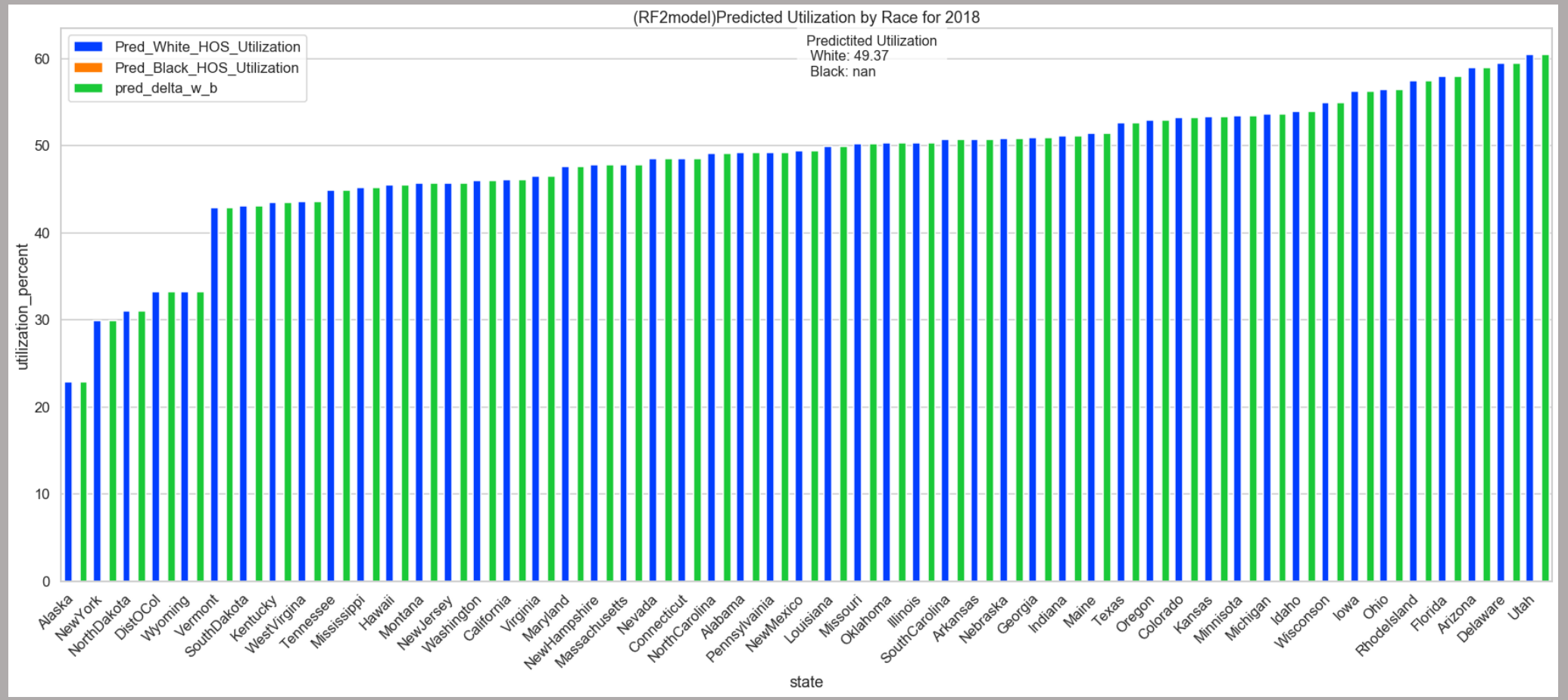


J Patrick Analytics

Can the records be classified by race?

- How model 2 was built built
  - Derived a training set of data
    - 2% of the data used for training
    - Race distribution represented in the data
    - Small sample used because of compute limitation
    - Train the model
    - Calculate the model accuracy
    - Determine the feature importance
  - Use the trained model on each states data
  - Plot the predicted hospice utilization

Predicted Hospice Utilization
(single RF model)



Model Training Accuracy: 0.794

J Patrick Analytics

Model Performance

- Both model performed poorly
  - Could augment the data for better performance
  - Increase the sample size for the training set

WAIT!!!!!

Review:

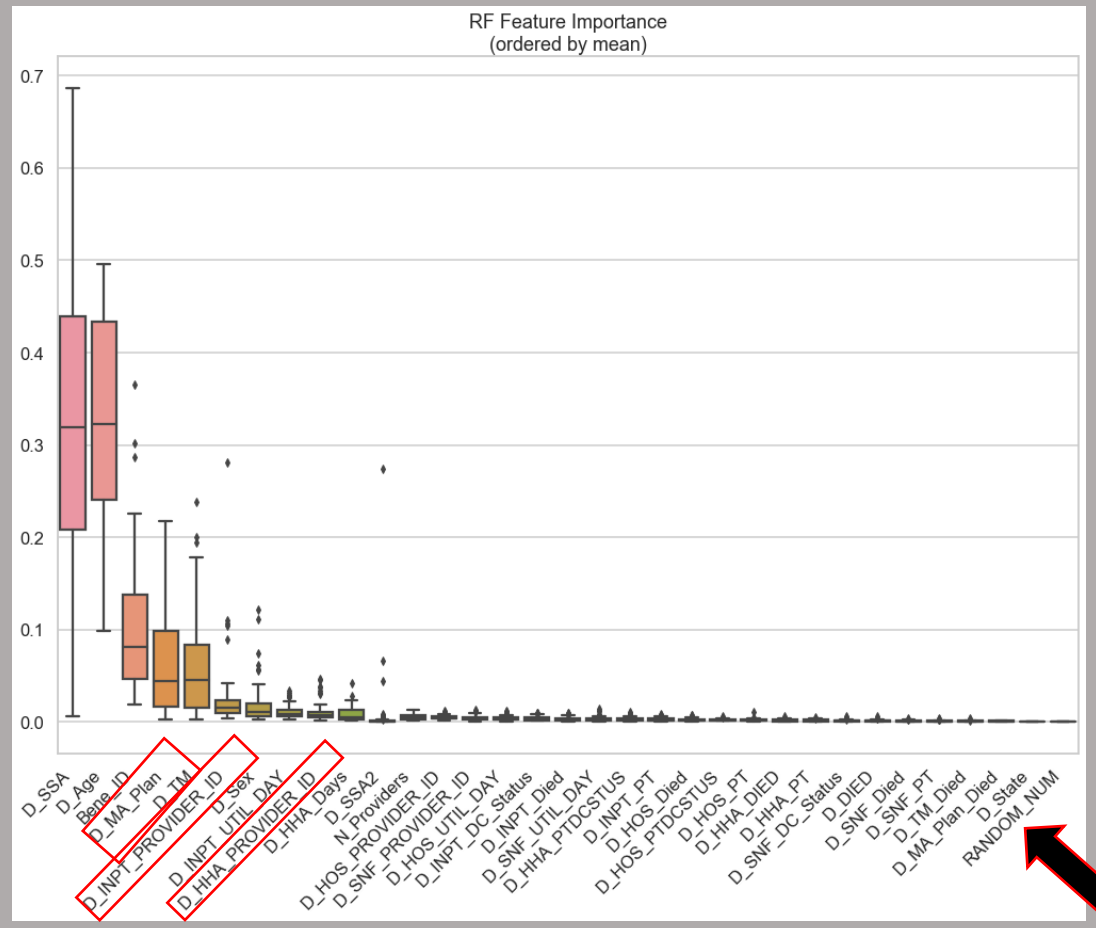
The purpose of this study is to identify racial disparity in the utilization of hospice care

The Models

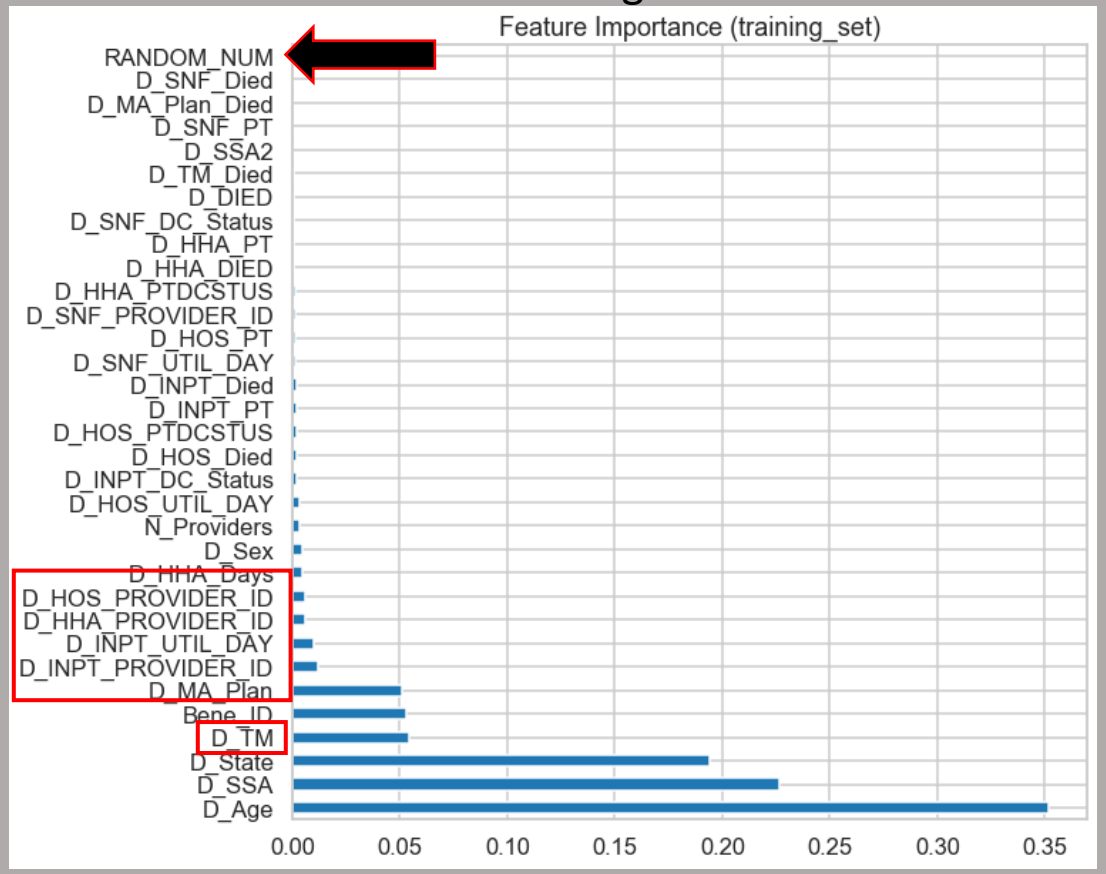
- Is it significant that two random forest model are having a difficult time classify race on the provided data?
- What does it mean if the model is having a hard time classifying the data?
- Random Forest models are very good classifiers and prediction algorithms
- So, It is significant that the algorithm is having a hard time classifying
- To me this in an indication that there is NOT a systemic component to hospice utilization.

Feature Importance

By state models

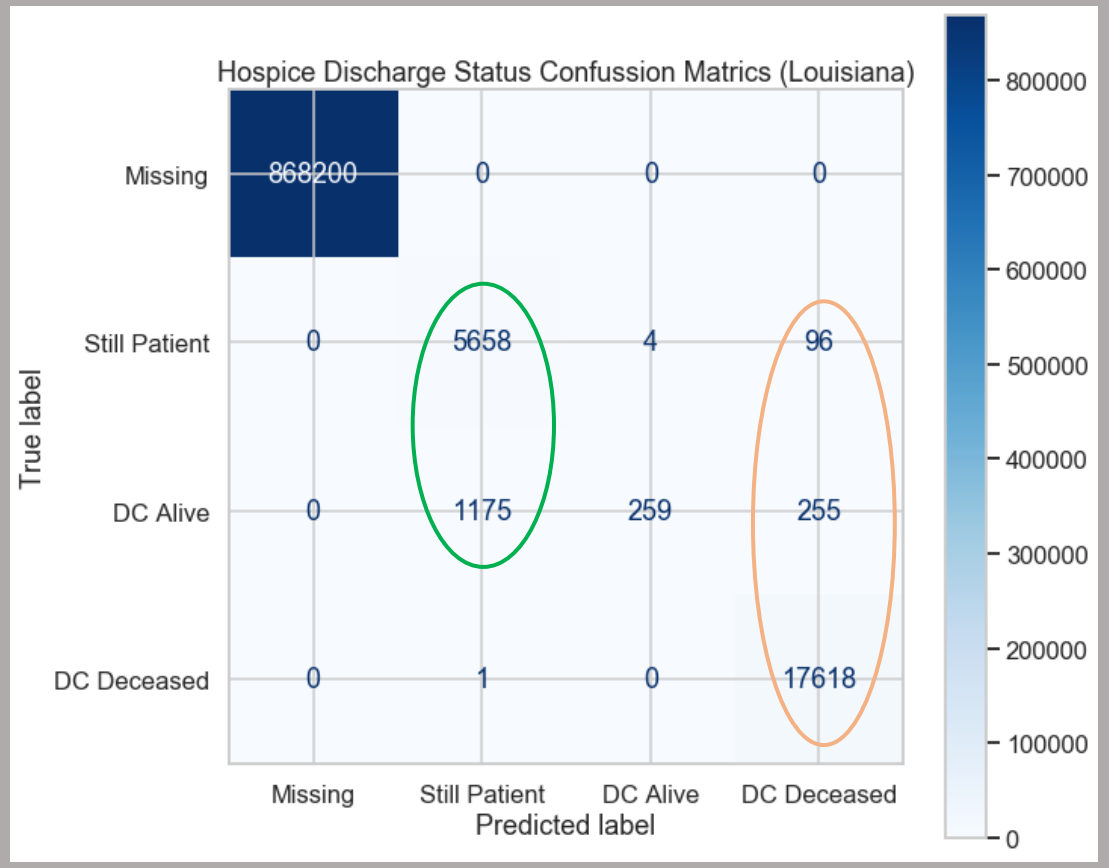


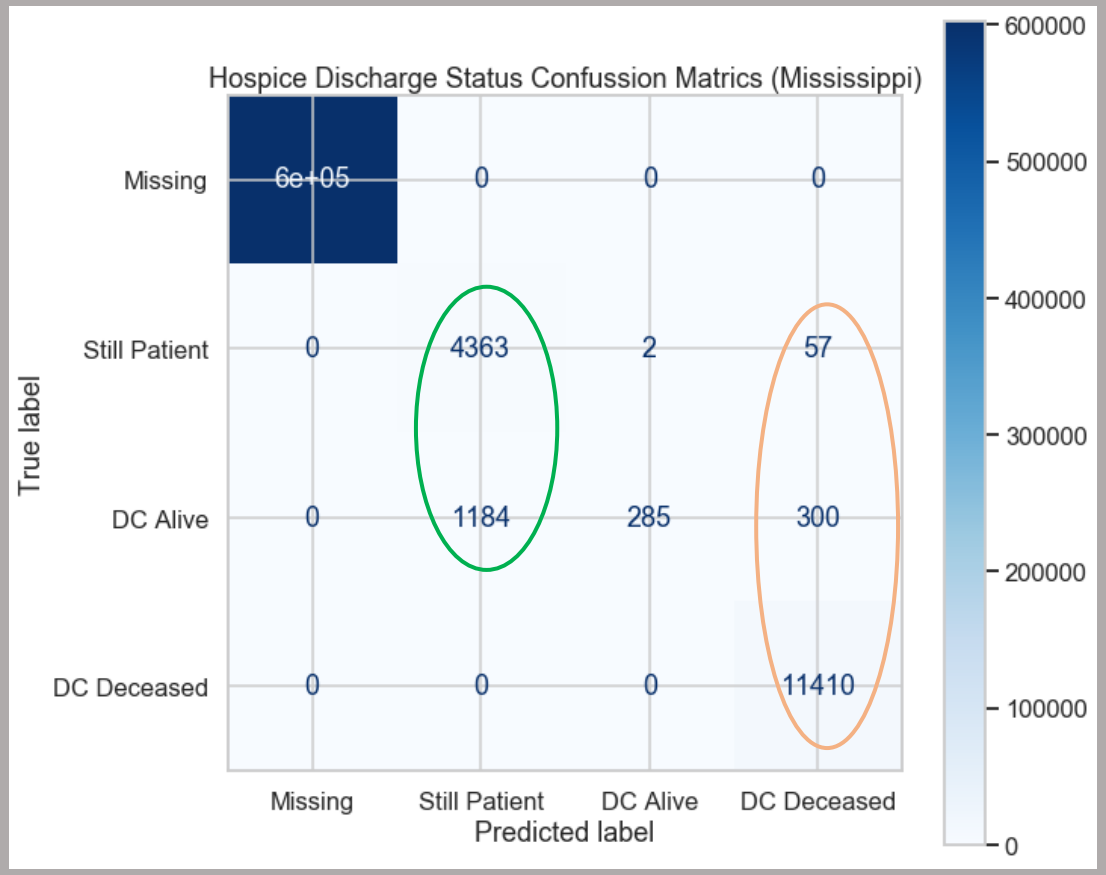
Single model



J Patrick Analytics

Verify the Classifier





J Patrick Analytics

Next Steps

- Get a larger compute environment to verify conclusions with a larger training set
- Try a different classifier
  - K-mean
  - T-pot
  - XGBoost
- Way to resources share
  - Better use of technologies like Dask
- Look into the difference in the Hospice discharge status

Lessons Learned

- Be careful of the rabbit holes that you can get caught going down.
  - when you find something interesting It may lead you away from your first objective.
  - Make a note of the idea so that you can come back to it.
- Keep the purpose of the study in mind as you build your models
- Use the Subject Matter Experts
  - Frequently Consult them for meaning and understanding of the data

Thank You

J Patrick Analytics