Don't Be Fooled by Accuracy: Understanding its Limitations in Machine Learning

Accuracy Using Confusion Matrix :

		Predicted Values: Positive	Predicted Values: Negative
Actual Values	Positive	True Positives (TP)	False Negatives (FN)
	Negative	False Positive (FP)	True Negatives (TN)

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

What is Accuracy?

Accuracy Using Confusion Matrix :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Key Components:

- True Positive (TP): Correctly predicted positive cases.
- True Negative (TN): Correctly predicted negative cases.
- False Positive (FP): Incorrectly predicted as positive (also known as Type I error).
- False Negative (FN): Incorrectly predicted as negative (also known as Type II error).

Another Definition:

- Accuracy is defined as the ratio of the number of correct predictions to the total number of predictions.
- Mathematically, it is expressed as:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

Calculating Accuracy Using a Confusion Matrix

Accuracy Using Confusion Matrix :

Example:

- Suppose we have the following confusion matrix for a model predicting whether an email is spam or not:

	Predicted Spam	Predicted Not Spam
Actual Spam	40	10
Actual Not Spam	5	45

Here:

- TP (True Positives) = 40
- TN (True Negatives) = 45
- FP (False Positives) = 5
- FN (False Negatives) = 10

Using the accuracy formula:

$$\text{Accuracy} = \frac{40 + 45}{40 + 45 + 5 + 10} = \frac{85}{100} = 0.85 \text{ or } 85\%$$

Balanced Dataset: When Accuracy is a Reliable Metric

What is a Balanced Dataset?

- A balanced dataset is one where the number of instances in each class is roughly equal.
- For example, if we are classifying whether an email is spam or not spam, a balanced dataset might have 500 spam emails and 500 non-spam emails.

Example of a Balanced Dataset:

- Scenario: We have a dataset of 1,000 emails, with 500 spam and 500 non-spam emails.
- Model Prediction: Let's say our model predicts the following:

	Predicted Spam	Predicted Not Spam
Actual Spam	450 (TP)	50 (FN)
Actual Not Spam	40 (FP)	460 (TN)

Calculating Accuracy:

$$\text{Accuracy} = \frac{450\ (\text{TP}) + 460\ (\text{TN})}{450 + 460 + 40\ (\text{FP}) + 50\ (\text{FN})} = \frac{910}{1000} = 0.91 \text{ or } 91\%$$

Imbalanced Dataset: When Accuracy Can Be Misleading

What is an Imbalanced Dataset?

- An imbalanced dataset is one where one class significantly outnumbers the other.
- For example, in a medical dataset predicting a rare disease, we might have 950 healthy patients and only 50 patients with the disease.

Example of an Imbalanced Dataset:

- Scenario: We have a dataset of 1,000 patients, with 950 healthy (negative class) and 50 with the disease (positive class).
- Model Prediction: Suppose our model predicts as follows:

	Predicted Disease	Predicted Healthy
Actual Disease	10 (TP)	40 (FN)
Actual Healthy	5 (FP)	945 (TN)

Calculating Accuracy:

$$\text{Accuracy} = \frac{10\ (\text{TP}) + 945\ (\text{TN})}{10 + 945 + 5\ (\text{FP}) + 40\ (\text{FN})} = \frac{955}{1000} = 0.955 \text{ or } 95.5\%$$

Imbalanced Dataset: When Accuracy Can Be Misleading

Feature	Balanced Dataset	Imbalanced Dataset
Class Distribution	Roughly equal number of instances in each class.	One class significantly outnumbers the other.
Accuracy as KPI	Reliable and meaningful.	Can be misleading.
Why?	Model's performance is evaluated equally across all classes.	High accuracy can be achieved by simply predicting the majority class, even if the model performs poorly on the minority class.
Other Metrics	N/A (Accuracy is sufficient)	Precision, Recall, F1-Score, ROC-AUC

Summary

- Accuracy is a commonly used metric in machine learning to measure the proportion of correct predictions.
- While useful in balanced datasets, accuracy can be misleading in imbalanced datasets, where other metrics like precision, recall, and F1-score provide a more comprehensive evaluation.

Special Thanks to ChatGPT and Gemini for Content support

in DineshLal