oduct is in a pre-release state and might change or have limited support. For more information, see the product laun
s (/products/#product-launch-stages).

After training a model, AutoML Video Intelligence Classification uses items from the TEST set (/video-intelligence/automl/docs/prepare) to evaluate the quality and accuracy of the new model.

AutoML Video Intelligence Classification provides an aggregate set of evaluation metrics indicating how well the model performs overall, as well as evaluation metrics for each category label, indicating how well the model performs for that label.

- **AuPRC** : Area under Precision/Recall curve (https://en.wikipedia.org/wiki/Evaluation_measures_(information_retrieval)#Average_precision), also referred to as "average precision." Generally between 0.5 and 1.0. Higher values indicate more accurate models.
- The **Confidence threshold curves** show how different confidence thresholds would affect precision, recall, true and false positive rates. Read about the relationship of precision and recall (https://en.wikipedia.org/wiki/Precision_and_recall).

Use this data to evaluate your model's readiness. Low AUC scores, or low precision and recall scores can indicate that your model needs additional training data or has inconsistent labels. A very high AUC score and perfect precision and recall can indicate that the data is too easy and may not generalize well.

If you're not happy with the quality levels, you can go back to earlier steps to improve the quality:

- You may need to add different types of videos (e.g. wider angle, higher or lower resolution, different points of view).
- Consider removing labels altogether if you don't have enough training videos.
- Remember that machines can’t read your label name; it's just a random string of letters to them. If you have one label that says "door" and another that says "door_with_knob" the machine has no way of figuring out the nuance other than the videos you provide it.
- Augment your data with more examples of true positives and negatives. Especially important examples are the ones that are close to the decision boundary.

Once you've made changes, train and evaluate a new model until you reach a high enough quality level.