Project 5: Predicting Appendicitis in Pediatric Patients

Introduction
The appendix is a small organ attached to the large intestine. Although its exact function is not fully understood, it is believed to contain lymphoid tissues that contribute to immune responses within the gut. The appendix itself can become inflamed and infected, leading to appendicitis, a condition that can be life-threatening if the organ ruptures and spreads infection. Appendicitis primarily affects children, and early diagnosis is critical to prevent complications. To better understand the factors that predict appendicitis and its severity, doctors at Regensburg Pediatric Hospital in Germany have collected patient data. The objective of this project is to use Naive Bayes classifiers to predict the presence of appendicitis, assess its severity, and determine the likely course of treatment based on clinical data.

Github: some45bucks/cs5830_project5 (github.com)
Presentation: Project 5 CS5830

Dataset
The dataset comprises 782 instances and 53 features of pediatric patients admitted to the hospital with suspected appendicitis. It can be accessed via the UCI Machine Learning Repository: Regensburg Pediatric Appendicitis. Preprocessing involved minimal data cleaning, including one-hot encoding and the removal of missing values where necessary.

Analysis
Before applying the Naive Bayes Classifier, we first examined which variables are most relevant to the target outcomes: “diagnosis,” “severity,” and “management.” To do this, we conducted non-parametric statistical tests. For numerical variables, which were not normally distributed, we used the Mann-Whitney U test and Kruskal-Wallis test. For categorical variables, we performed a Chi-Squared test of independence.

For the Naive Bayes Classifier, we employed the Gaussian Naive Bayes algorithm and encoded binary variables as 0s and 1s.

Results
To identify the key predictors of whether a patient had appendicitis, we analyzed the numerical variables using the Mann-Whitney U test Applying a Bonferroni correction with a significance threshold of 0.002, we identified several significant variables. For categorical variables, we applied the Chi-Squared test with a Bonferroni-corrected threshold of 0.001.

	feature	statistic	pvalue
0	Length_of_Stay	106349.5	1.266094e-41
1	Alvarado_Score	89690.5	2.108374e-28
2	Paedriatic_Appendicitis_Score	84649.0	4.363695e-20
6	BMI	58344.5	3.169961e-04
7	Appendix_Diameter	39363.0	1.708849e-49
8	WBC_Count	31394.5	2.640302e-17
9	Neutrophil_Percentage	26156.0	5.295038e-15
10	Body_Temperature	25740.5	6.376481e-05

	feature	chi2	pvalue
0	Diagnosis	775.860274	9.562521e-171
1	Management	311.122927	3.893230e-67
2	Appendix_on_US	146.366989	1.079226e-33
3	Peritonitis	108.238182	3.136010e-24
4	Severity	90.271775	2.075930e-21
5	Neutrophilia	67.169617	2.491240e-16
6	Lymph_Nodes_Location	56.788538	2.852986e-04
7	Target_Sign	35.216949	2.949454e-09
8	Free_Fluids	32.476397	1.206490e-08
9	Surrounding_Tissue_Reaction	31.390721	2.109869e-08

Key variables associated with appendicitis included:

- Length of hospital stay
- Appendix diameter
- Appendix visibility on ultrasound (US)
- Alvarado Score
- Presence of peritonitis

Notably, appendix diameter, appendix visibility in ultrasound, and Alvarado Score are critical diagnostic criteria used by physicians. Length of stay and peritonitis are associated with complications arising from appendicitis, where patients diagnosed with appendicitis tend to have longer hospital stays and face a higher risk of peritonitis if the appendix ruptures.

We next examined which variables predicted the severity of appendicitis, distinguishing between "uncomplicated" (non-ruptured) and "complicated" (ruptured) cases. The significant variables were determined using the same statistical tests as before.

	feature	statistic	pvalue
0	Length_of_Stay	453.947348	2.670459e-99
1	Alvarado_Score	124.883524	7.618812e-28
2	Appendix_Diameter	120.733048	6.069485e-27
3	WBC_Count	72.377256	1.920784e-16
4	Paedriatic_Appendicitis_Score	69.215519	9.333425e-16
5	Neutrophil_Percentage	68.624839	1.254023e-15
6	Body_Temperature	36.980209	9.329313e-09
7	Age	27.238371	1.216922e-06
8	Weight	18.854354	8.050613e-05
9	Height	15.683100	3.930593e-04

	feature	chi2	pvalue
0	Management	2343.000000	0.000000e+00
1	Diagnosis	311.122927	3.893230e-67
2	Peritonitis	276.073099	1.088275e-56
3	Severity	243.450085	1.707669e-52
4	Lymph_Nodes_Location	109.876814	2.210592e-06
5	Neutrophilia	92.536642	6.246545e-20
6	Appendix_Wall_Layers	70.175233	3.764103e-13
7	Free_Fluids	63.242792	1.191678e-13
8	Ketones_in_Urine	59.940926	1.376281e-09
9	Surrounding_Tissue_Reaction	41.295541	1.078418e-09

The top predictors for severe appendicitis included:

- Length of hospital stay
- Peritonitis
- Alvarado Score
- Pediatric Appendicitis Score

- Neutrophilia (elevated neutrophil count)

Peritonitis is a direct indicator of severe appendicitis, while the Alvarado Score, Pediatric Appendicitis Score, and neutrophilia are valuable clinical markers for assessing whether the appendix has ruptured.

Finally, we sought to predict the treatment course, which included conservative management, primary surgery, and secondary surgery. We used the Kruskal-Wallis test to compare the three treatment groups.

	feature	statistic	
0	Length_of_Stay	69243.5	2.310
1	Alvarado_Score	45647.5	2.560
2	Paedriatic_Appendicitis_Score	43044.0	8.767
3	Age	30675.5	1.203
4	BMI	28751.5	2.852
5	Weight	28154.0	9.343
6	Appendix_Diameter	17844.5	6.216
7	WBC_Count	17406.5	1.086
8	Body_Temperature	17173.0	4.185
9	Neutrophil_Percentage	14469.0	1.531

	feature	chi2	pvalue
0	Severity	773.276407	3.486379e-170
1	Management	243.450085	1.707669e-52
2	Peritonitis	97.993394	5.260231e-22
3	Diagnosis	90.271775	2.075930e-21
4	Neutrophilia	50.570661	1.149506e-12
5	Appendix_Wall_Layers	45.051691	9.021587e-10
6	Ketones_in_Urine	42.049269	3.916691e-09
7	Nausea	37.392456	9.659505e-10
8	Free_Fluids	35.066774	3.185907e-09
9	Loss_of_Appetite	34.108501	5.212306e-09

The significant variables influencing treatment decisions were:

- Length of hospital stay
- Peritonitis
- Alvarado Score
- Appendix diameter
- Neutrophilia

These variables were consistently significant across the different outcomes of interest.

Type	F1	Precision	Recall
Diagnosis	**.84**	.86	.84
Severity	**.90**	.92	.88
Management	**.80**	.82	.80

For each target (diagnosis, severity, and treatment), we applied the Naive Bayes Classifier using only the significant variables identified in the previous analysis. The F1 scores are a weighted

average between all categories scores. Interestingly, better F1 scores were achieved when excluding significant categorical variables from the model. In all cases, the F1 scores were above 0.8, indicating strong predictive performance.

Technical

We conducted data loading, preprocessing, statistical analysis, and model training to predict appendicitis outcomes using the Regensburg Pediatric Appendicitis dataset. Initially, we loaded the dataset, handled missing values, and performed one-hot encoding on categorical variables while converting boolean values to integers. Numerical and categorical features were identified separately, excluding the target variables ("Diagnosis," "Management," and "Severity"). Due to the non-normal distribution of numerical data, we employed non-parametric statistical tests for variable selection: the Mann-Whitney U test for binary targets and the Kruskal-Wallis test for multi-class targets. Categorical features were analyzed using the Chi-Squared test of independence, and significant features were determined by applying the Bonferroni correction to adjust for multiple comparisons.

Using the significant features identified, we trained Gaussian Naive Bayes classifiers for each target variable. The dataset was split into training and testing sets multiple times to ensure robust evaluation. For each split, the model was trained and its performance was assessed using metrics such as accuracy, precision, recall, and F1-score. We averaged the results over numerous iterations to obtain reliable performance metrics. The classifiers achieved high predictive accuracy, with F1-scores exceeding 0.8 for all targets, indicating strong predictive capability. This approach demonstrates the effectiveness of feature selection and Gaussian Naive Bayes classification in medical diagnosis applications, specifically for predicting appendicitis and its severity in pediatric patients.