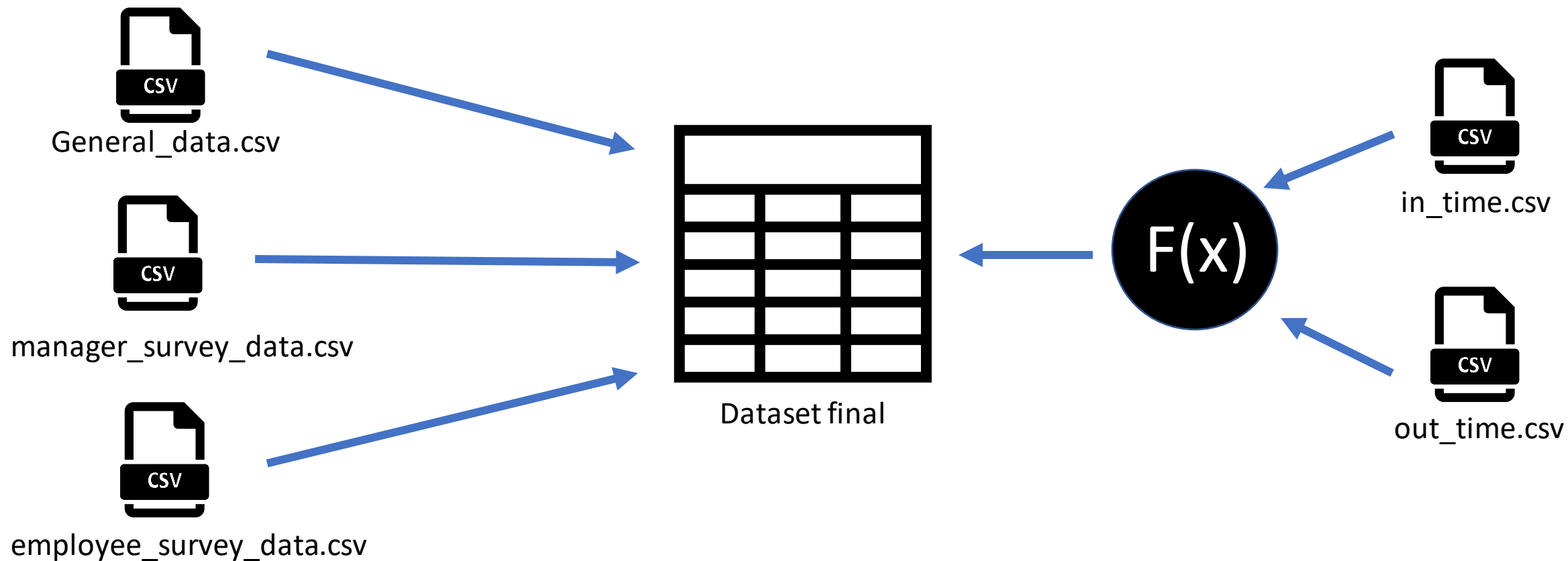# Intelligence artificielle

HumanForYou

Sommaire

- Préparation du dataset,

- Choix des Algorithmes retenus,

- Démarche mise en œuvre pour améliorer un modèle,

- Analyse des résultats obtenus,

- Choix du modèle,

- Propositions finales

Préparation du dataset

General_data.csv

manager_survey_data.csv

employee_survey_data.csv

Dataset final

F(x)

in_time.csv

out_time.csv

Préparation du dataset

```
#    Column                           #    Column
---  ------                          ---  ------
0    Age                              0    Attrition
1    Attrition                        1    BusinessTravel
2    BusinessTravel                   2    Department
3    Department                       3    DistanceFromHome
4    DistanceFromHome                 4    Education
5    Education                        5    EducationField
6    EducationField                   6    EmployeeID
7    EmployeeCount                    7    JobLevel
8    EmployeeID                       8    JobRole
9    Gender                           9    MonthlyIncome
10   JobLevel                         10   NumCompaniesWorked
11   JobRole                          11   PercentSalaryHike
12   MaritalStatus                    12   StockOptionLevel
13   MonthlyIncome                    13   TotalWorkingYears
14   NumCompaniesWorked               14   TrainingTimesLastYear
15   Over18                           15   YearsAtCompany
16   PercentSalaryHike                16   YearsSinceLastPromotion
17   StandardHours                    17   YearsWithCurrManager
18   StockOptionLevel                 18   EnvironmentSatisfaction
19   TotalWorkingYears                19   JobSatisfaction
20   TrainingTimesLastYear            20   WorkLifeBalance
21   YearsAtCompany                   21   JobInvolvement
22   YearsSinceLastPromotion          22   PerformanceRating
23   YearsWithCurrManager             23   AttendanceScore
24   EnvironmentSatisfaction
25   JobSatisfaction
26   WorkLifeBalance
27   JobInvolvement
28   PerformanceRating
29   AttendanceScore
```

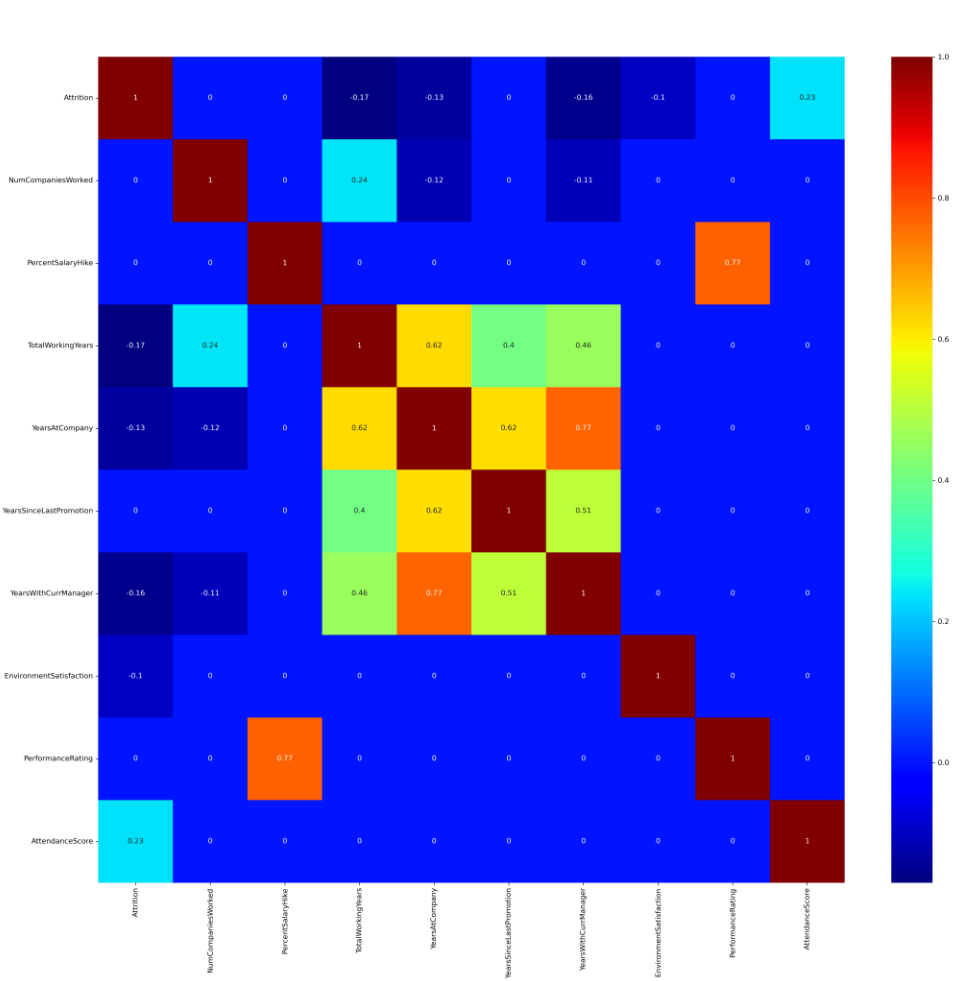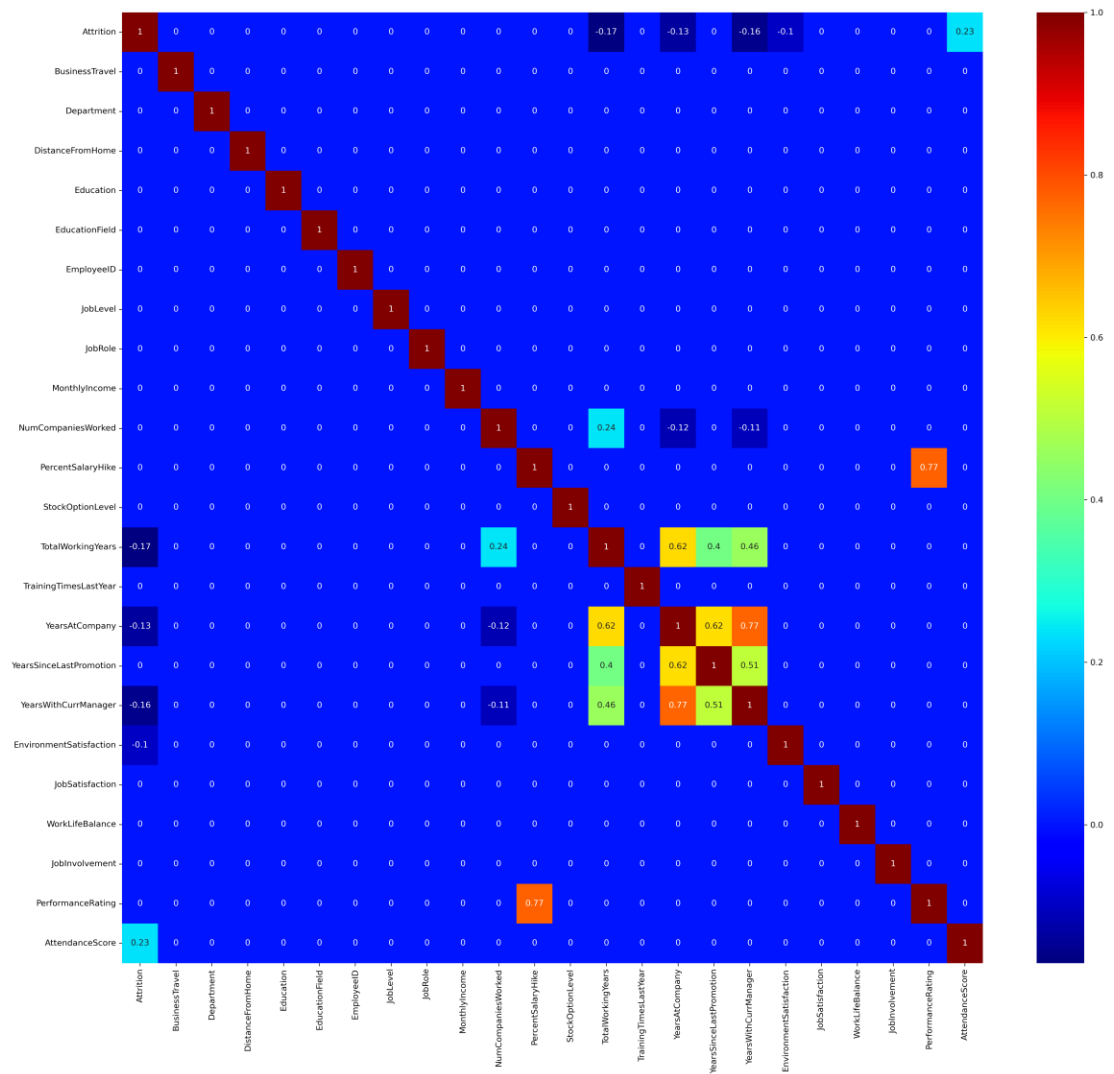Dataset après fusion

Après suppression
de colonnes

4

Préparation du dataset

Préparation du dataset

```
  #   Column
---   ------
  0   Age
  1   Attrition
  2   BusinessTravel
  3   Department
  4   DistanceFromHome
  5   Education
  6   EducationField
  7   EmployeeCount
  8   EmployeeID
  9   Gender
 10   JobLevel
 11   JobRole
 12   MaritalStatus
 13   MonthlyIncome
 14   NumCompaniesWorked
 15   Over18
 16   PercentSalaryHike
 17   StandardHours
 18   StockOptionLevel
 19   TotalWorkingYears
 20   TrainingTimesLastYear
 21   YearsAtCompany
 22   YearsSinceLastPromotion
 23   YearsWithCurrManager
 24   EnvironmentSatisfaction
 25   JobSatisfaction
 26   WorkLifeBalance
 27   JobInvolvement
 28   PerformanceRating
 29   AttendanceScore
```

Dataset après fusion

```
  #   Column
---   ------
  0   Attrition
  1   BusinessTravel
  2   Department
  3   DistanceFromHome
  4   Education
  5   EducationField
  6   EmployeeID
  7   JobLevel
  8   JobRole
  9   MonthlyIncome
 10   NumCompaniesWorked
 11   PercentSalaryHike
 12   StockOptionLevel
 13   TotalWorkingYears
 14   TrainingTimesLastYear
 15   YearsAtCompany
 16   YearsSinceLastPromotion
 17   YearsWithCurrManager
 18   EnvironmentSatisfaction
 19   JobSatisfaction
 20   WorkLifeBalance
 21   JobInvolvement
 22   PerformanceRating
 23   AttendanceScore
```

Après suppression
de colonnes

```
  #   Column
---   ------
  0   Attrition
  1   TotalWorkingYears
  2   YearsAtCompany
  3   YearsWithCurrManager
  4   EnvironmentSatisfaction
  5   AttendanceScore
```

Après corrélation

Choix des Algorithmes retenus

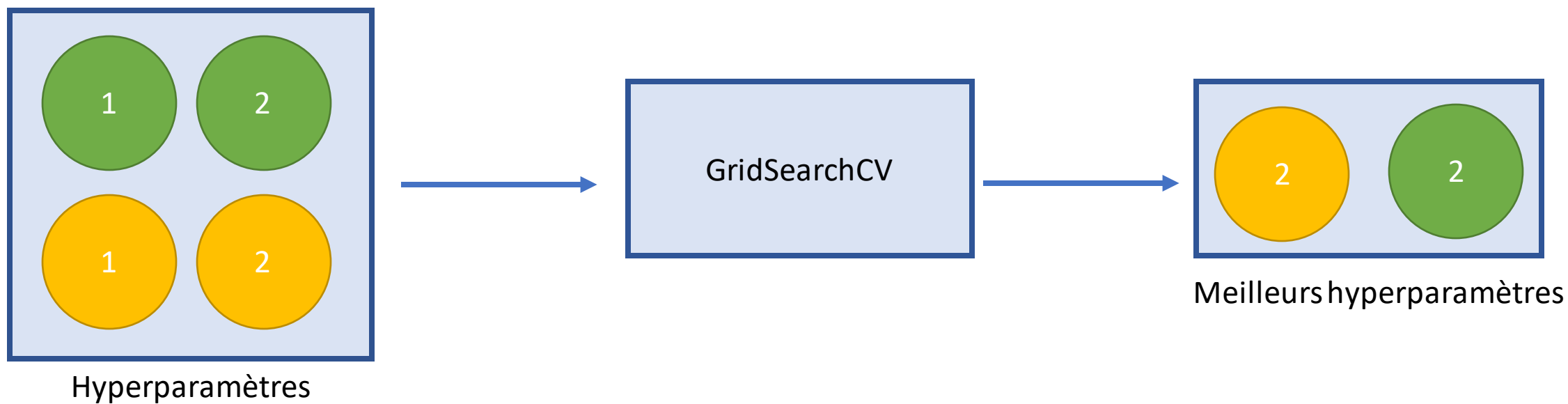DecisionTreeClassifier

RandomForestClassifier

Ridge

SGDClassifier
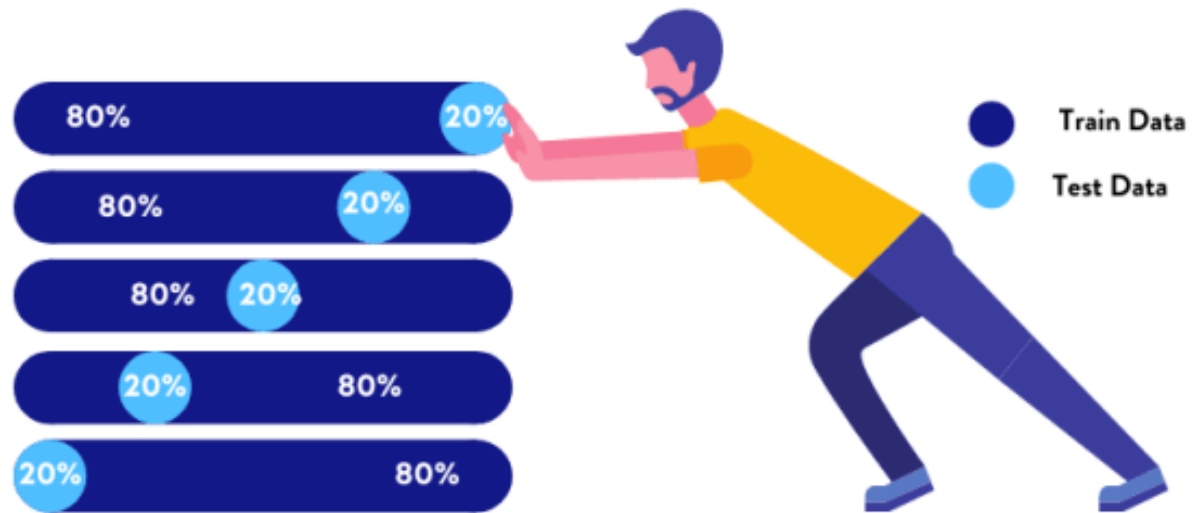
LogisticRegression

Perceptron

SVC

KNeighborsClassifier
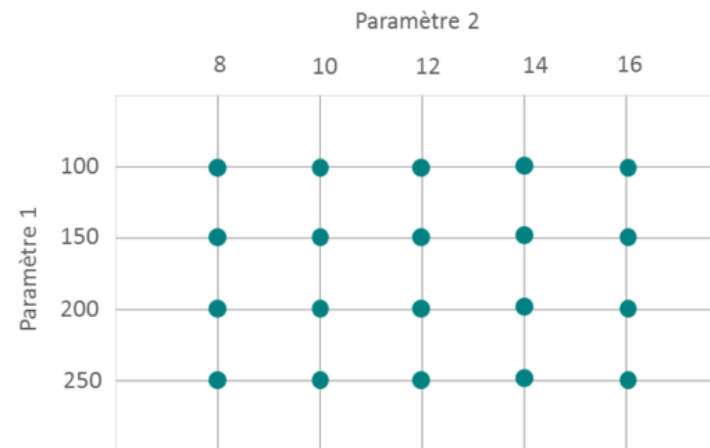
Démarche mise en œuvre pour entrainer un modèle

Hyperparamètres

GridSearchCV

Meilleurs hyperparamètres

La cross-validation

Grid Search

Démarche mise en œuvre pour entrainer un modèle

```
{
    'name': 'RandomForestClassifier',
    'method': RandomForestClassifier(random_state=100),
    "parameters": {
        "n_estimators": [100, 200, 300, 400] ,
        "max_features": [1, 2, 3, 4, 5, 6, 7, 8, 9] ,
        "bootstrap": [True, False],
    }
}
```

Analyse des résultats obtenus

	accuracy
RandomForestClassifier	0.938776
DecisionTreeClassifier	0.921769
KNeighborsClassifier	0.883220
LogisticRegression	0.862812
Ridge	0.848073
SVC	0.848073
Perceptron	0.757370
SGDClassifier	0.744898

	precision
RandomForestClassifier	0.850877
LogisticRegression	0.809524
DecisionTreeClassifier	0.730496
KNeighborsClassifier	0.701299
Perceptron	0.305825
SGDClassifier	0.286385
Ridge	NaN
SVC	NaN

	recall
DecisionTreeClassifier	0.768657
RandomForestClassifier	0.723881
Perceptron	0.470149
SGDClassifier	0.455224
KNeighborsClassifier	0.402985
LogisticRegression	0.126866
Ridge	0.000000
SVC	0.000000

	f1-score
RandomForestClassifier	0.782258
DecisionTreeClassifier	0.749091
KNeighborsClassifier	0.511848
Perceptron	0.370588
SGDClassifier	0.351585
LogisticRegression	0.219355
Ridge	NaN
SVC	NaN

Choix final du modèle

Choix final du modèle

Choix final du modèle

Temps d'éxécution en ms
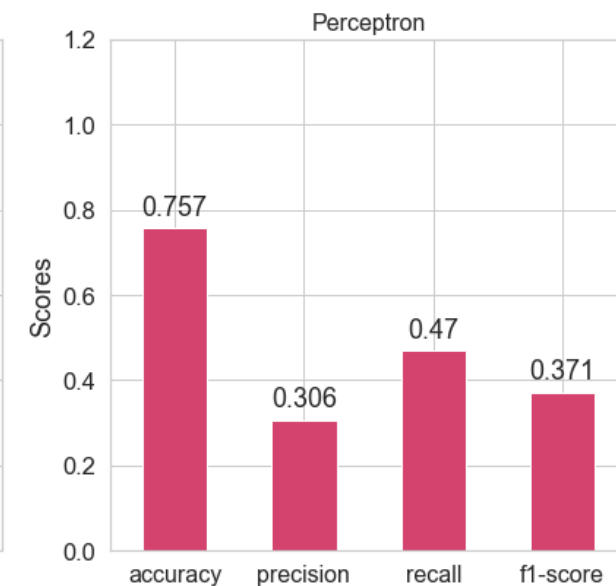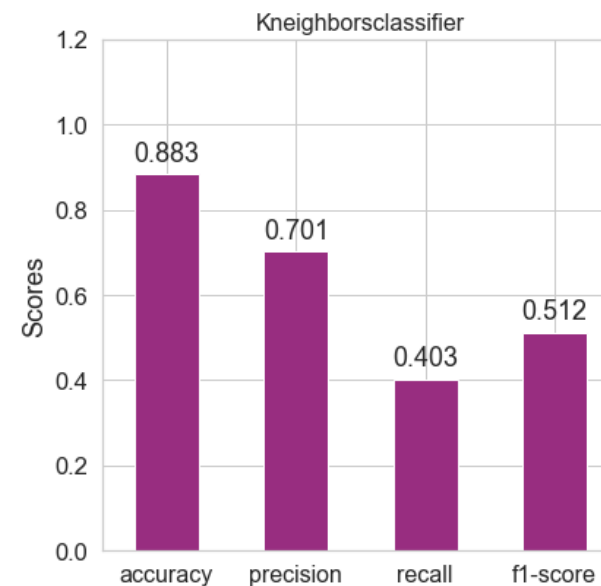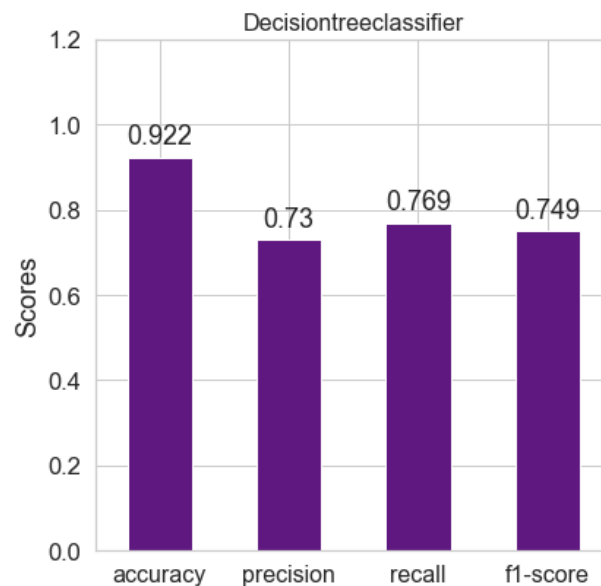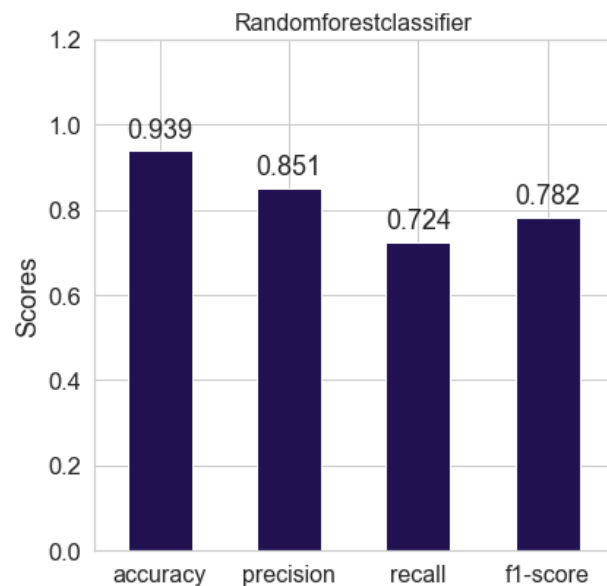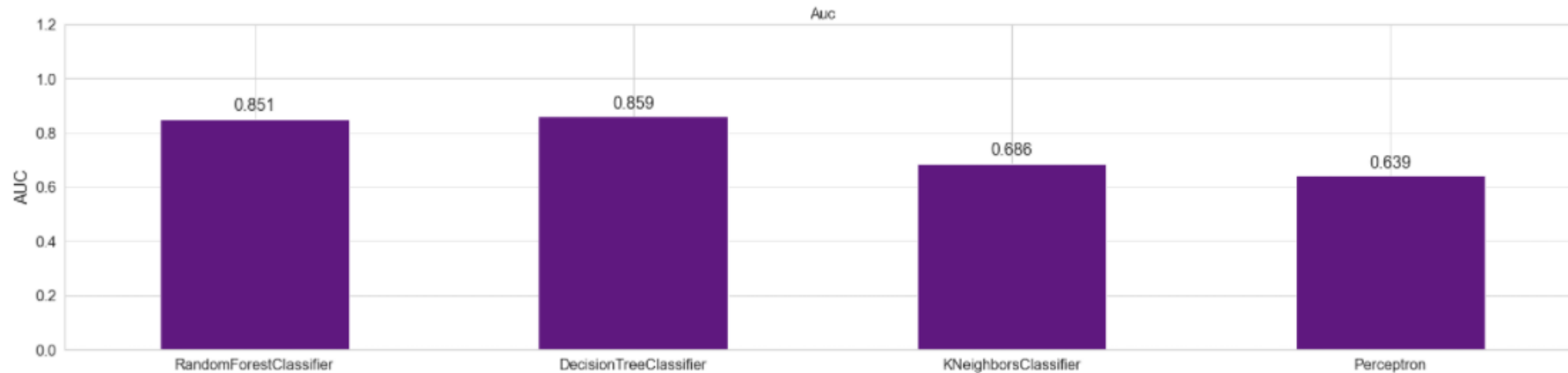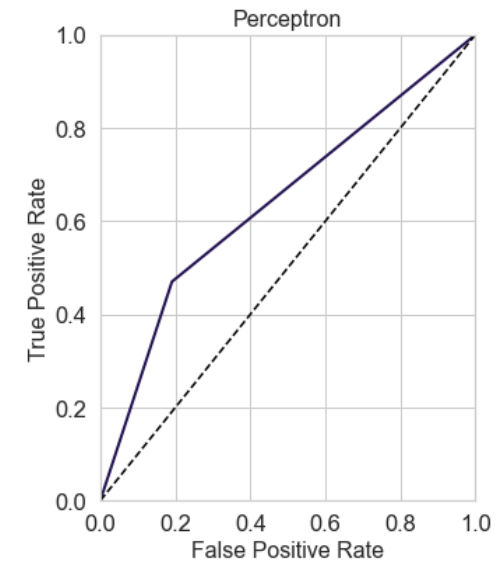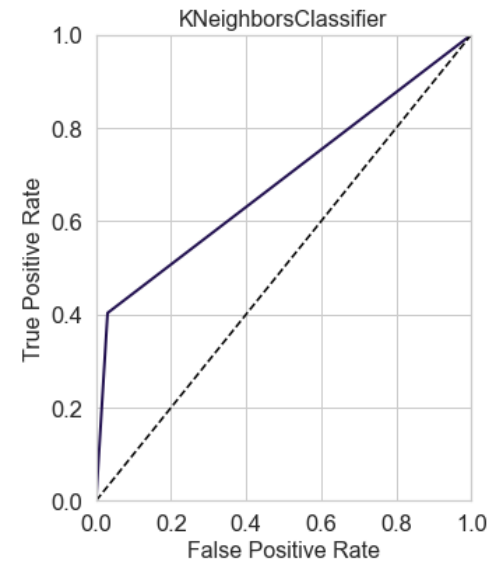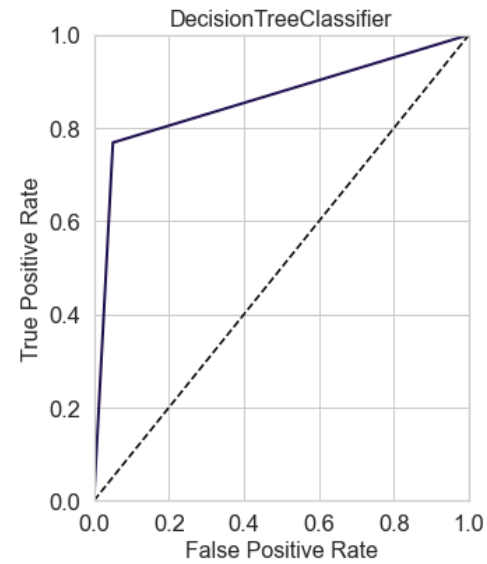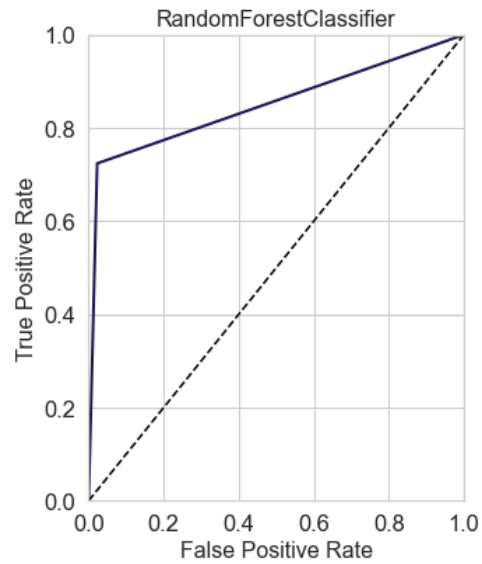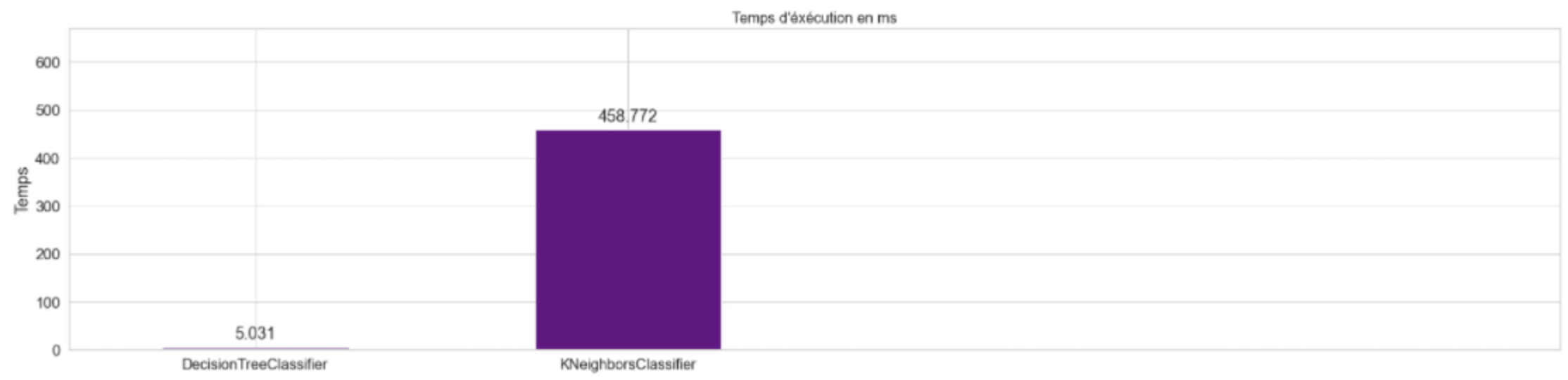
Propositions finales
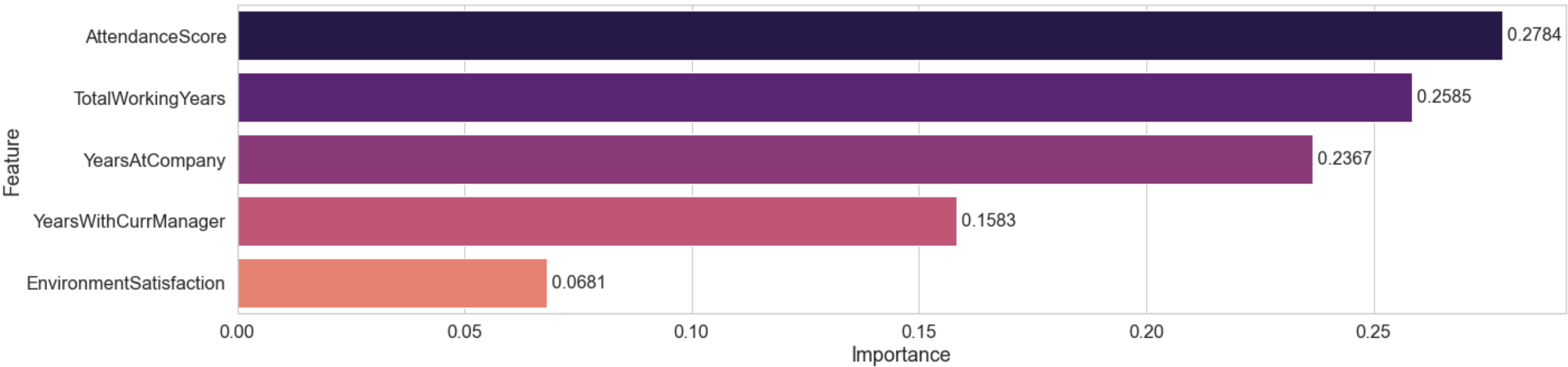
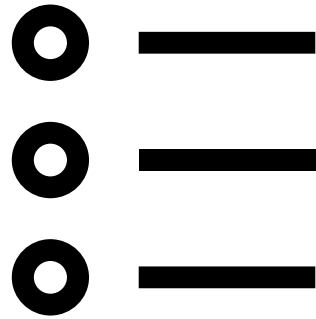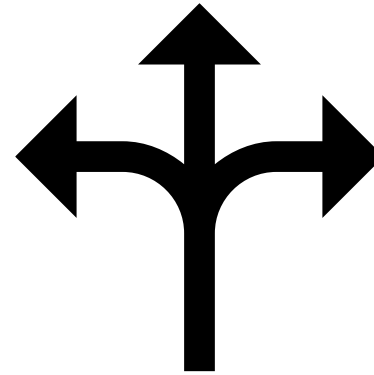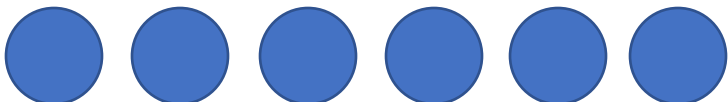# Propositions finales

Sondages
semestriels

Diversification des
tâches

Réorientation

Récompenses

FIN

Merci de votre attention