

PART 2

RATIONAL STATEMENT

Help determine if the patients have leukemia or not. Finding a better algorithm and model for the leukemia remission dataset (leuananalysisNew). With focus on analyzing output/results and making better analysis. In this experiment, we used Naïve Bayes (NB) and Supply Vector Machine (SVM).

This power point will address;

- Three insights from the classification report (Recall, Precision, F1) for SVM and NB
- Comparing SVM and NB classification report with (2) insights and possible examples to follow
- Suggesting a model from the two used in our code for the dataset, and reasons why it will be good to use that model. Why it is better and how it performs best than the other.

THREE INSIGHTS FROM THE CLASSIFICATION REPORT

Firstly looking at support (6+2=8), shows it is a very small sample, and an imbalance dataset we are working with. Below, you will notice an unsimilar metrics for SVM and I believe do determine a good model for predicting leukemia as compared to NB

CONCERNING SVM

Precision

- Looking at precision, weighted average stands at (56), this comes high to the others but still low in predicting who has and who does not have leukemia. The high score of (67) does a better job at predicting patients who do not have leukemia

Recall

- Here, weighted average comes at (50) not good for predicting leukemia. Putting this into practice will give a 50 to 40 chance of me predicting they do or do not have leukemia

F1

- Analyzing F1's weighted average at (51), similar range with recall, they both do not give a better score at predicting leukemia. Looking at the top scores, it predicts at (57) for patients not having leukemia.

THREE INSIGHTS FROM THE CLASSIFICATION REPORT (continuation)

CONCERNING NB

Below, we will notice varying metrics for this model, with recall being a better predictor at (67) though not really good. NB comes with an accuracy of (67). This looks like a better model than SVM, but poses a problem at predicting those who have leukemia.

NB can only predict one class(Remission cannot be predicted)

Precision

- At (44), due to the imbalance nature, I will be looking at the recall and F1 scores for better analysis.

Recall

- At (67) the weighted average. Looking at the top scores for the classes, recall does better at (1.00) predicting patients who do not have leukemia. But does not give any score for patients having. So using this model for predicting patients without leukemia is better.

F1

- At (53) for the weighted average. F1 does a good job at (80) and is a good metric for predicting patients who do not have leukemia. If we have to use this same metric for predicting who has leukemia, I advise us to forget and not use it.

Comparing SVM and Naïve Bayes classification reports

- SVM looks better model at predicting both classes for patients who have and do not have leukemia, but fails to be accurate helping us make a decision. It fails to give a high score at predicting the condition of patients
 - The weights stay within the 50% mark but do not go above for better determination
 - Precision and F1 looks a better metrics
-
- NB is better at predicting patients who do not have leukemia as it comes with an accuracy of (67) and better metrics at (67),(1.00) and (80) as compared to SVM which floats at a score of (50).
 - NB has a somewhat increasing and fluctuating weight, yet looks better than SVM
 - Recall and F1 seem the better metrics to use.

Recommend one (1) model

- Depending on what John wants to do with the dataset, SVM I do not recommend.
- I will recommend Naïve Bayes as the better model to use for John Hughes. NB has better scores and clearly predicts patients who do not have leukemia. This algorithm serve him better in predicting patients who do not have leukemia
- SVM will definitely land him in trouble if he uses it to predict who has leukemia, given it is a small and an imbalance dataset, SVM still does not do justice with it. At a (50%) prediction rate, patients might take treatments when they should not, due to a low prediction scores.
- **I conclude, NB is better to use for predicting patients who do not have leukemia, but problematic for predicting who has. SVM looks a better score at averagely predicting who has and who does not.**