DRP analysis model evaluation

**Inhoud**

[Model results 3](#__RefHeading___Toc124_599662797)

[Statistics explanation 3](#__RefHeading___Toc126_599662797)

[SVC Model: 3](#__RefHeading___Toc128_599662797)

[SVC Model with same sized data samples: 4](#__RefHeading___Toc130_599662797)

[SVC Model with same sized data samples and simplified categories: 4](#__RefHeading___Toc132_599662797)

[What is the performance? 4](#__RefHeading___Toc134_599662797)

[Why is the performance so low? 5](#__RefHeading___Toc136_599662797)

# Model results

When evaluation the performance of a model we usually look at the evaluation metrics. The ones I used are: accuracy, precision, recall and the f1 score.

### Statistics explanation

* **Accuracy:** or Classification Accuracy as it is also known. Is the ratio of number of correct predictions to the total number of input samples.
* **Precision:** is the number of correct positive results divided by the number of positive results predicted by the classifier.
* **Recall:** is the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive).
* **F1 score:** The F1 score can be interpreted as a harmonic mean of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall to the F1 score are equal.

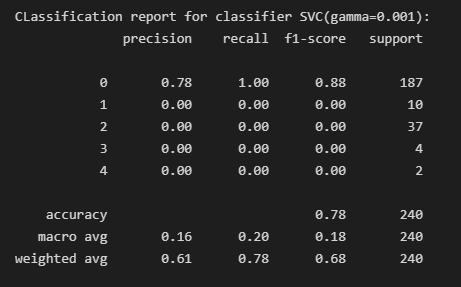
In most cases scoring a higher percentile means more correctly guessed images, and a lower score means less correctly guessed. This is not always necessarily a good thing, for example a high accuracy could mean that your data set is biased and requires cross validation.

However generally scoring high in all statistics is a positive. Here were the evaluation scores for the recent models.

## SVC Model:

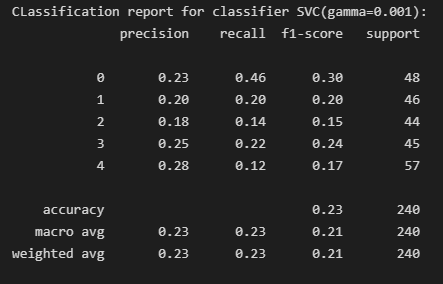
The following models used 800 total images.

SVC made by taking random images.



## SVC Model with same sized data samples:

SVC made by taking the same amount of images for each category.



## SVC Model with same sized data samples and simplified categories:

SVC made by taking the same amount of images for each category but labeling 1-4 as 1s.

After consulting the stakeholder this idea was dropped as without the details of each category the model has no value.

Afbeelding met tekst

Automatisch gegenereerde beschrijving

# Model Performance

## What is the performance?

As you can see in the images above, the models are basically guessing. An accuracy of 53% for 2 categories or 23% for 5 just means the model is guessing and gets lucky a few times. This performance is somewhat disappointing as I was hoping for a score of at least 70% for the initial results.

## Why is the performance so low?

There could be many different explanations as to why the score is this low. Low amount of data, low quality of data and the data is too similar to each other are a few that come to mind.

The reason I think the data is to similar comes from the evaluation matrix having the highest precision on category 4, which is the most different from the other categories. While category 1 and 2