Is Combining Classifiers with Stacking Better than Selecting the Best One?

Base Classifier :

i. Support Vector Machine
ii. Random Forest
iii. Decision Tree
iv. Naive Bayes
v. Xgboost

Stacking Method Implemented

i. Voting

ii. Generalize stacking

Voting

- Collect the output of each Base Classifier and combing it.
- Predict class which has a maximum votes from Base Classifier

Generalize Stacking

- Split the train data into two equal parts
- Use first part for training and second part for prediction
- Create meta train data and meta test data
- Apply Logistic regression for prediction on test data.
- Base Classifier → Metadata → LogisticRegression → output

Output

Average Error rate on Base Classifier:

- Support Vector Machine: 17.2132865846
- Random Forest: 12.5656493768
- Decision Tree: 16.276822636
- Naive Bayes: 28.3806293607
- Xgboost: 12.2396629821

Output

Average Error rate on Stack Classifier:

- Voting: 11.0162346319
- Generalize stacking:11.1330153

Conclusion

Combining Classifiers with Stacking Better than Selecting the Best One !!