A Comparative Analysis of Feature Ensembles for ICU Mortality Prediction

By: Sachin Mohandas

For: Dr. Eugene Pinsky

MET CS677 – Data Science with Python: Term Project

The MIMIC-III ICU Dataset

- Subset of a larger database system that stores clinical data for patients of Beth Israel Deaconess Medical Center
- Obtained from Kaggle, this dataset contains 1177 tidy observations spanning 49 columns
  - Minimal cleaning and preprocessing required
- Target column ‘outcome’ contains either ‘0’ indicating alive, or ‘1’ indicating death

Motivation and Methods

- High traffic hospital ER/ICU units stand to benefit significantly from the streamlining and optimization of triage scenarios and care allotment
- Aim is to develop a reliable tool for the prediction of mortality in ICU patients
- We implement well known classification algorithms to this end
  - Logistic Regression
  - Gaussian Naïve Bayesian
  - Random Forest

Methods (cont.)

- First, compute the accuracy of the classifier when all relevant columns are considered
- Then, individually analyze the predictive power of all 7770 unique combinations of 3 columns chosen from the 37 non-categorical columns
  - Intention is to assess if there are ensembles having higher predictive power than the base model
  - Conduct EDA on such ensembles to attempt to find a reason for their increased accuracy

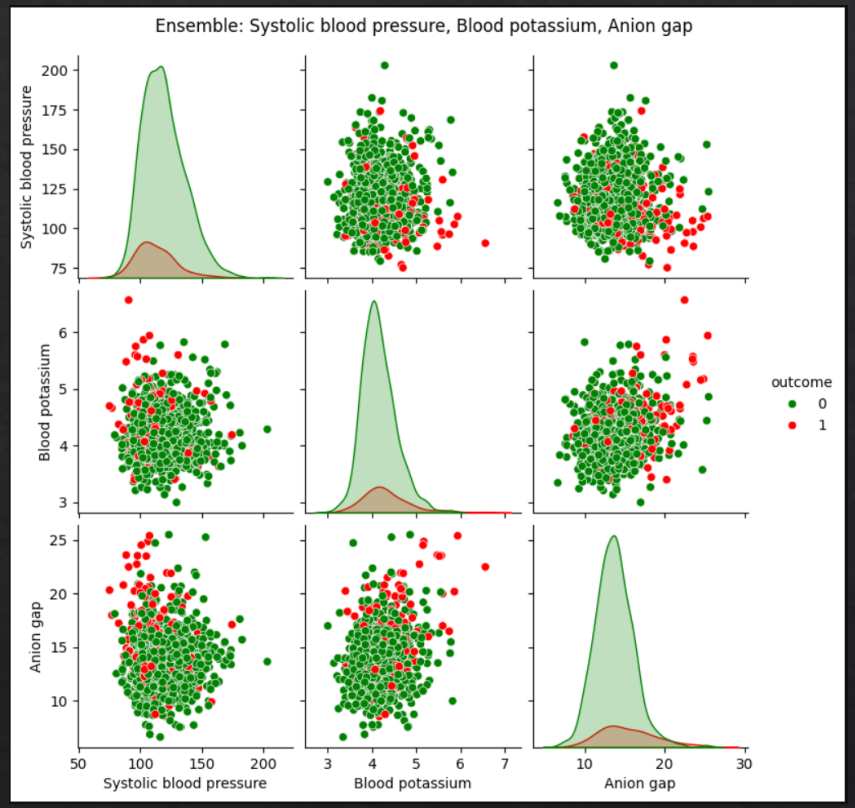
Accuracies and Collinearities

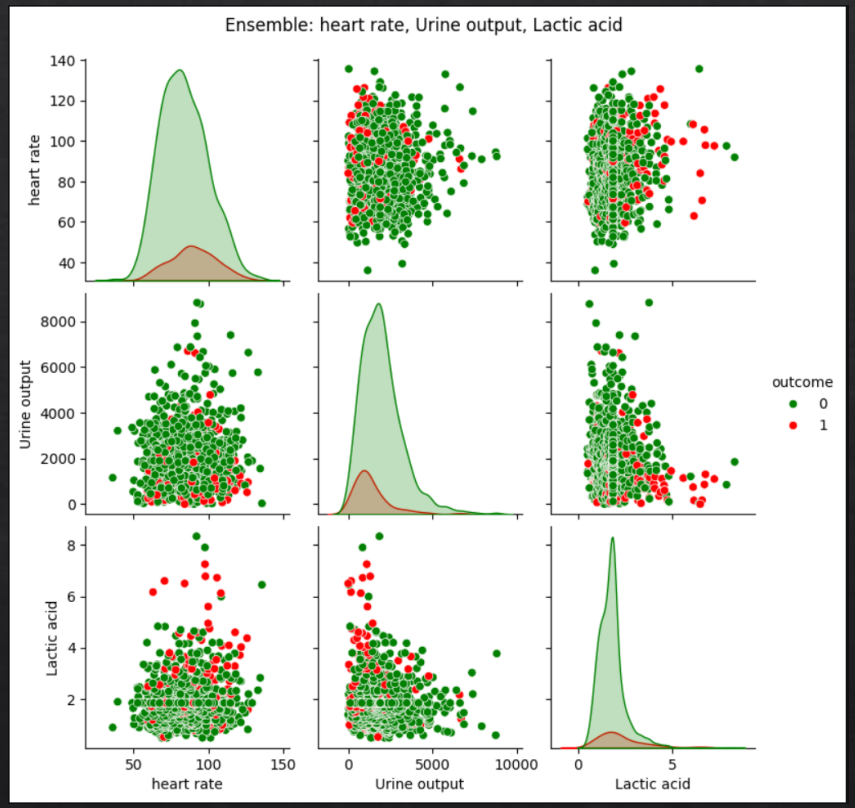
- Accuracy when considering all columns (base model): **85.6%**
- Number of ensembles having higher accuracy than the base model: **2607**
- Most accurate ensemble ('Blood calcium', 'Anion gap', 'Lactic acid'): **87.4%**

```
Summary of collinearities of all ensembles:
count    7770.000000
mean        0.289442
std         0.206335
min         0.008343
25%         0.151333
50%         0.236090
75%         0.362102
max         1.706955
```

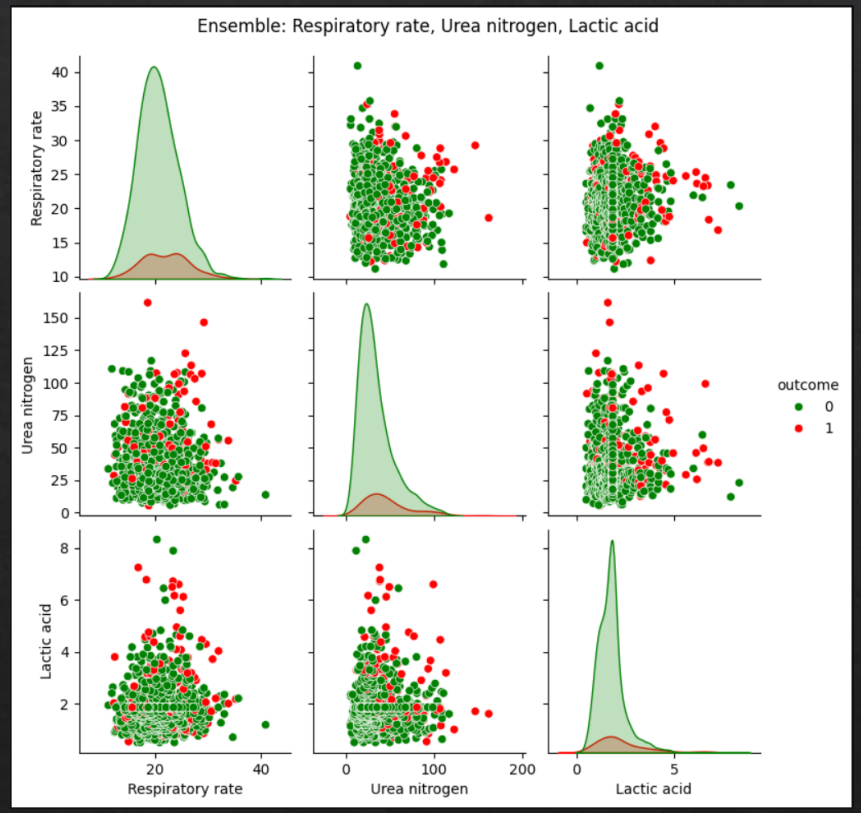
```
Summary of collinearities of top 50 ensembles:
count    50.000000
mean      0.405981
std       0.285188
min       0.049614
25%       0.256983
50%       0.330061
75%       0.499965
max       1.706955
```

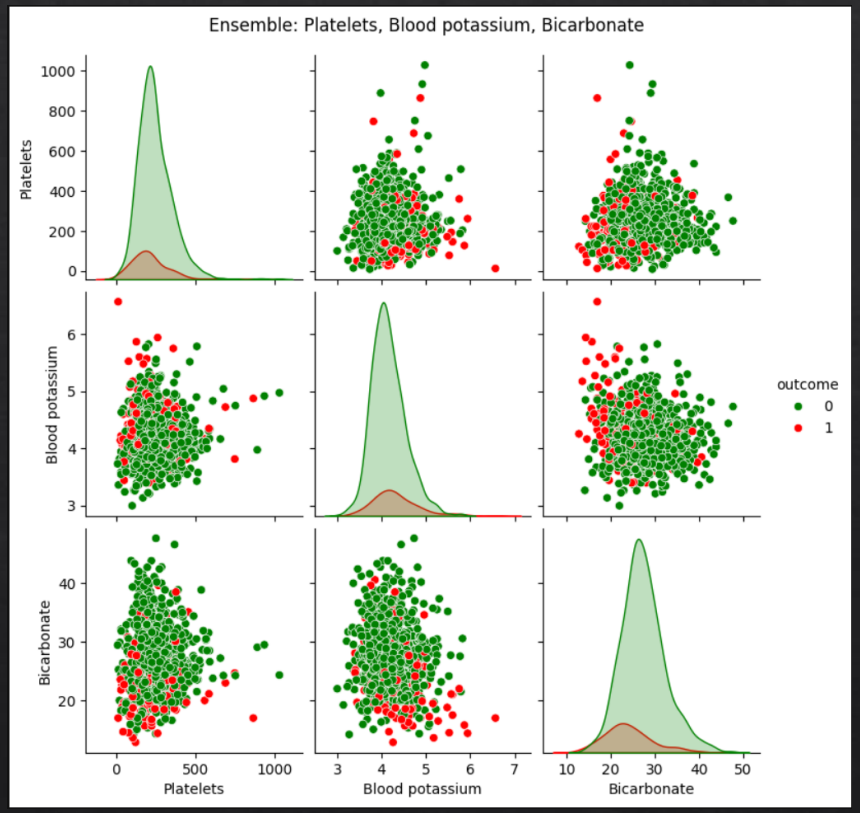
Pair Plots


Ensemble: Systolic blood pressure, Blood potassium, Anion gap
Systolic blood pressure
Blood potassium
Anion gap
outcome
0
1


Ensemble: heart rate, Urine output, Lactic acid
heart rate
Urine output
Lactic acid
outcome
0
1

Pair Plots


Ensemble: Respiratory rate, Urea nitrogen, Lactic acid
Respiratory rate
Urea nitrogen
Lactic acid
outcome
0
1

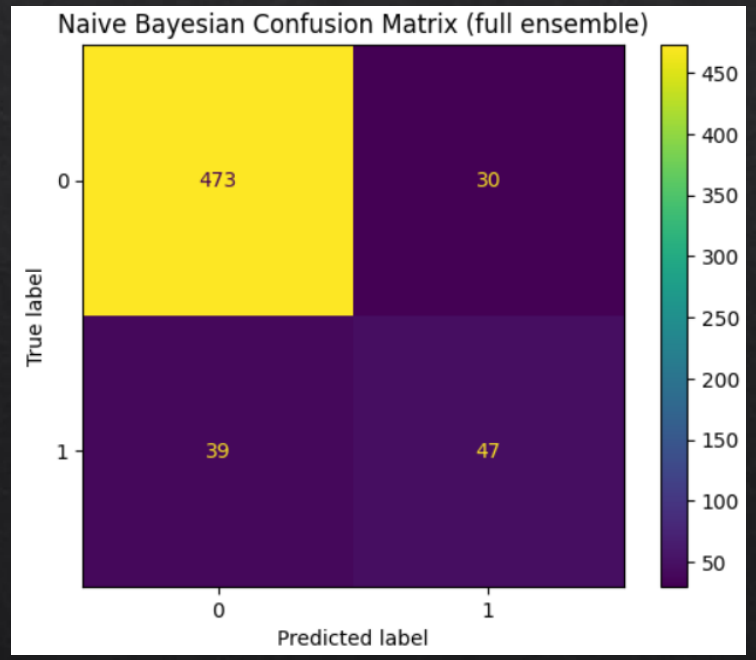

Ensemble: Platelets, Blood potassium, Bicarbonate
Platelets
Blood potassium
Bicarbonate
outcome
0
1

Model Coefficients

```
Model Coefficients for the top 10 ensembles

Columns: ('Blood calcium', 'Anion gap', 'Lactic acid')
Coefficients: [[-1.03252433  0.21422091 -0.00405508]]

Columns: ('heart rate', 'Urea nitrogen', 'Lactic acid')
Coefficients: [[ 0.03091895  0.02504876 -0.06982544]]

Columns: ('heart rate', 'Urine output', 'Lactic acid')
Coefficients: [[ 0.02371812 -0.00047218 -0.08191985]]

Columns: ('Systolic blood pressure', 'RDW', 'Anion gap')
Coefficients: [[-0.01542157  0.17020151  0.02361099]]

Columns: ('Systolic blood pressure', 'Blood potassium', 'Anion gap')
Coefficients: [[-0.01884787  1.21328944  0.03758682]]

Columns: ('Respiratory rate', 'Urine output', 'Lactic acid')
Coefficients: [[ 0.08800663 -0.00049753 -0.08100689]]

Columns: ('Respiratory rate', 'Urea nitrogen', 'Lactic acid')
Coefficients: [[ 0.08404689  0.02114094 -0.07632598]]

Columns: ('SP O2', 'Urea nitrogen', 'Lactic acid')
Coefficients: [[-0.18260112  0.02034342 -0.09224466]]

Columns: ('hematocrit', 'Blood potassium', 'Bicarbonate')
Coefficients: [[ 0.02004818  1.08578749 -0.96090244]]

Columns: ('Platelets', 'Blood potassium', 'Bicarbonate')
Coefficients: [[-0.0029852   1.13367    -0.81068625]]
```

Naïve Bayesian Analysis

- Accuracy when considering all columns (base model): **88.3%**
- Number of ensembles having higher accuracy than the base model: **0**

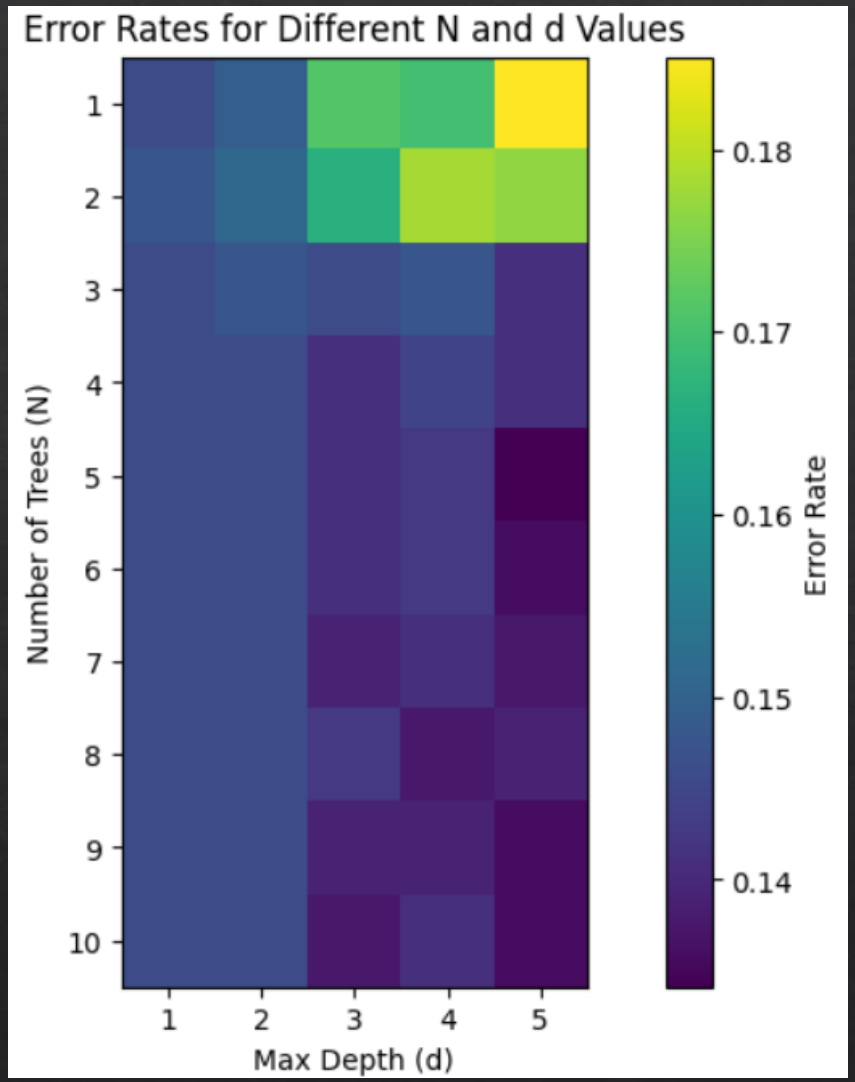


TPR = 54.7%

TNR = 94.0%

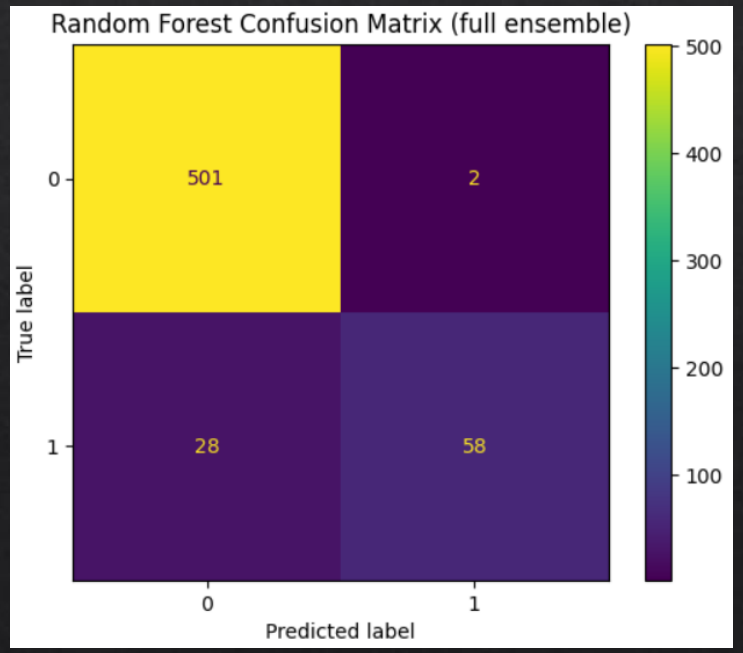
Random Forest Classification

- Iterating through different numbers of estimators (1,…, 10) and different depths (1,…, 5), we find that the optimal configuration of N and d for the classifier (as tested on the base model) is **N = 5** and **d = 5**, as shown pictorially in the error rate heat map



Random Forest Analysis

- Accuracy when considering all columns (base model): **94.9%**
- Number of ensembles having higher accuracy than the base model: **0**



TPR = 67.4%

TNR = 99.6%

Considering the Results

- The implementation of the random forest classifier yielded the best results, outperforming the Naïve Bayesian classifier and the best ensemble from the Logistic Regression classifier by a significant margin.

- This is likely owing to the Random Forest classifier's aptitude in modeling complex non-linear relationships, whereas Naïve Bayesian classifiers assume independence among features (as is often not the case when considering medical comorbidities) and Logistic Regression classifiers assume linear relationships.

- Additionally, Random Forest classifiers naturally handle categorical features (several of which are included in the full ensemble) much more effectively than Naïve Bayesian and Logistic Regression classifiers

Considering the Results (cont.)

- It is also interesting to note how similar the accuracies of the best individual ensembles are across each of the classifiers. This may indicate that the properties of the data which most heavily influence accuracy lie instead in the features that are not included in the ensemble assembly.

- Accuracy ranges for the top 10 ensembles for each classifier:
  - Logistic Regression: 86.9% - 87.4%
  - Naïve Bayesian: 86.9% - 87.1%
  - Random Forest (N = d = 5): 86.9% - 87.4%

Closing Remarks and Future Work

- A rudimentary predictive tool has been developed
  - Although the best accuracy (94.9%) indicates high reliability, the high quantity of false negatives this classifier produced is worrisome, especially in a critical medical environment (a false positive is better than a false negative in such scenarios)
  - Patient cases are often volatile and complex and may present to the ER/ICU under a very broad spectrum of circumstances that the models may not be able to adequately account for
  - Clinical interpretability of the models is not lucid, forcing medical professionals to stake faith in a system without necessarily understanding its underlying methodologies
- EDA was relatively inconclusive in finding trends in the individual ensembles, but much can still be done!
  - Assess additional/alternative machine learning algorithms/neural networks and fine tune parameters therein
  - Consider additional features and observations

Thank you for your attention!

Additional notes pertaining to some slides are included in the following appendix.

Appendix

- Slide 2
  - Columns contain a litany of relevant medical data gleaned from tests conducted upon the patient's admittance to the ICU
  - Cleaning only entailed imputing NA's with the means of their respective columns
- Slide 4
  - The goal is to assess if there is a way to provide consistently valuable insight into patient cases for which only a truncated amount of feature data may be available
- Slide 5
  - A t-test was conducted on the means of the collinearities and confirmed that there was no statistically significant difference between them

Appendix (cont.)

- Slides 6 and 7
  - Low blood pressure, low urine output, low platelet count, and low bicarbonate concentration could all be indicators of kidney dysfunction, though a biostatistician would need to confirm this
- Slide 8
  - High blood potassium could indicate higher risk of mortality, while high blood calcium and/or high bicarbonate levels could indicate a lower risk of mortality, though a biostatistician would need to confirm both this assertion and those listed for slides 6 and 7
- Slide 11
  - These results may be slightly biased since the N and d values used were those that had been proven to be the best for the base model, and no other combinations were implemented when testing the individual ensembles

Appendix (cont.)

- Slide 12
  - In the same vein as the better handling of categorical features by Random Forests, we complementarily see that the top results for the individual ensembles as computed by the Random Forest classifier is much more similar to the results of the other two classifiers, indicating that the source of the increased accuracy could in fact be these categorical features.