# Comparison of Stacked Classifiers Using Predictions and Scores

Giovanni Mantovani Andrea Palmieri

March 22, 2024

**Abstract**

This report presents an analysis of stacked classifier configurations, focusing on the performance impact of training the meta-classifier using Predictions versus Scores, and evaluating the implications of training on split versus non-split datasets. We assessed the performance of five base classifiers—SVM with Gaussian kernel, SVM with Polynomial kernel, Decision Tree, Naive Bayes, and Ensemble of Decision Trees—followed by two meta-classifiers in a stacked arrangement.

## 1 Introduction

Stacked classification, a robust machine learning technique, combines multiple model predictions to improve overall accuracy. After computing the Predictions and the Scores of the first five base classifiers, two meta-classifiers are trained: the first is trained on the Scores, the second on the Predictions. This process is performed first using a split training data approach, then without splitting the training data. Finally their performances in terms of accuracy in predicting classes are analysed.

## 2 Methodology

The following are the five base classifiers:

- SVM with Gaussian kernel
- SVM with Polynomial kernel
- Decision Tree
- Naive Bayes
- Ensemble of Decision Trees

The 2 stacked classifiers are:

- Meta-Classifier trained on Scores of the five base classifiers
- Meta-Classifier trained on Predictions of the five base classifiers

Assignment 1: Training with Split Data

We split the data and labels into two folds, with each fold containing equal proportions of samples from each class. We first train the level 1 classifiers on Fold 1 of the data. Subsequently, predictions are made on Fold 2 using these trained base classifiers. These predictions are then utilized to train two stacked classifiers: one on the Scores and the other on the Predictions.

Assignment 2: Training without Split Data

In a parallel experiment, we train the base classifiers directly on the entire training dataset without employing data splitting. Then the stacked classifiers are trained on the Scores and Predictions as in the first approach. This allows us to investigate the impact of data partitioning on the performance of both base classifiers and stacked classifiers.

## 2.1 Model Evaluation

The performance of the classifiers was assessed on a test set, with accuracy as the key performance metric. We compared the efficacy of training the meta-classifier on Predictions (the categorical outputs of the base classifiers) versus Scores (the confidence measures or probabilities associated with each class). Additionally, we evaluated the impact of using split and non-split datasets for training the model layers.

# 3 Results

The performance of the stacked classifier under different training conditions yielded the accuracy results shown in Table 1 and Table 2.

These results demonstrate that training the meta-classifier on Scores resulted in higher accuracy compared to using Predictions. Training both base classifiers and meta-classifiers on the same dataset (non-split approach) slightly decreased performance, indicating potential overfitting issues. The difference between training scores and predictions is higher in the split case but much less significant when we trained the model on all the data.

| **Classifier** | **Accuracy** |
|---|---|
| SVM (Gaussian) | 0.8683 |
| SVM (Polynomial) | 0.6250 |
| Decision Tree | 0.9483 |
| Naive Bayes | 0.9783 |
| Ensemble (Decision Trees) | 0.9533 |
| Meta-Classifier (Scores) | 0.9900 |
| Meta-Classifier (Predictions) | 0.9700 |

Table 1: Accuracy of classifiers splitting the training data.

| **Classifier** | **Accuracy** |
|---|---|
| SVM (Gaussian) | 0.9000 |
| SVM (Polynomial) | 0.6333 |
| Decision Tree | 0.9667 |
| Naive Bayes | 0.9917 |
| Ensemble (Decision Trees) | 0.9683 |
| Meta-Classifier (Scores) | 0.9700 |
| Meta-Classifier (Predictions) | 0.9683 |

Table 2: Accuracy of classifiers without splitting training data.

# 4 Conclusion

The results highlight a difference in accuracy between meta-classifiers trained on non-split data (97 % and 96.83%) and those trained on data divided into two folds (99% and 97%). This is expected as the common rule to prevent information leak for stacked classifiers is to originate level one predictions from an independent subset of the training data, thus preventing potential information leakage.

Moreover, the first approach shows that the accuracy of the meta-classifiers increases when their training is performed using Scores rather than Predictions of the level one classifiers. This findings highlight the importance of the training strategy adopted in stacked classifiers: splitting the training data for the two levels and utilizing Scores as a training base for the meta-classifier enhances performance by leveraging detailed confidence information from the base models. In contrast, employing the same dataset for both level one and meta-classifiers training can risk overfitting, as evidenced by the slight decrease in accuracy.

In summary, these results show that stacked classifiers trained using a split training data approach and on the predictions of the base classifiers possess better performances than the other methodologies described in this report.