THE DATA

- UC Irvine Machine Learning Repository
- Regensburg Pediatric Appendicitis - Children's Hospital St. Hedwig in Regensburg, Germany
- 782 observations
- 58 variables

OBJECTIVE

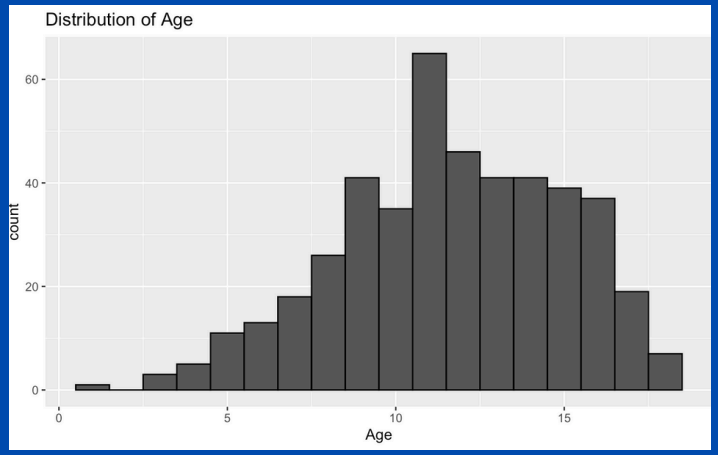
To figure out the factors that might be associated with the diagnosis of pediatric appendicitis

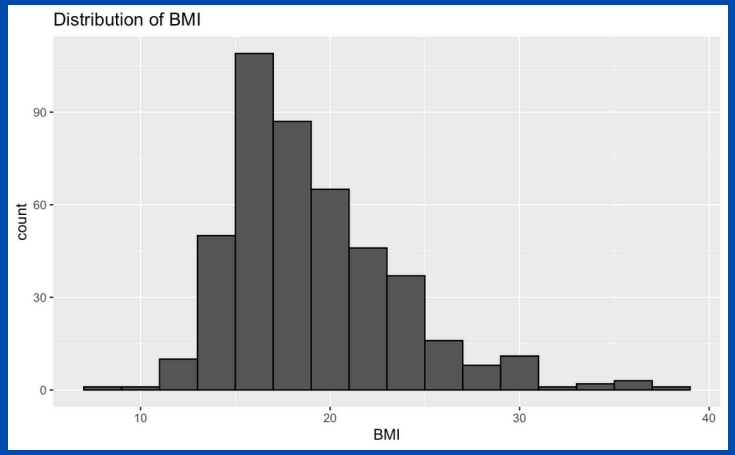
Build models to predict whether someone has appendicitis or not using classification

VARIABLES

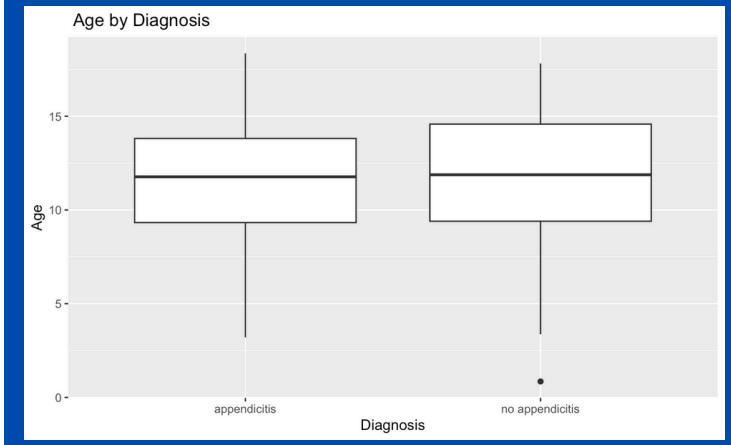
6 Variable Groups:
Demographic / Other: Age, Sex, BMI
Scoring: Pediatric Appendicitis Score, Alvarado Score
Clinical: Coughing Pain, Abdominal Pain, Body Temperature
Laboratory: Hemoglobin, WBC, RBC
Ultrasound: Appendix diameter, Appendix perfusion
Diagnosis / Management / Severity: Diagnosis (Response) - (Appendicitis/No Appendicitis)

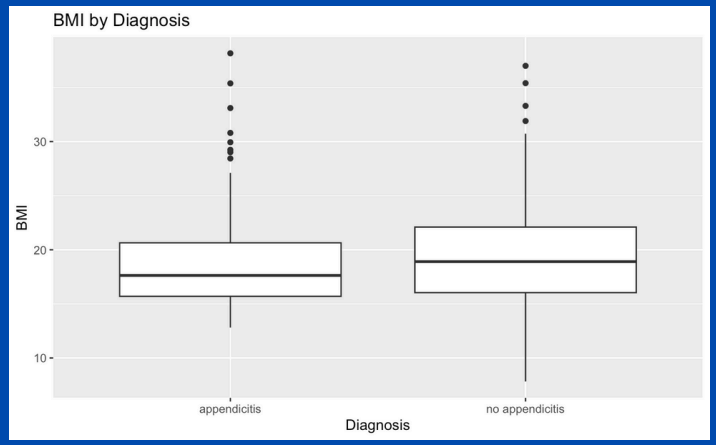
EXPLORATORY DATA ANALYSIS



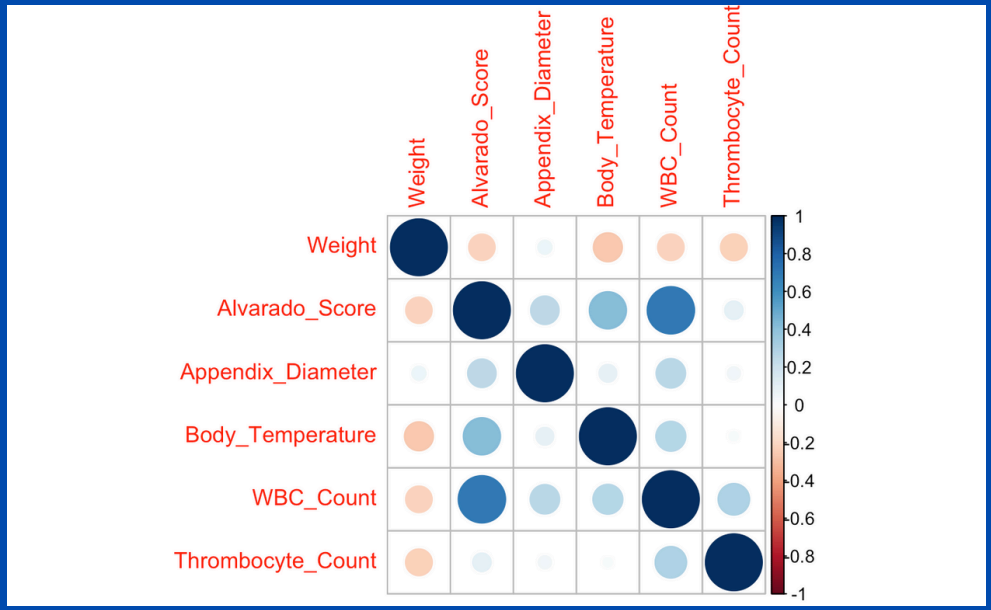


EXPLORATORY DATA ANALYSIS





EXPLORATORY DATA ANALYSIS



SUBSET SELECTION

- Used stepwise regression to choose variables
- Used AIC and BIC
- “Best” AIC model had 11 predictors
- “Best” BIC model had 4 predictors
- AIC was used to build ML models

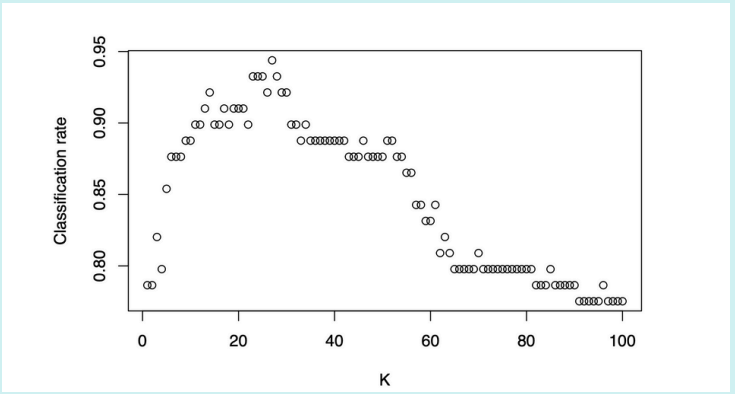
SUBSET SELECTION

- BIC Model: Diagnosis ~ Management + Alvarado_Score + Appendix_on_US + Appendix_Diameter

- AIC Model: Weight + Management + Alvarado_Score + Appendix_on_US + Appendix_Diameter + Coughing_Pain + Nausea + Body_Temperature + WBC_Count + Thrombocyte_Count + Neutrophilia

MODEL 1: KNN

- First, set k = 6 and the classification rate was approx. 85%
- Then, found the optimal k which was 27
- There was an increase in the classification rate
- Classification rate: 94.38%



MODEL 2: LOGISTIC REGRESSION

- Using AIC Model
- Classification Rate: 21.35%
- Much lower than KNN model

	appendicitis	no appendicitis
appendicitis	0	40
no appendicitis	30	19

[1] 0.2134831

MODEL 3: LDA

- Using AIC Model
- Classification Rate: 90.84%

	appendicitis	no appendicitis
appendicitis	172	31
no appendicitis	10	235

[1] 0.9084

MODEL 4: QDA

- Using AIC Model
- Classification Rate: 65.63%

	appendicitis	no appendicitis
appendicitis	128	75
no appendicitis	79	166

[1] 0.65625

RESULTS AND RECOMMENDATIONS

- Prediction accuracy rate
- KNN: 94.38%
- Logistic Regression: 21.35%
- LDA: 90.85%
- QDA: 65.63%
- KNN has the highest correct prediction accuracy rate; therefore, KNN is recommended for this data

REFERENCES

- Regensburg pediatric appendicitis. UCI Machine Learning Repository. (n.d.). https://archive.ics.uci.edu/dataset/938/regensburg+pediatric+appendicitis