Antepartum Cardiotocography Analysis

Thinkful Data Science
Corey Coole

Presentation Contents:

- Data Source and Background Information
- Dataset Description and Research Goal
- FIGO Target Class Guidelines
- Feature Descriptions
- Target Class Distribution and Predictive Model Descriptions
- Feature Missingness and Multicollinearity
- Logistic Regression Performance and Metrics
- Support Vector Classifier Performance and Metrics
- Gradient Boost Machine Performance and Metrics
- Final Model Selection and Summary
- Data Challenges and Resolutions

Data Source and Background Information:

The purpose of this study is to assist in the use and interpretation of intrapartum cardiotocography (CTG), as well as in the clinical management of specific CTG patterns.

Unexpected complications may occur during labor, even in patients without prior evidence of risk, so maternity hospitals need to ensure the presence of trained staff, as well as appropriate facilities and equipment for an expedite delivery.

Unnecessary obstetric intervention confers additional risks for the mother and newborn and the former may result from poor CTG interpretation, limited knowledge of the pathophysiology of fetal oxygenation, and inadequate clinical management.

FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography, Diogo Ayres-de-Campos, et al., 2015

Data Source and Background Information continued:

Cardiotocography has well-documented limitations, and it is necessary to be aware of these for safe use of the technology.

The main aspects that are prone to observer disagreement are the identification and classification of decelerations, the evaluation of variability, and the classification of tracings as suspicious and pathological.

The subjectivity of observer analysis has also been demonstrated in retrospective audit of tracings, where CTG features are frequently assessed to be more abnormal in cases with known adverse neonatal outcome.

FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography, Diogo Ayres-de-Campos, et al., 2015

Dataset Description and Research Goal:

- 2126 fetal cardiotocograms (CTGs) were processed and the respective diagnostic features measured. The CTGs were also classified by three expert obstetricians and a consensus classification label assigned to each of them. Classification was both with respect to a fetal state (N, S, P).

- The following analysis asks what predictive model best encompasses the features and target data within the CTG dataset.
    - A successful model will be highly accurate with minimal required processing time, as well as minimal difference between the training and testing datasets by resulting type one and type two errors.

Ayres de Campos et al. (2000) SisPorto 2.0 A Program for Automated Analysis of Cardiotocograms. J Matern Fetal Med 5:311-318

The International Federation of Gynecology and Obstetrics, Characteristic Guidelines:

Table 1
Cardiotocography classification criteria, interpretation, and recommended management.[a]

	Normal	Suspicious	Pathological
Baseline	110–160 bpm	Lacking at least one characteristic of normality, but with no pathological features	<100 bpm
Variability	5–25 bpm	Lacking at least one characteristic of normality, but with no pathological features	Reduced variability, increased variability, or sinusoidal pattern
Decelerations	No repetitive[b] decelerations	Lacking at least one characteristic of normality, but with no pathological features	Repetitive[b] late or prolonged decelerations during >30 min or 20 min if reduced variability, or one prolonged deceleration with >5 min
Interpretation	Fetus with no hypoxia/acidosis	Fetus with a low probability of having hypoxia/acidosis	Fetus with a high probability of having hypoxia/acidosis
Clinical management	No intervention necessary to improve fetal oxygenation state	Action to correct reversible causes if identified, close monitoring or additional methods to evaluate fetal oxygenation [49]	Immediate action to correct reversible causes, additional methods to evaluate fetal oxygenation [49], or if this is not possible expedite delivery. In acute situations (cord prolapse, uterine rupture, or placental abruption) immediate delivery should be accomplished.

[a] The presence of accelerations denotes a fetus that does not have hypoxia/acidosis, but their absence during labor is of uncertain significance.
[b] Decelerations are repetitive in nature when they are associated with more than 50% of uterine contractions [29].

FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography, Diogo Ayres-de-Campos, et al., 2015

Feature Descriptions (all numeric):

AC: accelerations
Mode: histogram mode
Mean: histogram mean
mLTV: mean value of long term variability
Median: histogram median
UC: uterine contractions
Tendency: histogram tendency: -1=left asymmetric; 0=symmetric; 1=right asymmetric
mSTV: mean value of short term variability
Width: histogram width
Max: high freq. of the histogram
Nmax: number of histogram peaks
Nzeros: number of histogram zeros
DL: light decelerations
Min: low freq. of the histogram
FM: foetal movement
DS: severe decelerations
LB: baseline value
Variance: histogram variance
ALTV: percentage of time with abnormal long term variability
ASTV: percentage of time with abnormal short term variability
DP: prolonged decelerations

N.S.P. Correlation:

AC	-0.364066
Mode	-0.250412
Mean	-0.226985
MLTV	-0.226797
Median	-0.205033
UC	-0.204894
Tendency	-0.131976
MSTV	-0.103382
Width	-0.068789
Max	-0.045265
Nmax	-0.023666
Nzeros	-0.016682
DL	0.058870
Min	0.063175
FM	0.088010
DS	0.131934
LB	0.148151
Variance	0.206630
ALTV	0.426146
ASTV	0.471191
DP	0.484859

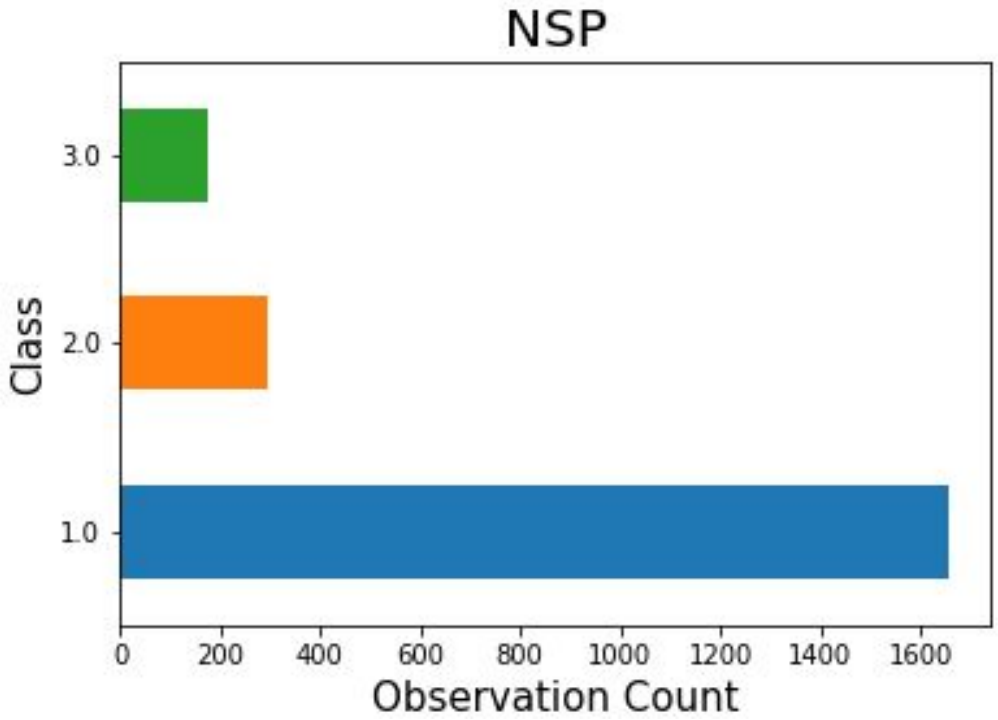
Target Class Distribution and Model Descriptions:

- Categorical target class requires classification predictive models.

- Logistic Regression: predictive analysis algorithm based on the concept of probability

- Support Vector Classifier: analysis algorithm that finds a hyperplane in an N-dimensional space that distinctly classifies the data points

- Gradient Boost Machine: create decision trees one at a time, where each new tree helps to correct errors made by previously trained tree



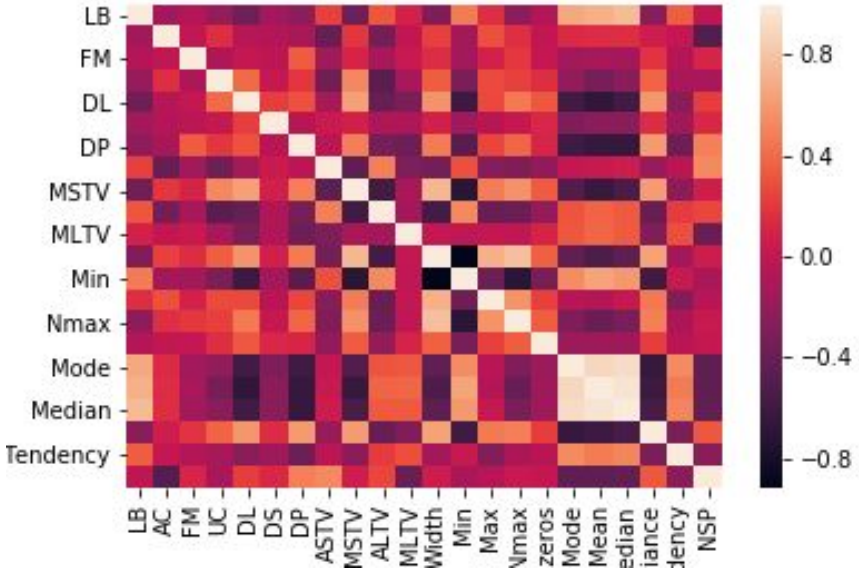
Feature Missingness and Multicollinearity:

	LB	AC	FM	UC	DL	DS	DP	ASTV	MSTV	ALTV	MLTV	Width	Min	Max	Nmax	Nzeros	Mode	Mean	Median	Variance	Tendency	NSP
2126	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2127	NaN	NaN	NaN	NaN	0.000	0.000	0.000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2128	NaN	NaN	0.481	0.015	0.015	0.001	0.005	87.0	7.0	91.0	50.7	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Data Missingness: 3 rows removed as they were found to encapsulate all missing values.

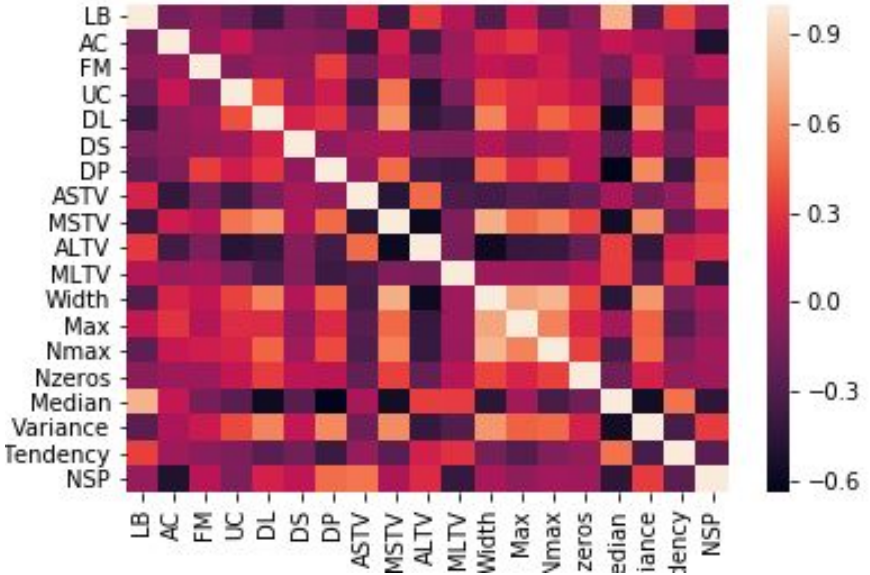
Multicollinearity: 2 sets of features were found to have high multicollinearity (|c|>.85)

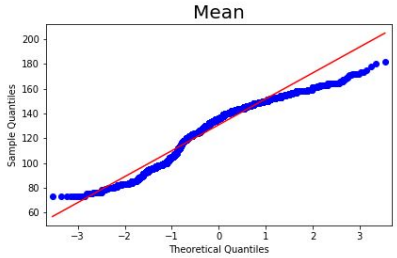
- Min, Width: corr. = -0.917
- Mean, Mode: corr. = 0.898
- Median, Mode: corr. = 0.921
- Median, Mean: corr. = 0.956

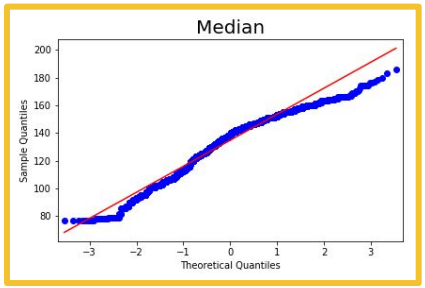


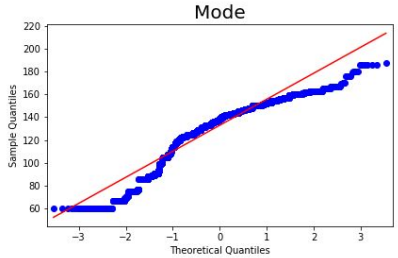
Feature Multicollinearity and Selection Continued:

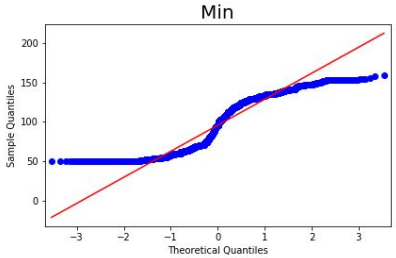
- Features from each set were assessed for their data distribution and selected for their closeness to a normal distribution's shape.

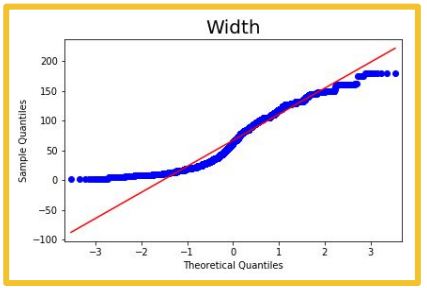












Logistic Regression Model:

- Baseline: Accuracy(recall): 86.4%, 20 fold Average Cross Validation: 86%, AVG. AUC: 92.7%

Feature Selection Methods:

- selectKbest: ANOVA F-value between label/feature for classification tasks.
  - Accuracy(recall): 86.9%, 20 fold Average Cross Validation: 86%, AUC: 92.7 %

- L1 Based Feature Elimination: each non-zero coefficient adds to the penalty, it forces weak features to have zero as coefficients.
  - Accuracy(recall): 85.7%, 20 fold Average Cross Validation: 85.6%, AUC: 92.4%

- L2 Based Feature Elimination: forces the coefficient values to be spread out more equally. For correlated features, it means that they tend to get similar coefficients.
  - Accuracy(recall): 85.4%, 20 fold Average Cross Validation: 83.9%, AUC: %92.1

Support Vector Classifier Model:

- Baseline: Accuracy(recall): 81.7%, 20 fold Average Cross Validation: 79.6%, AUC: 94.4%

Feature Selection Methods:

- selectKbest:
  - Accuracy(recall): 87.3%, 20 fold Average Cross Validation: 88.6%, AUC: 94.3%
- L1 Based Feature Elimination:
  - Accuracy(recall): 88.3%, 20 fold Average Cross Validation: 88.1%, AUC: 94.4%
- L2 Based Feature Elimination:
  - Accuracy(recall): 83.8%, 20 fold Average Cross Validation: 80.4%, AUC: 90.7%

Gradient Boost Machine Model:

- Baseline: Accuracy(recall): 96.7%, 20 fold Average Cross Validation: 90.9%, AUC: 98.4%

Feature Selection Methods:

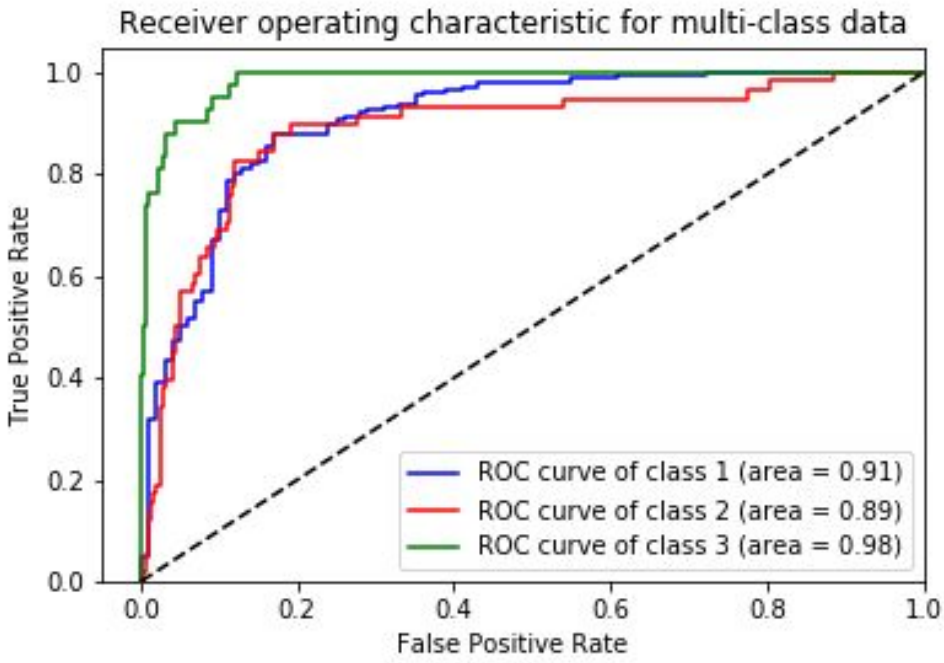
- selectKbest:
  - Accuracy(recall): 95.1%, 20 fold Average Cross Validation: 94.2%, AUC: 98%

- Recursive Feature Elimination: removes the weakest feature(s), features are ranked by the model's coef_ or feature importance attributes.
  - Accuracy(recall): 95.3%, 20 fold Average Cross Validation: 90.6%, AUC: 98.6%

- F-score Feature Elimination: ANOVA F-beta feature analysis
  - Accuracy(recall): 95.3%, 20 fold Average Cross Validation: 91.2%, AUC: 98.1%

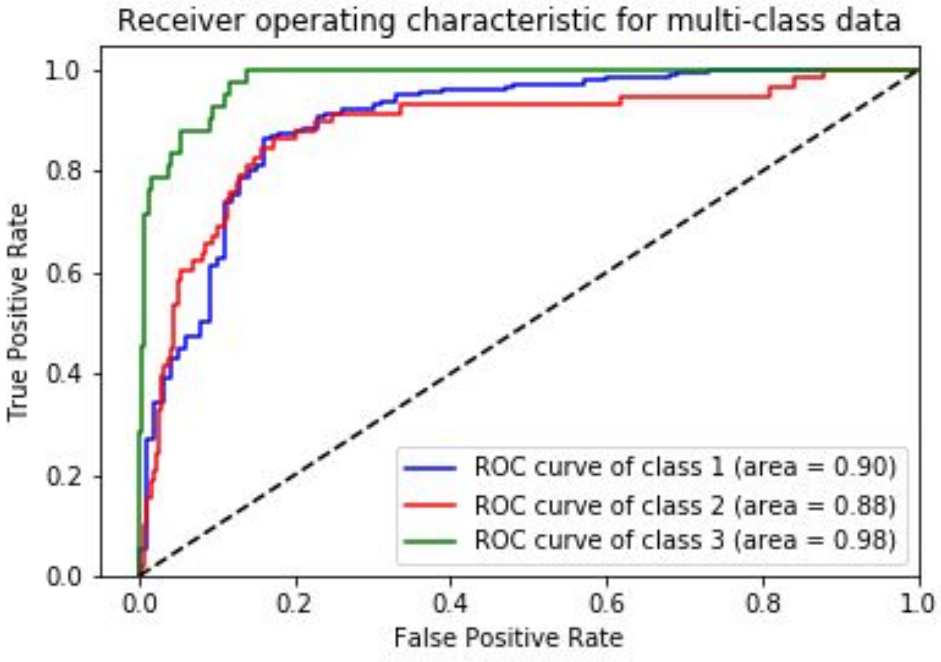
Final Logistic Regression Model:

Feature Selection Method:	Drop NA, Drop MultiColl.
Hyper Parameters:	solver: ‘Newton-cg’
Accuracy (recall)	87.8%
20 fold Average Cross Validation Score:	88.9%
Runtime:	7.64 seconds
Training set type[1,2] errors:	.1818, .2666
Testing set type[1,2] error:	.1982, .2852
Type[1,2] error difference:	.0163, .0186



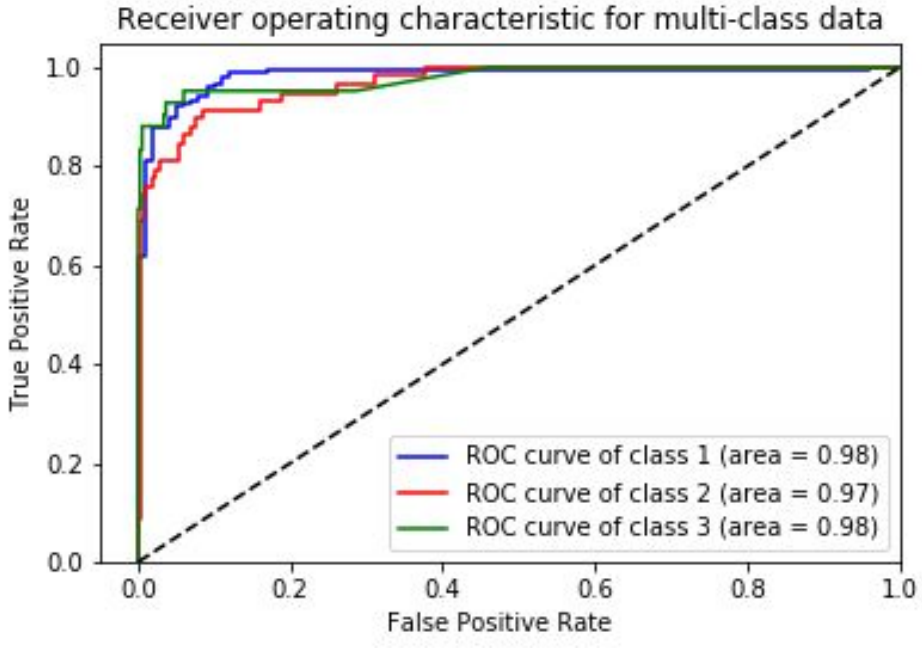
Final Support Vector Machine Model:

Feature Selection Method:	Drop NA, Drop MultiColl., ‘L1’ F.S.
Hyper Parameters:	kernel:’linear’
Accuracy (recall)	88.3%
20 fold Average Cross Validation Score:	86.25%
Runtime:	3.7035 seconds
Training set type[1,2] errors:	.1812, .208
Testing set type[1,2] error:	.2142, .2743
Type[1,2] error difference:	.033, .066



Final Gradient Boost Machine Model:

Feature Selection Method:	Drop NA, Drop MultiColl., ‘F-score” F.S.
Hyper Parameters:	n_estimators: 1000, max_depth: 6, loss: ‘deviance’, learning _rate: .2
Accuracy (recall)	96.4%
20 fold Average Cross Validation Score:	93%
Runtime:	35.93 seconds
Training set type[1,2] errors:	.0016, .0016
Testing set type[1,2] error:	.0435, .1078
Type[1,2] error difference:	.0419, .1062



Final Model Selection and Summary:

In final analysis, all models predicted equally well with respect to the most severe target class (pathological). However the Gradient Boost Machine model had the least resulting predictive ROC variance. Though we must recognize this model did require the longest runtime and did exhibit the most amount of type 1 and type 2 error difference between testing and training datasets, indicating the more overfitting error than the Support Vector or the Logistic Models. In practical terms, the qualifications that determine the success of a predictive model will always be subject to the modalities of target characteristics.

- With respect to the serious nature of the dataset's predictive class, the Gradient Boost Machine Model is chosen as most effective for its accuracy.
- Although the GBM's resulting figures outlined a measure of overfitting error, the SVC and Logistic models were less accurate on validation datasets than the GBM model.

Data Processing Challenges and Resolutions:

Data Leakage: information from outside the training dataset is used to create the model. Data leakage can cause you to create overly optimistic if not completely invalid predictive models.

- Causes: Data Standardization, Up/Down sample dataset balancing
  - "The data rescaling process that you performed had knowledge of the full distribution of data in the training dataset when calculating the scaling factors (like min and max or mean and standard deviation). This knowledge was stamped into the rescaled values and exploited by all algorithms in your cross validation test harness."
- Corrections: Perform data preparation within cross-validation folds, hold back a validation dataset
  - "Essentially the only way to really solve this problem is to retain an independent test set and keep it held out until the study is complete and use it for final validation."

Data Leakage in Machine Learning, by Jason Brownlee

Notebook link:

https://github.com/coreycoole/Thinkful-Capstone-3-Model-Comparison