<Model Development>

Lecture Link:  
 https://www.cloudskillsboost.google/paths/17/course\_templates/593/video/508328

Q:

A hospital uses the machine learning technology of Google to help pre-diagnose cancer by feeding historical patient medical data to the model. The goal is to identify as many potential cases as possible. Which metric should the model focus on?  Please pick from the following choices.

Feature importance

Precision

Recall

Confusion matrix

A: Recall

**Explanation:**

|  |
| --- |
| Confusion Matrix:   From Lecture:  @3:36 https://www.cloudskillsboost.google/paths/17/course\_templates/593/video/508328  From <https://encord.com/glossary/confusion-matrix/> |

In this scenario, the **goal is to identify as many potential cases of cancer as possible** (i.e., minimizing missed cases or false negatives). Among the given choices, the most appropriate metric to focus on is **Recall**.

• **Recall** (Sensitivity or True Positive Rate): Measures the proportion of actual positive cases (cancer cases) correctly identified by the model. A high recall ensures that the model captures as many true cancer cases as possible, even if it means accepting more false positives.

**Recall = True Positives / (True Positives} + False Negatives)**

Since missing a cancer diagnosis can have severe consequences, recall is critical in this context.

**Why not the other metrics?**

1. **Feature Importance**: This is useful for understanding which features (medical data) contribute most to the predictions, but it is not a performance metric for the model itself.

2. **Precision**: Precision measures the proportion of correctly identified positive cases out of all predicted positive cases. While precision is important, in this scenario, prioritizing recall is more critical because false negatives (missed cancer cases) are more harmful than false positives (which can be ruled out with follow-up testing).

**Precision = True Positives / (True Positives + False Positives)**

3. **Confusion Matrix**: This provides a comprehensive view of model performance (True Positives, True Negatives, False Positives, and False Negatives), but it is not a single metric. It is a tool for calculating metrics like precision and recall.

Q:   
A farm uses the machine learning technology of Google to detect defective apples in their crop, like those with irregular sizes or scratches. The goal is to identify only the apples that are actually bad so that no good apples are wasted. Which metric should the model focus on? Choose from one of the following:

Feature importance

Precision

Recall

Confusion matrix

A: Precision

**Explanation:**

• **Precision** measures the proportion of correctly identified defective apples (true positives) out of all apples predicted as defective. A high precision ensures that most of the apples labeled as defective are actually bad, minimizing the waste of good apples.

**Precision = True Positives / (True Positives + False Positives)**

By prioritizing precision, the farm reduces the risk of incorrectly discarding good apples (false positives).

**Why not the other metrics?**

1. **Feature Importance**: This helps understand which features (e.g., size, shape, or scratch depth) are most influential in predictions, but it does not measure model performance directly.

2. **Recall**: Recall focuses on identifying all defective apples (minimizing false negatives). However, in this scenario, the emphasis is on avoiding false positives (discarding good apples), which makes precision more relevant.

**Recall = True Positives / (True Positives} + False Negatives)**

3. **Confusion Matrix**: This provides a full breakdown of the model’s performance (true positives, false positives, true negatives, and false negatives) but is not a specific metric. It is useful for calculating metrics like precision and recall.