

Programa de Pós-graduação em Sistemas de Informação

SIN5007 - Reconhecimento de Padrões (2023)

Censo da Educação Superior - Cursos 2021

MSc. Leonardo Cunha dos Santos
Gabriel Francisco dos Santos Silva

São Paulo / 2023



Agenda

- ➊ Utilizando Stacking Classifier
- ➋ Overview do código
- ➌ Comparativo do experimento (amostra de teste)
- ➍ Comparativo do experimento (calibração)

Utilizando Stacking Classifier

- Classificadores:
 - Naive Bayes
 - Decision Tree
 - Neural Network (Keras)
- Classificador final:
 - Logistic Regression

Overview do código

Utilizando Python, Scikit-learn e Keras

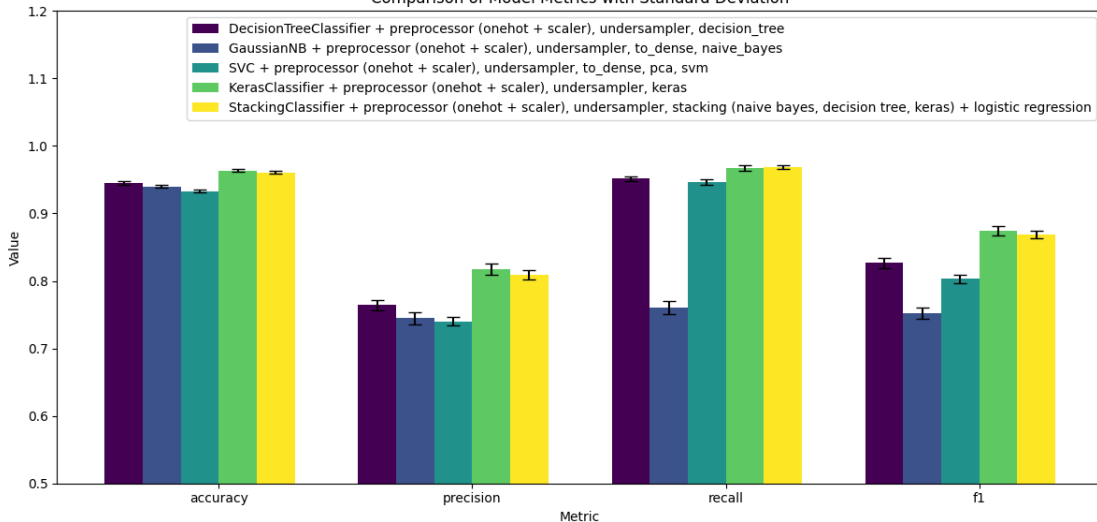
```
def stacking(additional_steps, additional_params={}):
    stacking_prefix = 'stacking (naive bayes, decision tree, keras) + logistic regression'

    return {
        "pipeline": Pipeline(steps=[
            *additional_steps,
            (stacking_prefix, StackingClassifier(estimators=[
                ("naive_bayes", Pipeline(steps=[("to_dense", to_dense), ("naive_bayes", GaussianNB()))]),
                ("decision_tree", DecisionTreeClassifier(random_state=seed_value)),
                ("keras", KerasClassifier(get_model, verbose=0))
            ], final_estimator=LogisticRegression(random_state=seed_value))
        )),
        "params": {
            ('%s__naive_bayes__naive_bayes__var_smoothing' % stacking_prefix): [1e-9, 1e-7],
            ('%s__decision_tree__max_depth' % stacking_prefix): [30],
            ('%s__decision_tree__min_samples_split' % stacking_prefix): [5, 10],
            ('%s__decision_tree__min_samples_leaf' % stacking_prefix): [3, 5],
            ('%s__keras__model__hidden_layer_dim' % stacking_prefix): [100],
            ('%s__keras__optimizer' % stacking_prefix): ["adam"],
            ('%s__keras__loss' % stacking_prefix): ["binary_crossentropy"],
            ('%s__keras__epochs' % stacking_prefix): [20],
            ('%s__final_estimator__C' % stacking_prefix): [0.1],
            ('%s__final_estimator__solver' % stacking_prefix): ['saga'],
            ('%s__final_estimator__max_iter' % stacking_prefix): [200],
            ('%s__final_estimator__penalty' % stacking_prefix): ['l1'],
            **additional_params
        }
    }
```

Comparativo do experimento

Resultados – Amostra de teste + resample

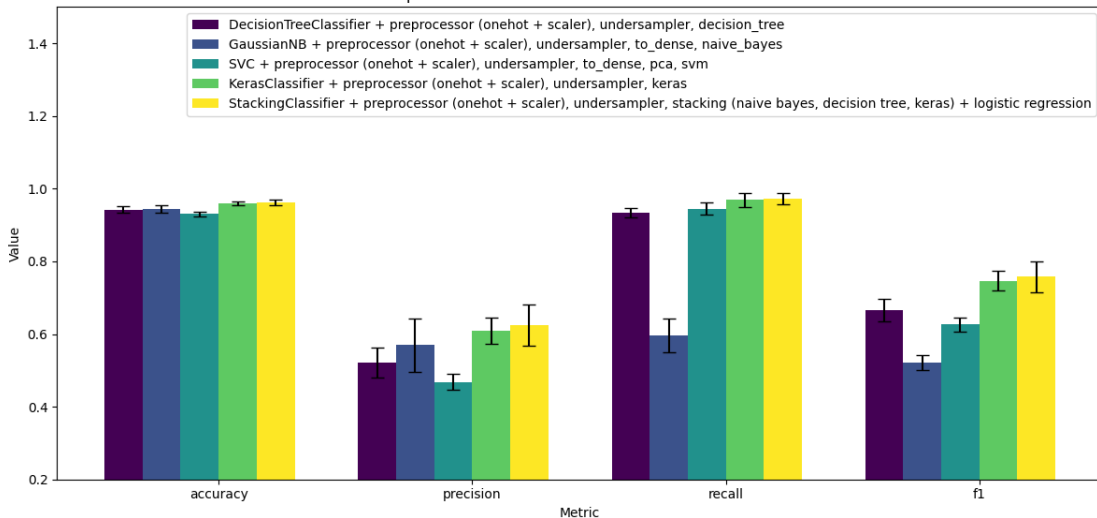
Comparison of Model Metrics with Standard Deviation



Comparativo do experimento

Resultados – Amostra de calibração do cross validation

Comparison of Model Metrics with Standard Deviation



Obrigado!

Thanks! / ¡Gracias!

Leonardo Cunha dos Santos

`lattes.cnpq.br/5620610314140397`

`leonardo.cunha.santos@usp.br`

Gabriel Francisco dos Santos Silva

`gabfssilva@gmail.com`