

Developing predictive models for COVID-19 diagnosis in paediatric patients

A case study about the potentials of Machine Learning in Public Health

By Anna Mas-Casadesús

15th July 2020

1



WHAT IS THE PROBLEM?



1 → WHAT IS THE PROBLEM?



1 → WHAT IS THE PROBLEM?

Children appear to have a **similar prevalence** of COVID-19 antibodies to adults
But most of them tend to present **mild or no symptoms**
Making **COVID-19 diagnosis in children particularly difficult**

<https://www.sjdhospitalbarcelona.org/en/kidscorona>



1 → WHAT IS THE PROBLEM?

Children appear to have a **similar prevalence** of COVID-19 antibodies to adults

But most of them tend to present **mild or no symptoms**

Making **COVID-19 diagnosis in children particularly difficult**



<https://www.sjdhospitalbarcelona.org/en/kidscorona>

Dataset on paediatric patients with and without COVID-19 but similar symptoms



1 → WHAT IS THE PROBLEM?

Children appear to have a **similar prevalence** of COVID-19 antibodies to adults

But most of them tend to present **mild or no symptoms**

Making **COVID-19 diagnosis in children particularly difficult**



<https://www.sjdhospitalbarcelona.org/en/kidscorona>

Dataset on paediatric patients with and without
COVID-19 but similar symptoms

Contact with **Dr. Antoni Soriano-Arandes**



1 → WHAT IS THE PROBLEM?

Children appear to have a **similar prevalence** of COVID-19 antibodies to adults

But most of them tend to present **mild or no symptoms**

Making **COVID-19 diagnosis in children particularly difficult**



<https://www.sjdhospitalbarcelona.org/en/kidscorona>

Dataset on paediatric patients with and without
COVID-19 but similar symptoms

Contact with Dr. Antoni Soriano-Arandes



Machine Learning classification models to
predict COVID-19 diagnose in children



2 → WHAT DO I NEED TO DO TO USE THE DATA?



2

WHAT DO I NEED TO DO TO USE THE DATA?

BASIC DATA CLEANING &
MANIPULATIONS

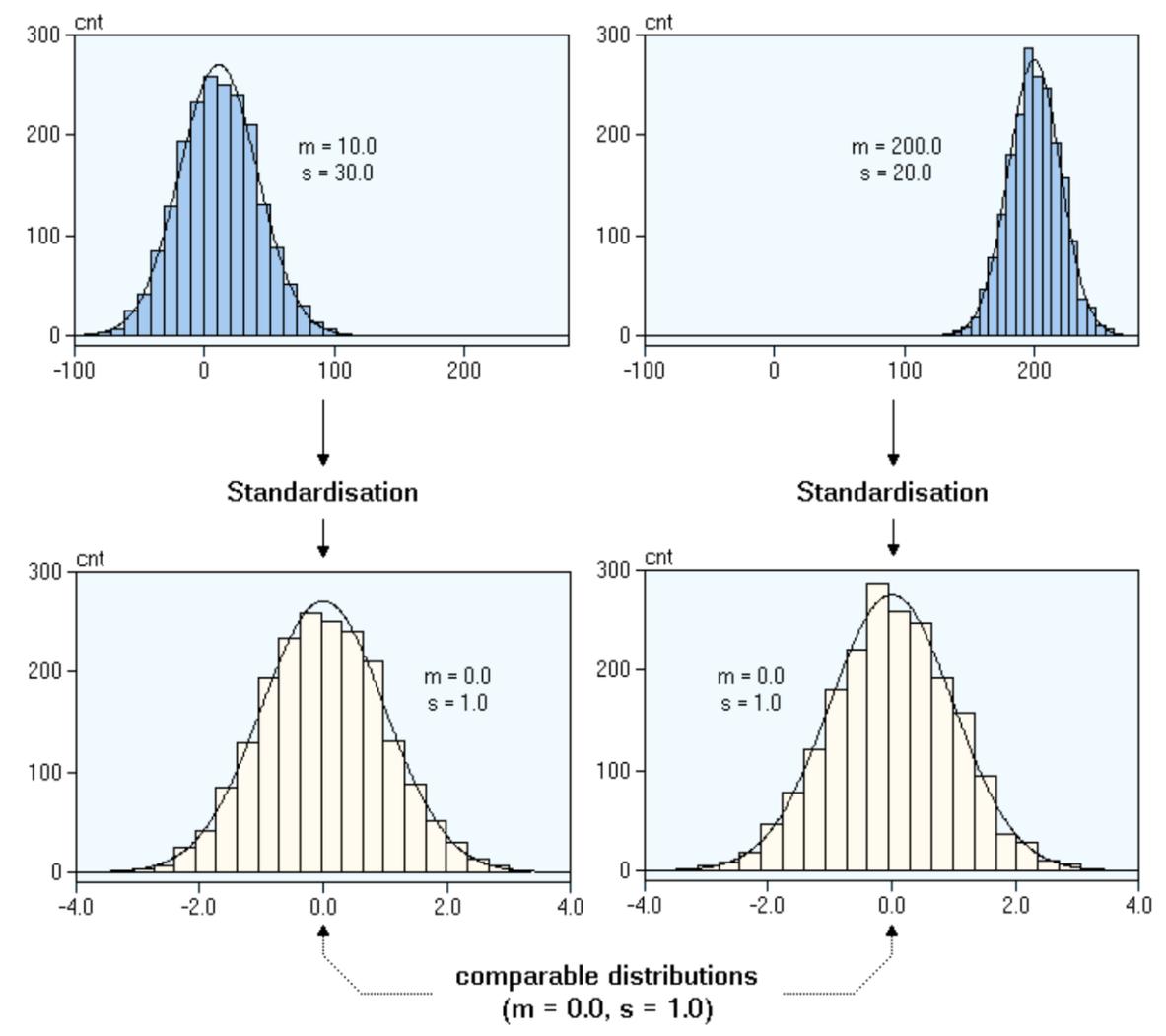
2

WHAT DO I NEED TO DO TO USE THE DATA?

BASIC DATA CLEANING & MANIPULATIONS



STANDARDISATION



2

WHAT DO I NEED TO DO TO USE THE DATA?

BASIC DATA CLEANING & MANIPULATIONS



STANDARDISATION



MISSING DATA MANAGEMENT

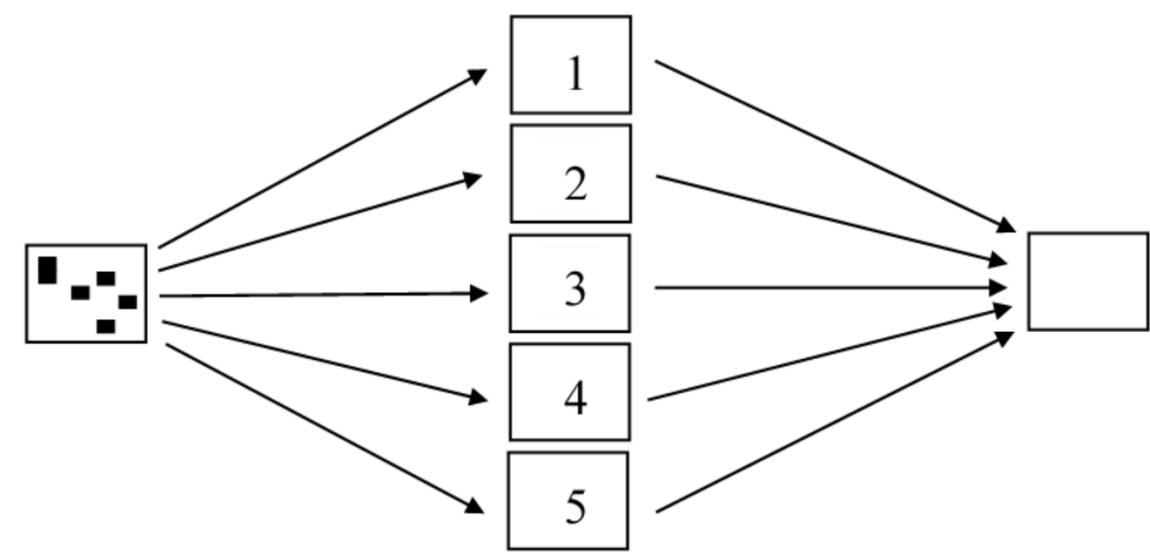
MICE METHOD

Incomplete Dataset

Imputation

Completed data analysis

Combining results



2

WHAT DO I NEED TO DO TO USE THE DATA?

BASIC DATA CLEANING & MANIPULATIONS



STANDARDISATION



MISSING DATA MANAGEMENT



FEATURE SELECTION

All Features



Feature Selection



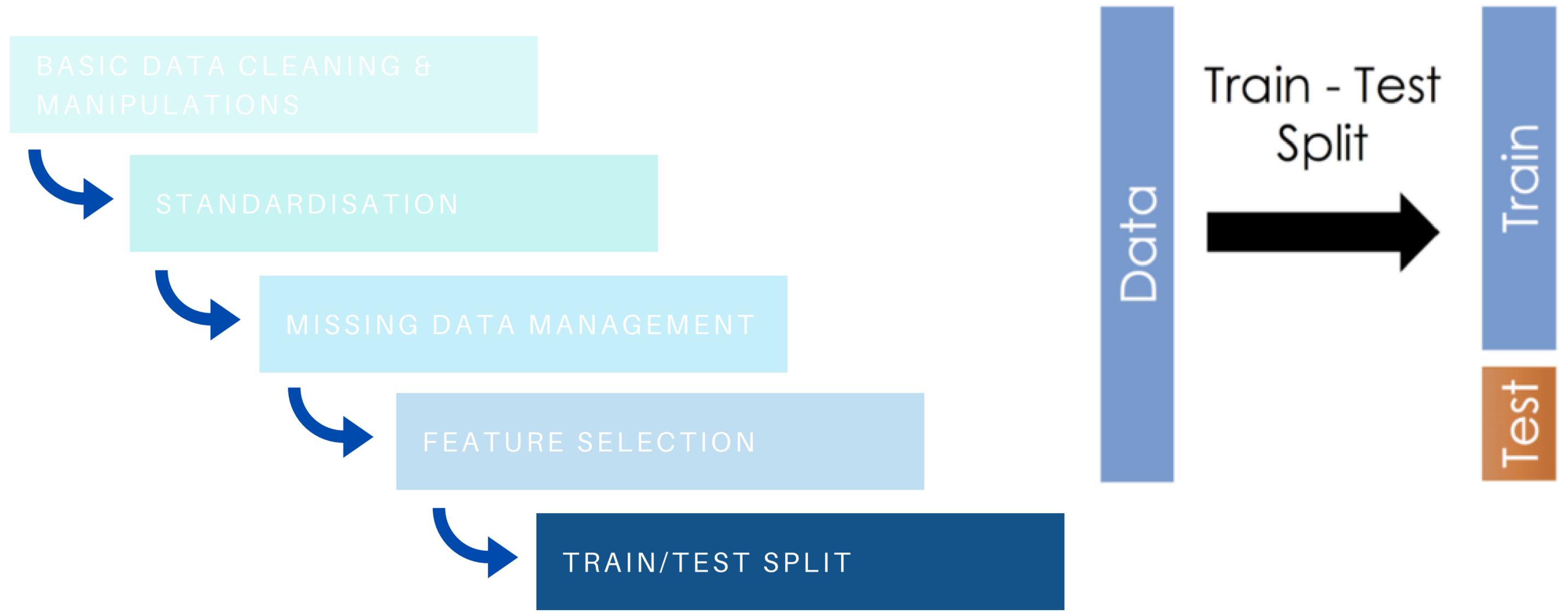
Final Features



E.G. PCA

2

WHAT DO I NEED TO DO TO USE THE DATA?



2 → WHAT DO I NEED TO DO TO USE THE DATA?

BASIC DATA CLEANING & MANIPULATIONS



STANDARDISATION



MISSING DATA MANAGEMENT



FEATURE SELECTION

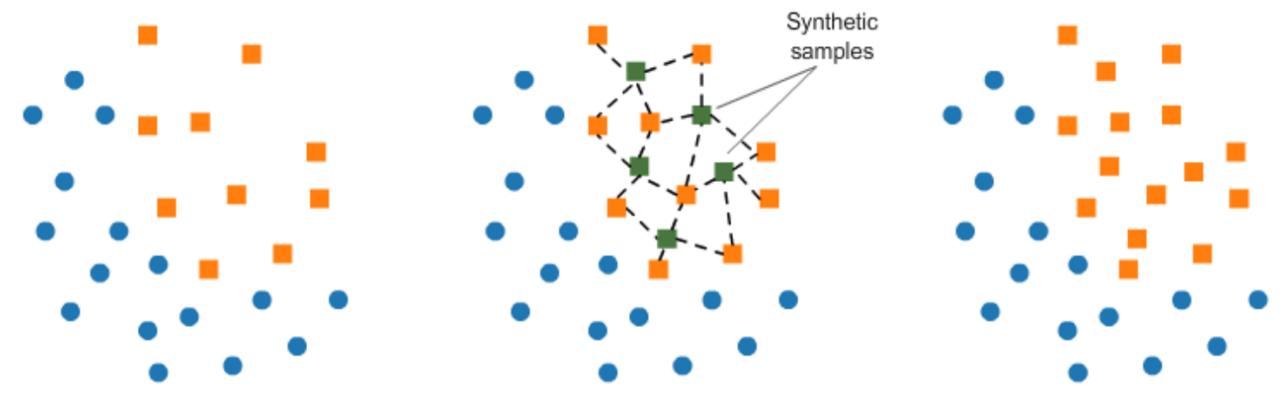
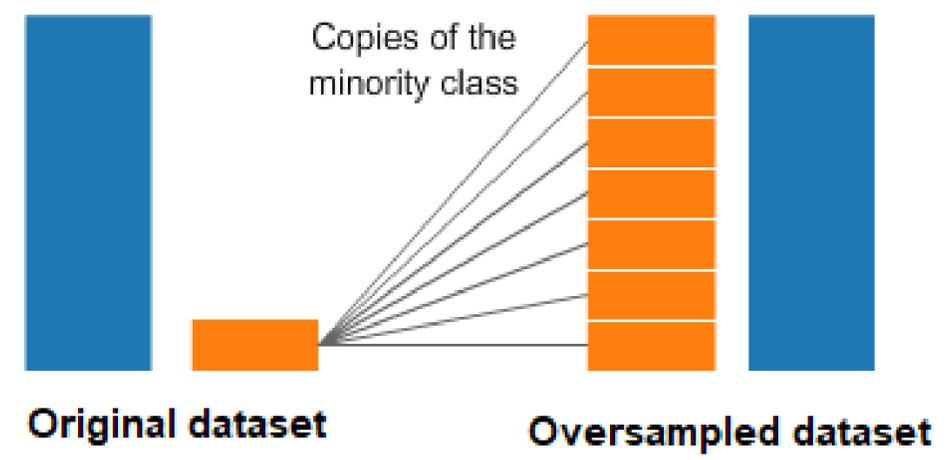


TRAIN/TEST SPLIT



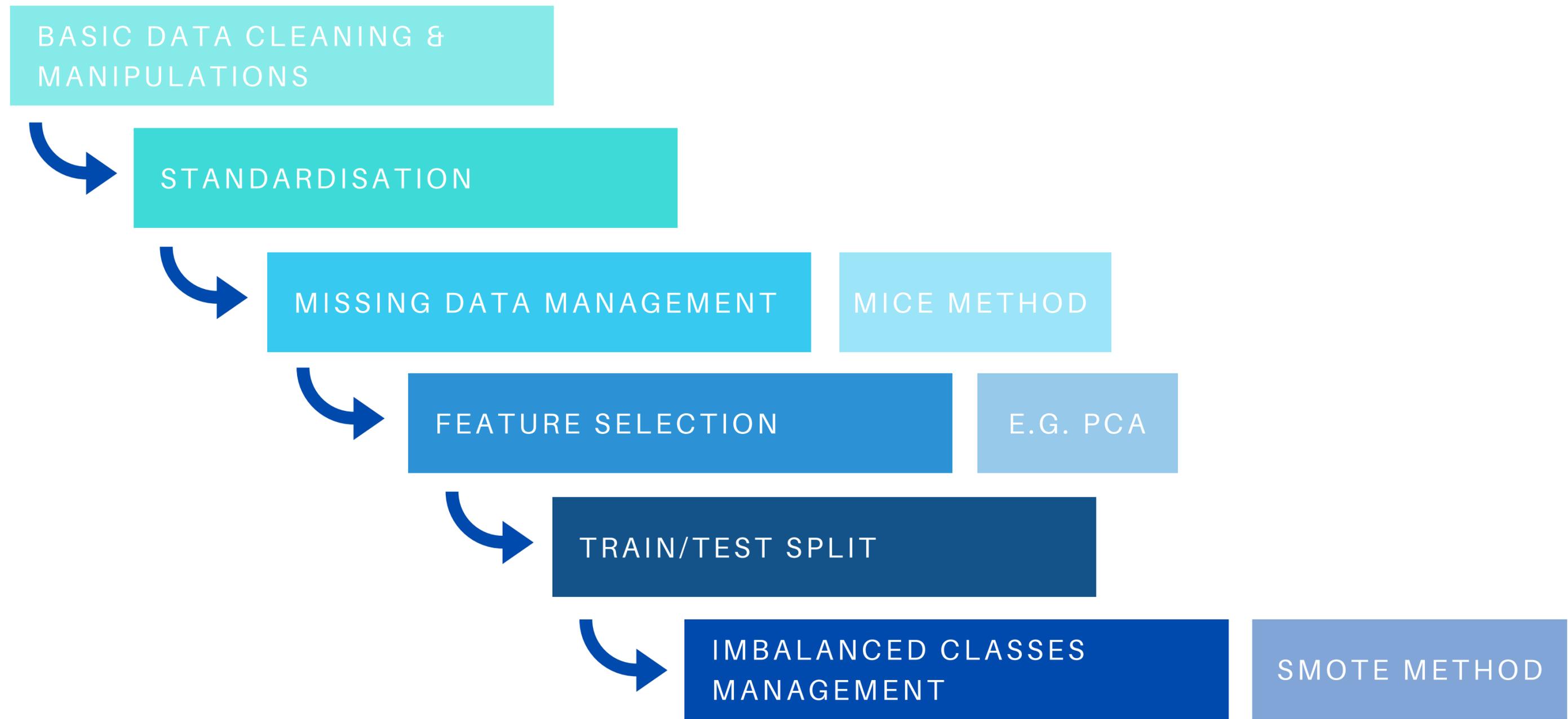
IMBALANCED CLASSES MANAGEMENT

SMOTE METHOD



2

WHAT DO I NEED TO DO TO USE THE DATA?



3



HOW DOES THE DATA LOOK LIKE?



3



HOW DOES THE DATA LOOK LIKE?

86 cases

19 CovidPos

67 CovidNeg

18 features

3

HOW DOES THE DATA LOOK LIKE?

86 cases

19 CovidPos

67 CovidNeg

18 features

Demographics

Symptoms

Previous conditions

Treatments

Tests

Evolution

Age

Gender

Respiratory symptoms

Gastrointestinal symp.

Pneumonia

Fever

Immunosuppression

Immunodeficiency

Other previous cond.

Oxygen therapy

Antibiotics

Corticoids

Chest x-ray

Neutrocytes

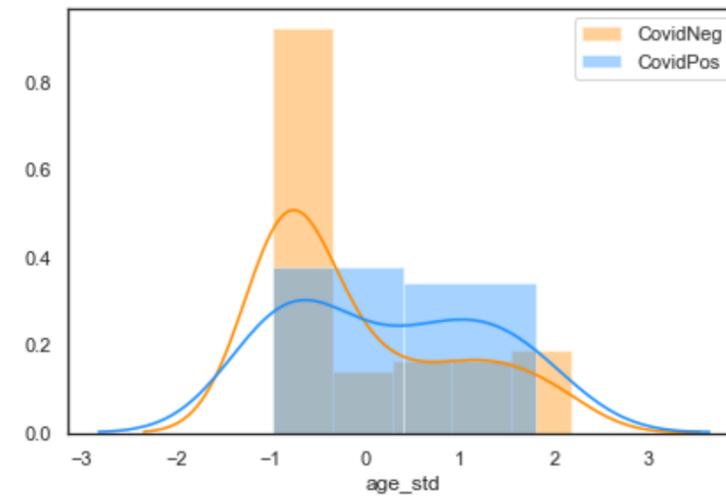
Lymphocytes

C-reactive protein

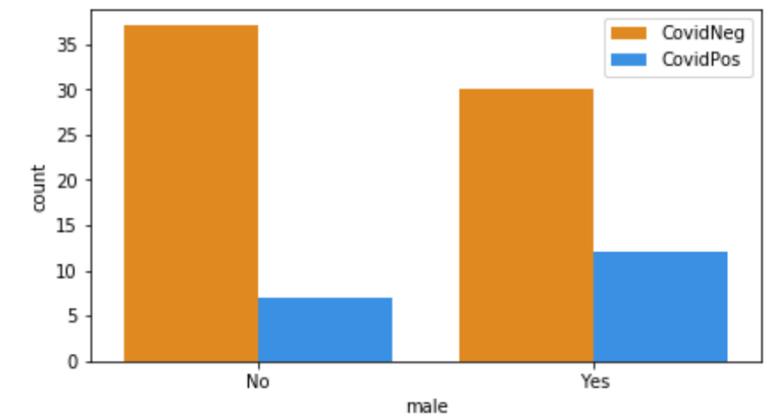
Diagnosis delay

Adm. Intensive Care Unit

AGE



GENDER



3

HOW DOES THE DATA LOOK LIKE?

86 cases

19 CovidPos

67 CovidNeg

18 features

Demographics

Symptoms

Previous conditions

Treatments

Tests

Evolution

Age

Gender

Respiratory symptoms

Gastrointestinal symp.

Pneumonia

Fever

Immunosuppression

Immunodeficiency

Other previous cond.

Oxygen therapy

Antibiotics

Corticoids

Chest x-ray

Neutrocytes

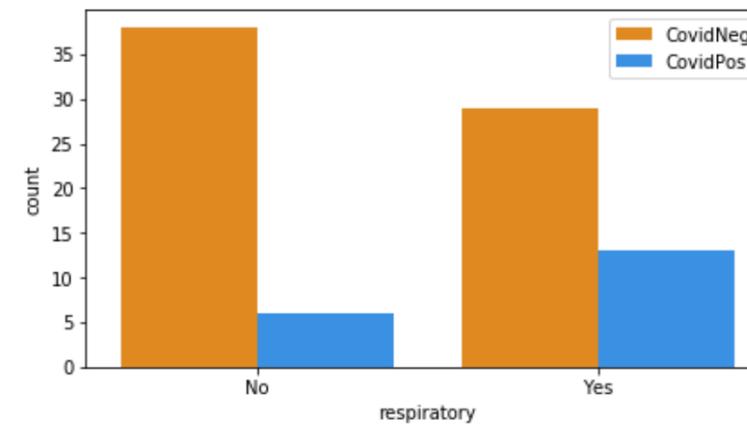
Lymphocytes

C-reactive protein

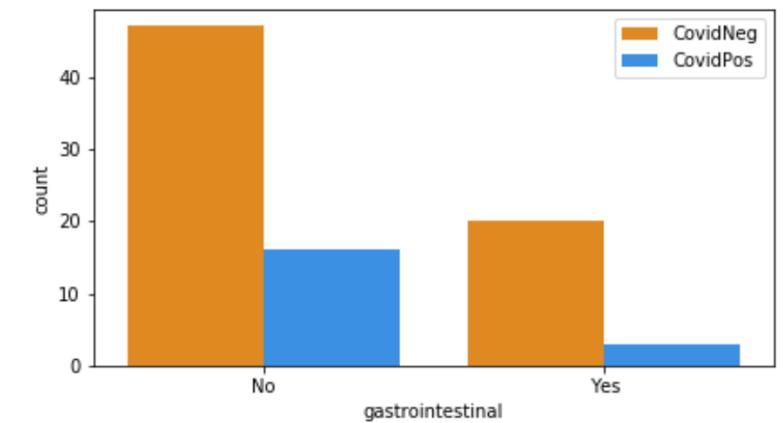
Diagnosis delay

Adm. Intensive Care Unit

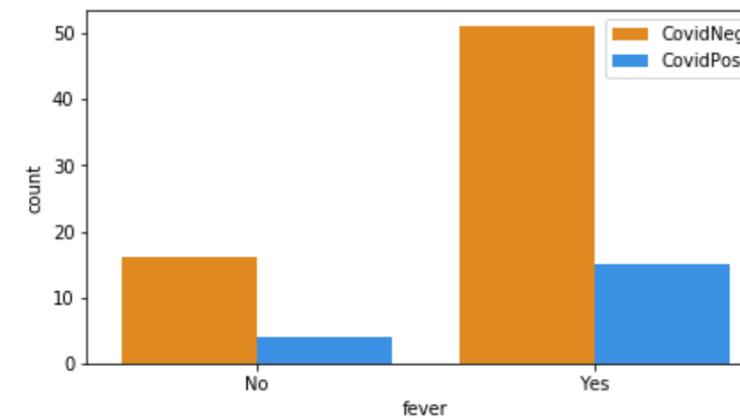
RESPIRATORY



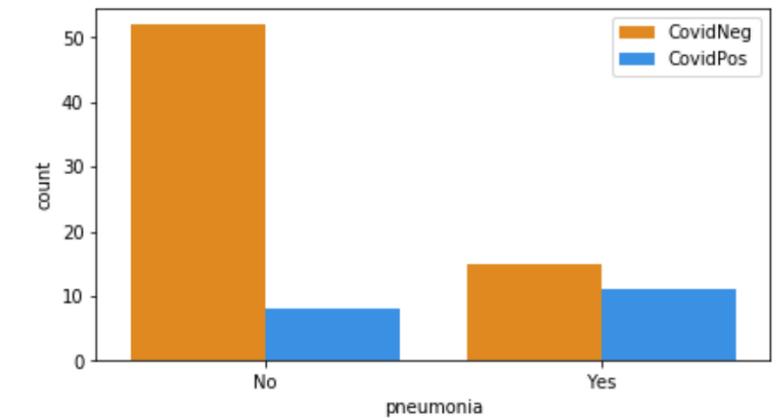
GASTROINTESTINAL



FEVER



PNEUMONIA



3

HOW DOES THE DATA LOOK LIKE?

86 cases

19 CovidPos

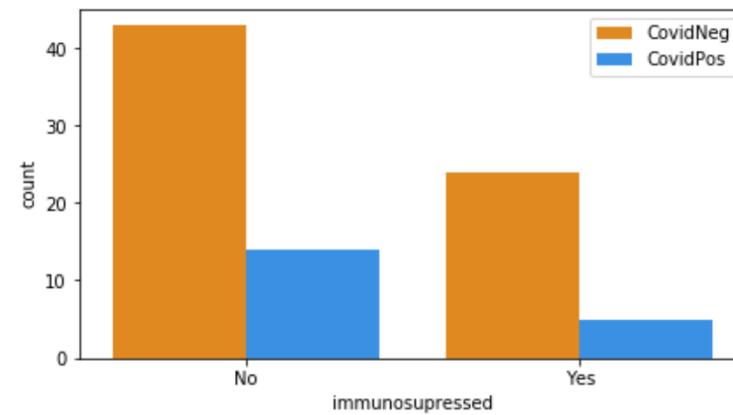
67 CovidNeg

18 features

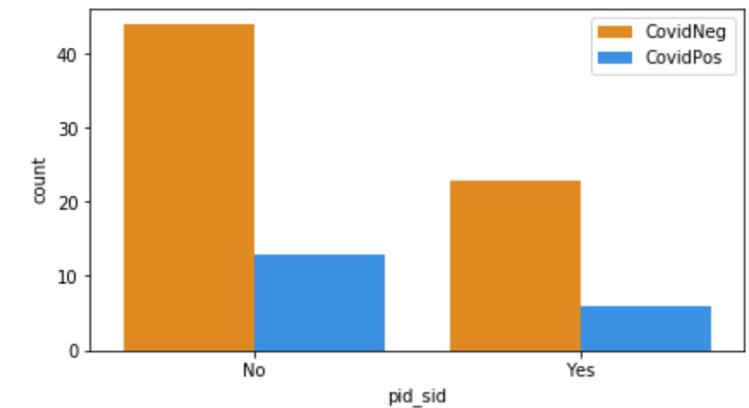
Demographics
Symptoms
Previous conditions
Treatments
Tests
Evolution

Age
Gender
Respiratory symptoms
Gastrointestinal symp.
Pneumonia
Fever
Immunosuppression
Immunodeficiency
Other previous cond.
Oxygen therapy
Antibiotics
Corticoids
Chest x-ray
Neutrocytes
Lymphocytes
C-reactive protein
Diagnosis delay
Adm. Intensive Care Unit

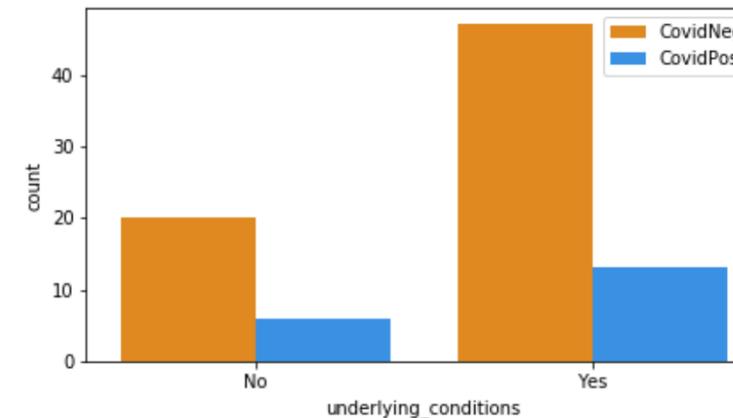
IMMUNOSUPPRESSION



IMMUNODEFICIENCY



OTHER PREVIOUS CONDITIONS



3

HOW DOES THE DATA LOOK LIKE?

86 cases

19 CovidPos

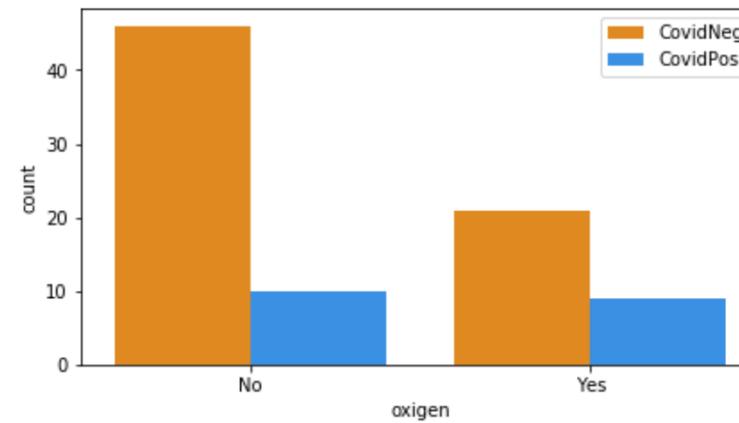
67 CovidNeg

18 features

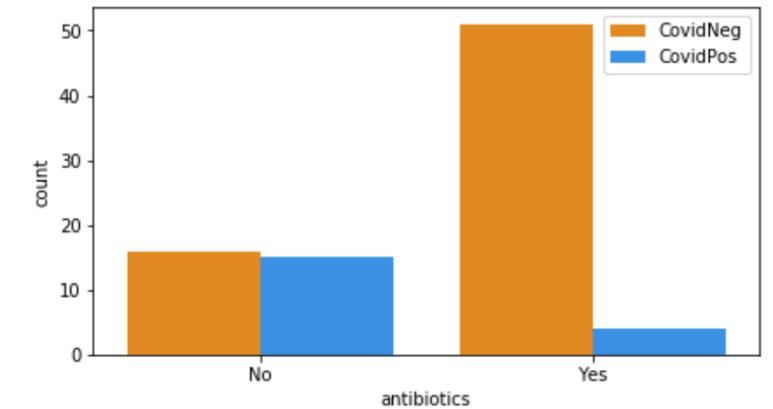
Demographics
Symptoms
Previous conditions
Treatments
Tests
Evolution

Age
Gender
Respiratory symptoms
Gastrointestinal symp.
Pneumonia
Fever
Immunosuppression
Immunodeficiency
Other previous cond.
Oxygen therapy
Antibiotics
Corticoids
Chest x-ray
Neutrocytes
Lymphocytes
C-reactive protein
Diagnosis delay
Adm. Intensive Care Unit

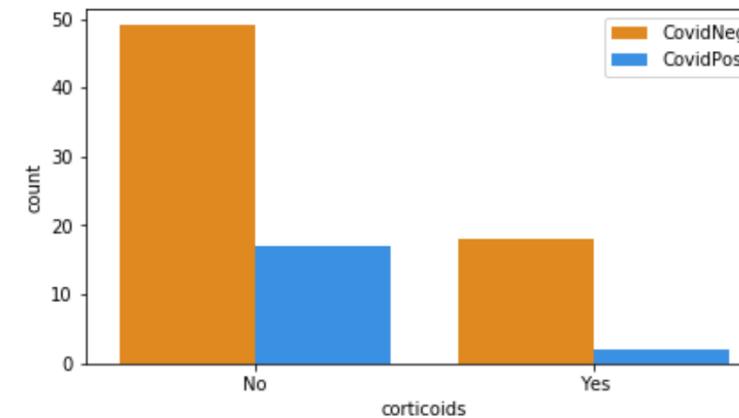
OXYGEN THERAPY



ANTIBIOTICS



CORTICOIDS



3

HOW DOES THE DATA LOOK LIKE?

86 cases

19 CovidPos

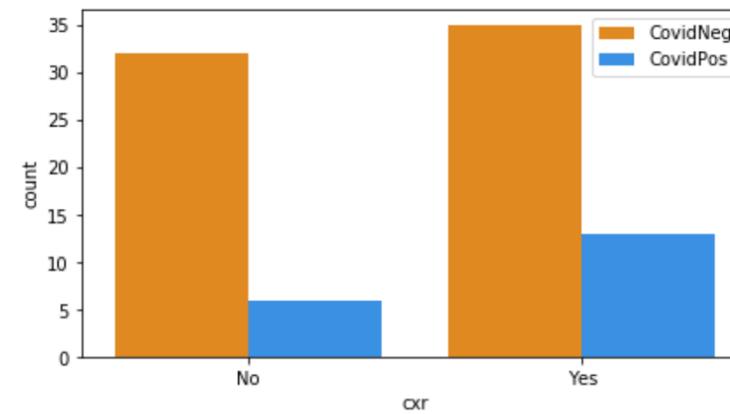
67 CovidNeg

18 features

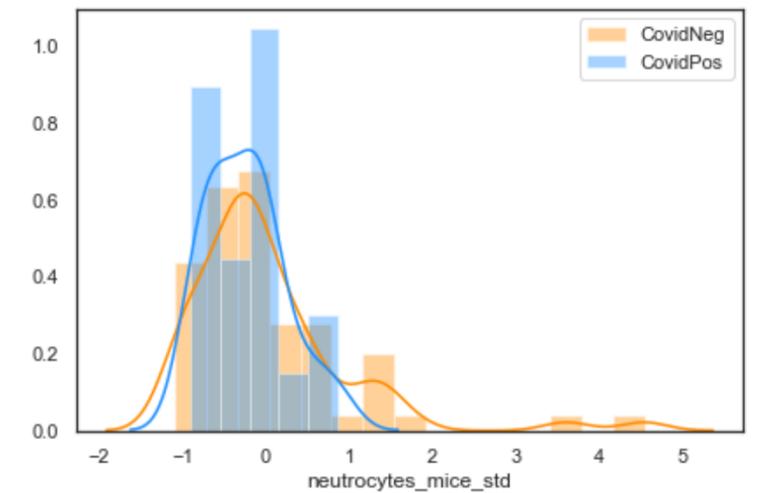
- Demographics
- Symptoms
- Previous conditions
- Treatments
- Tests
- Evolution

- Age
- Gender
- Respiratory symptoms
- Gastrointestinal symp.
- Pneumonia
- Fever
- Immunosuppression
- Immunodeficiency
- Other previous cond.
- Oxygen therapy
- Antibiotics
- Corticoids
- Chest x-ray**
- Neutrocytes**
- Lymphocytes**
- C-reactive protein**
- Diagnosis delay
- Adm. Intensive Care Unit

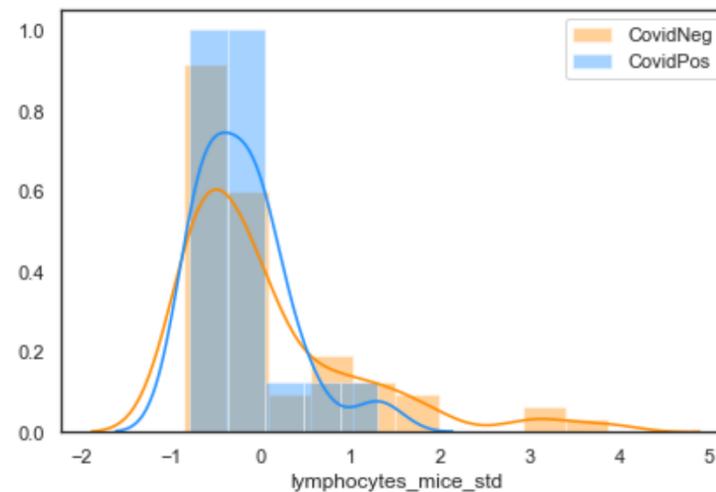
CHEST X-RAY



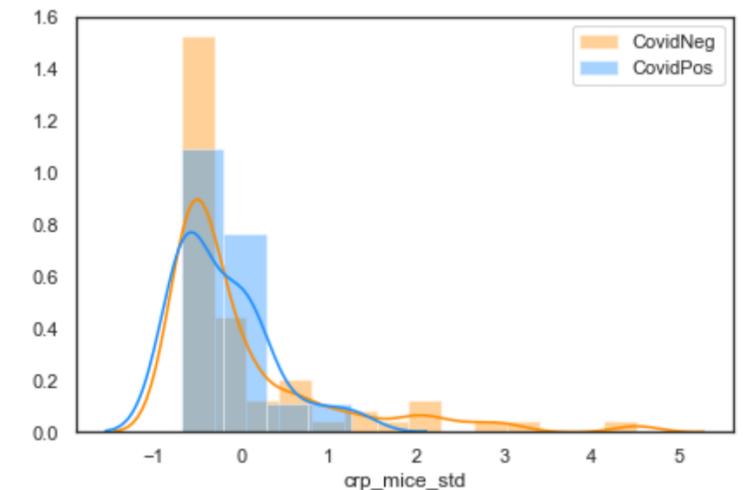
NEUTROCYTES



LYMPHOCYTES



C-REACTIVE PROTEIN



3

HOW DOES THE DATA LOOK LIKE?

86 cases

19 CovidPos

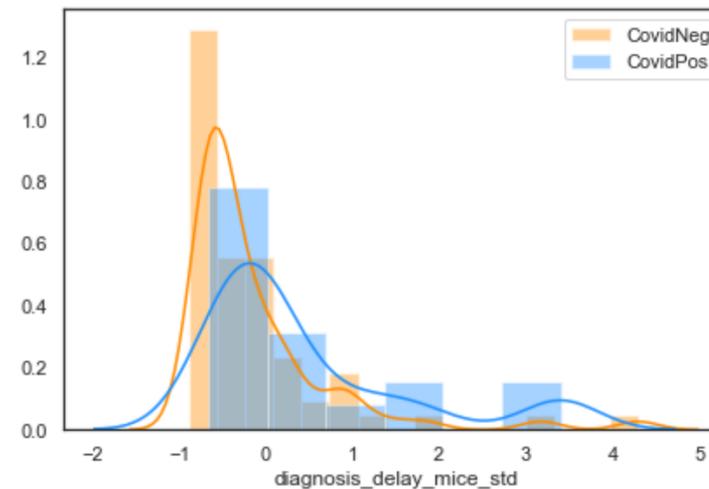
67 CovidNeg

18 features

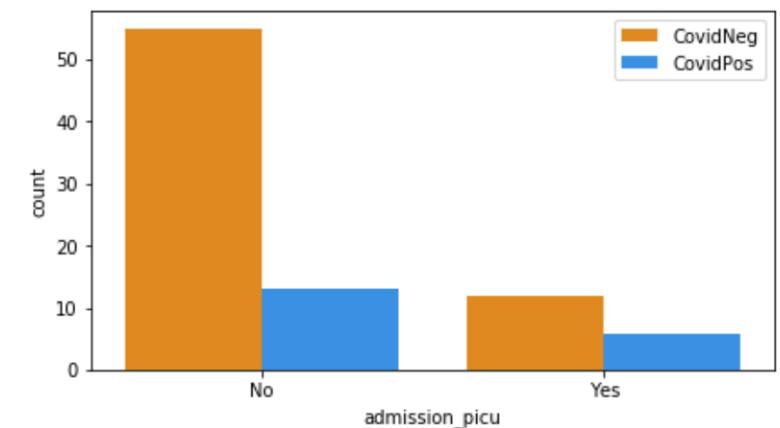
Demographics
Symptoms
Previous conditions
Treatments
Tests
Evolution

Age
Gender
Respiratory symptoms
Gastrointestinal symp.
Pneumonia
Fever
Immunosuppression
Immunodeficiency
Other previous cond.
Oxygen therapy
Antibiotics
Corticoids
Chest x-ray
Neutrocytes
Lymphocytes
C-reactive protein
Diagnosis delay
Adm. Intensive Care Unit

DIAGNOSIS DELAY



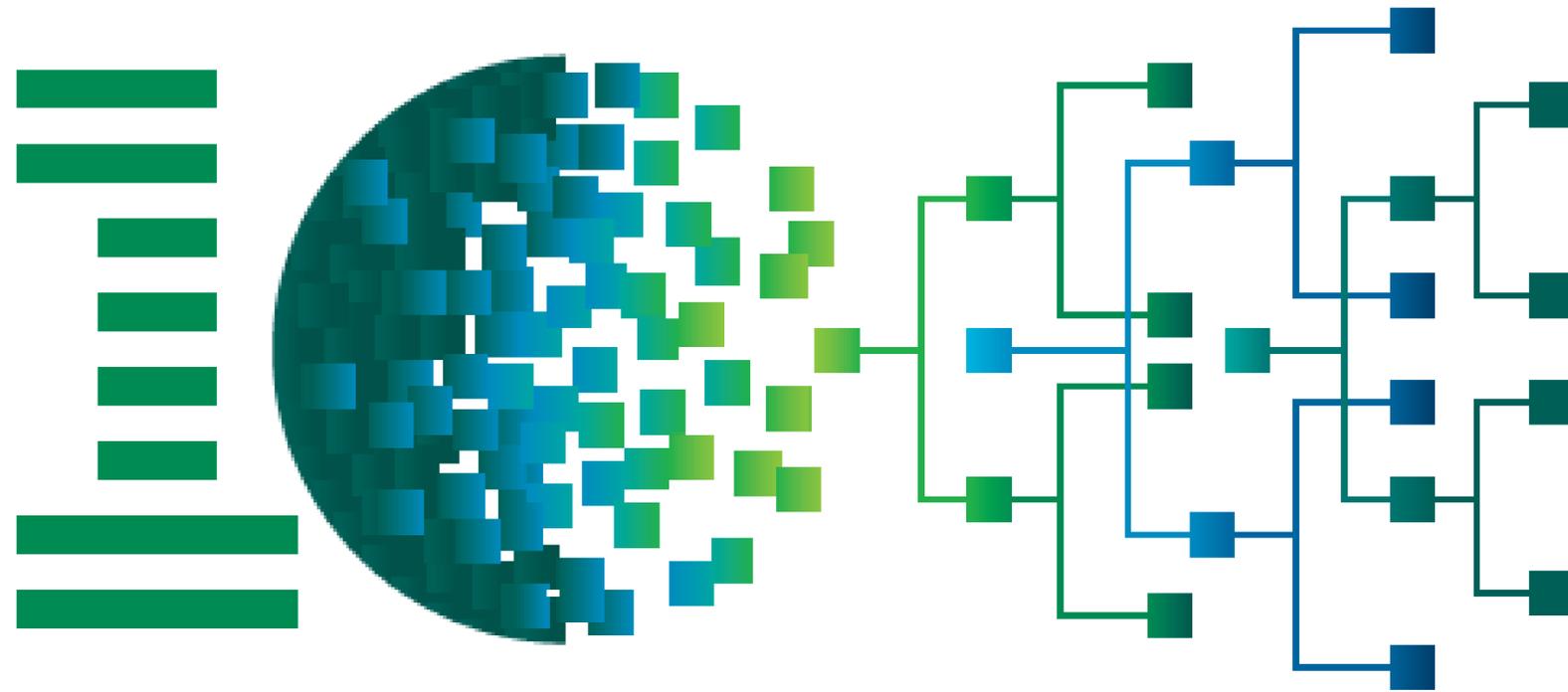
ADMISSION TO INTENSIVE CARE UNIT



4

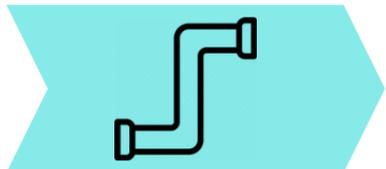


WHAT CAN I DO WITH THE DATA?



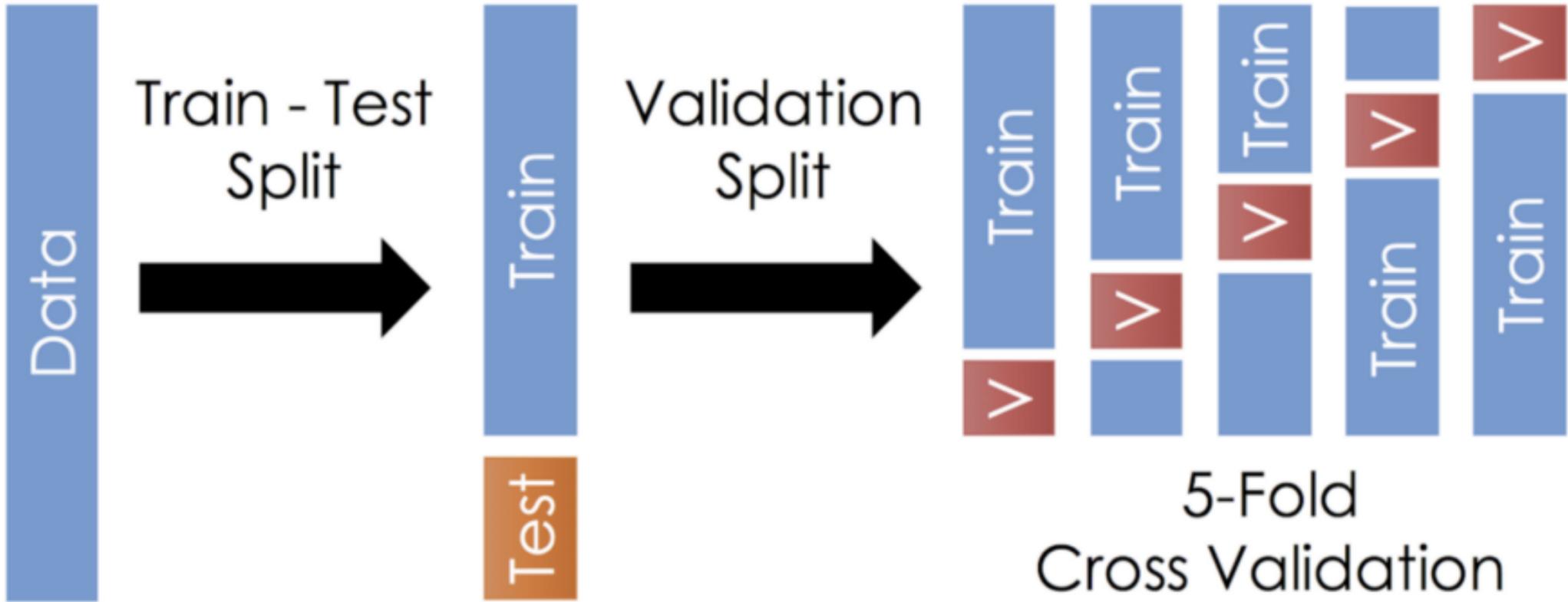
4

WHAT CAN I DO WITH THE DATA?



1) TRAIN DIFFERENT MACHINE LEARNING MODELS

Pipeline with different classification models (Stratified KFold cross-validation, RandomizedSearch strategy)



4

WHAT CAN I DO WITH THE DATA?



1) TRAIN DIFFERENT MACHINE LEARNING MODELS

Pipeline with different classification models (Stratified KFold cross-validation, RandomizedSearch strategy)

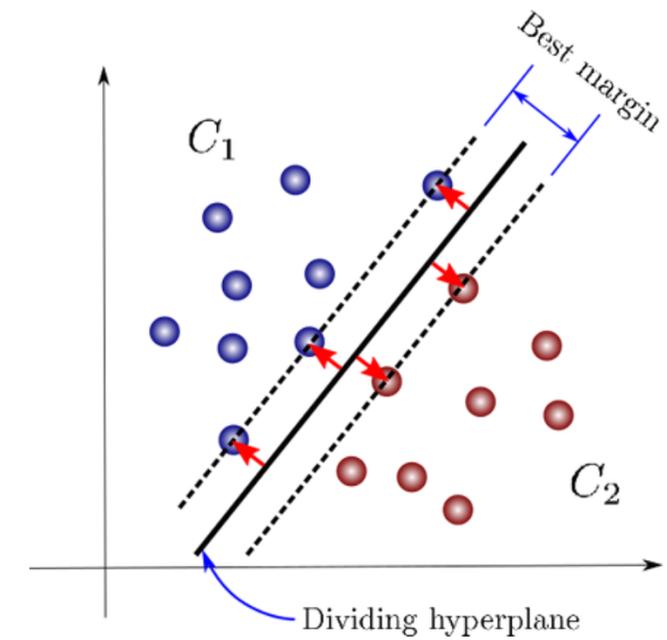
4

WHAT CAN I DO WITH THE DATA?



1) TRAIN DIFFERENT MACHINE LEARNING MODELS

Pipeline with different classification models (Stratified KFold cross-validation, RandomizedSearch strategy)



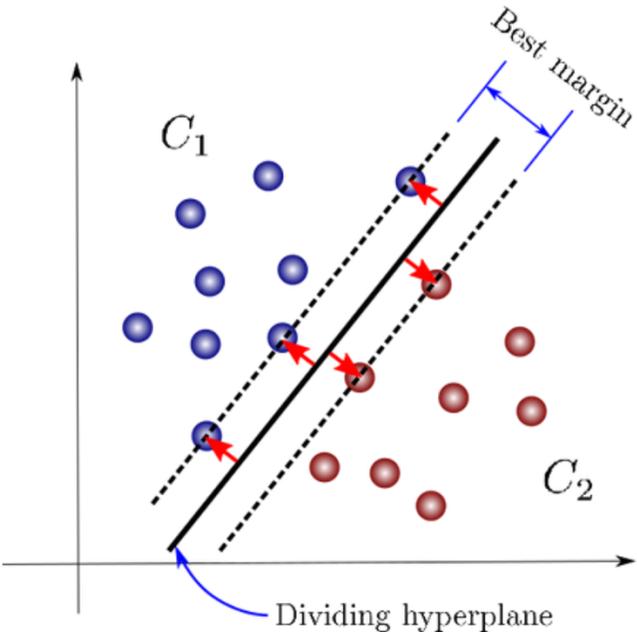
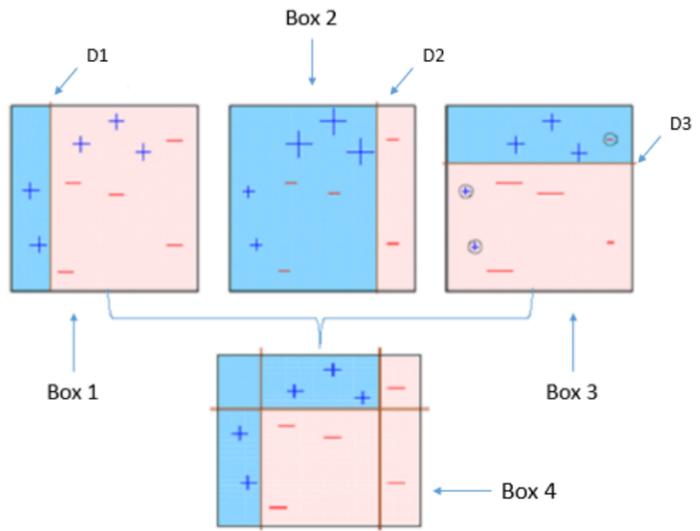
4

WHAT CAN I DO WITH THE DATA?



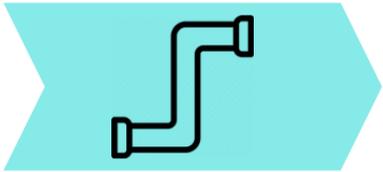
1) TRAIN DIFFERENT MACHINE LEARNING MODELS

Pipeline with different classification models (Stratified KFold cross-validation, RandomizedSearch strategy)



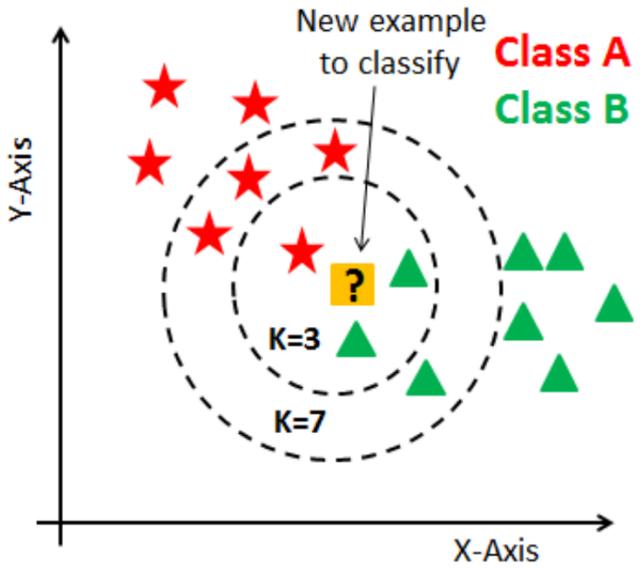
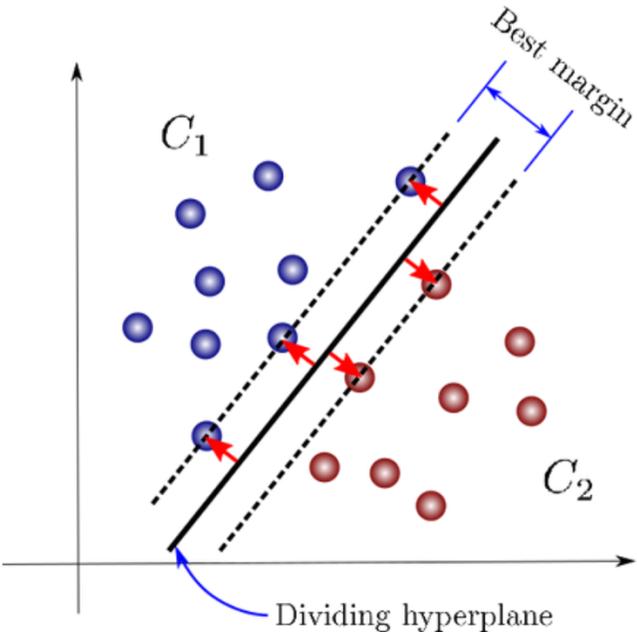
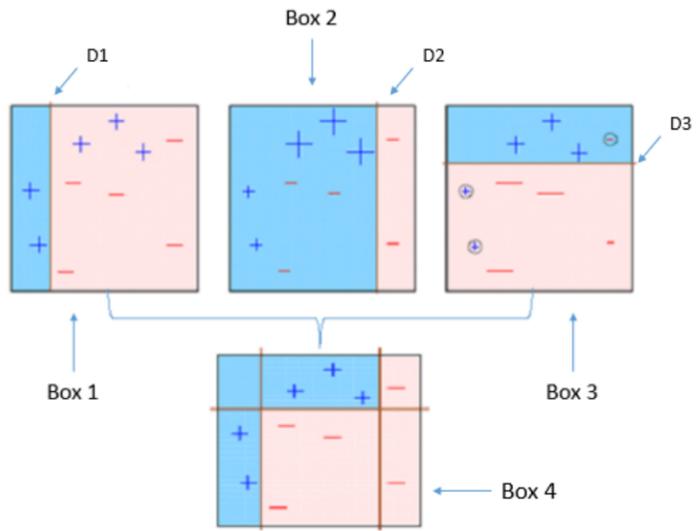
4

WHAT CAN I DO WITH THE DATA?



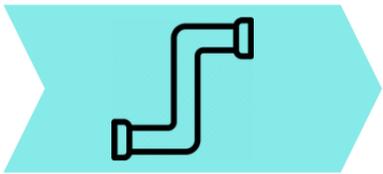
1) TRAIN DIFFERENT MACHINE LEARNING MODELS

Pipeline with different classification models (Stratified KFold cross-validation, RandomizedSearch strategy)



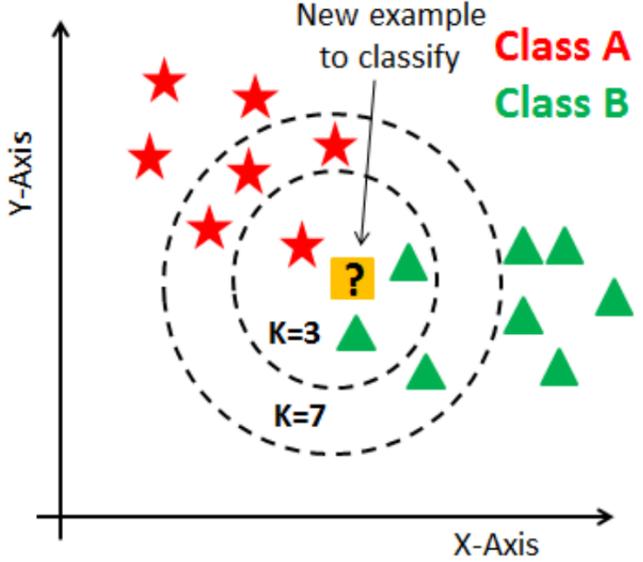
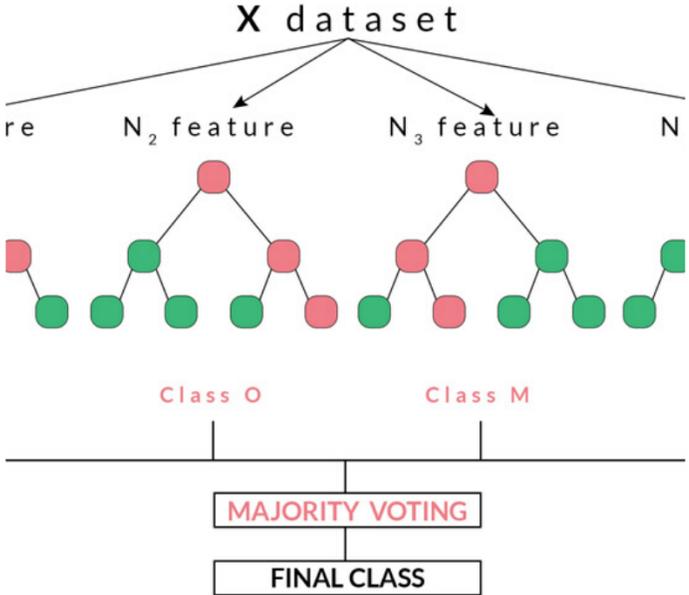
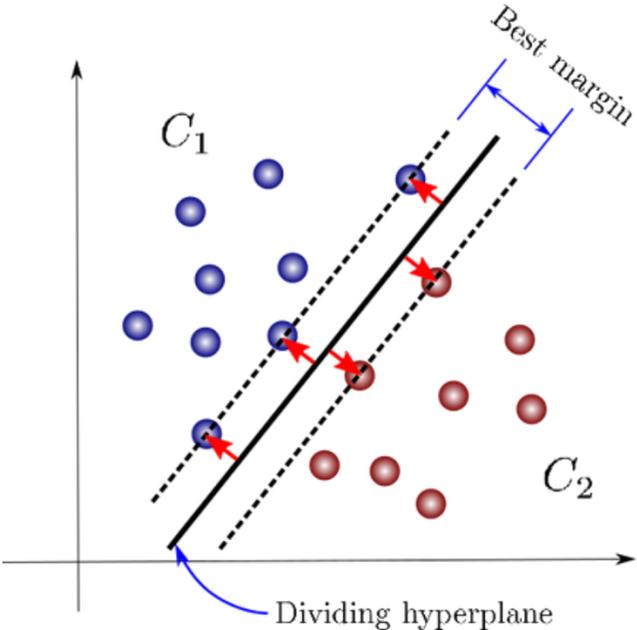
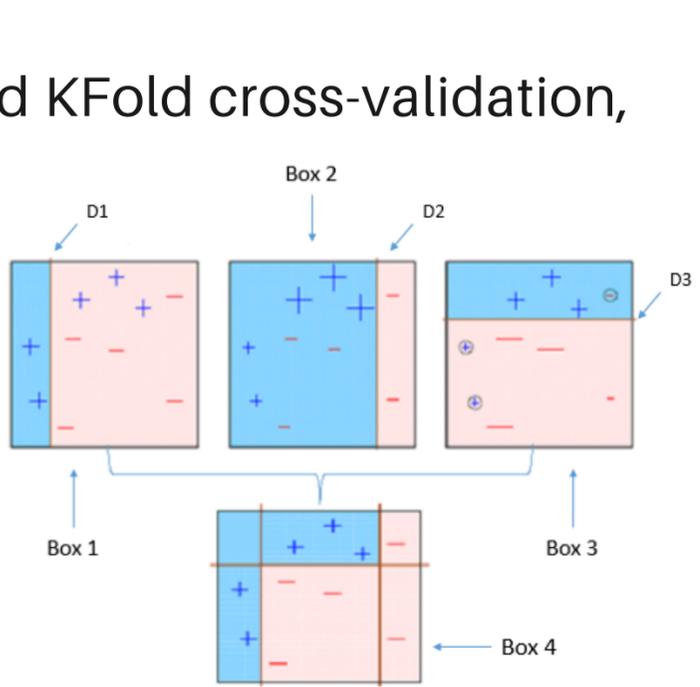
4

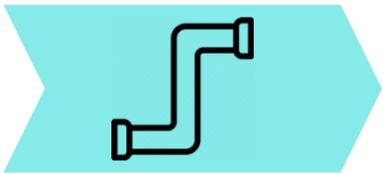
WHAT CAN I DO WITH THE DATA?



1) TRAIN DIFFERENT MACHINE LEARNING MODELS

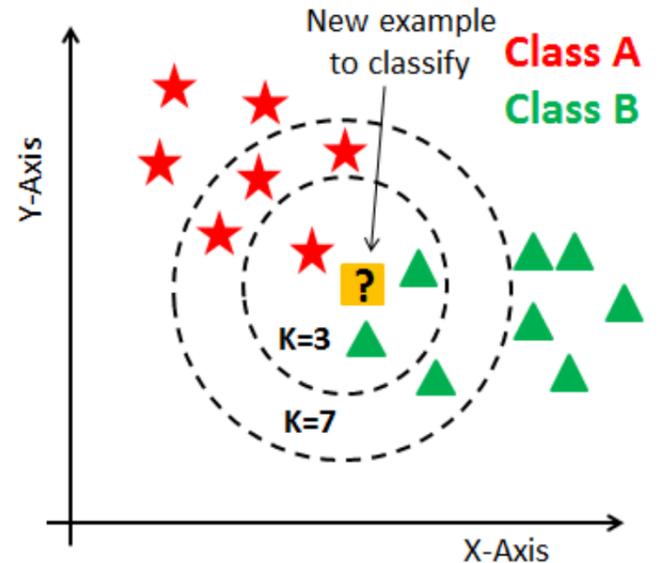
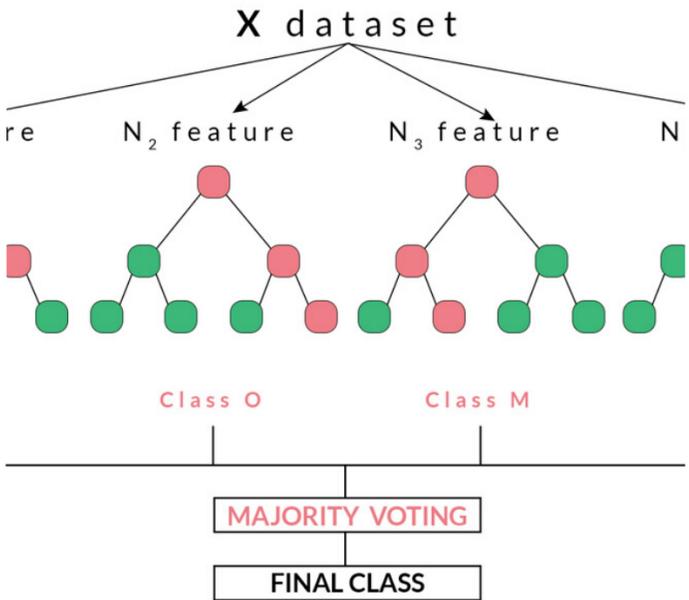
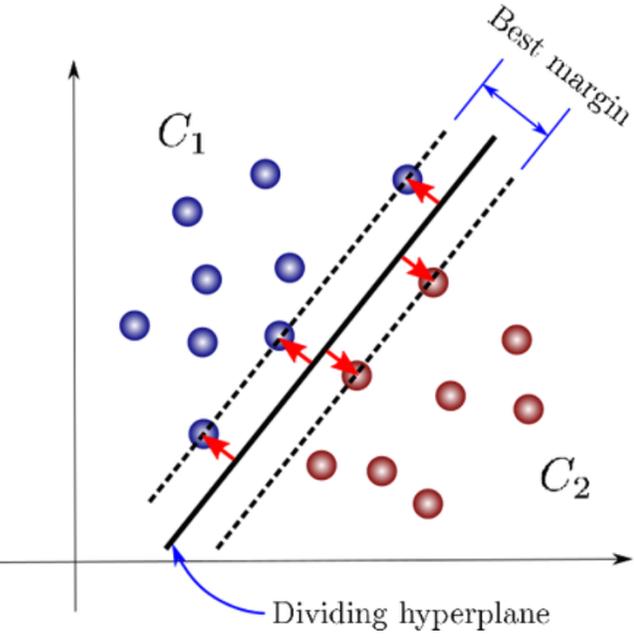
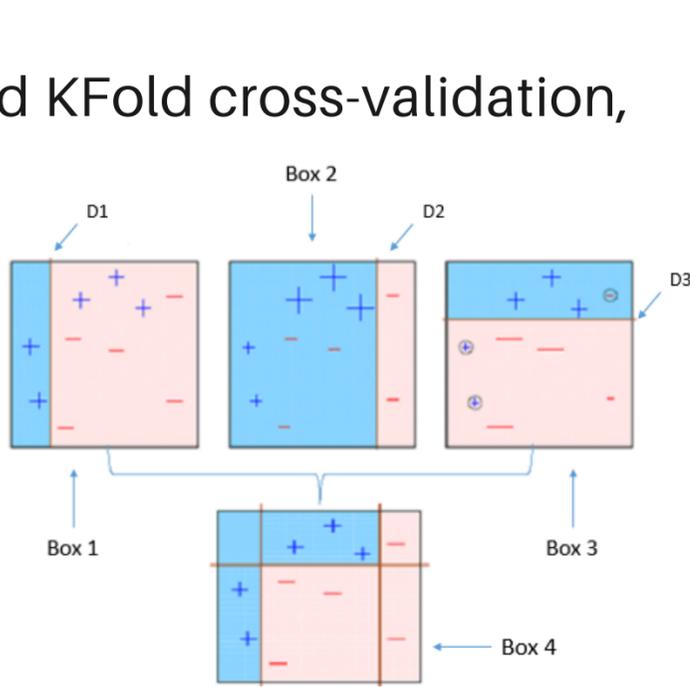
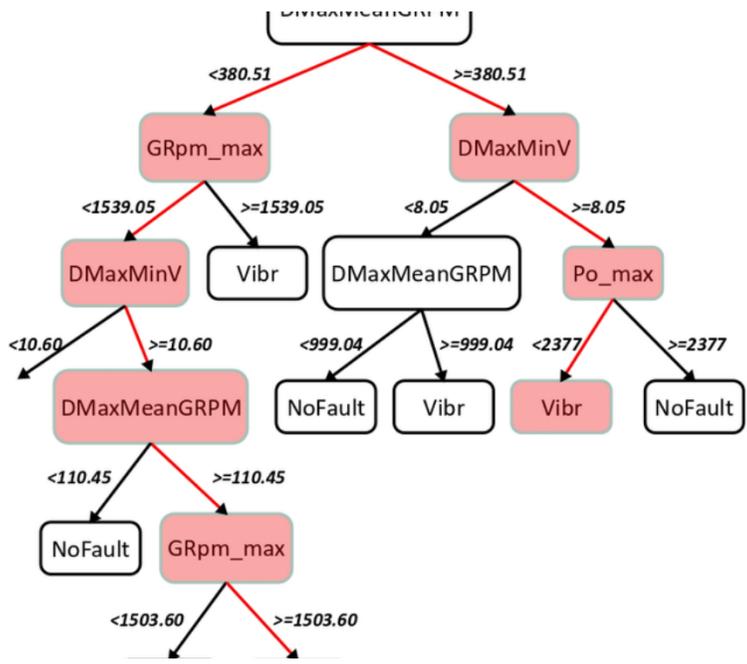
Pipeline with different classification models (Stratified KFold cross-validation, RandomizedSearch strategy)

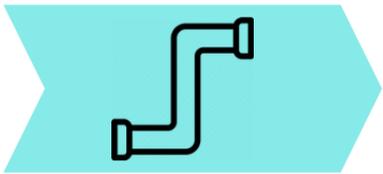




1) TRAIN DIFFERENT MACHINE LEARNING MODELS

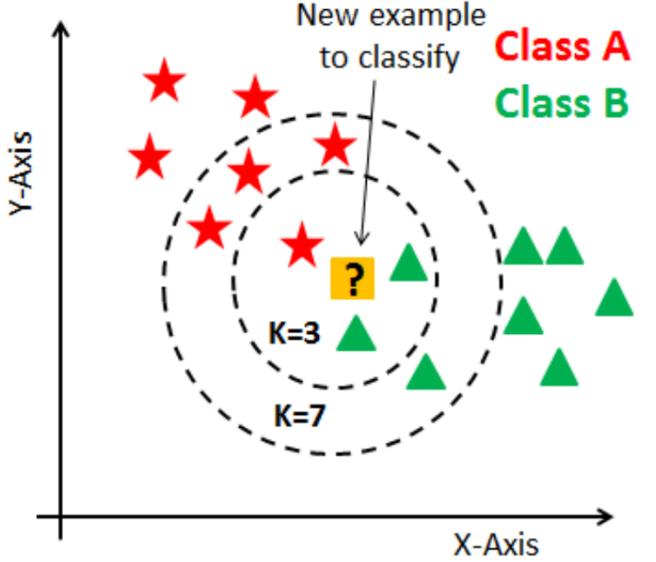
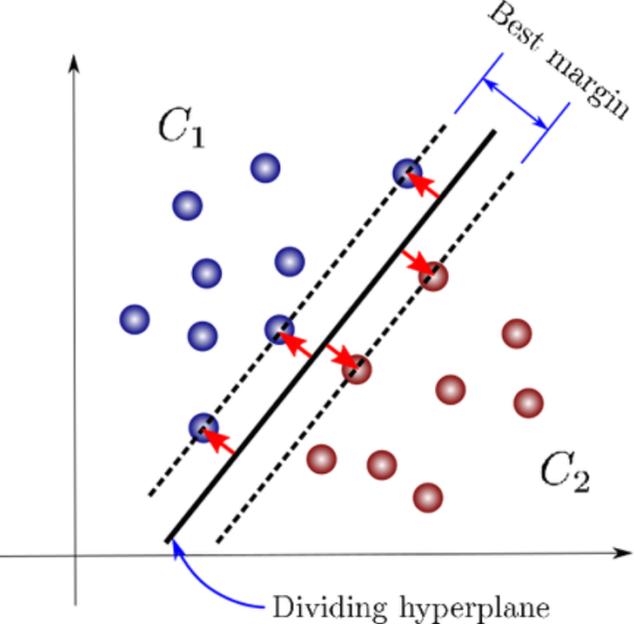
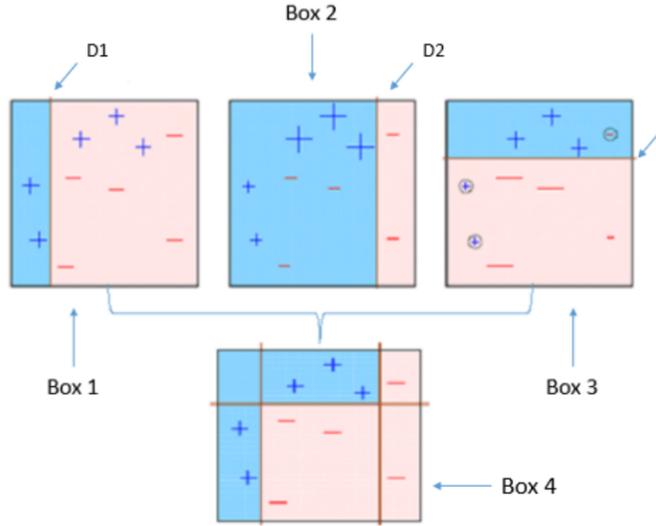
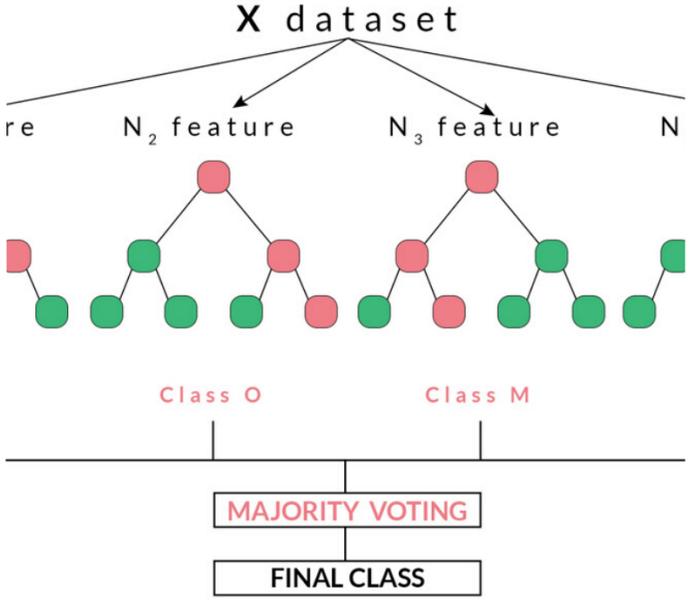
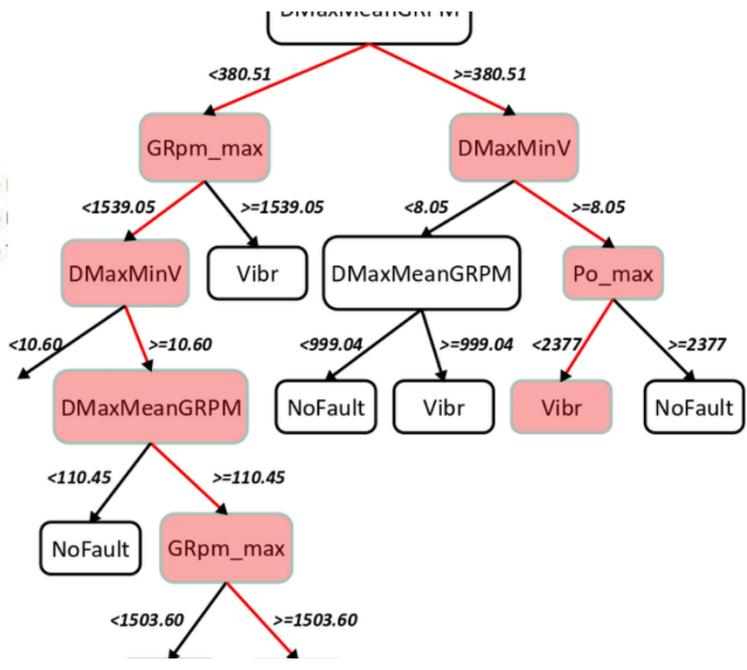
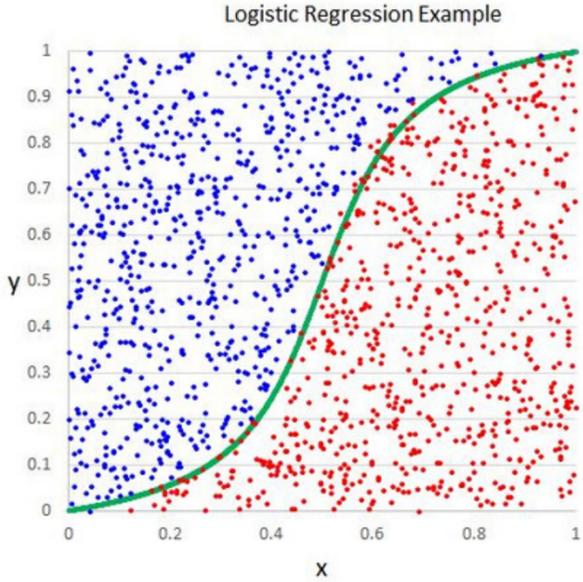
Pipeline with different classification models (Stratified KFold cross-validation, RandomizedSearch strategy)





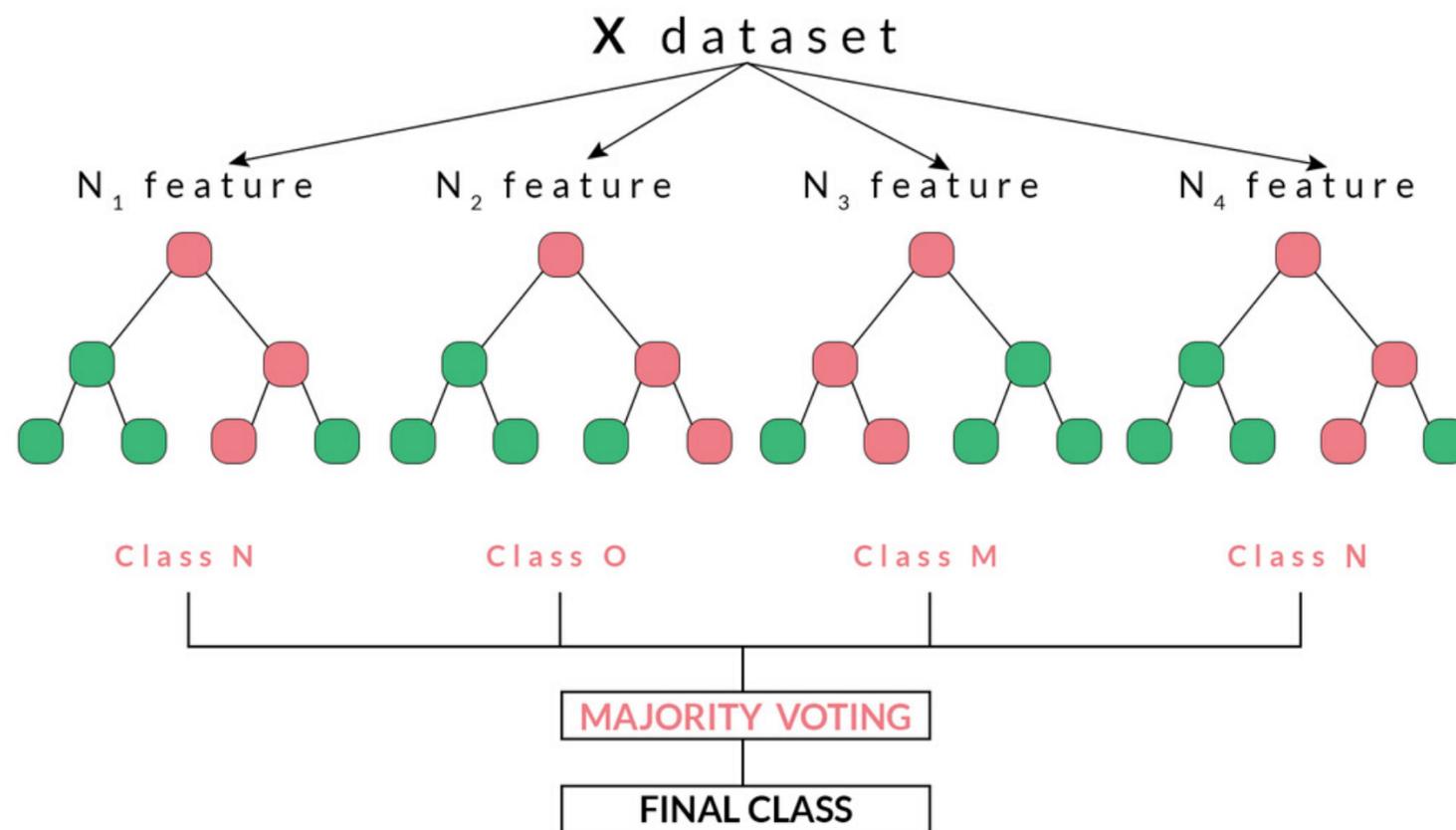
1) TRAIN DIFFERENT MACHINE LEARNING MODELS

Pipeline with different classification models (Stratified KFold cross-validation, RandomizedSearch strategy)



2) TRAIN BEST PERFORMING MODELS

Reduced pipeline with best performing models: Random Forests (Stratified KFold cross-validation, GridSearch strategy)



Best parameters output

Original model

```
{'clf__criterion': 'gini',
'clf__max_depth': 2,
'clf__min_samples_leaf': 2,
'clf__min_samples_split': 2}
```

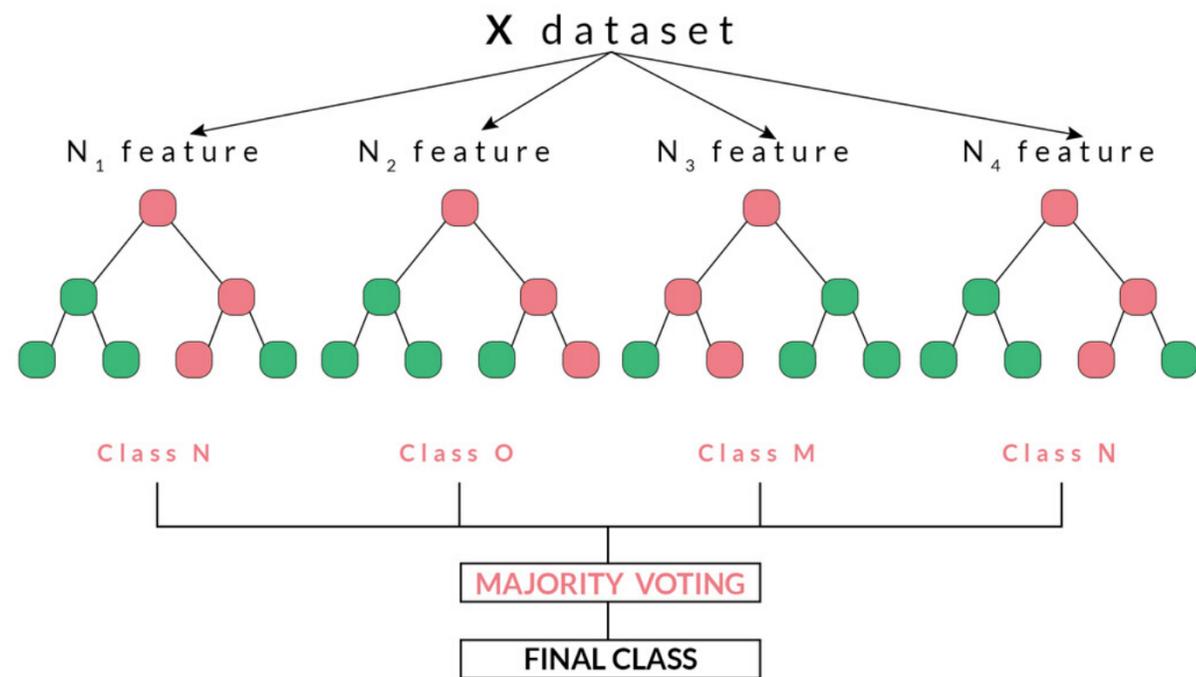
Oversampled model

```
{'clf__criterion': 'entropy',
'clf__max_depth': 3,
'clf__min_samples_leaf': 3,
'clf__min_samples_split': 8}
```



3) GLOBAL SURROGATE METHOD

Modelling of Decision Trees on Random Forest model predictions



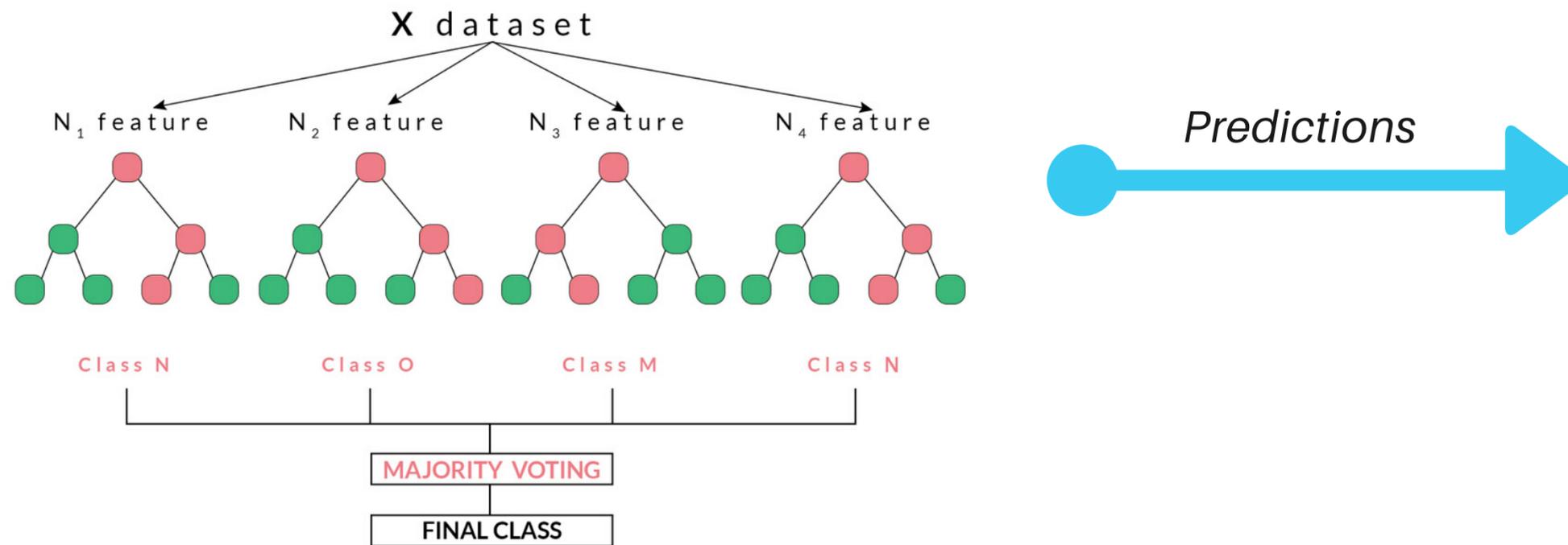
4

WHAT CAN I DO WITH THE DATA?



3) GLOBAL SURROGATE METHOD

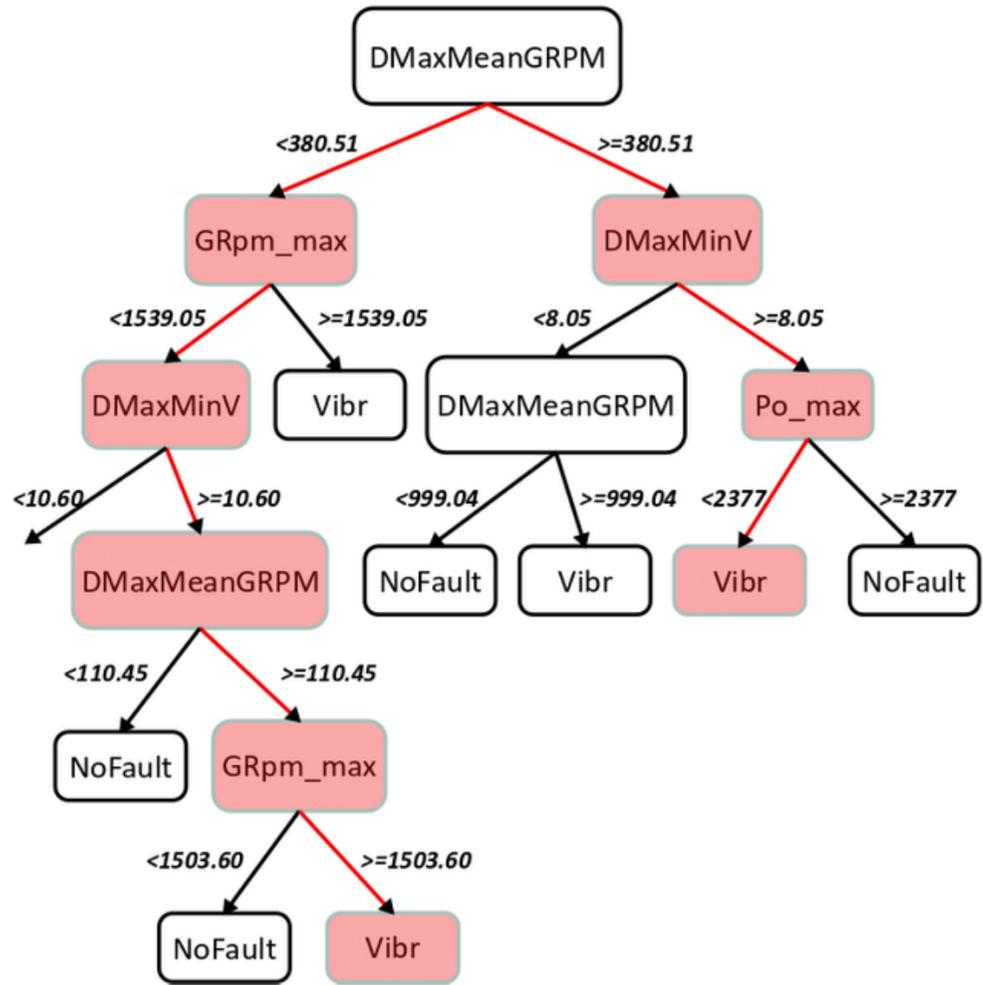
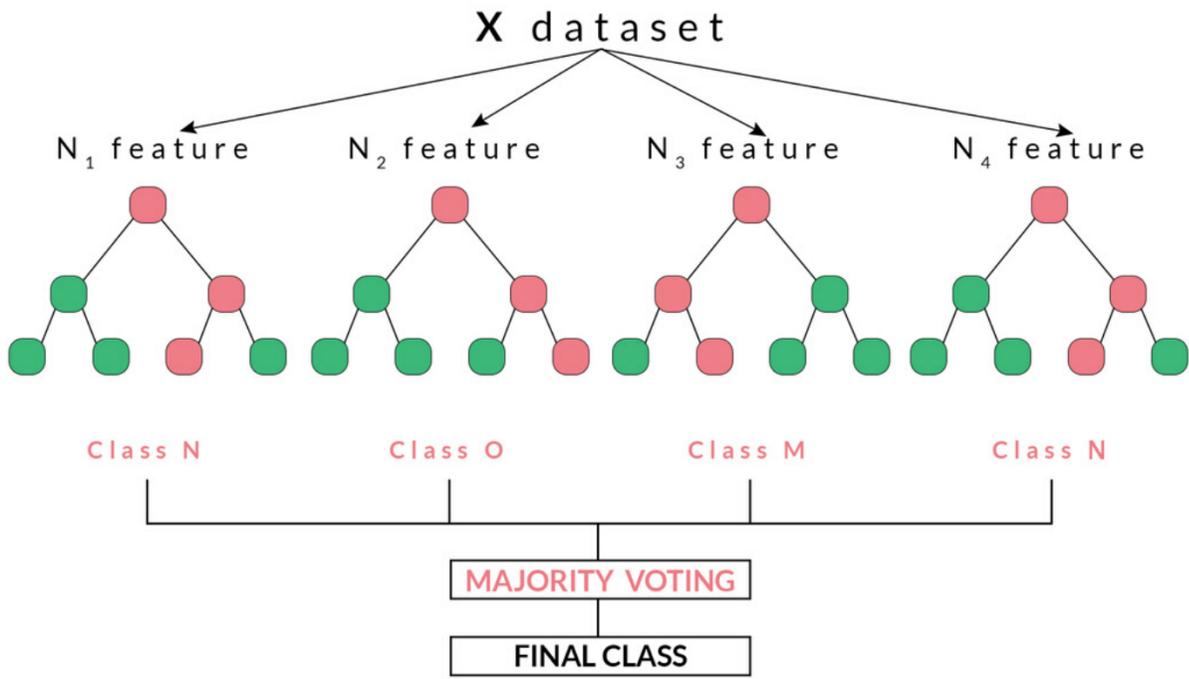
Modelling of Decision Trees on Random Forest model predictions





3) GLOBAL SURROGATE METHOD

Modelling of Decision Trees on Random Forest model predictions



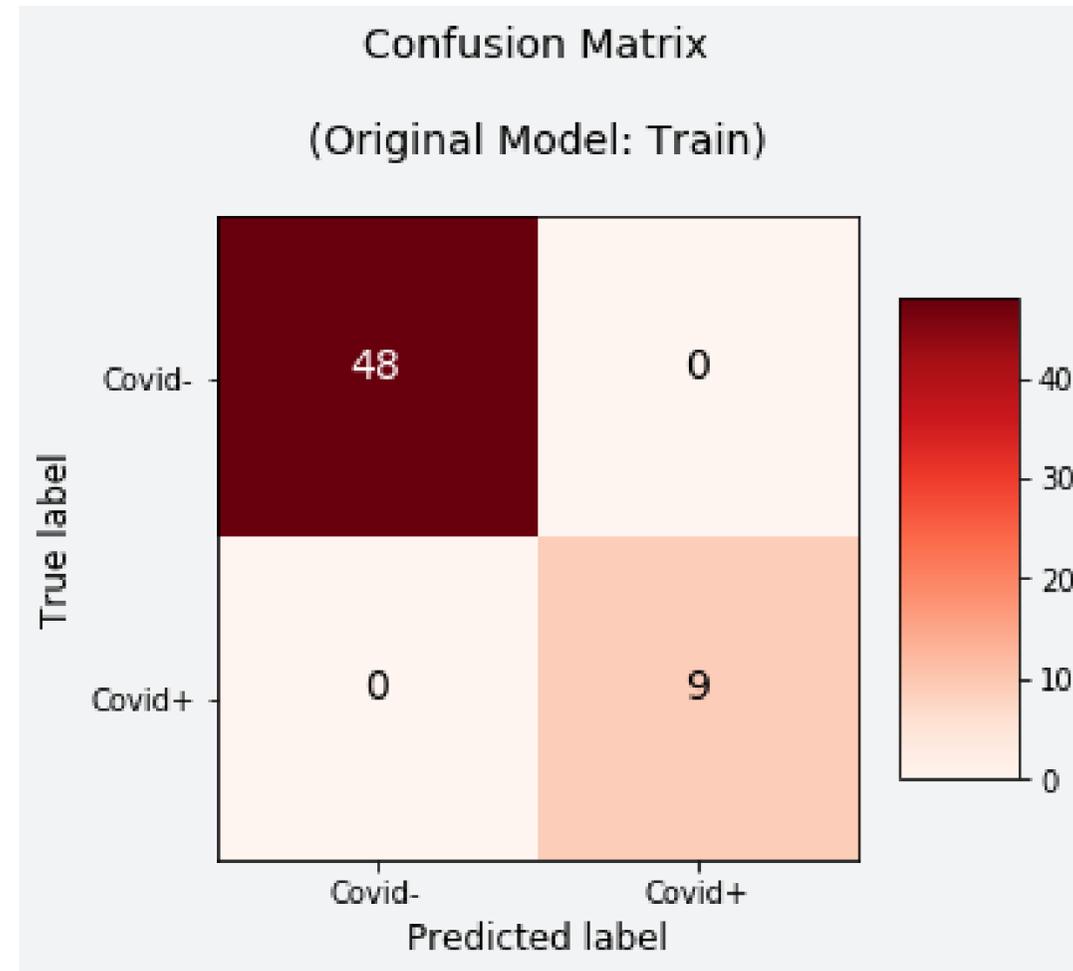
4

WHAT CAN I DO WITH THE DATA?



4) MODEL PERFORMANCE EVALUATION

Accuracy, Precision, Recall, F1 Score, AUC Score, Confusion Matrix



4

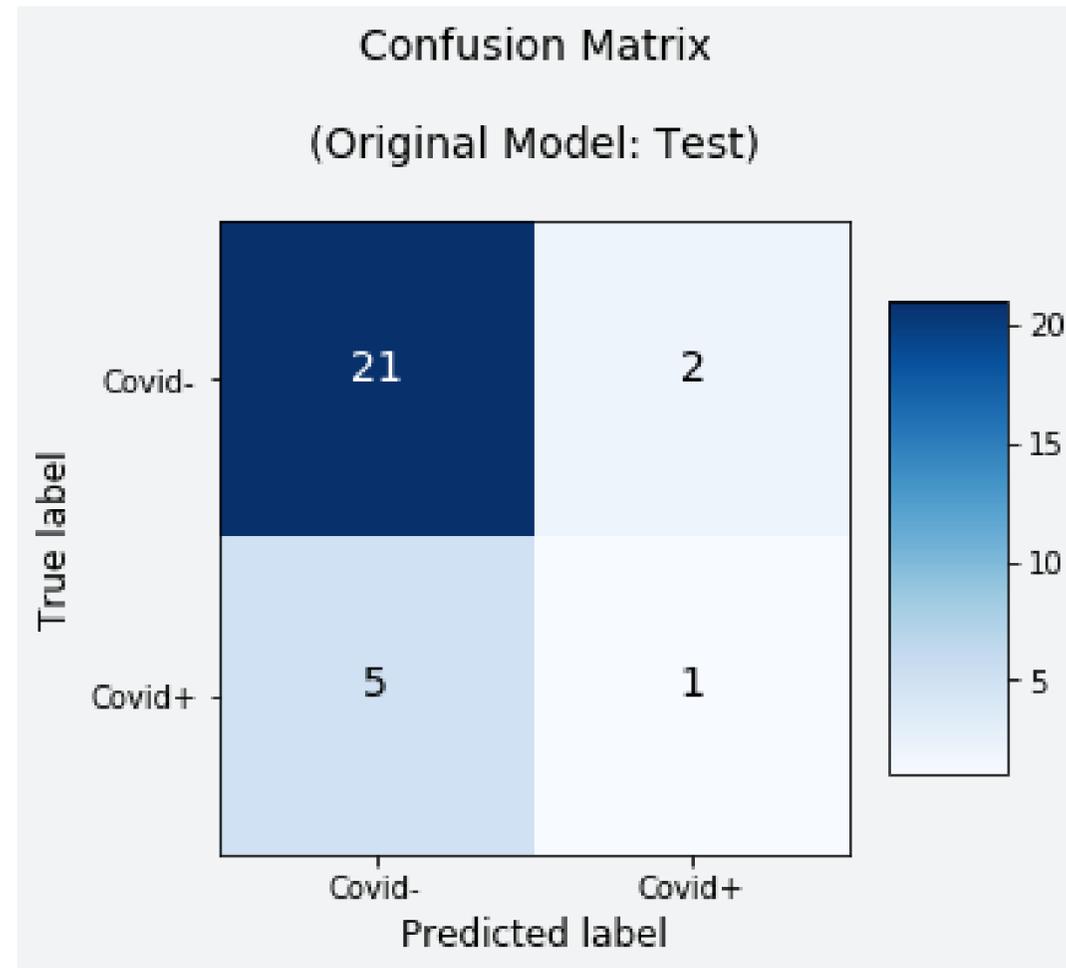


WHAT CAN I DO WITH THE DATA?



4) MODEL PERFORMANCE EVALUATION

Accuracy, Precision, Recall, F1 Score, AUC Score, Confusion Matrix



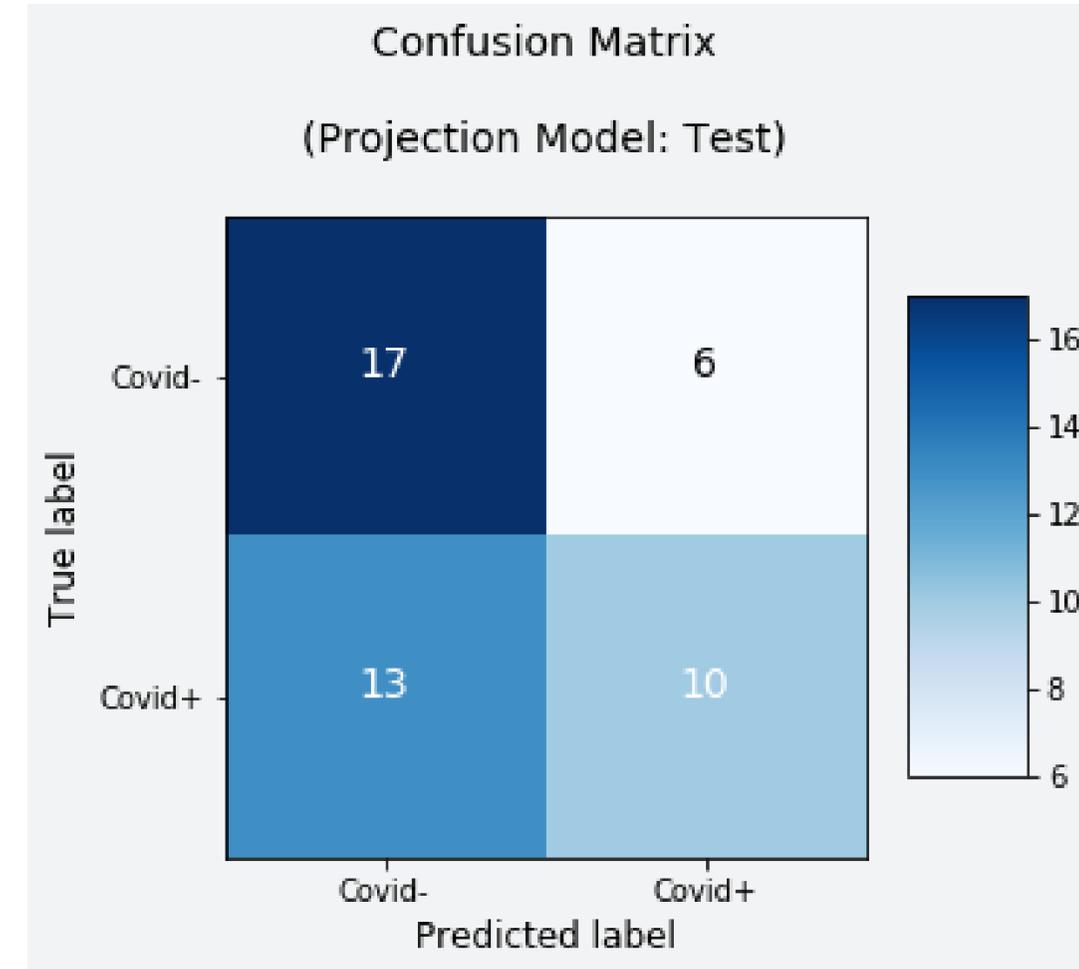
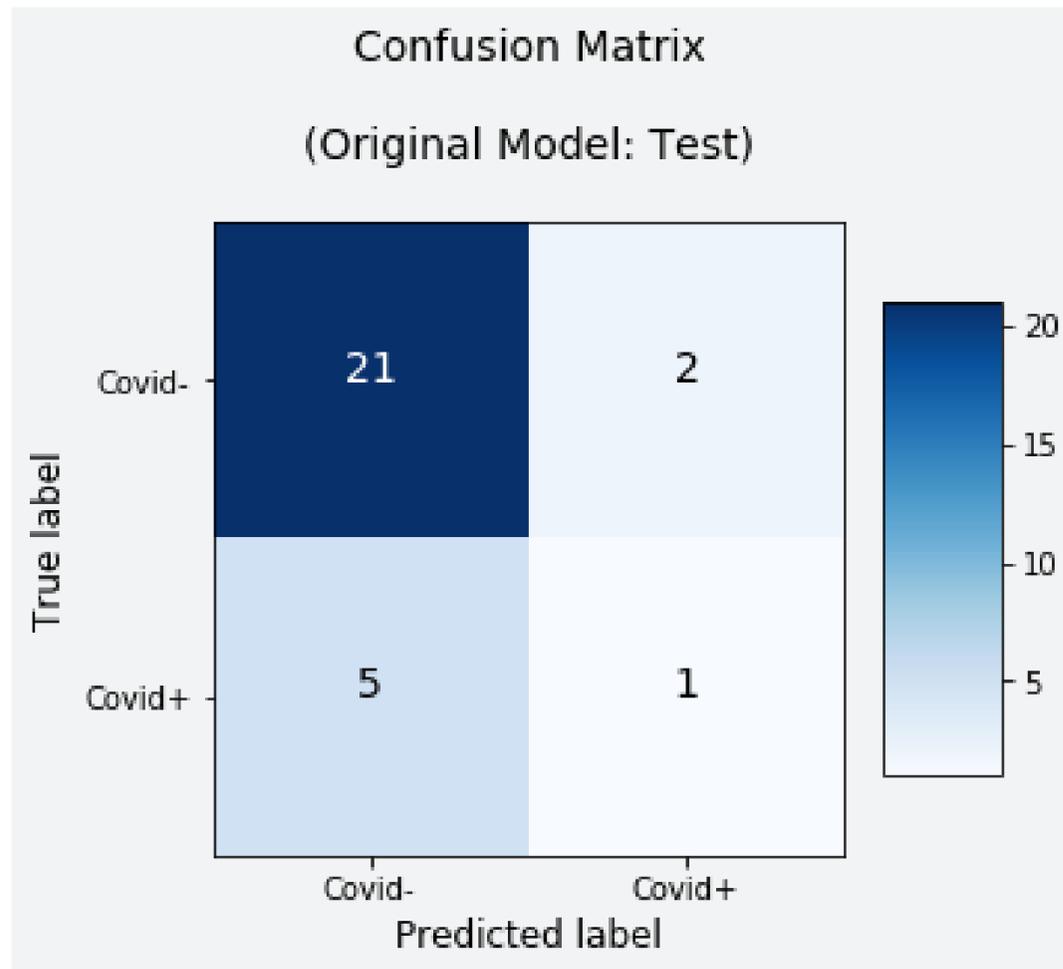
4

WHAT CAN I DO WITH THE DATA?



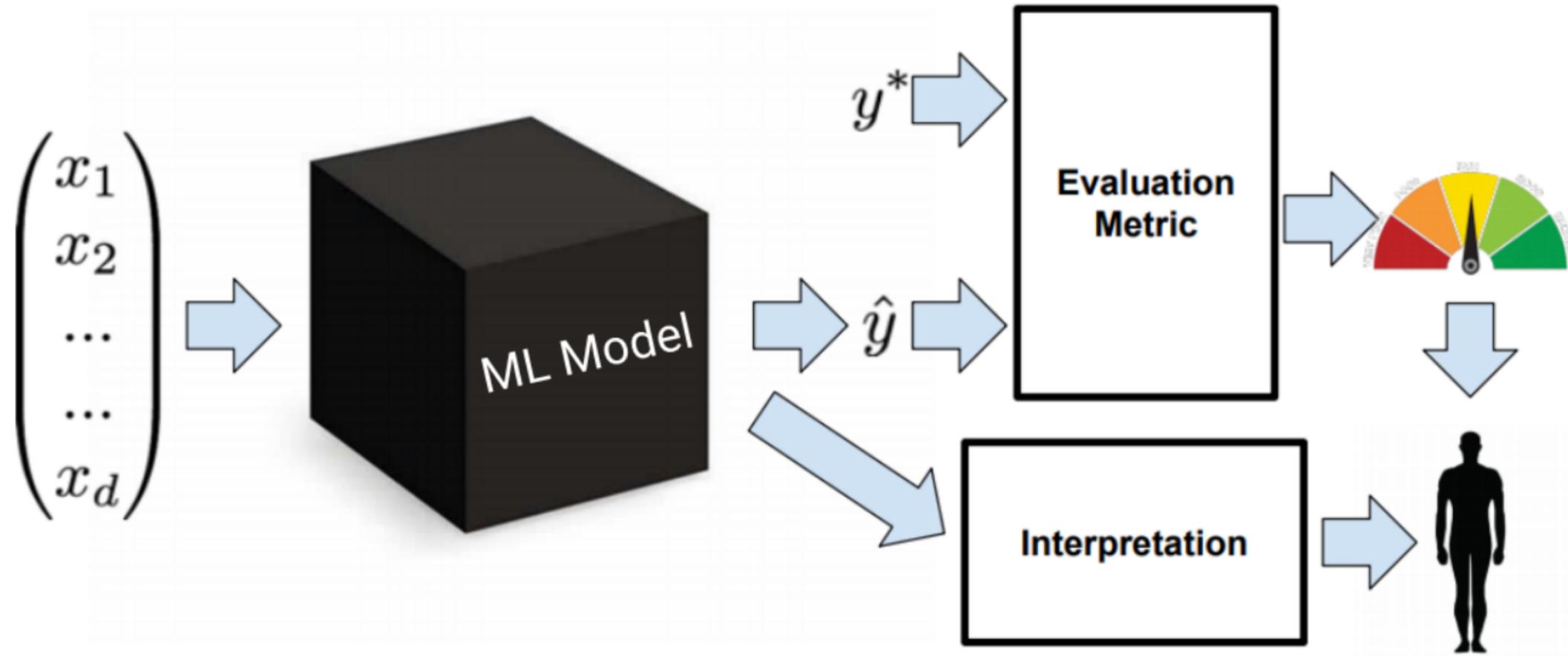
4) MODEL PERFORMANCE EVALUATION

Accuracy, Precision, Recall, F1 Score, AUC Score, Confusion Matrix



5

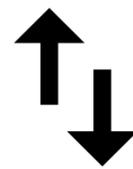
WHAT IS THE DATA TELLING ME?



5

WHAT IS THE DATA TELLING ME?

Decision Tree
visualisations

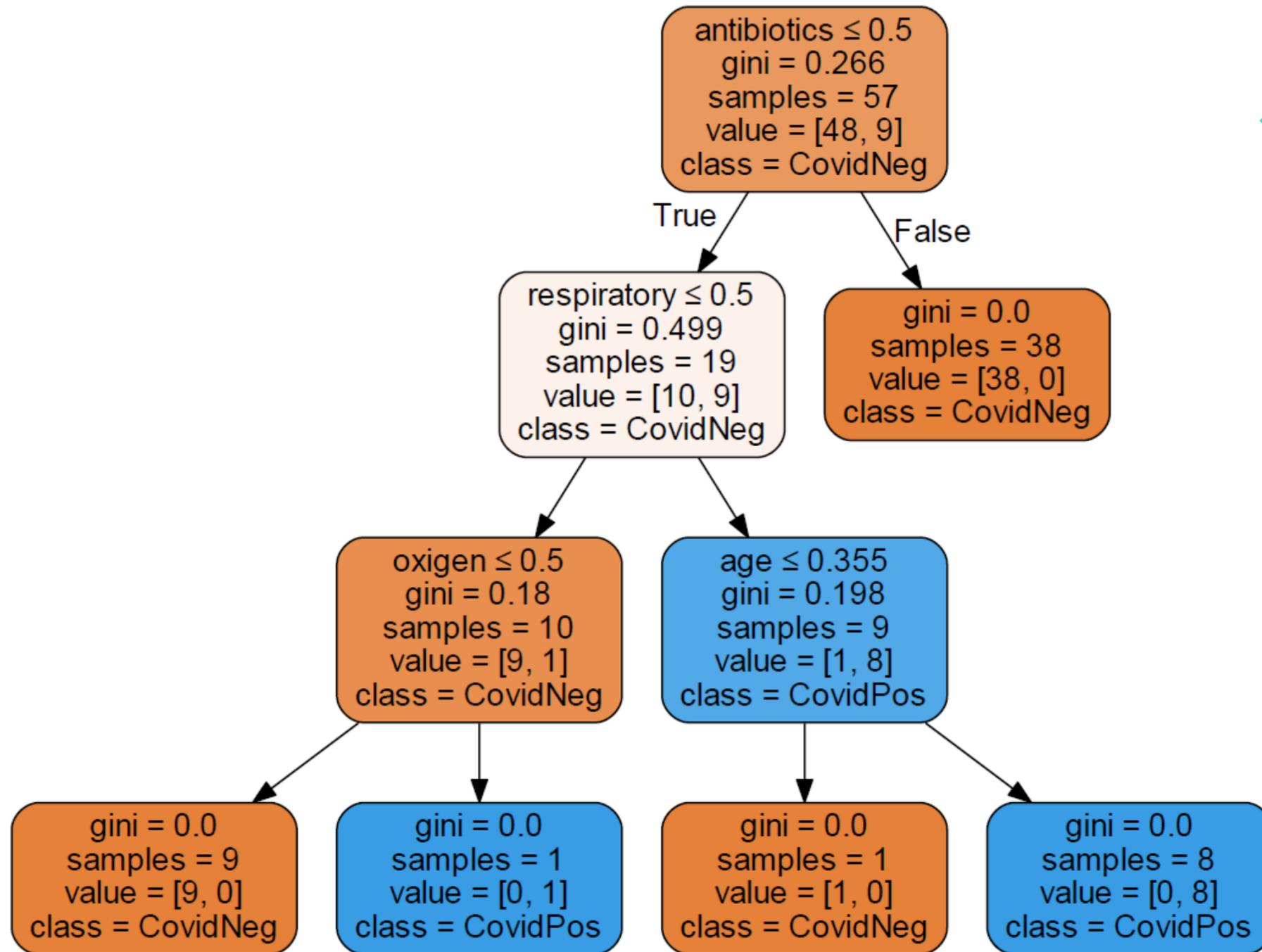


Feature importance
analyses &
visualisations

5

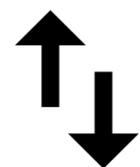
WHAT IS THE DATA TELLING ME?

Decision Tree
visualisations

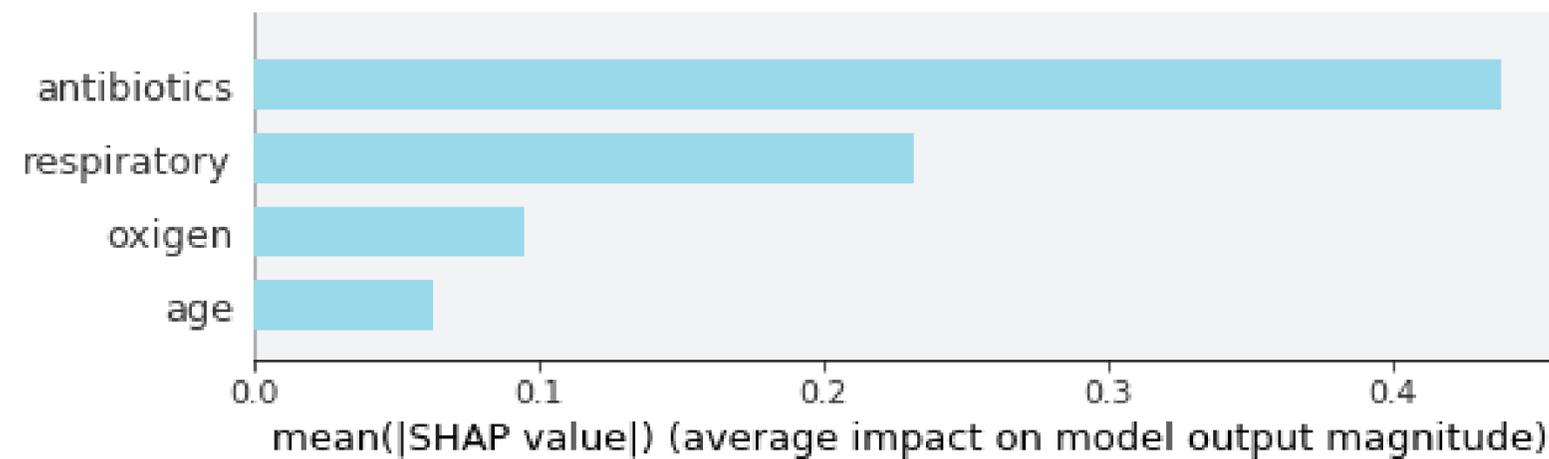


5

WHAT IS THE DATA TELLING ME?



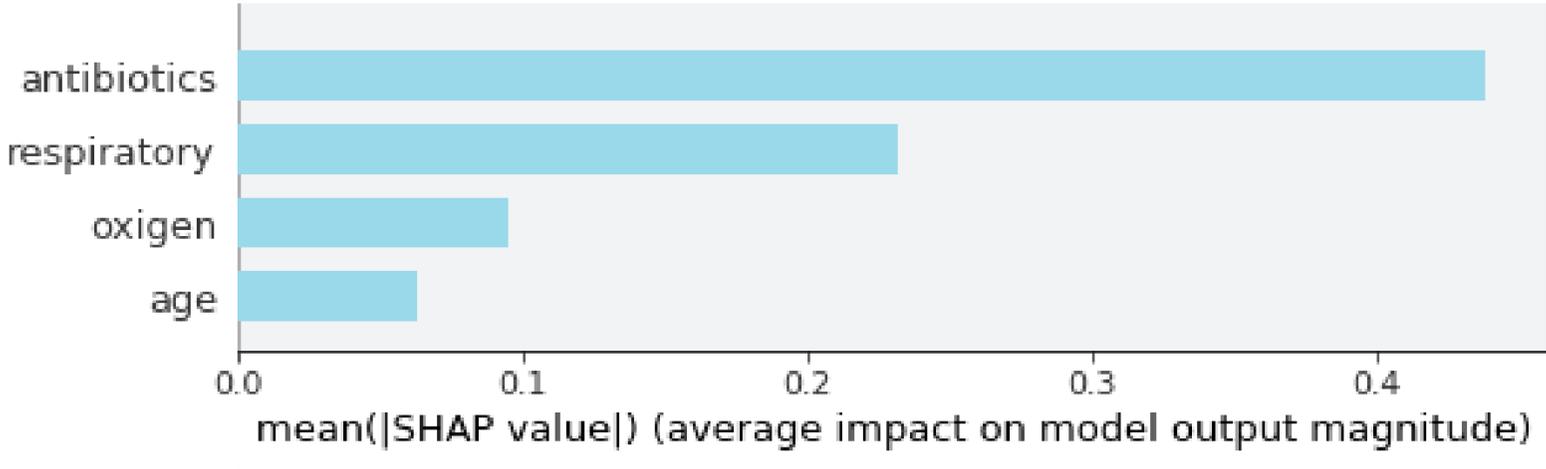
Feature importance
analyses &
visualisations



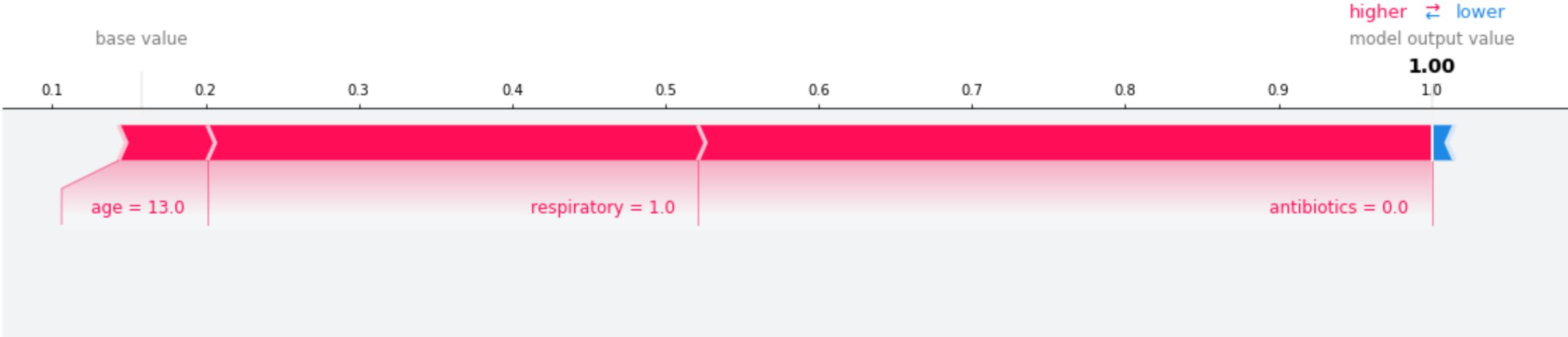
5

WHAT IS THE DATA TELLING ME?

Feature importance analyses & visualisations

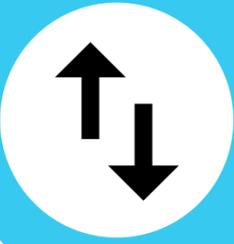


COVID POSITIVE CASE

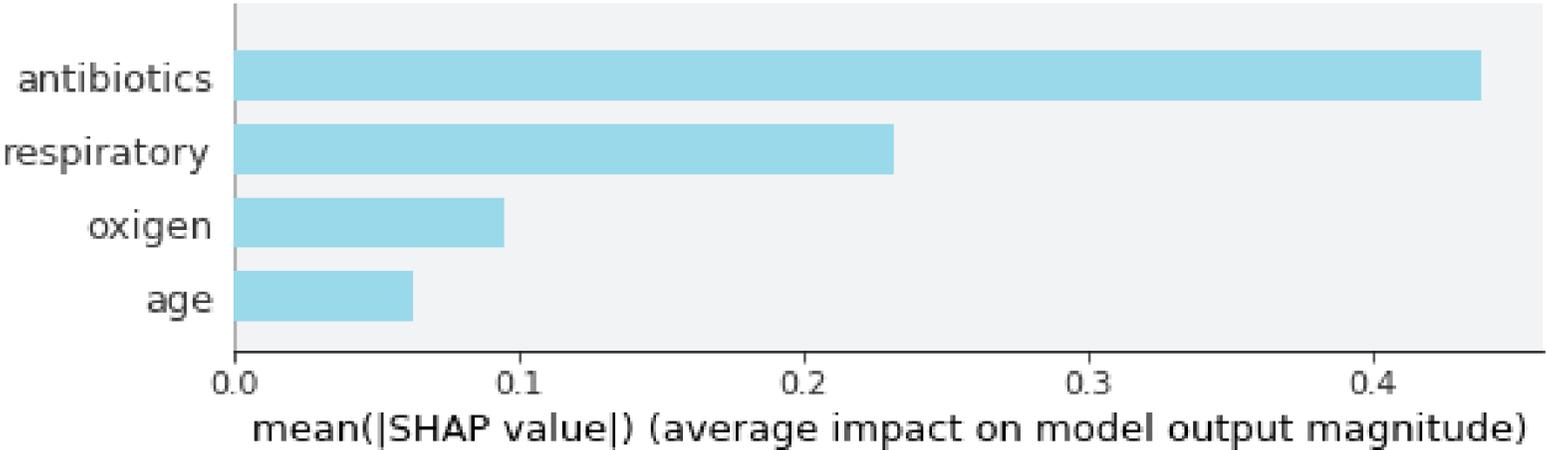


5

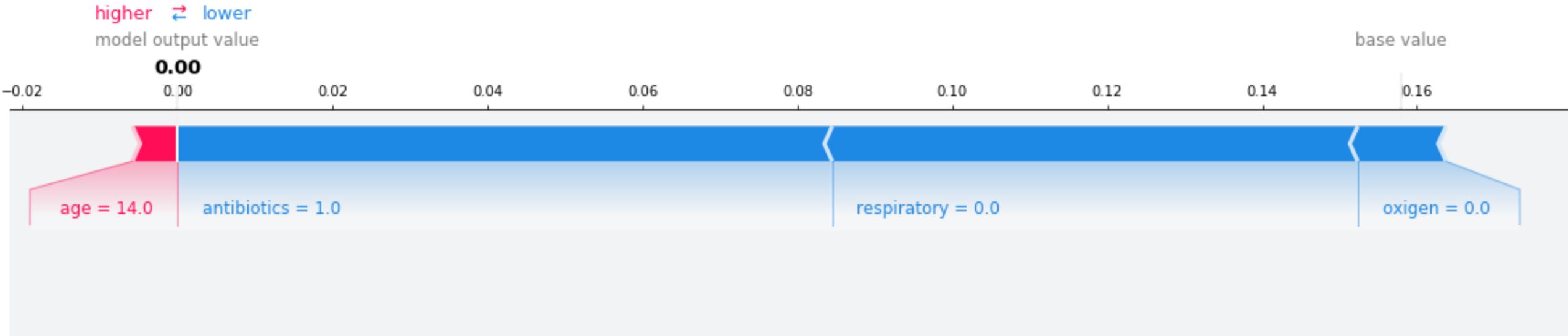
WHAT IS THE DATA TELLING ME?



Feature importance
analyses &
visualisations



COVID NEGATIVE CASE



6



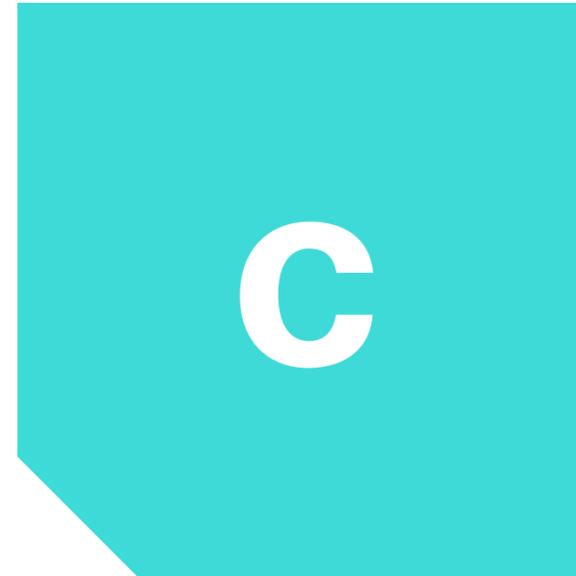
CONCLUSIONS & NEXT STEPS

6

CONCLUSIONS & NEXT STEPS

CONCLUSIONS

- Intricacy and difficulties of real data
- Need to be resourceful and find solutions to data shortcomings
- Promising results despite small sample size
- **Potential of Machine Learning techniques in clinical settings and public health**



CONCLUSIONS

- Intricacy and difficulties of real data
- Need to be resourceful and find solutions to data shortcomings
- Promising results despite small sample size
- **Potential of Machine Learning techniques in clinical settings and public health**



NEXT STEPS

- **Gather more data!**
- Improve model
- Develop predictive models of paediatric COVID-19 severity
- Get in touch with stakeholders that could make good use of the generated knowledge
- Further projects and collaborations in the area (e.g. *in talks*: develop predictive models of adult COVID-19 diagnosis of a Barcelona GP)

THANKS!

amascasadesus@gmail.com

GitHub: [/amascasadesus](https://github.com/amascasadesus)

LinkedIn: [/amascasadesus](https://www.linkedin.com/company/amascasadesus)

