

Diabetes Story

| | | | | | |
|---------------------|---------------------|-------------------|---------------------|---------------------------|------------|
| Predicting Diabetes | Correlation Heatmap | Insulin x Glucose | BMI x SkinThickness | RF: Full vs Reduced Model | Conclusion |
|---------------------|---------------------|-------------------|---------------------|---------------------------|------------|

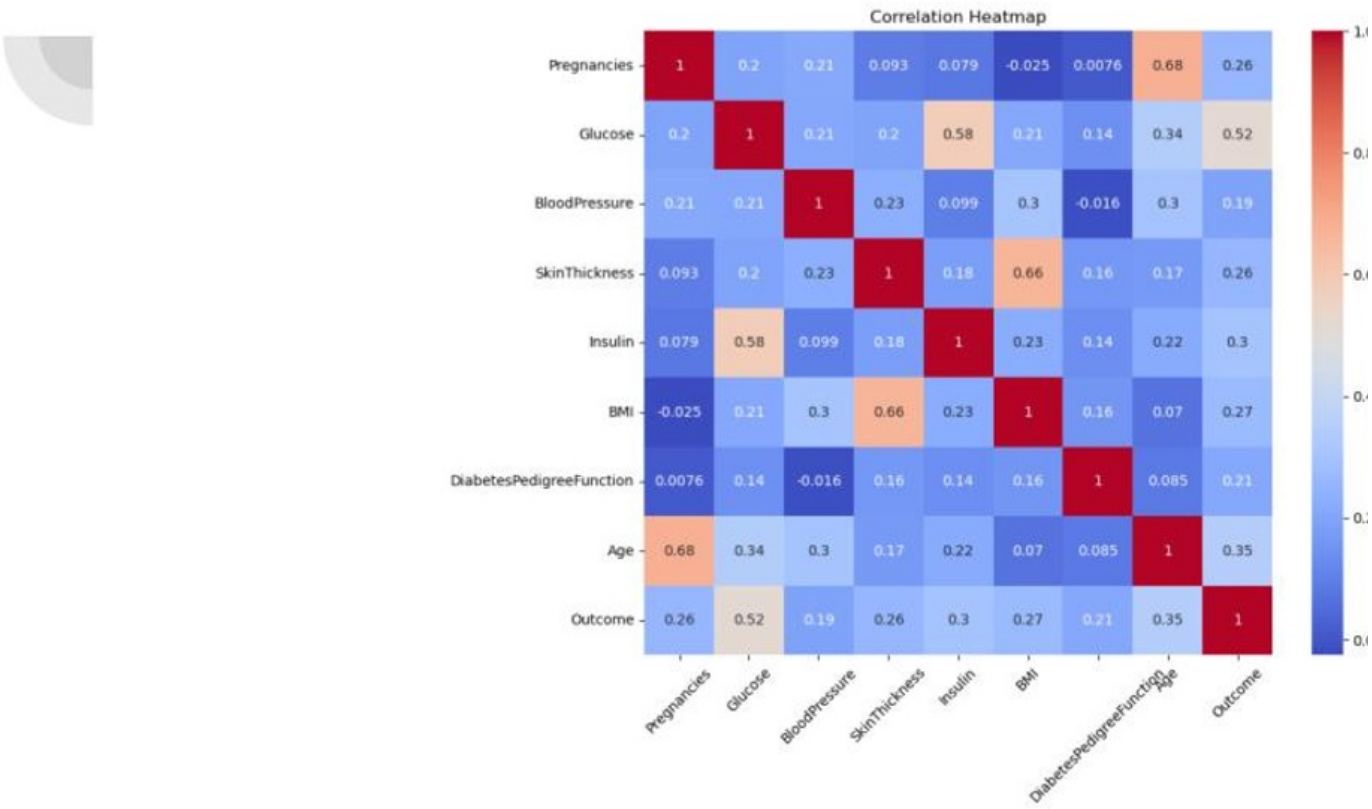


Predicting Diabetes
Through a Supervised Learning Model

Diabetes Story

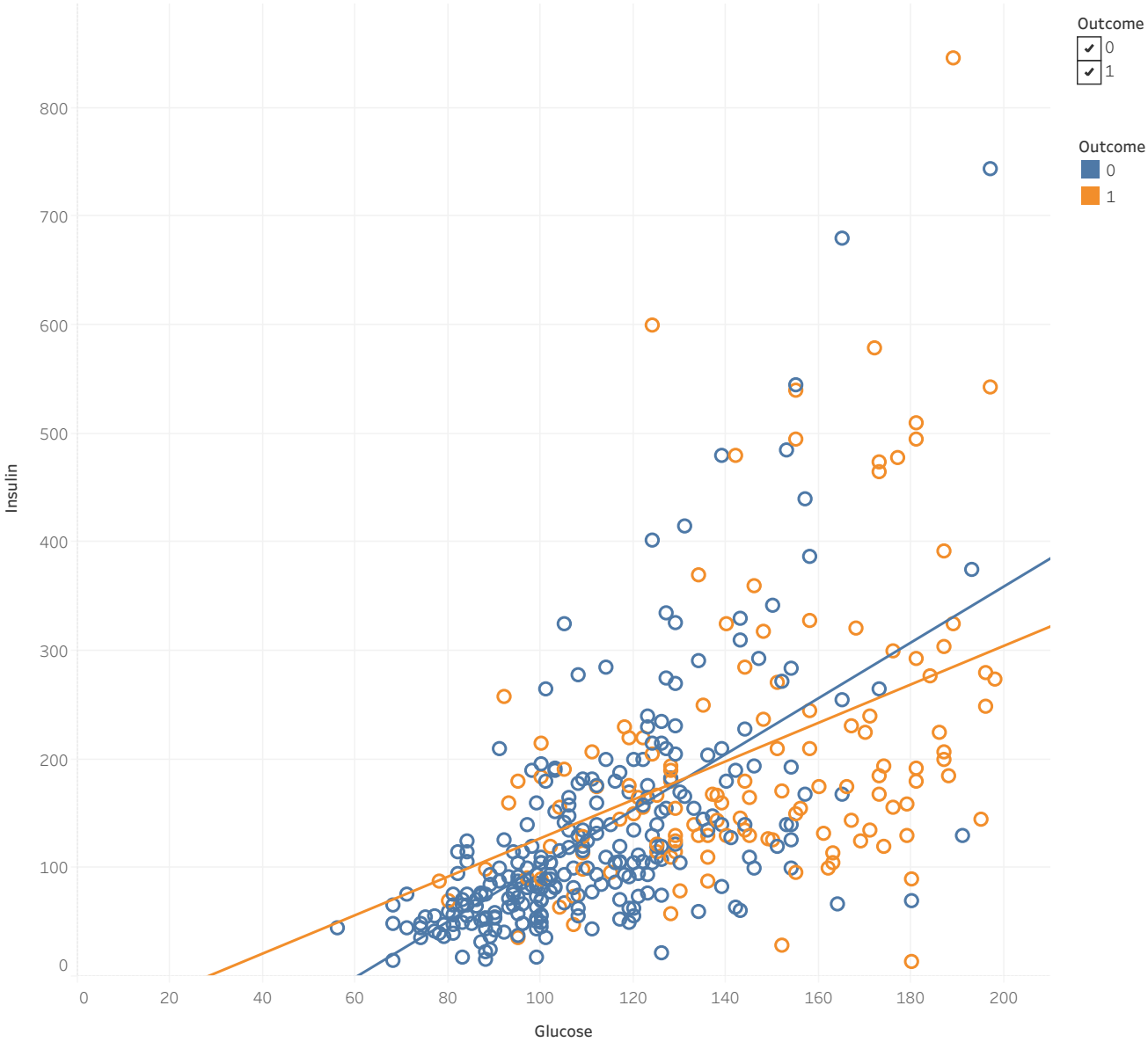
| | | | | | |
|---------------------|---------------------|-------------------|---------------------|---------------------------|------------|
| Predicting Diabetes | Correlation Heatmap | Insulin x Glucose | BMI x SkinThickness | RF: Full vs Reduced Model | Conclusion |
|---------------------|---------------------|-------------------|---------------------|---------------------------|------------|

Correlation Matrix



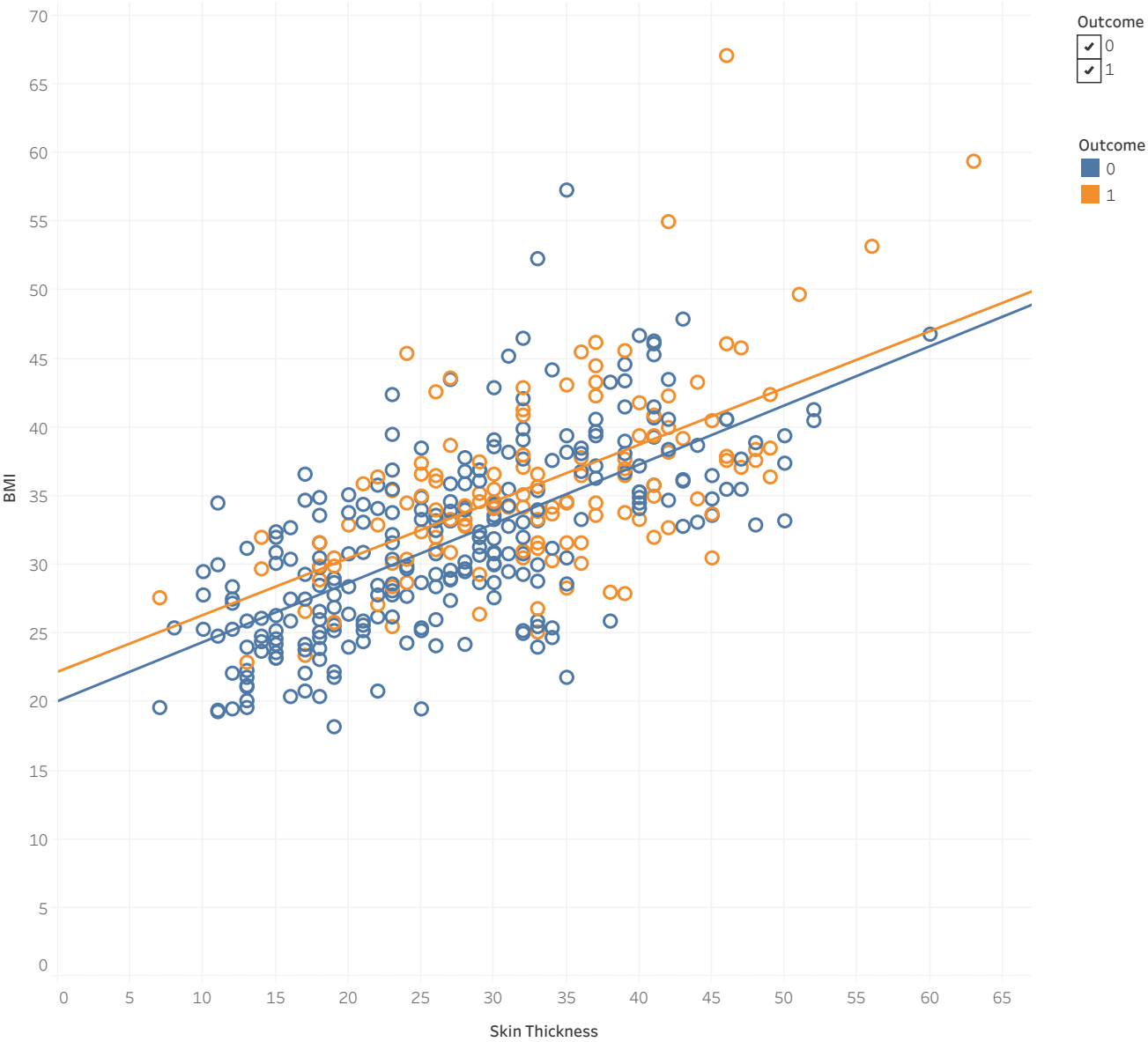
Diabetes Story

| | | | | | |
|---------------------|---------------------|-------------------|---------------------|---------------------------|------------|
| Predicting Diabetes | Correlation Heatmap | Insulin x Glucose | BMI x SkinThickness | RF: Full vs Reduced Model | Conclusion |
|---------------------|---------------------|-------------------|---------------------|---------------------------|------------|



Diabetes Story

| | | | | | |
|---------------------|---------------------|-------------------|---------------------|---------------------------|------------|
| Predicting Diabetes | Correlation Heatmap | Insulin x Glucose | BMI x SkinThickness | RF: Full vs Reduced Model | Conclusion |
|---------------------|---------------------|-------------------|---------------------|---------------------------|------------|



Diabetes Story

| | | | | | |
|---------------------|---------------------|-------------------|---------------------|---------------------------|------------|
| Predicting Diabetes | Correlation Heatmap | Insulin x Glucose | BMI x SkinThickness | RF: Full vs Reduced Model | Conclusion |
|---------------------|---------------------|-------------------|---------------------|---------------------------|------------|



Random Forest: Full Model vs. Reduced

- 8 Features
- Accuracy = 0.822
- Precision = 0.788
- Recall = 0.650
- Confusion Matrix

```
[[71 7]
 [14 26]]
```

- ROC = 0.89

- 5 Features (no BP, SkinThickness, or DBF)
- Accuracy = 0.82
- Precision = 0.744
- Recall = 0.725
- Confusion Matrix

```
[[68 10]
 [11 29]]
```

- ROC = 0.89

Diabetes Story

| | | | | | |
|---------------------|---------------------|-------------------|---------------------|---------------------------|------------|
| Predicting Diabetes | Correlation Heatmap | Insulin x Glucose | BMI x SkinThickness | RF: Full vs Reduced Model | Conclusion |
|---------------------|---------------------|-------------------|---------------------|---------------------------|------------|



Key Takeaways

- The Reduced model with 5 most relevant features, as selected through the chi2 process, had the same accuracy and ROC as the full model. It also notably improved in the Recall score with fewer False Negatives, a misclassification of clinical concern.
- Random Forest was the ensemble method used here with both Full and Reduced models. With either model, it resulted in the best evaluation metrics compared to the non-ensemble methods. Its multiple, parallel training does appear to have delivered a better model for making predictions about new data.