Data Mining and Decision Systems  
600092  
Assigned Coursework Report

Student ID: 201601628  
Date: 07 October 2019

## Due Date: 12 December 2019

**Report must be within 8 page maximum. Strict page limits will be enforced. Any extra pages will be ignored and no marks awarded for any work on these. Exclusions to this limit are the front page, the references section, and any appendices. Please keep to the given section headings and format; subsections are permitted.**

# Methodology

Provide details on the methodology applied towards the data mining analysis undertaken, providing rationale for these steps.

This should detail how you went from the raw data provided to the chosen model(s), choice of model, and how this methodology helps address the problem domain.

Evidence to support the following of this methodology should be presented, especially any cases which required moving backwards in the process to readdress issues.

**Introduction**

**CRISP DM**

1. **Business Understanding**
2. **Data Understanding**
3. **Data Preparation**
4. **Modelling**

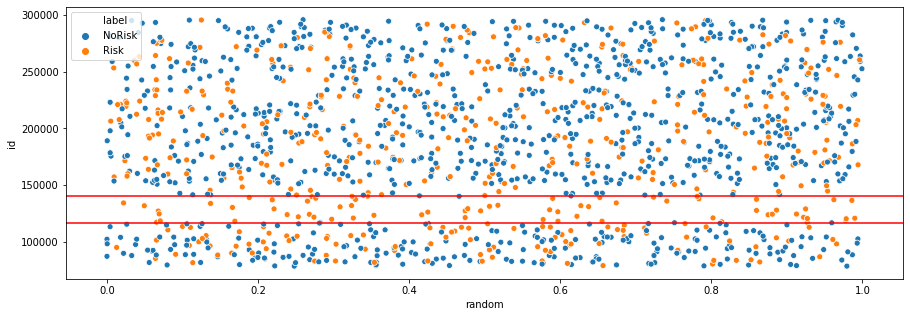
# Results

Results should include tables showing model performance with appropriately selected metrics. No rationale should be provided for this section - simply results of evaluative processes.

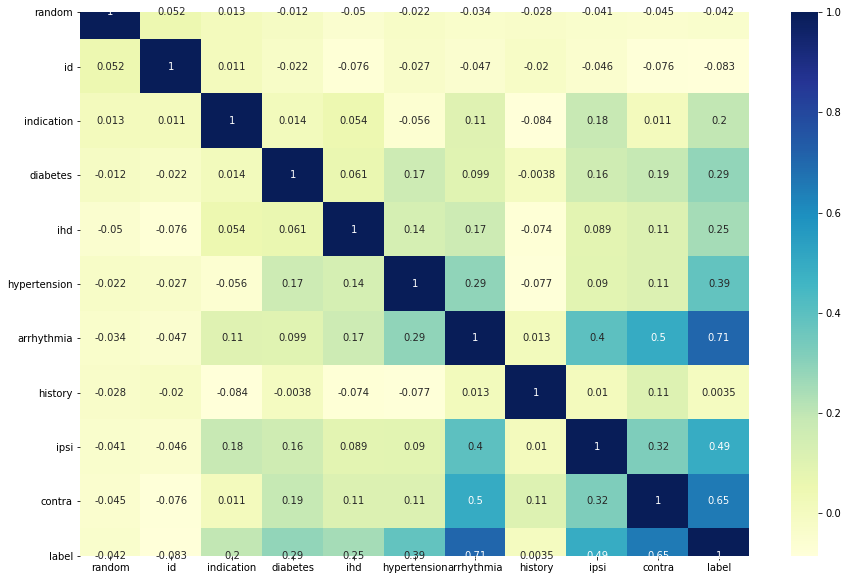
If using modified variants of the dataset, these should be clearly identified in the tables with appropriate naming. The justification and description of modification is not for this section.

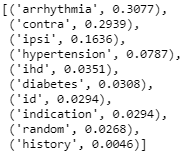
Additional figures may be used as appropriate, in support of discussion points in the Evaluation & Discussion section, or as evidence for methodology following above.

**ID Cluster**

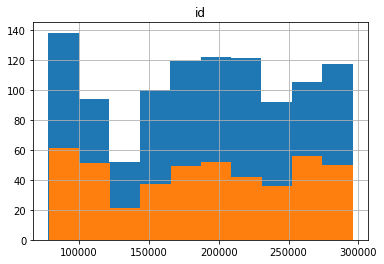


**Feature Heatmap**

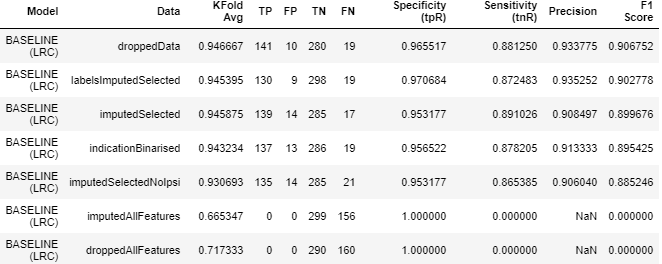


****

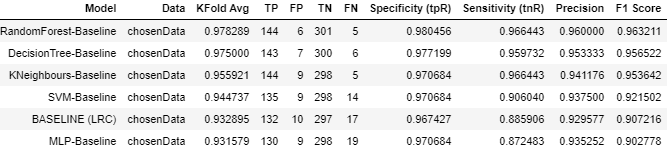
**Stratification**



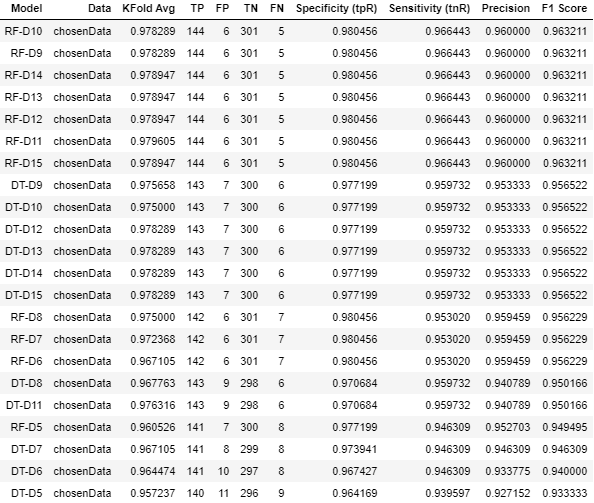
**Baseline Model with Data Variants**

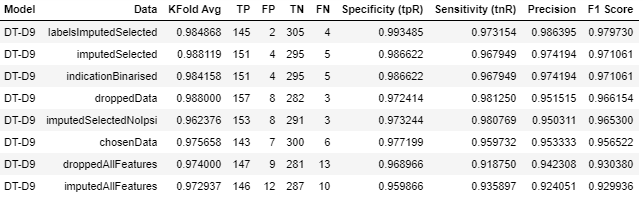


**Additional Models with Chosen Data**



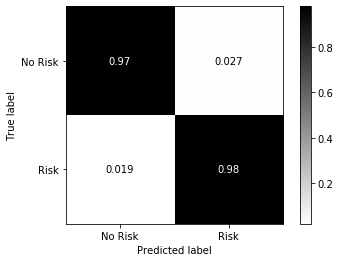
**Hyperparameter Tuning**

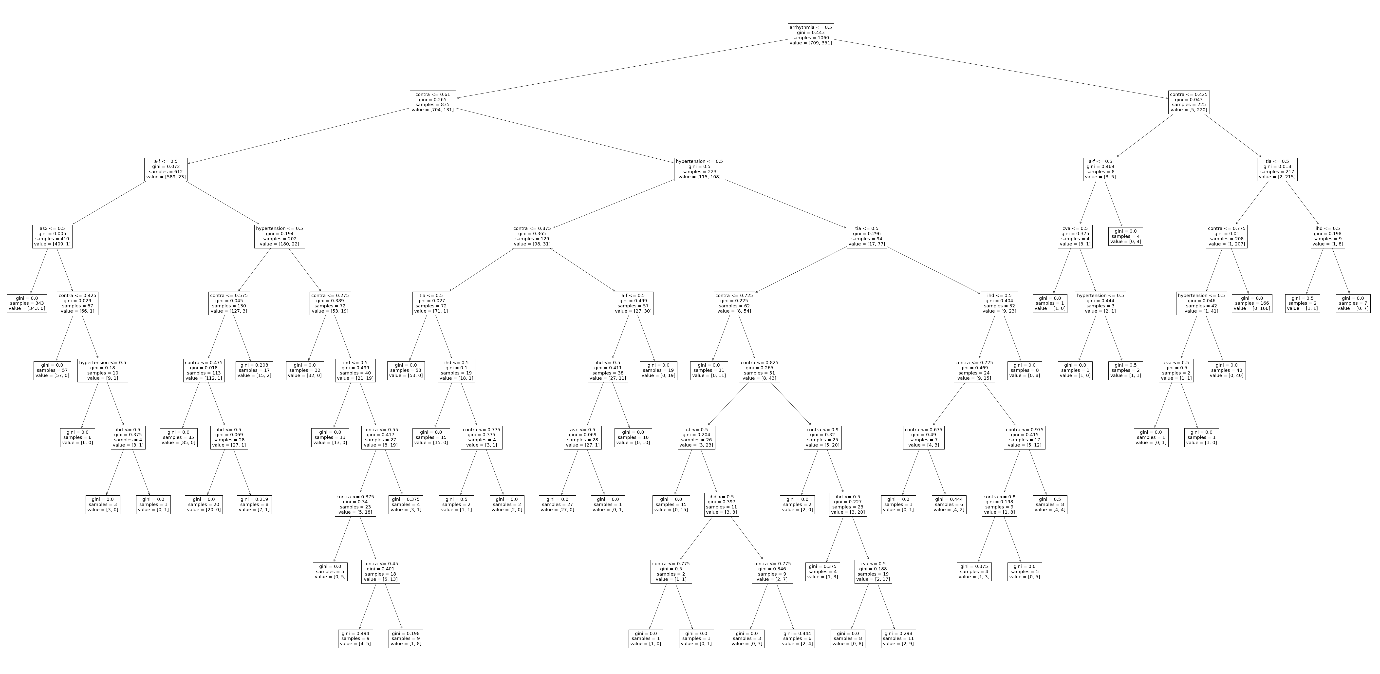




**Final Model**







# Evaluation & Discussion

Evaluation methodology used for generating the results provided in the previous section. How were these evaluated? Why was this selected? What metrics were used and why?

Discussion of the results should be presented with appropriate evidence and rationale. E.g Which is the best model, and why?

Consider each stage in the methodology, and reflect on any improvements which could have been made. Could any techniques have been used which may have improved performance? Why?

1. **Evaluation**
2. **Deployment**

# References

Any references used throughout the report should be included here in Hull Harvard Style. If no references used, remove this section.