



Fetal Health Prediction

BY ABDULAZIZ & IBRAHIM

Problem - Importance

- 211 death per 100,000 live birth (2017).
- Most could have been **PREVENTED**.



Problem - Goal

Classify **fetal health** in order to prevent child and maternal mortality.



Normal



Suspect



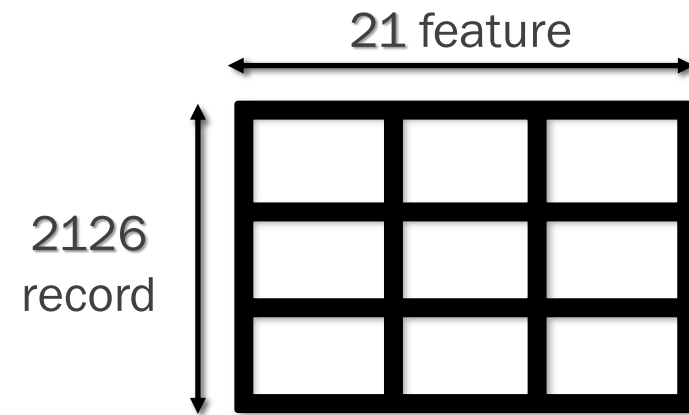
Pathological

Data

- Collected from Cardiotocograms (CTGs).
- Labeled by 3 experts.

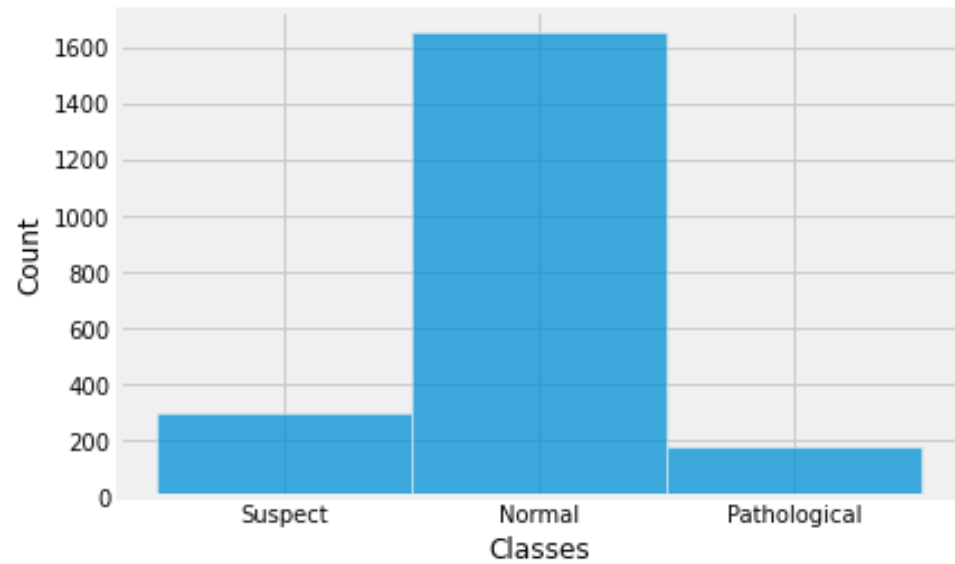


Data

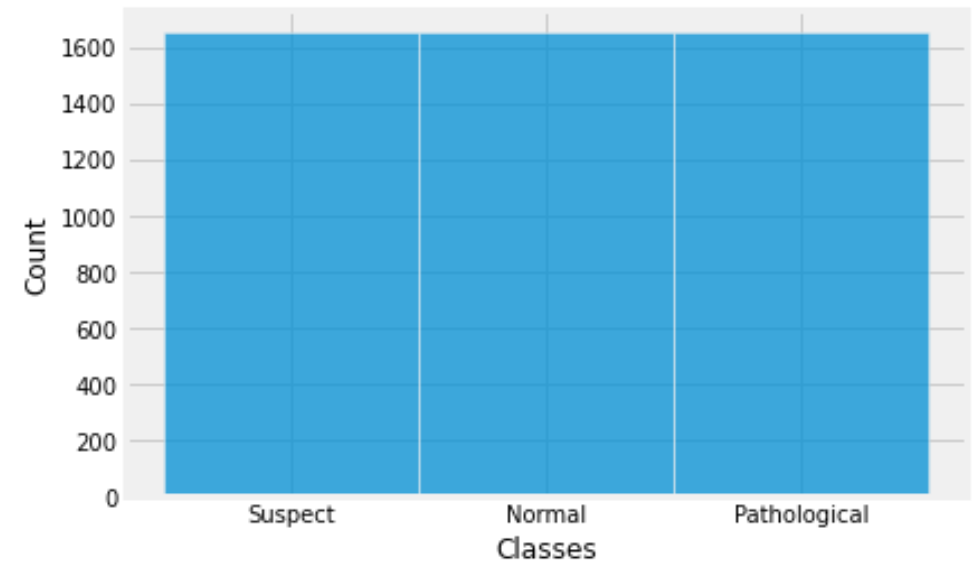


EDA – Balance Data (SMOTE)

ORIGINAL Y



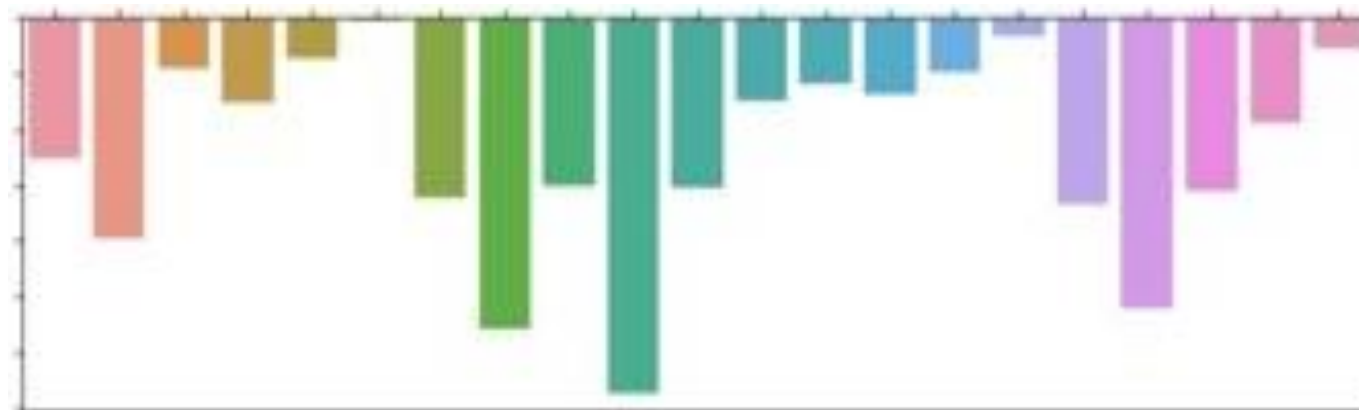
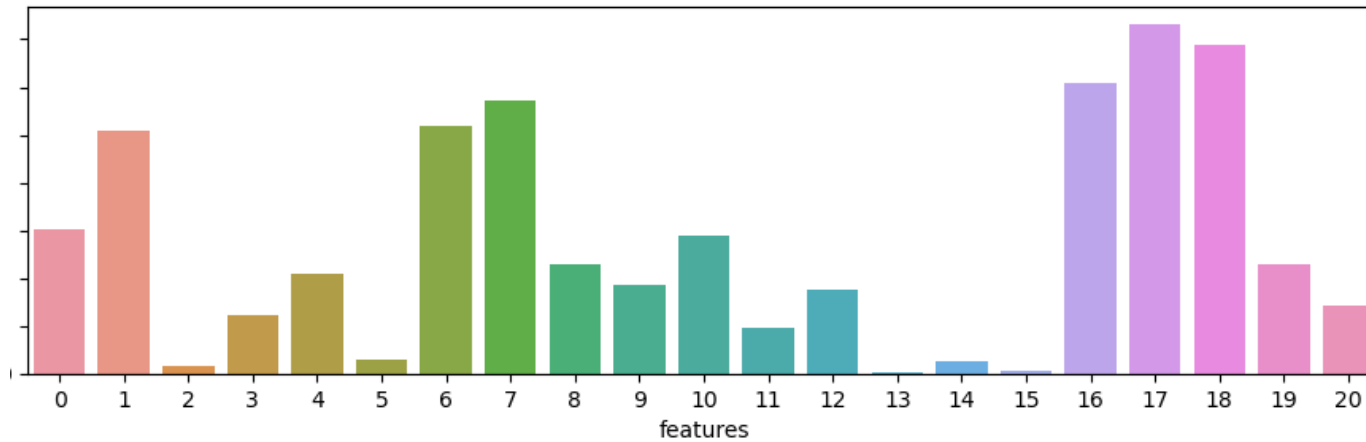
BALANCED Y



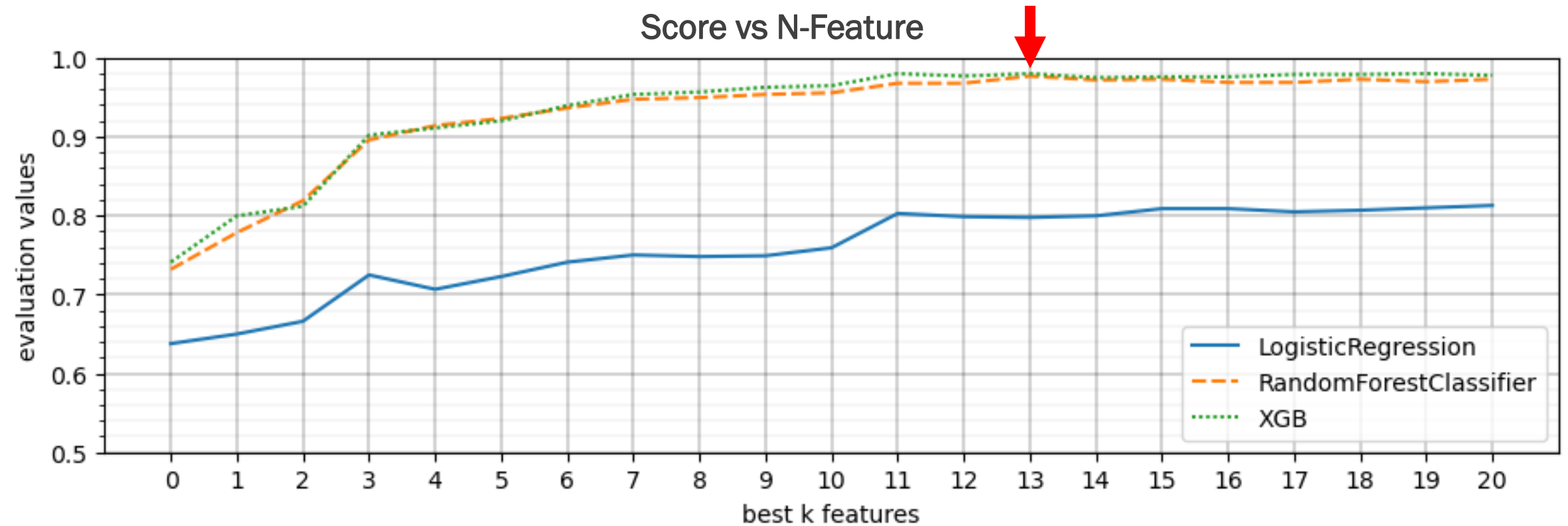
EDA – Cleaning ?

Duplicated	24 rows
Missing Values	None
Outliers	None

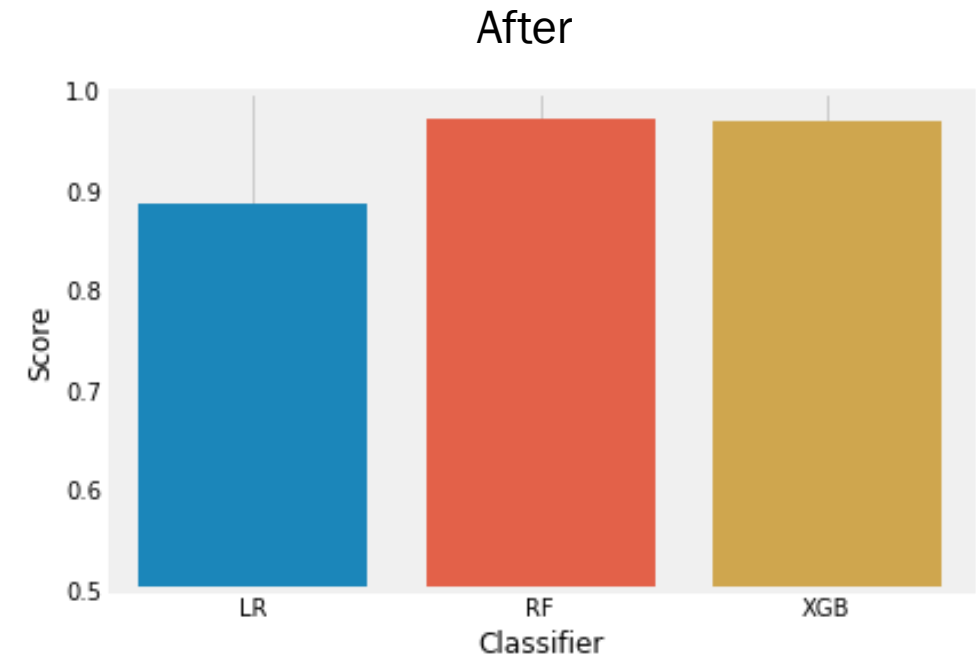
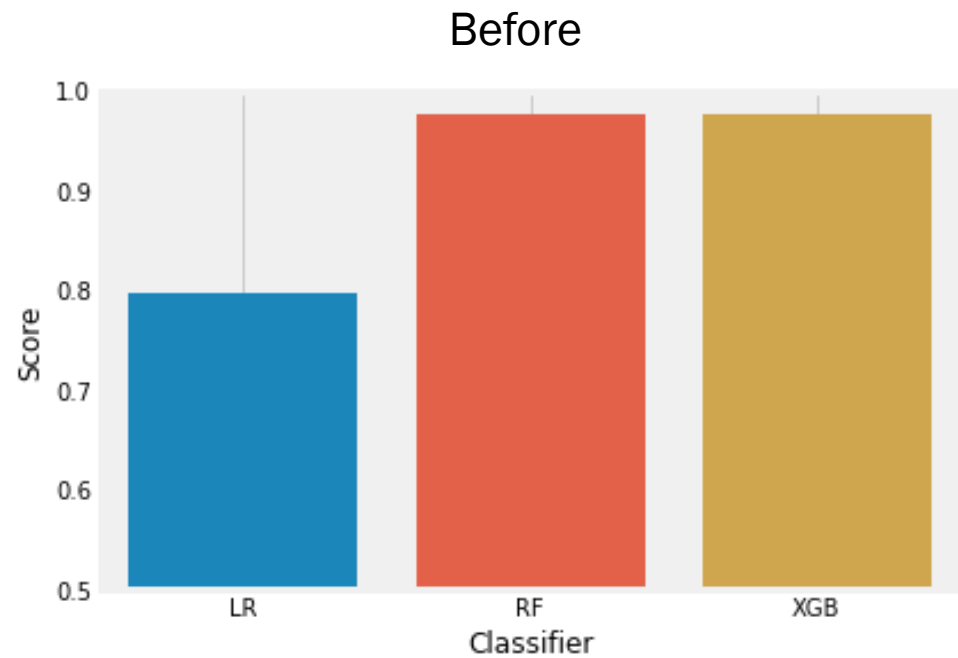
Feature Selection



Feature Selection



Finetuning - GridSearchCV



Results

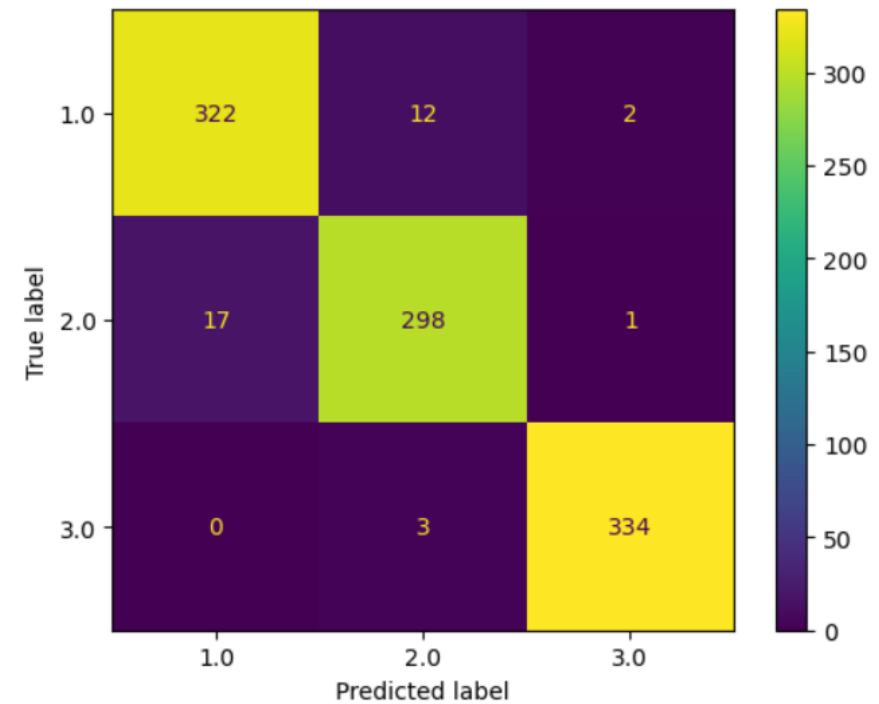
XGBoost classifier 14 Features

Precision: 96.434

Recall 96.415

Accuracy: 96.461

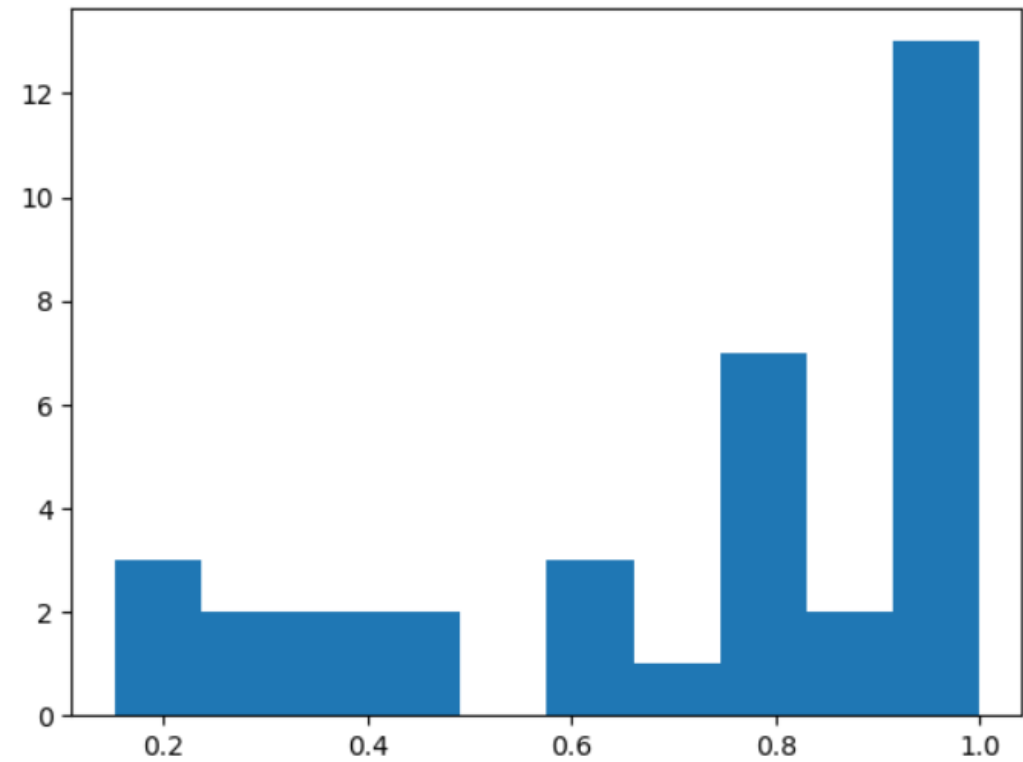
ROC over AUC: 99.622



Results

Examining Misclassified probabilities differences

About 13 predictions were incorrectly classified with a very high confidence



Minimum difference of probability to other classes

AutoML

AutoML is google cloud service that automatically find the best model and hyper parameter to use.

Target	Feature columns	Optimized for	AUC PR ?	AUC ROC ?	Precision ?	Recall ?	Log loss ?
fetal_health	21 included 200 test rows	Log loss	0.984	0.990	96.0%	95.0%	0.163

→ EXPORT PREDICTIONS ON TEST DATASET TO BIGQUERY

You have up to 30 days to export your test dataset

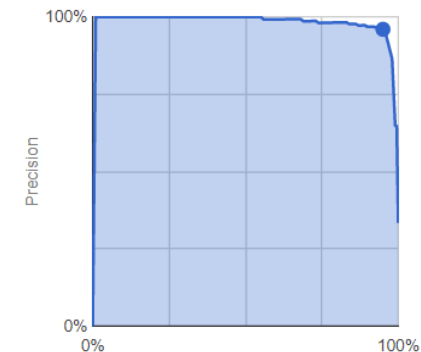
Filter	Filter labels
All	
1	
2	
3	

All

Score threshold 0.50

F1 score ?	0.955
Precision ?	96.0% (190/198)
True positive rate (Recall) ?	95.0% (190/200)
False positive rate ?	0.020 (8/400)

The score threshold determines the minimum level of confidence needed to make a prediction positive. [Learn more about model evaluation](#)



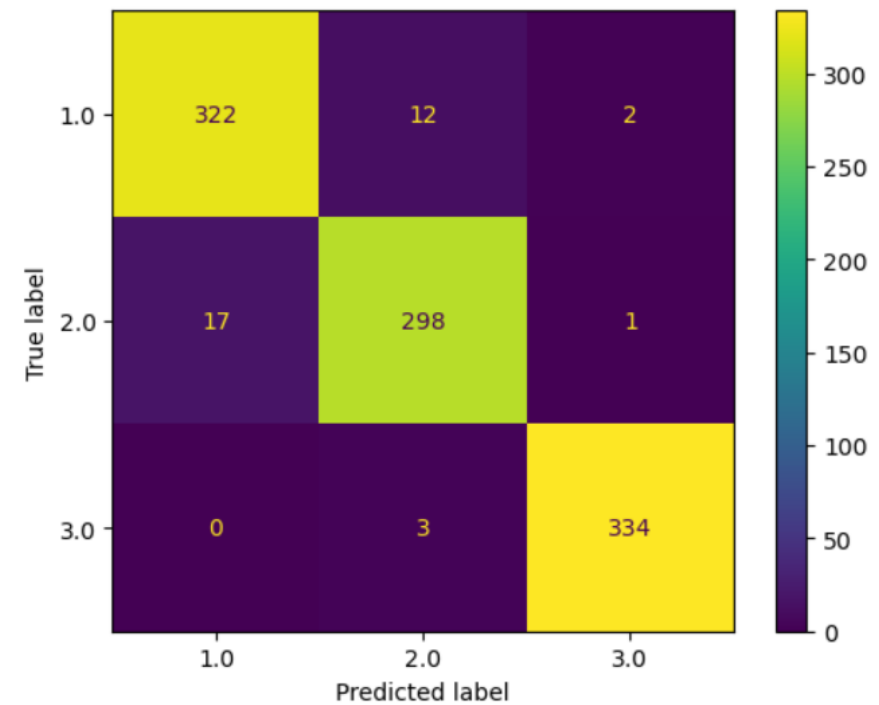
AutoML

Confusion matrix

A confusion matrix summarizes how successful a classification model's predictions are. The rows are the true labels and the columns are the predicted labels. Scan each row to determine how many accurate predictions for the label (true positives or negatives) while the columns help you learn more.

True labels	Predicted labels		
	1	2	3
1	98%	1%	1%
2	15%	85%	-
3	8%	-	92%

Google Cloud AutoML



Our XGBoost model

Thank You
