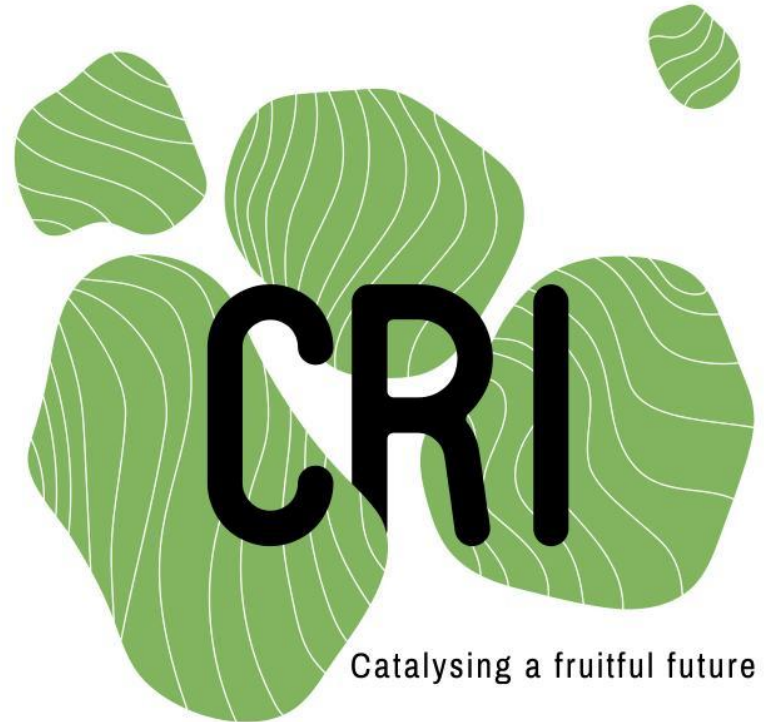


Survival Modelling for Titanic passengers

SIDIBE Moussa
CULPIN Alexis



SOMMAIRE

1

Feature
engineering

2

Data splitting

3

Modelisation

4

Model
evaluation

Feature engineering

Age

Missing values

Filling with nearest
value by interpolation

Feature engineering

Embarked

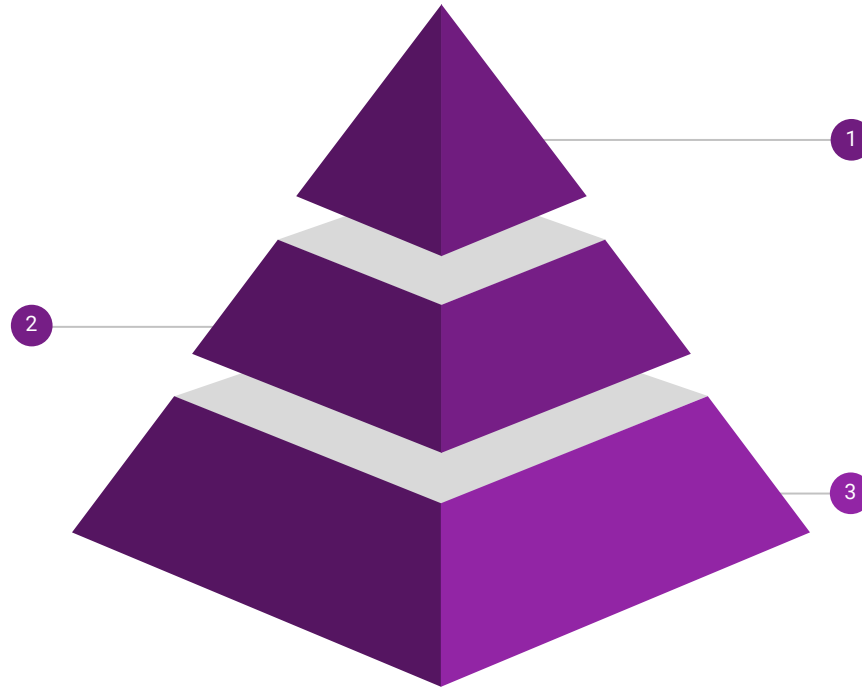
Missing values

Filling with high mode
value

Cabin

Get Cabin labels for each unique values of Fare

Cabin is calculated with based on the most neighbor fare for defined cabin



Converting Fare to integer and unique values

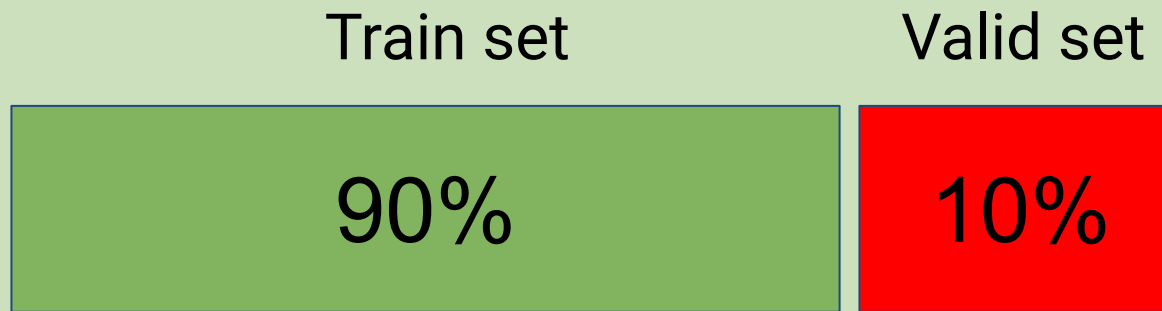
1 Separate Fares with Cabin label and others without label

Taking the first character of cabin label

3 To avoid high dimensionality and mismatching with test cabin labels



Data split



Method

LDA

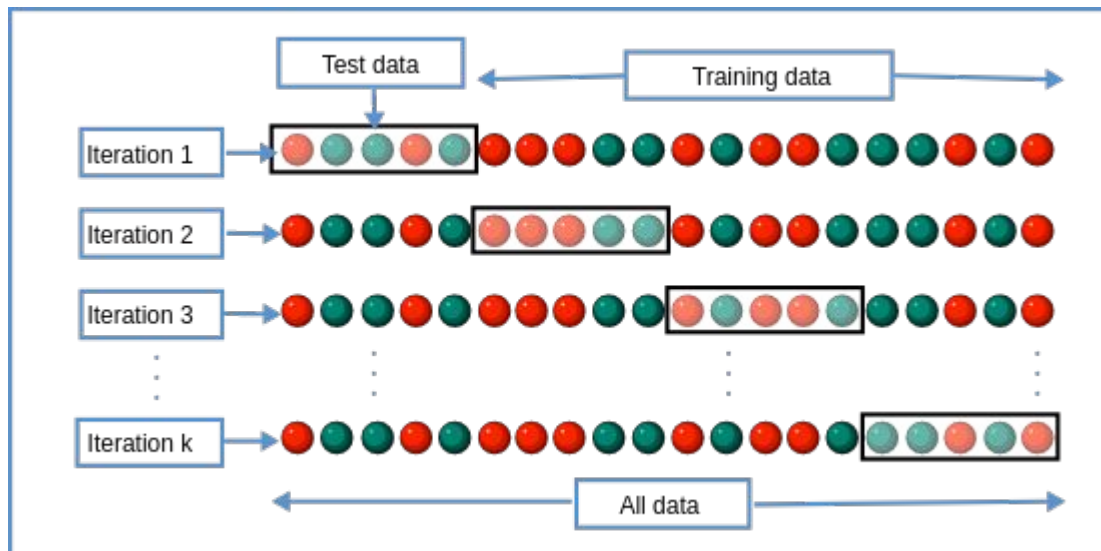
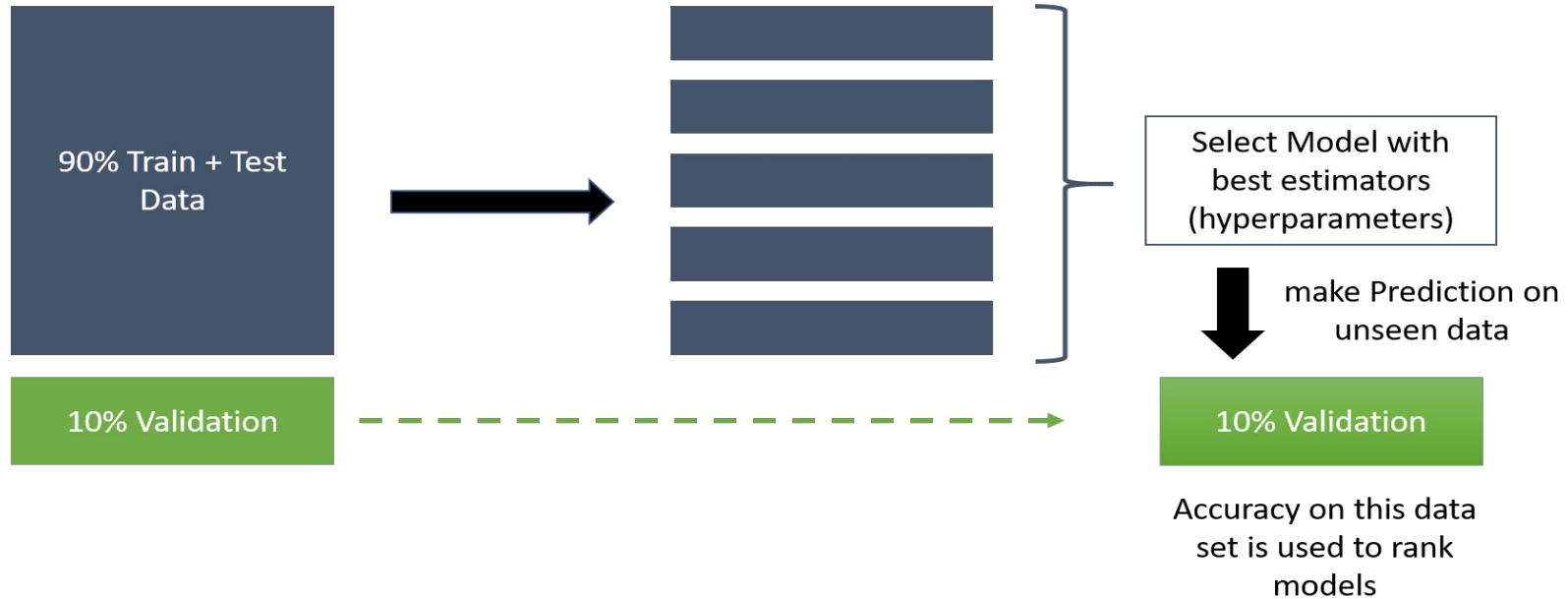
Logistic Regression

KNN

GridSearchCV

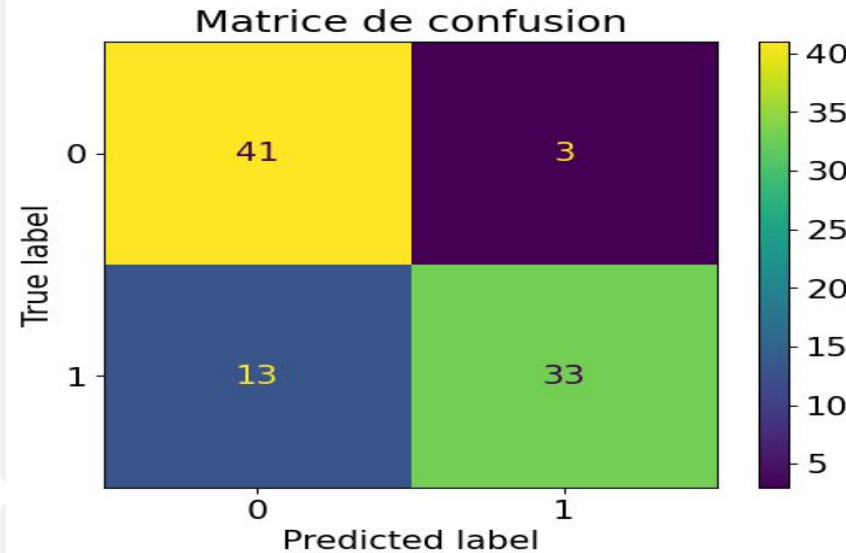
`Train_test_split()`

5-fold cross-validation



Model evaluation

Logistic regression



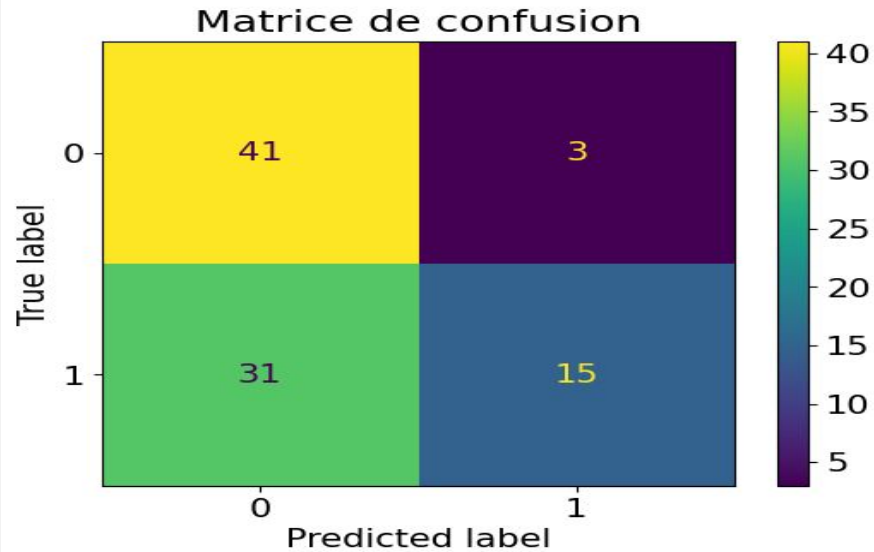
```
y_predict=logreg_cv.predict(X_valid)
print(classification_report(y_valid,y_predict))
```

✓ 0.7s

	precision	recall	f1-score	support
0	0.76	0.93	0.84	44
1	0.92	0.72	0.80	46
accuracy			0.82	90
macro avg	0.84	0.82	0.82	90
weighted avg	0.84	0.82	0.82	90

Model evaluation

KNN



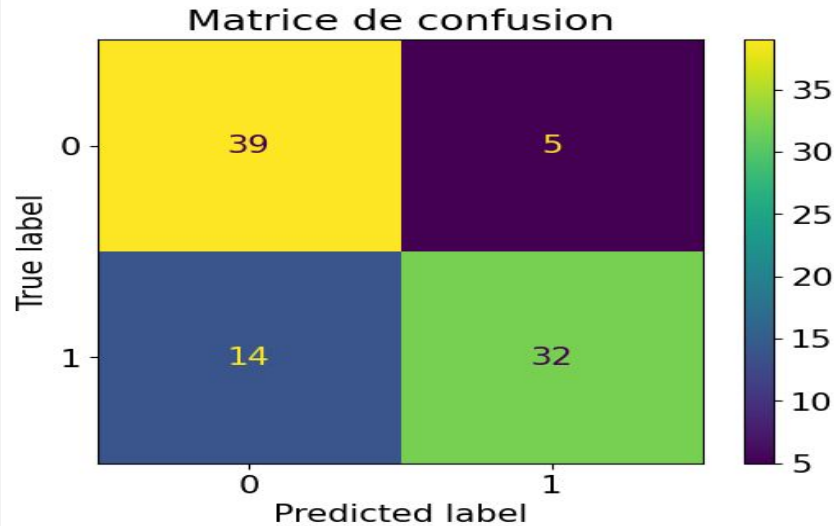
```
y_predict=grid_search_knn.predict(X_valid)  
print(classification_report(y_valid,y_predict))
```

✓ 0.8s

	precision	recall	f1-score	support
0	0.57	0.93	0.71	44
1	0.83	0.33	0.47	46
accuracy			0.62	90
macro avg	0.70	0.63	0.59	90
weighted avg	0.70	0.62	0.59	90

Model evaluation

LDA



```
y_predict=lda_search.predict(X_valid)
print(classification_report(y_valid,y_predict))
```

✓ 0.1s

	precision	recall	f1-score	support
0	0.74	0.89	0.80	44
1	0.86	0.70	0.77	46
accuracy			0.79	90
macro avg	0.80	0.79	0.79	90
weighted avg	0.80	0.79	0.79	90

Model evaluation

Summary

LR get an accuracy of 79.44% with STD 7.05%

KNN get an accuracy of 63.33% with STD 7.11%

LDA get an accuracy of 79.17% with STD 10.52%

Important variable

SelectKBest

– score_func=f_classif

- Fare
- Pclass
- Sexe
- Embarked (S,C)
- Cabin (C,E,F)

Model evaluation

Summary

LR get an accuracy of 76.315%

KNN get an accuracy of 62.918%

LDA get an accuracy of 77.751%

Thank you for
your attention

