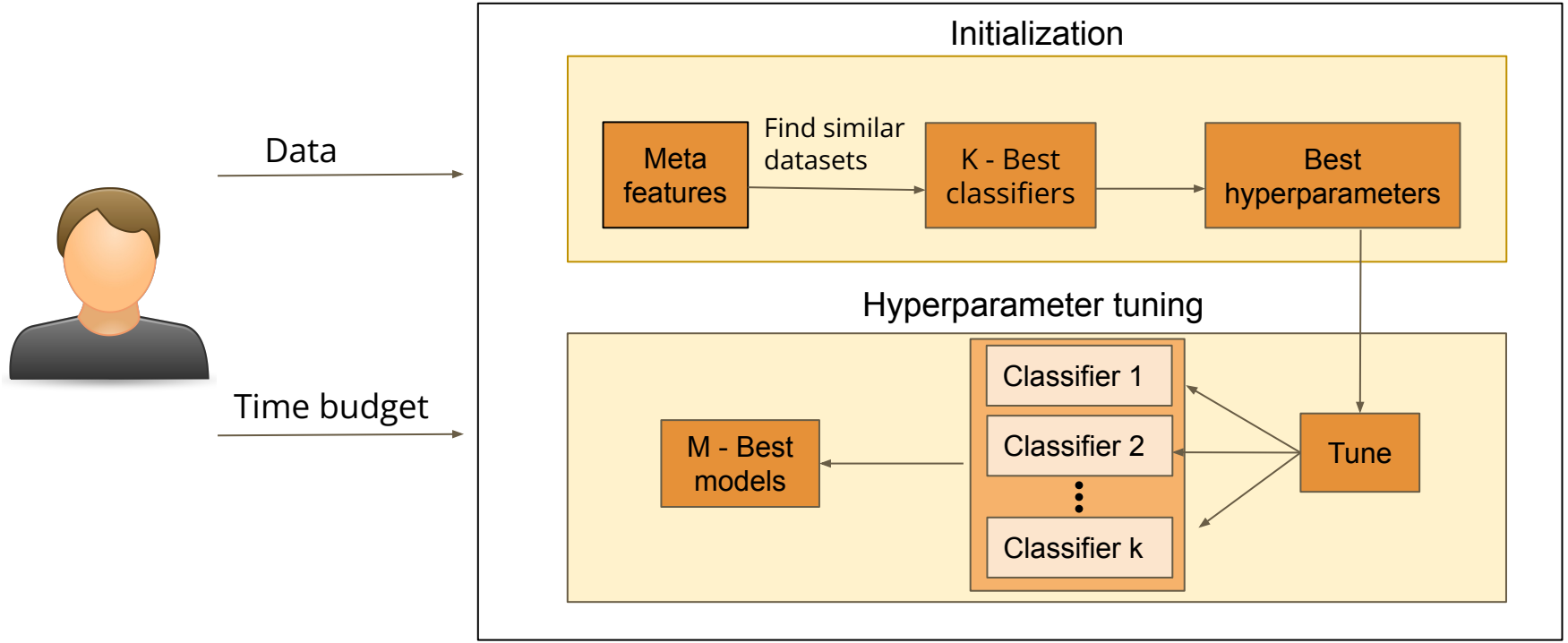
A Multi-Objective Decision Support for AutoML Frameworks

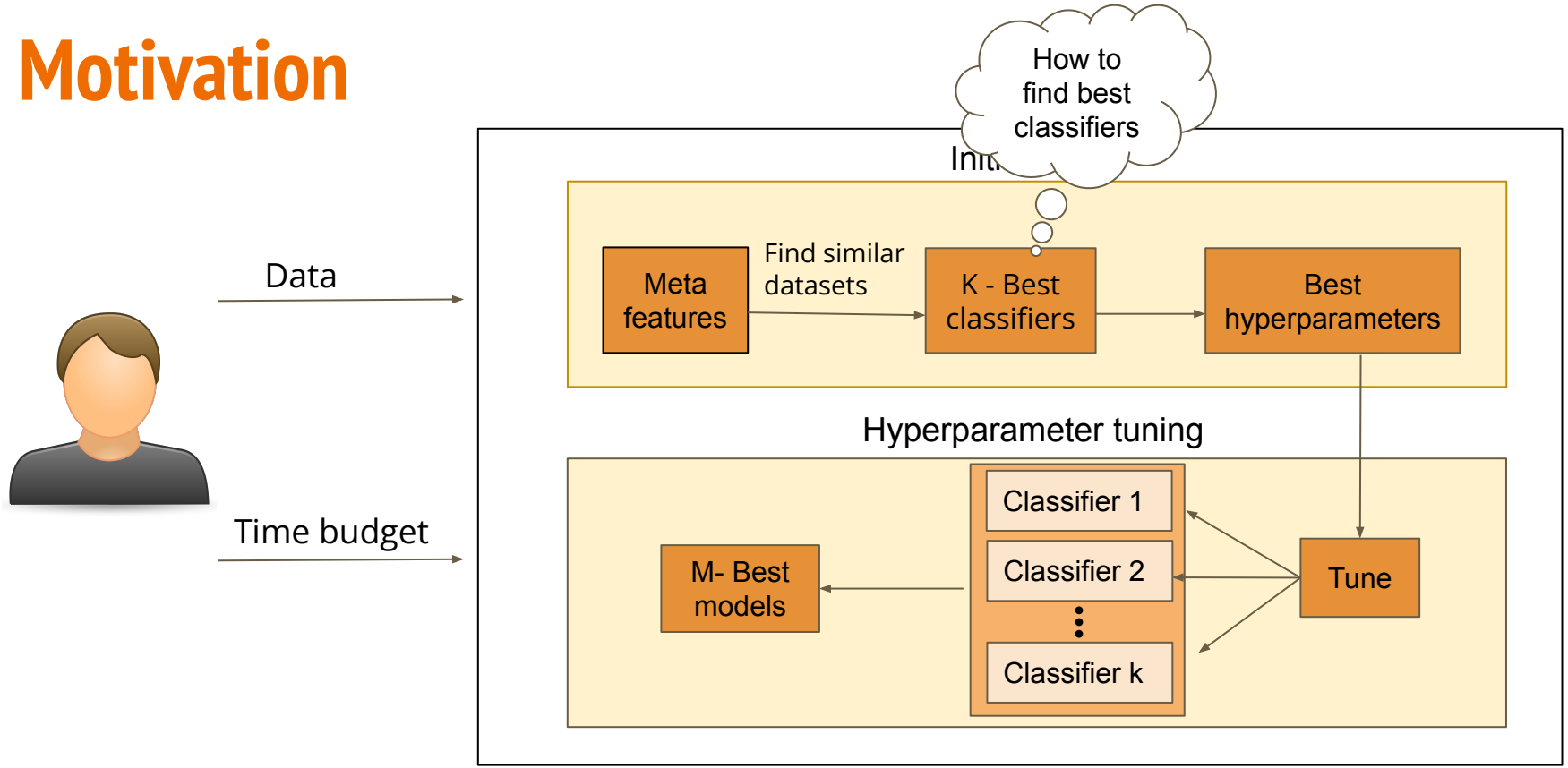
Data Systems Group

Table of contents

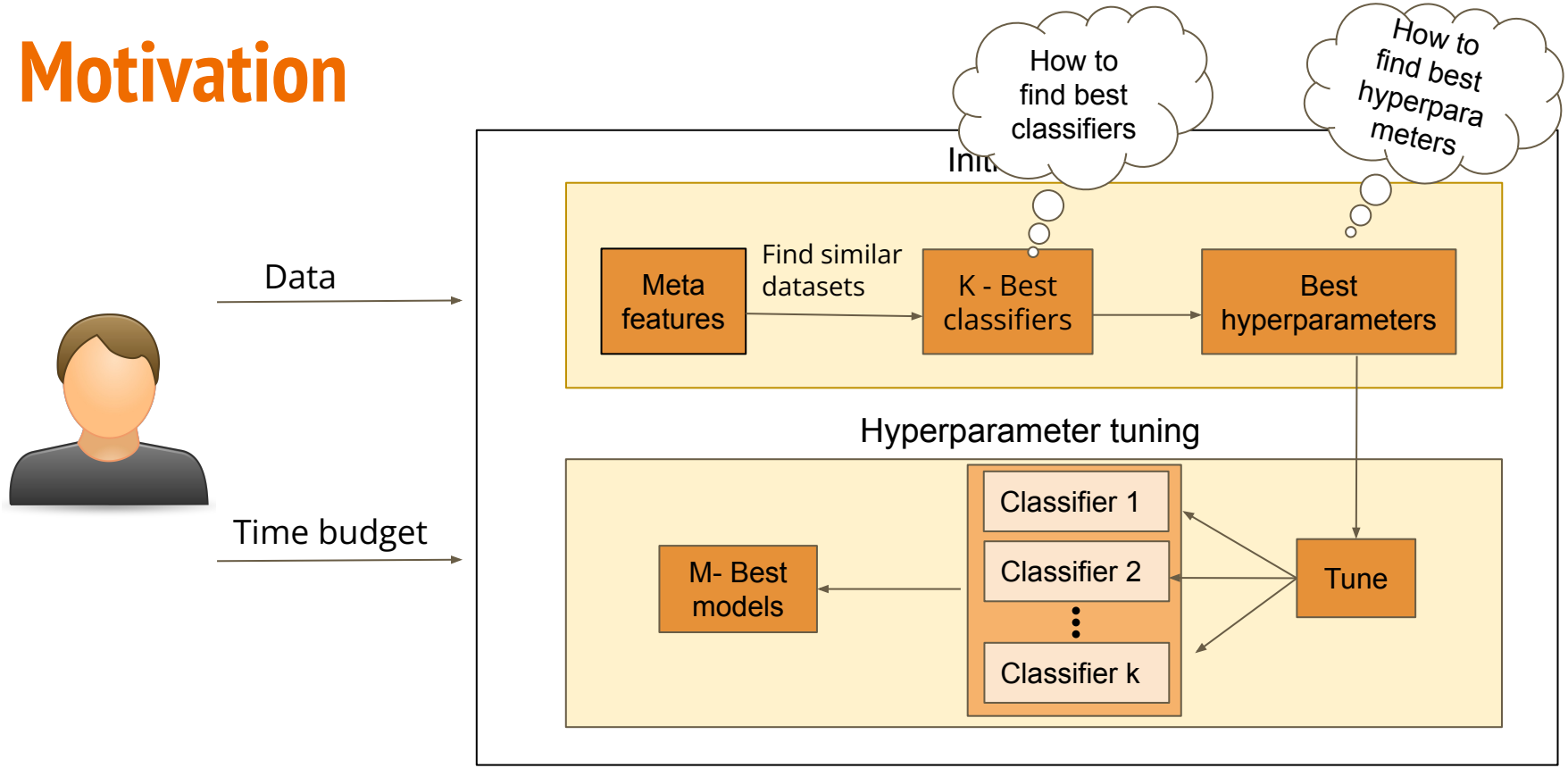
Motivation


Data
Time budget
Initialization
Meta features
Find similar datasets
K - Best classifiers
Best hyperparameters
Hyperparameter tuning
Classifier 1
Classifier 2
Classifier k
M - Best models
Tune

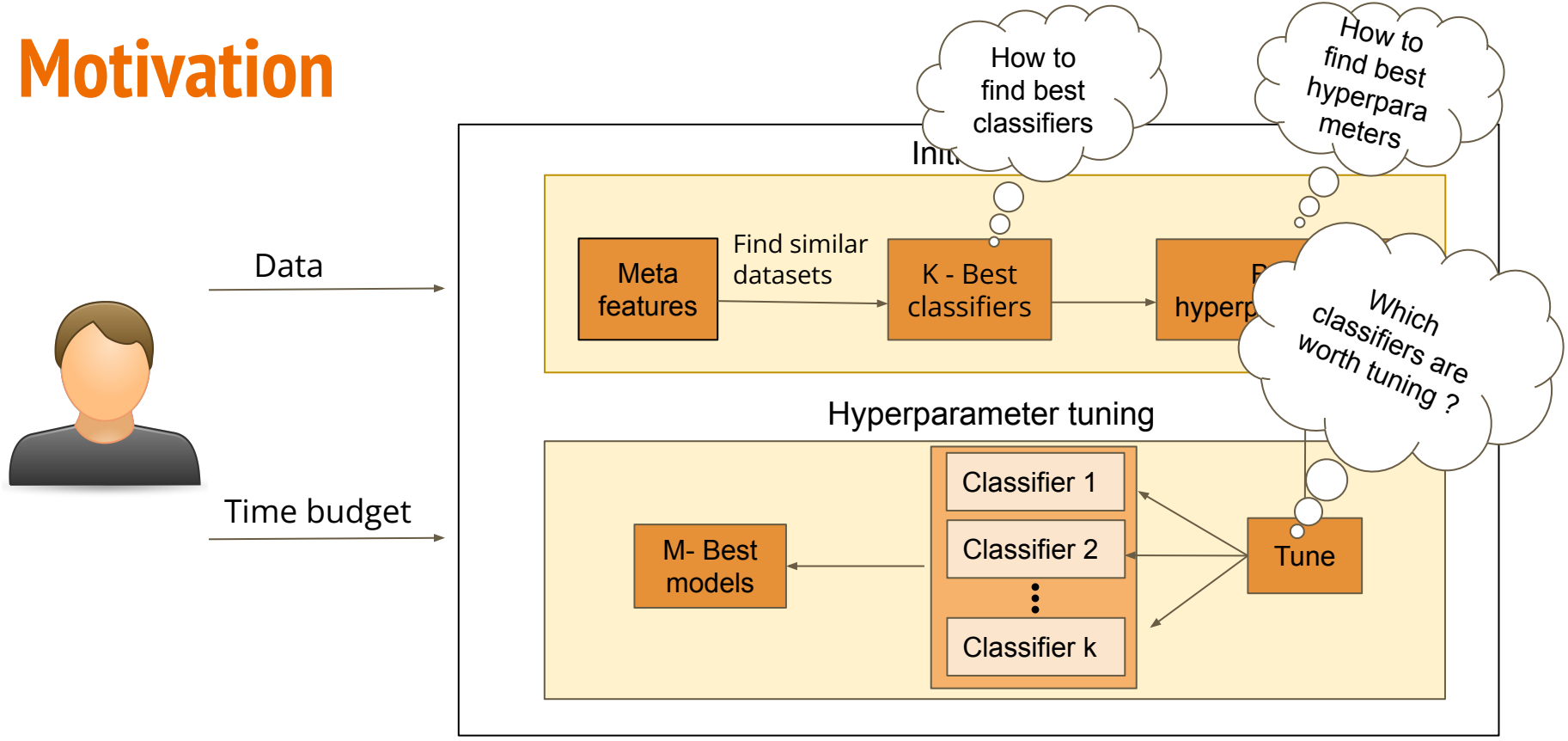
Motivation


How to find best classifiers
Data
Time budget
Meta features
Find similar datasets
K - Best classifiers
Best hyperparameters
Hyperparameter tuning
Classifier 1
Classifier 2
Classifier k
M- Best models
Tune

Motivation


How to find best classifiers
How to find best hyperpara meters
Data
Time budget
Meta features
Find similar datasets
K - Best classifiers
Best hyperparameters
Hyperparameter tuning
Classifier 1
Classifier 2
Classifier k
M- Best models
Tune

Motivation


Data
Time budget
How to find best classifiers
How to find best hyperpara meters
Meta features
Find similar datasets
K - Best classifiers
Which classifiers are worth tuning ?
Hyperparameter tuning
Classifier 1
Classifier 2
Classifier k
M- Best models
Tune

Experimental setup

200 datasets from OpenML :

- 54% binary and 46 % multiclass classification tasks
- 5 Kb to 350 Mb

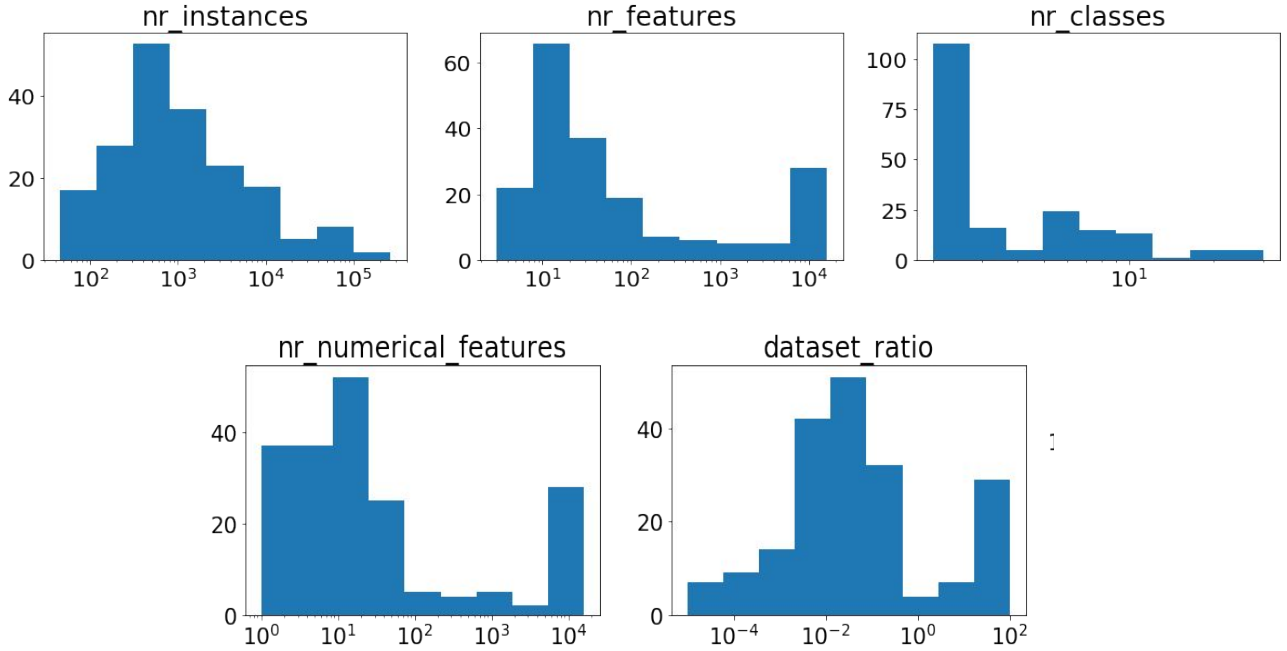
30 classifiers :

- 13 from Scikit-Learn (Python)
- 17 from Weka (R)

27 meta features:

- simple
- statistical
- information-theoretic types.

Meta features distribution



Experimental setup

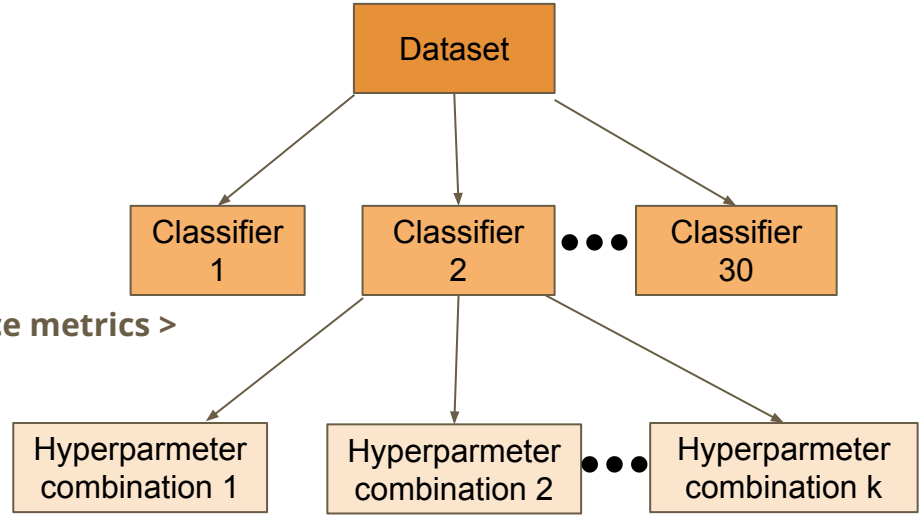
Different combinations of hyperparameters (ceiling 40)

No preprocessing

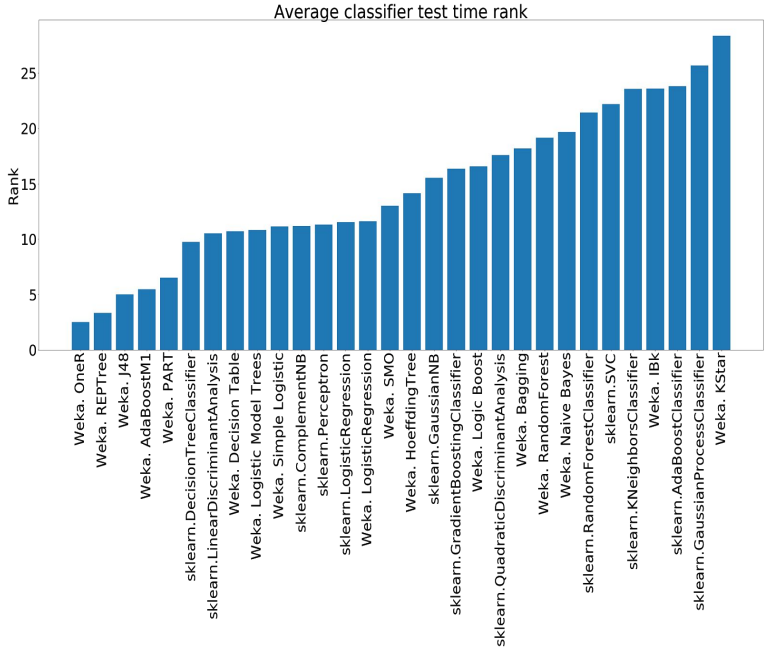
New dataset = 132K records

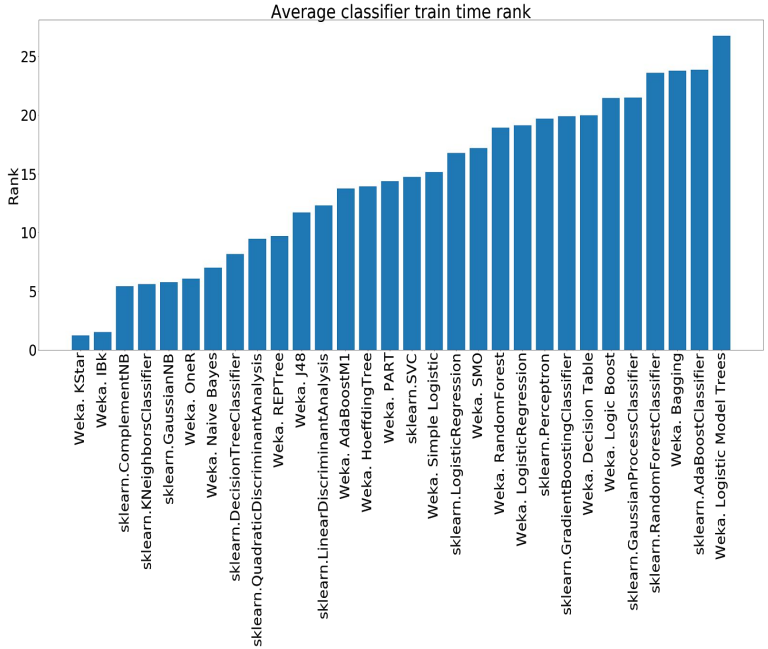
< Meta features, classifier, hyperparameters, performance metrics >

Time out to prevent long runtime

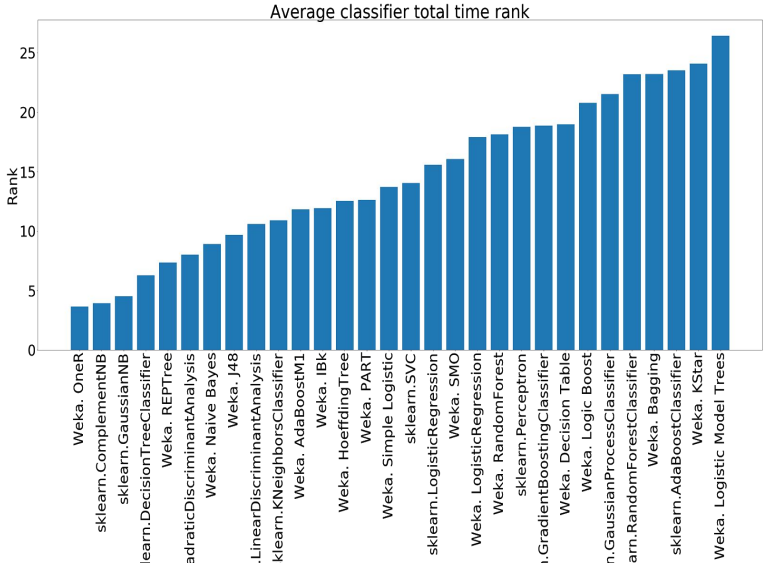


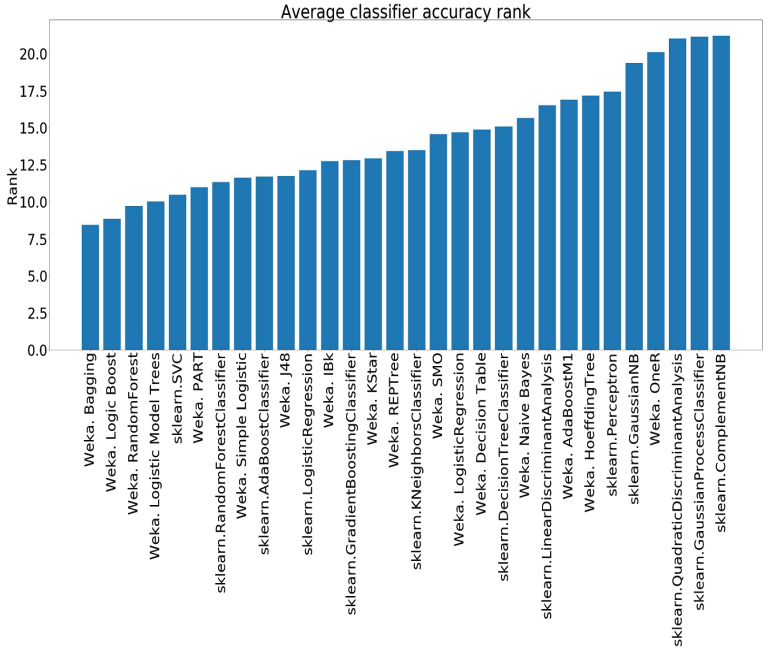
Classifier performance



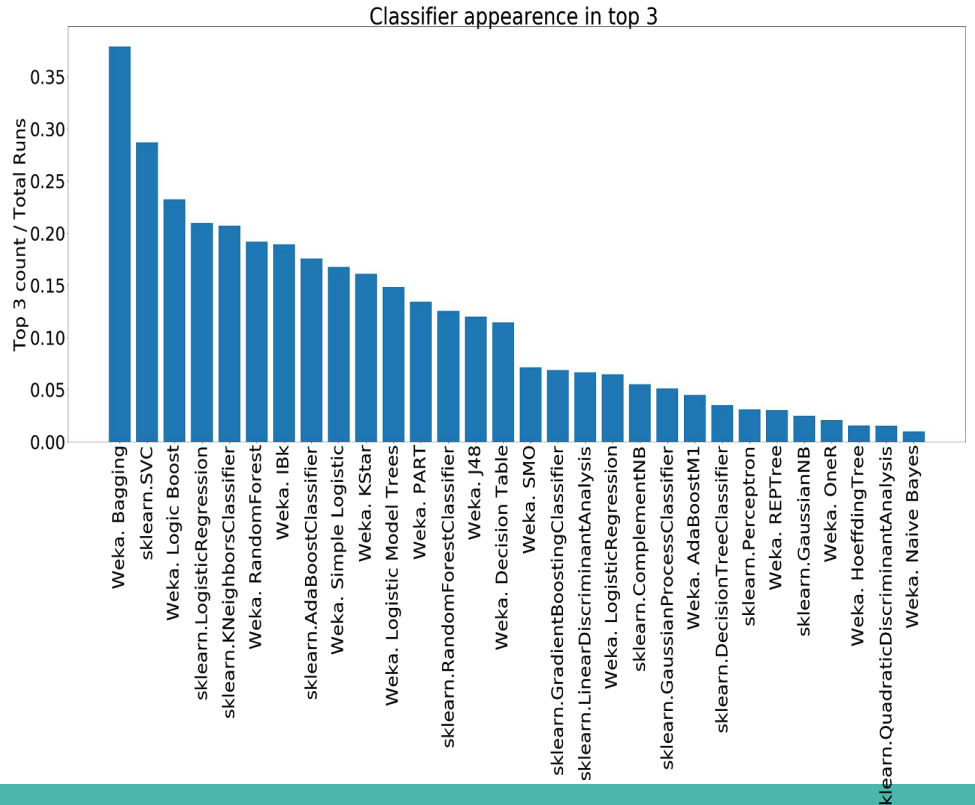


Classifier performance





Classifier performance



Weka vs Scikit-learn

SMO is solving the quadratic programming (QP) of SVM

AdaBoost :

- Scikit-learn uses SAMME algorithm
- Weka uses M1 algorithm

Decision Tree base classifier:

- Scikit-learn uses DecisionTreeClassifier(max_depth=1)
- Weka uses DecisionStump

Scikit-Learn	Weka
AdaBoost	AdaBoostM1
SVM	SMO
KNeighborsClassifier	IBk
RandomForestClassifier	RandomForest
Logistic Regression	Logistic Regression

Same , same but different !

Weka vs Scikit-learn

Random Forest base classifier :

- Scikit-Learn uses Decision Tree
- Weka uses Random Tree

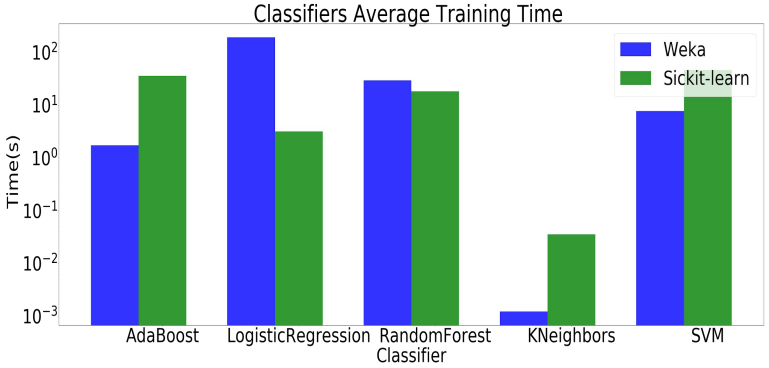
The **Logistic Regression** in Scikit-Learn supports much more hyper parameters and optimizations than Weka.

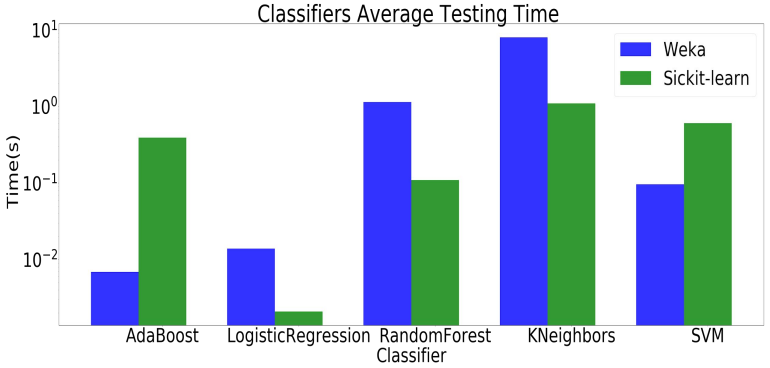
KNeighbors in Scikit-Learn supports different algorithms for computing the nearest neighbors.

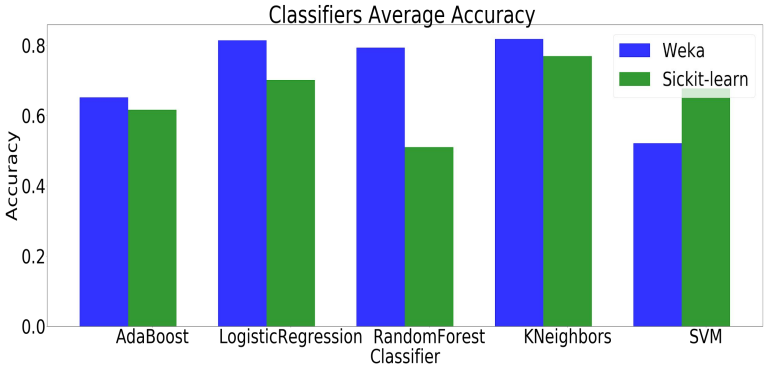
Scikit-Learn	Weka
AdaBoost	AdaBoostM1
SVM	SMO
KNeighborsClassifier	IBk
RandomForestClassifier	RandomForest
Logistic Regression	Logistic Regression

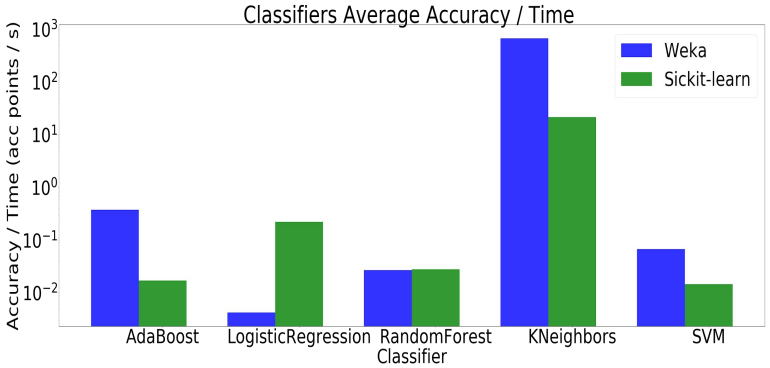
Same , same but different !

Weka vs Scikit-learn

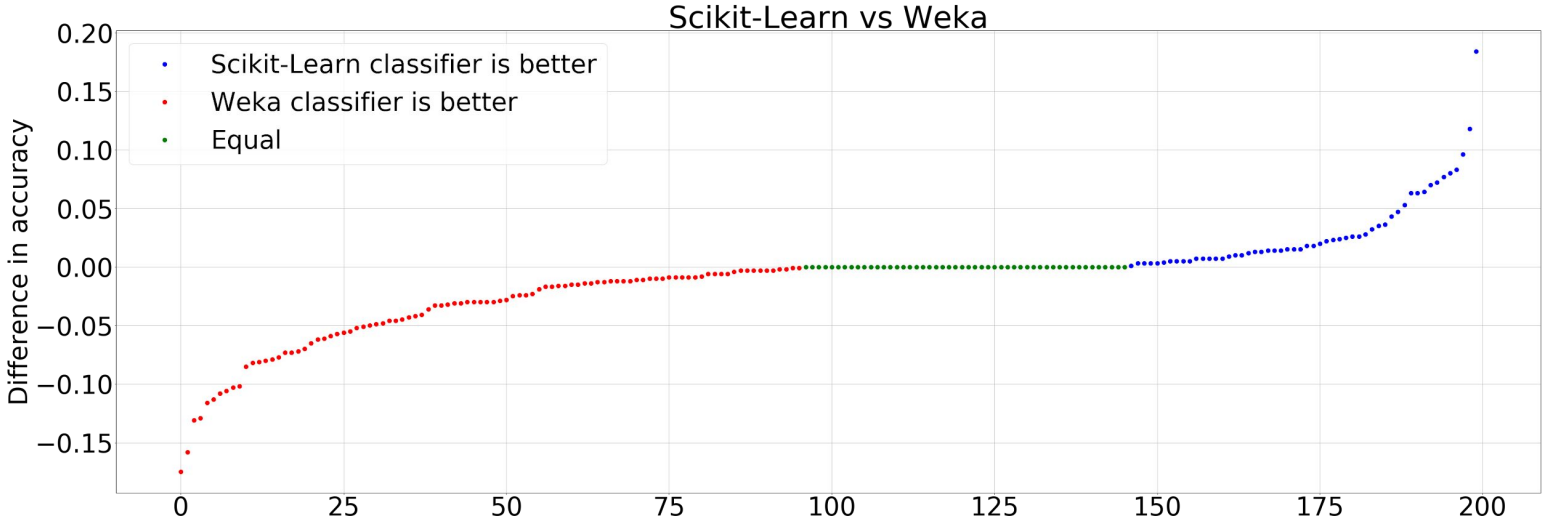






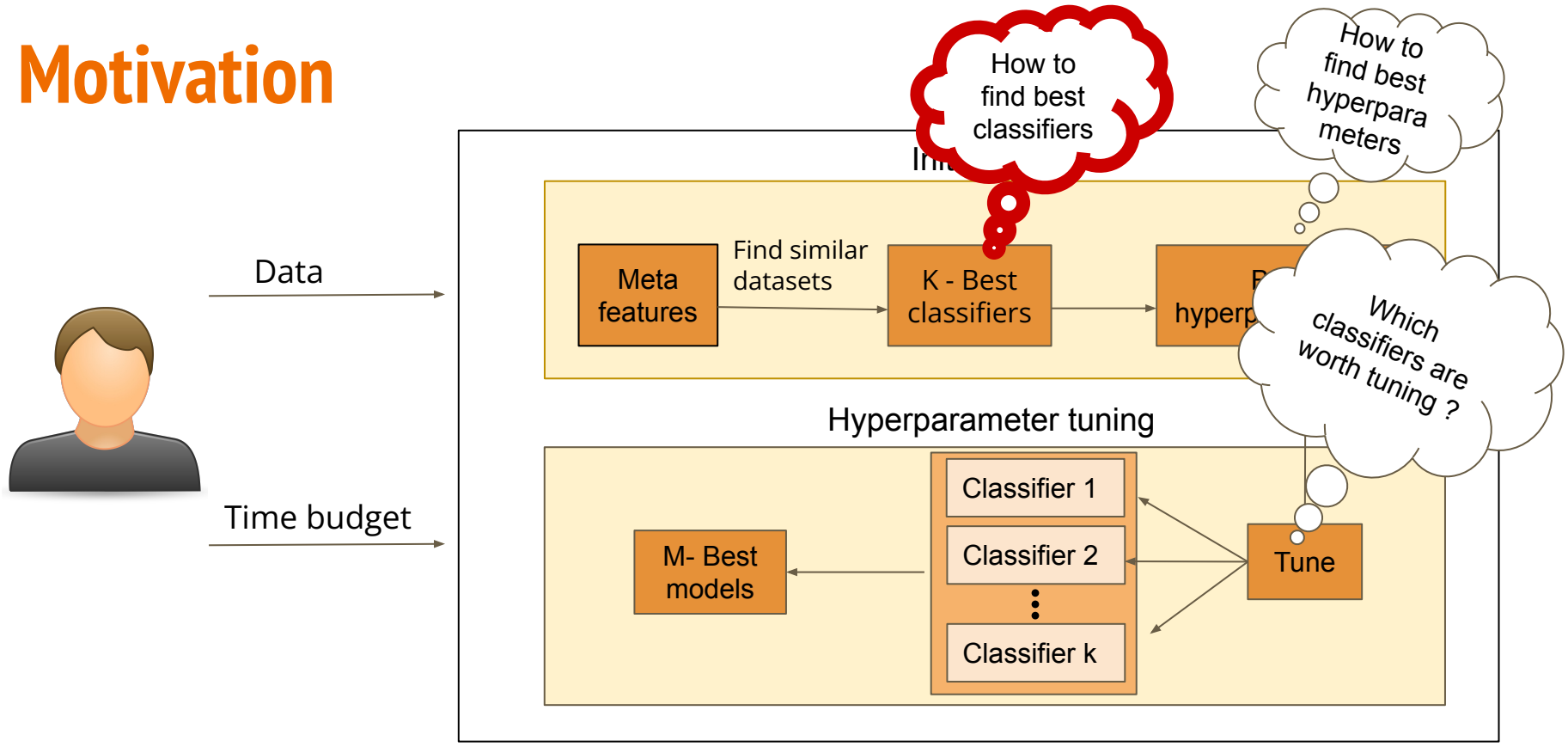


Weka vs Scikit-learn



Scikit-Learn classifier is better : **50**
Weka classifier is better : **96**
Equal : 54

Motivation


How to find best classifiers
How to find best hyperpara meters
Data
Time budget
Meta features
Find similar datasets
K - Best classifiers
Which classifiers are worth tuning ?
Hyperparameter tuning
Classifier 1
Classifier 2
Classifier k
M- Best models
Tune

Predicting best classifiers

Select classifier best result and create a new dataset (5432 instances)

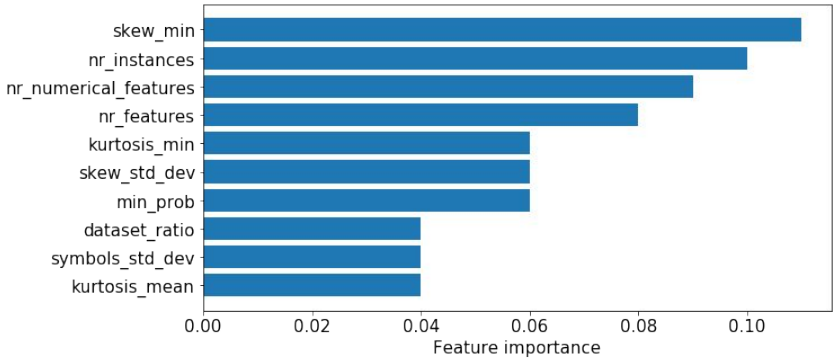
Binary classifications :

- **Class 0** - Best classifiers
- **Class 1** - Not good classifier

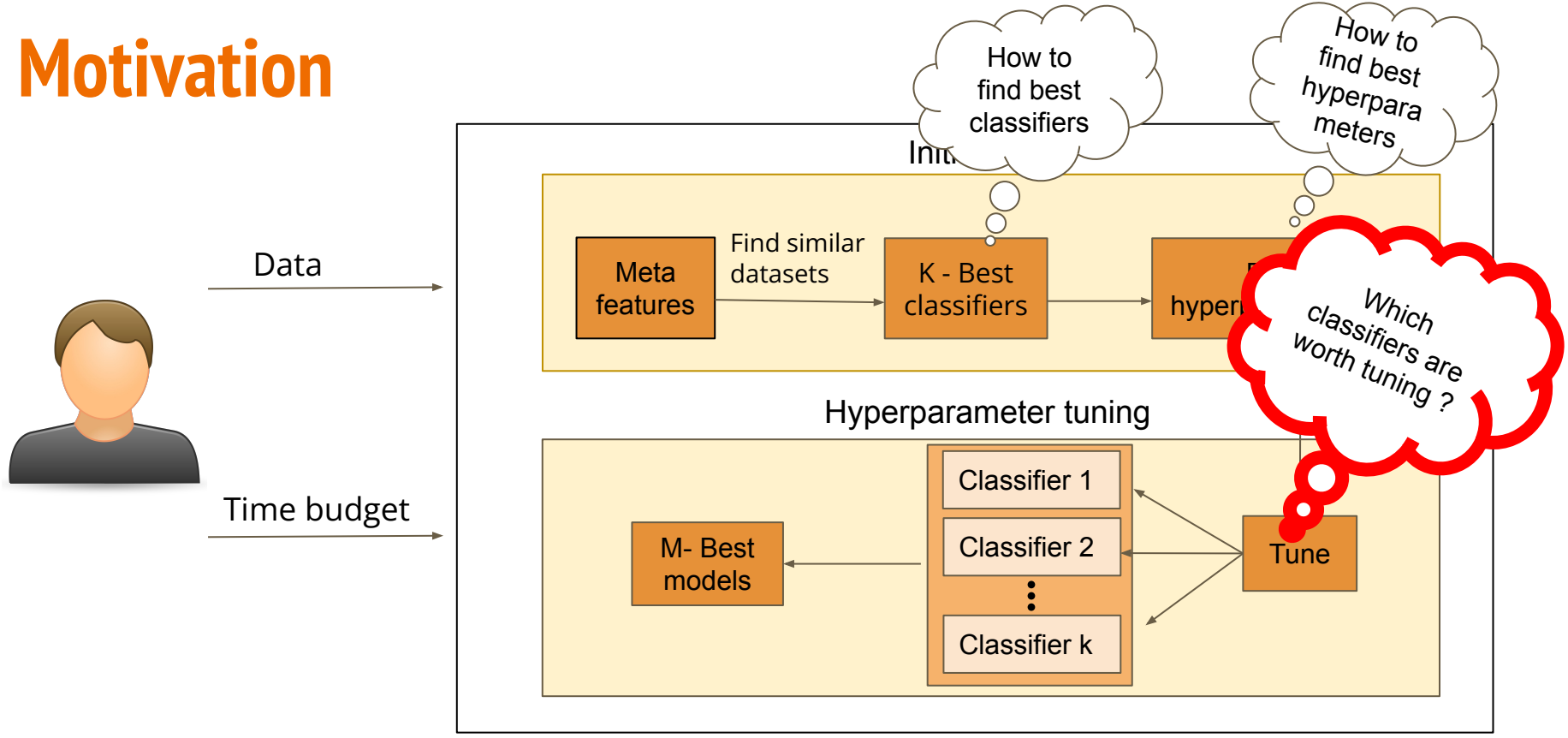
Take into consideration the accuracies standard deviation

Fit a decision tree after performing Randomized Search

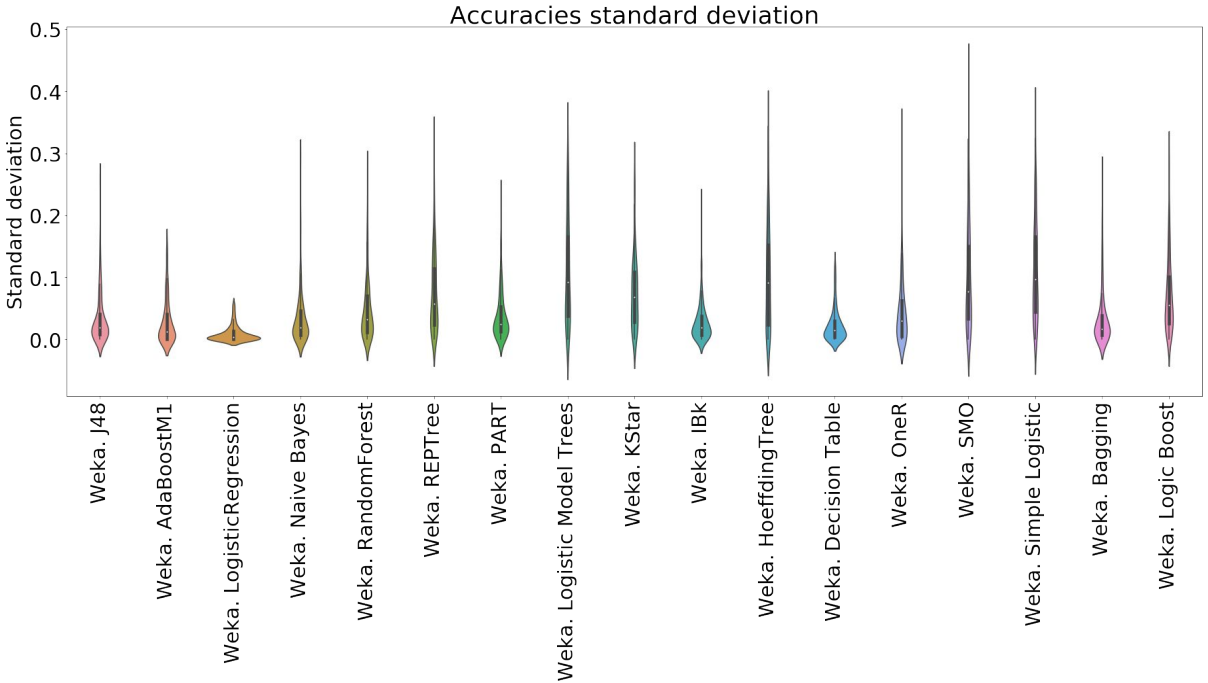
84 % recall on Class 0



Motivation


How to find best classifiers
How to find best hyperpara meters
Data
Time budget
Meta features
Find similar datasets
K - Best classifiers
Which classifiers are worth tuning ?
Hyperparameter tuning
Classifier 1
Classifier 2
Classifier k
M- Best models
Tune

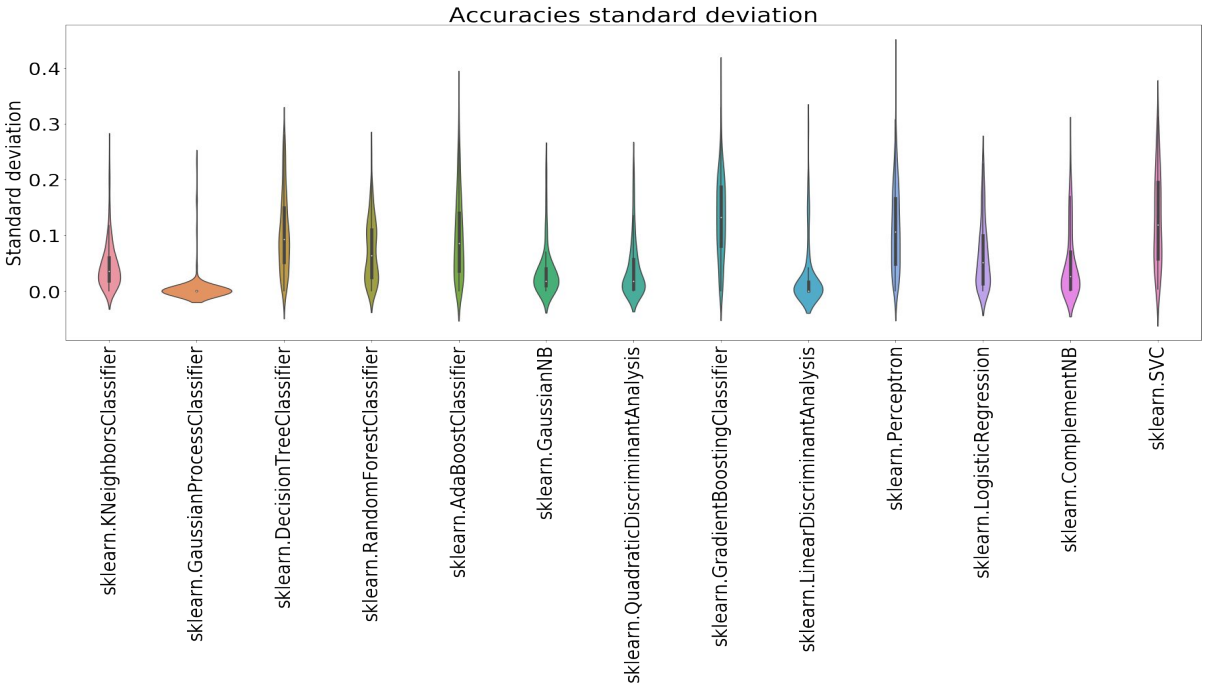
To tune or not to tune ?



No difference between best and worst run

AdaBoostM1	23.5 %
LogisticRegression	22.4 %
OneR	18.2 %
Decision Table	11.5 %

To tune or not to tune ?



No difference between best and worst run

GaussianProcessClassifier	63.5 %
LinearDiscriminantAnalysis	47.7 %
ComplementNB	20.7 %
QDA	19.1 %

To tune or not to tune ?

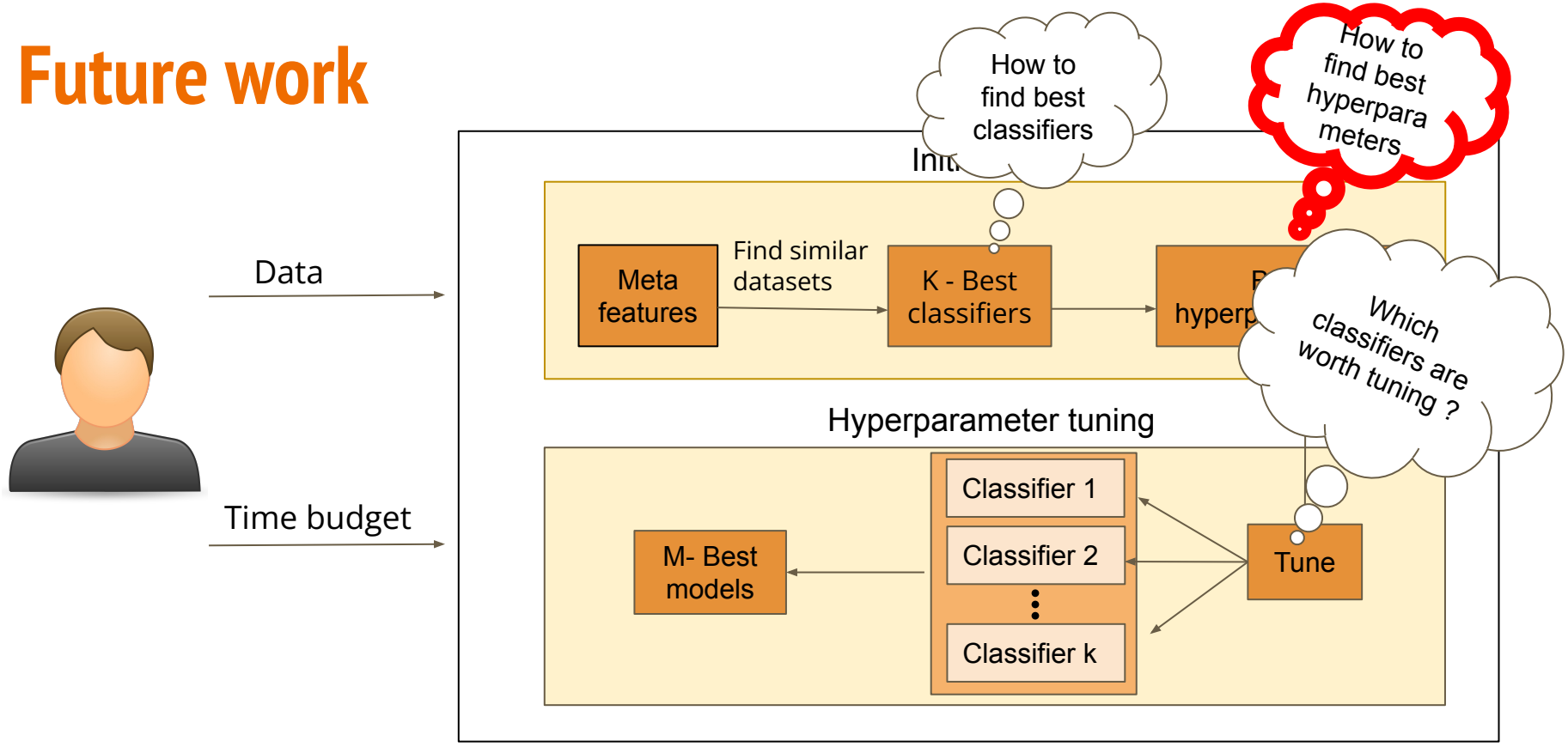
Define a tuning score s as :

$$s = \mu + \frac{T \cdot \sigma}{{c_t}^2}$$

Use autoSklearn Regressor to predict :

- Training accuracy — RMSE : 0.12
- Accuracy standard deviation — RMSE : 0.07
- Training time — RMSE : 154.8 MAE : 23.98
- Confidence interval — RMSE : 0.06

Future work


How to find best classifiers
How to find best hyperpara meters
Data
Time budget
Meta features
Find similar datasets
K - Best classifiers
Which classifiers are worth tuning ?
Hyperparameter tuning
Classifier 1
Classifier 2
Classifier k
M- Best models
Tune

Conclusions

Not all classifiers are worth including in AutoML pipeline

Not all classifiers are worth tuning

There is always going to be a trade-off between time and accuracy