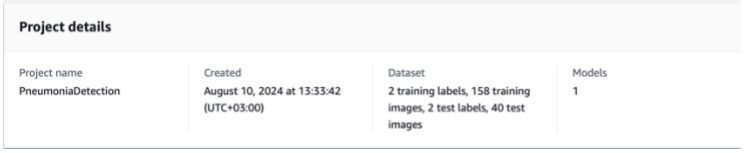
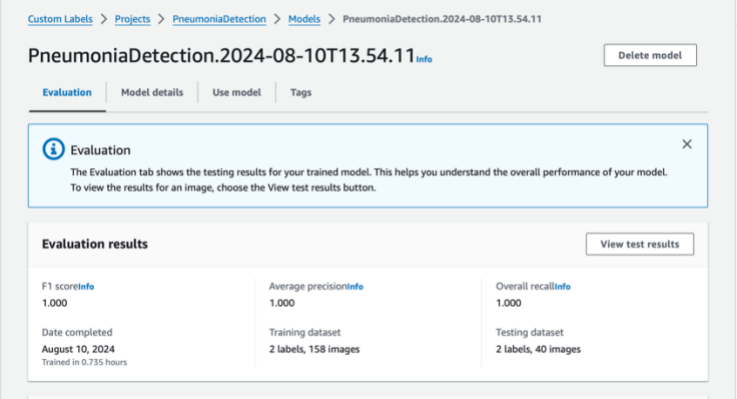
AutoML Modeling Report

Nouv B. Al-Qahtani

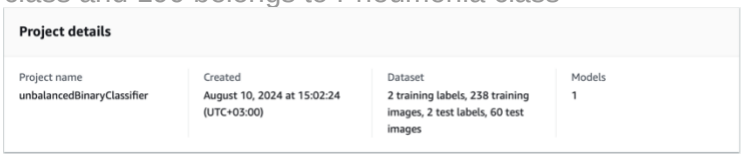
Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	Initially, 198 images were used. Out of which 99 to Normal class and remaining 99 belongs to pneumonia class to create a Clean/Balanced data.  In which the 80% of the images are used for Training and 20% are used for Testing.
Confusion Matrix What do each of the sections in the confusion matrix describe? What values did you observe (include a screenshot)? What is the true positive rate for the “pneumonia” class? What is the false positive rate for the “normal” class?	A **confusion matrix** is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. • True positives (TP): These are cases in which we predicted yes for the data • True negatives (TN): We predicted no for the data • False positives (FP): We predicted yes, but it’s not true (Also known as a "Type I error.") • False negatives (FN): We predicted no, but it’s true (Also known as a "Type II error.")

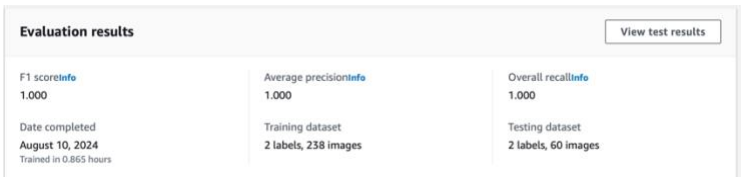
	EvaluationResultSummary-PneumoniaDetection-PneumoniaDetection.2024-08-10T13.54.11 〉 No Selection
	{
	"AggregatedEvaluationResults": {
	"ConfusionMatrix": [
	{
	"GroundTruthLabel": "normal",
	"PredictedLabel": "normal",
	"Value": 0.95
	},
	{
	"GroundTruthLabel": "normal",
	"PredictedLabel": "pneumonia",
	"Value": 0.05
	},
	{
	"GroundTruthLabel": "pneumonia",
	"PredictedLabel": "normal",
	"Value": 0.0
	},
	{
	"GroundTruthLabel": "pneumonia",
	"PredictedLabel": "pneumonia",
	"Value": 1.0
	}
	],
	"F1Score": 1.0,
	"Precision": 1.0,
	"Recall": 1.0
	The true positive rate for the "pneumonia" class is 1.0 (the model has a 100% success rate in detecting pneumonia cases)
	The false positive rate for the "normal" class is 0.05 (the model mistakenly identified 5% of the normal cases as having pneumonia)

Precision and Recall What does precision measure? What does recall measure? What precision and recall did the model achieve (report the values for a score threshold of 0.5)?	 The **precision** is the proportion of relevant results in the list of all returned search results. The **recall** is the ratio of the relevant results returned by the search engine to the total number of the relevant results that could have been returned. The model achieved a precision of 100% and recall of 100%.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	298 images were used, in which 99 belongs to Normal class and 199 belongs to Pneumonia class  **Project details** Project name: unbalancedBinaryClassifier Created: August 10, 2024 at 15:02:24 (UTC+03:00) Dataset: 2 training labels, 238 training images, 2 test labels, 60 test images Models: 1 80% of the images are used for Training, 20% used for Testing.

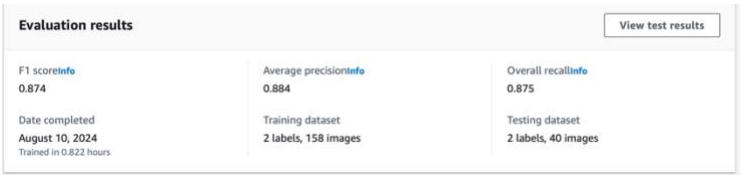
Confusion Matrix
How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix summary
	EvaluationResultSummary-unbalancedBinaryClassifier-unbalancedBinaryClassifier.2024-08-10T15.37.41 〉 No Selection
	{
	"AggregatedEvaluationResults": {
	"ConfusionMatrix": [
	{
	"GroundTruthLabel": "normal",
	"PredictedLabel": "normal",
	"Value": 0.95
	},
	{
	"GroundTruthLabel": "normal",
	"PredictedLabel": "pneumonia",
	"Value": 0.05
	},
	{
	"GroundTruthLabel": "pneumonia",
	"PredictedLabel": "normal",
	"Value": 0.0
	},
	{
	"GroundTruthLabel": "pneumonia",
	"PredictedLabel": "pneumonia",
	"Value": 1.0
	}
	],
	"F1Score": 1.0,
	"Precision": 1.0,
	"Recall": 1.0
	},
	The confusion matrix shows that the model has a high true positive rate for the "pneumonia" class (1.0) but a relatively high false positive rate for the "normal" class (0.05). This indicates that the model is biased towards the majority "pneumonia" class due to the unbalanced dataset, where there are more pneumonia samples than normal samples in the training data.

Precision and Recall How have the model's precision and recall been affected by the unbalanced data?	 **Evaluation results** View test results F1 score: 1.000 Average precision: 1.000 Overall recall: 1.000 Date completed: August 10, 2024 Trained in 0.865 hours Training dataset: 2 labels, 238 images Testing dataset: 2 labels, 60 images The model's precision and recall are both 1.0, indicating excellent overall performance. However, this is likely due to the unbalanced nature of the dataset, where the model has learned to perform well on the majority "pneumonia" class, while the performance on the

	minority "normal" class may be less reliable.
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	Unbalanced classes in a dataset can significantly bias a machine learning model towards the majority class, leading to poor performance on the minority class. This affects evaluation metrics and makes it challenging for the model to converge during training.

Binary Classifier with Dirty/Balanced Data

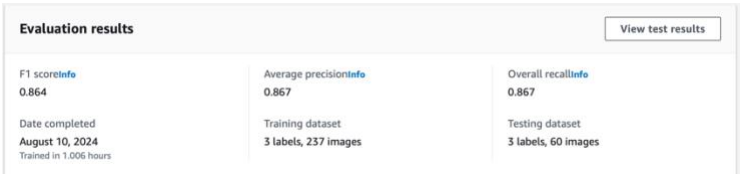
Confusion Matrix
How has the confusion matrix been affected by the dirty data? Include a screenshot of the new confusion matrix information.
	EvaluationResultSummary-binaryClassifierWithDirtyData-binaryClassifierWithDirtyData.2024-08-10T16.13.21 〉 No Selectio
	{
	"AggregatedEvaluationResults": {
	"ConfusionMatrix": [
	{
	"GroundTruthLabel": "normal",
	"PredictedLabel": "normal",
	"Value": 0.4
	},
	{
	"GroundTruthLabel": "normal",
	"PredictedLabel": "pneumonia",
	"Value": 0.6
	},
	{
	"GroundTruthLabel": "pneumonia",
	"PredictedLabel": "normal",
	"Value": 0.0
	},
	{
	"GroundTruthLabel": "pneumonia",
	"PredictedLabel": "pneumonia",
	"Value": 1.0
	}
	],
	"F1Score": 0.8742928975487115,
	"Precision": 0.8836317135549872,
	"Recall": 0.875
	},
	with a high false positive rate for the "normal" class (60% of normal instances misclassified as pneumonia), while perfectly classifying the "pneumonia" class. This indicates the model has become biased towards the majority "pneumonia" class, likely due to noisy or mislabeled data in the training set.

Precision and Recall How have the model's precision and recall been affected by the dirty data. Of the binary classifiers, which has the highest precision? Which has the highest recall?	 The Precision and Recall of the model went down to 0.884 and 0.875. Of the binary classifiers, the balanced, unbalanced data has the highest Precision and recall of 1.00.

Dirty Data From what you have observed, how does dirty data affect a machine learning model?	Dirty data leads to: • Trade-offs between precision and recall for different classes • Unbalanced performance across classes • Overall degradation in model metrics like F1-score

3-Class Model

Confusion Matrix
Summarize the 3-class confusion matrix. Which classes is the model most likely to confuse? Which class(es) is the model most likely to get right? Why might you do to try to remedy the model's "confusion"? Include a screenshot of the new confusion matrix information.
	EvaluationResultSummary-three-class-three-class.2024-08-10T15.57.39 ⟩ No Selection
	{
	"AggregatedEvaluationResults": {
	"ConfusionMatrix": [
	{
	"GroundTruthLabel": "bacterial_pneumonia",
	"PredictedLabel": "bacterial_pneumonia",
	"Value": 0.85
	},
	{
	"GroundTruthLabel": "bacterial_pneumonia",
	"PredictedLabel": "normal",
	"Value": 0.1
	},
	{
	"GroundTruthLabel": "bacterial_pneumonia",
	"PredictedLabel": "viral_pneumonia",
	"Value": 0.05
	},
	{
	"GroundTruthLabel": "normal",
	"PredictedLabel": "bacterial_pneumonia",
	"Value": 0.0
	},
	{
	"GroundTruthLabel": "normal",
	"PredictedLabel": "normal",
	"Value": 1.0
	},
	{
	"GroundTruthLabel": "normal",
	"PredictedLabel": "viral_pneumonia",
	"Value": 0.0
	},
	{
	"GroundTruthLabel": "viral_pneumonia",
	"PredictedLabel": "bacterial_pneumonia",
	"Value": 0.4
	},
	{
	"GroundTruthLabel": "viral_pneumonia",
	"PredictedLabel": "normal",
	"Value": 0.15
	},
	{
	"GroundTruthLabel": "viral_pneumonia",
	"PredictedLabel": "viral_pneumonia",
	"Value": 0.45
	}
	],
	"F1Score": 0.8636811229834486,
	"Precision": 0.867306052855925,
	"Recall": 0.8666666666666667
	The model is most likely to confuse "bacterial pneumonia" and "viral pneumonia" classes, with 5% of "bacterial pneumonia" predicted as "viral pneumonia" and 40% of "viral pneumonia" predicted as "bacterial pneumonia".
	The model performs best on the "normal" class, correctly predicting 100% of the samples.
	To improve the model's performance, you could:
	1. Collect more training data for the pneumonia classes
	2. Explore better feature engineering or model architectures to capture differences between

	pneumonia types 3. Use class-weighted loss functions to focus the model on the more challenging pneumonia classes
Precision and Recall What are the model's precision and recall? How are these values calculated?	 The model's overall precision is 0.867 and recall is 0.867. Precision measures the fraction of true positive predictions out of all positive predictions. Recall measures the fraction of true positive predictions out of all actual positive instances
F1 Score What is this model's F1 score?	The F1 score of the model is 0.864.