

Do you have Covid-19 based on your symptoms?

AGENDA

1. Experiments:
preprocessing and
methods
2. Analysis of results
3. Comparison with original
study

kaggle

	gender	age_year	fever	cough	runny_nose	muscle_soreness	pneumonia	diarrhea	lung_infection	travel_history	isolation_treatment	SARS-CoV-2 Positive
0	male	89	1	1	0	0	0	0	0	1	0	0
1	male	68	1	0	0	0	0	0	0	0	0	0
2	male	68	0	0	0	0	0	0	0	1	0	0
3	male	68	1	1	0	0	0	0	0	1	1	1
4	male	50	1	1	1	0	1	0	0	1	0	1
...
6507	female	44	1	1	0	0	0	0	0	1	0	1
6508	female	44	1	1	0	0	0	0	0	0	0	0
6509	female	58	0	0	0	0	0	0	0	0	0	0
6510	female	58	1	1	0	0	0	0	0	0	0	0
6511	male	12	1	1	0	0	0	0	0	0	1	0

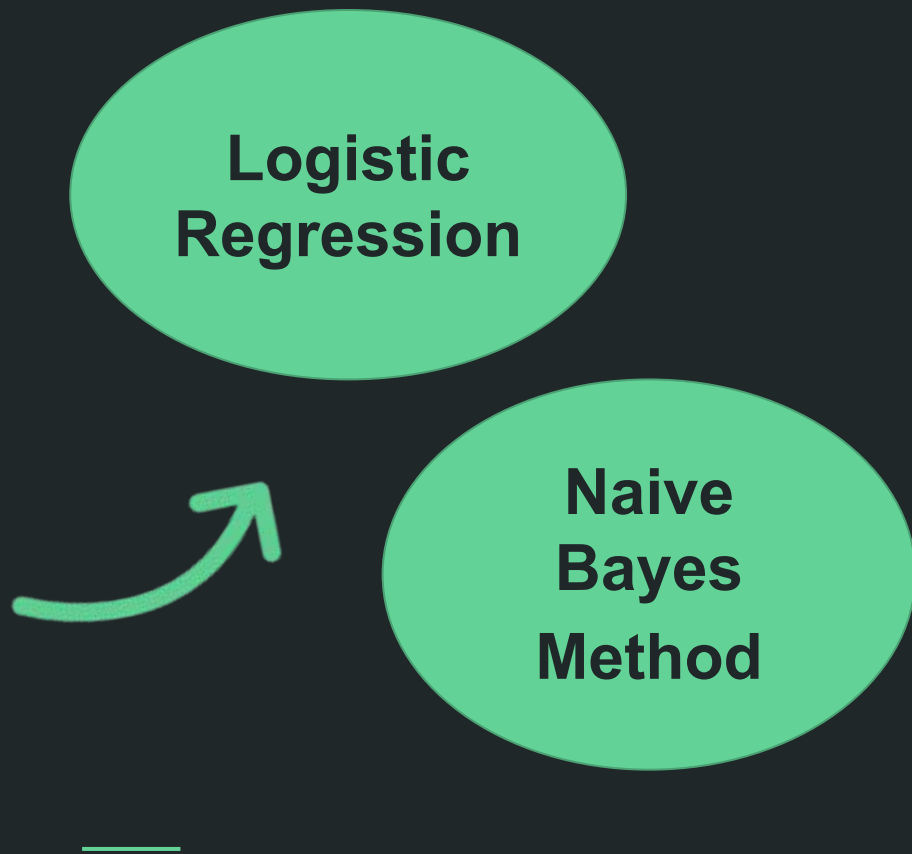
6512 rows × 12 columns

Early stage symptoms of COVID-19 patients
6,512 samples | Kaggle

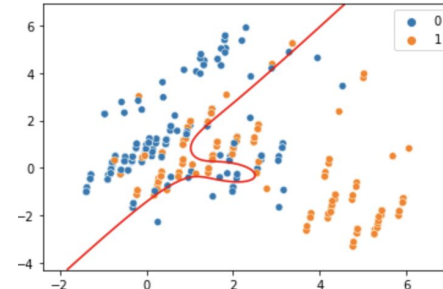
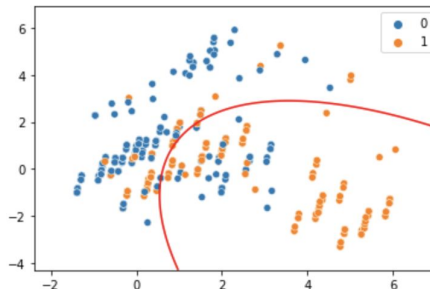
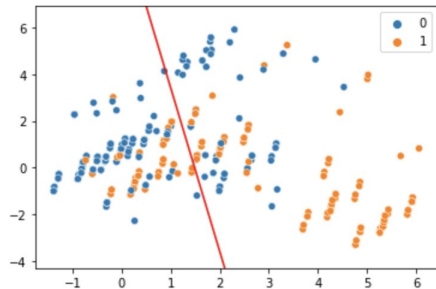
<https://www.kaggle.com/martuza/early-stage-symptoms-of-covid19-patients>

Preprocessing of the Dataset

- Encoding
- Normalization
- Reduction of dimensions :
PCA



Results



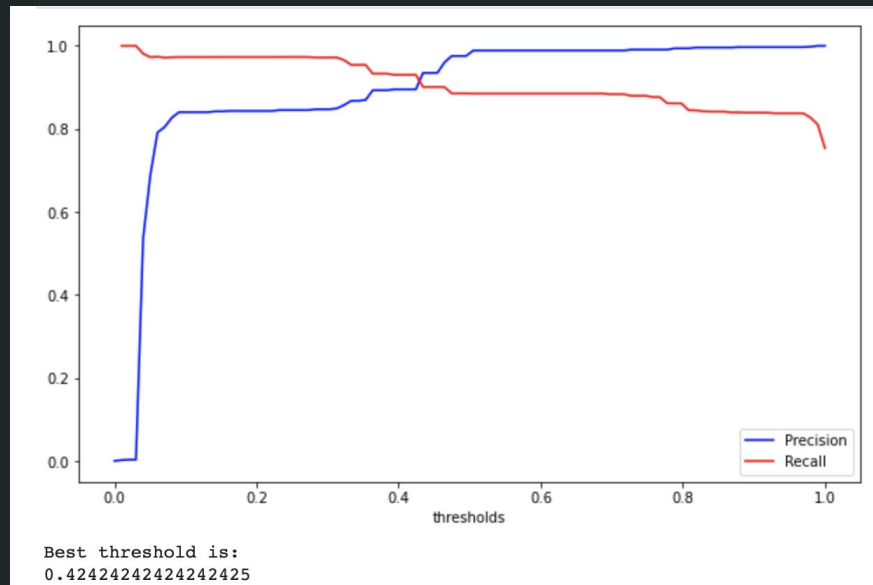
	Logistic Regression	LR lower threshold	polyRegression deg = 2	polyRegression deg = 3	Naive Bayes
TP	958	879	956	964	860
FP	24	103	26	18	122
FN	124	66	170	186	91
TN	197	255	151	135	230
Precision	0.97556	0.895112	0.973523	0.98167	0.875764
Recall	0.885397	0.930159	0.849023	0.838261	0.904311
Accuracy	0.886416	0.870299	0.849578	0.843438	0.836531
Specificity	0.891403	0.712291	0.853107	0.882353	0.653409

Comparison with original study

	XGBoost	GBM	SVM	Random Forest	DTree
Precision	0.91	0.93	0.95	0.91	0.90
Recall	0.92	0.88	0.87	0.90	0.90
Accuracy	0.88	0.89	0.93	0.88	0.89

	Logistic Regression	LR lower threshold
Precision	0.97	0.90
Recall	0.88	0.93
Accuracy	0.88	0.87

Trade-off between Precision and Recall



Thank you for your attention

Team members:

Guglielmo Felici, Alice Palmucci, Aurelie Pommier