From network-level measurements to expected Quality of Experience

the Skype use case

2015 IEEE International Workshop on Measurements & Networking (M&N)

13 OCTOBER 2015

INTRODUCTION

Quality of Experience

of Internet Services

INTRODUCTION

Quality of Experience

of Internet Services

"A measure of user performance based on both objective and subjective psychological measures of using an ICT service or product."

[ETSI TR 102 643 V1.0.1]

INTRODUCTION

Quality of Experience

INTRODUCTION

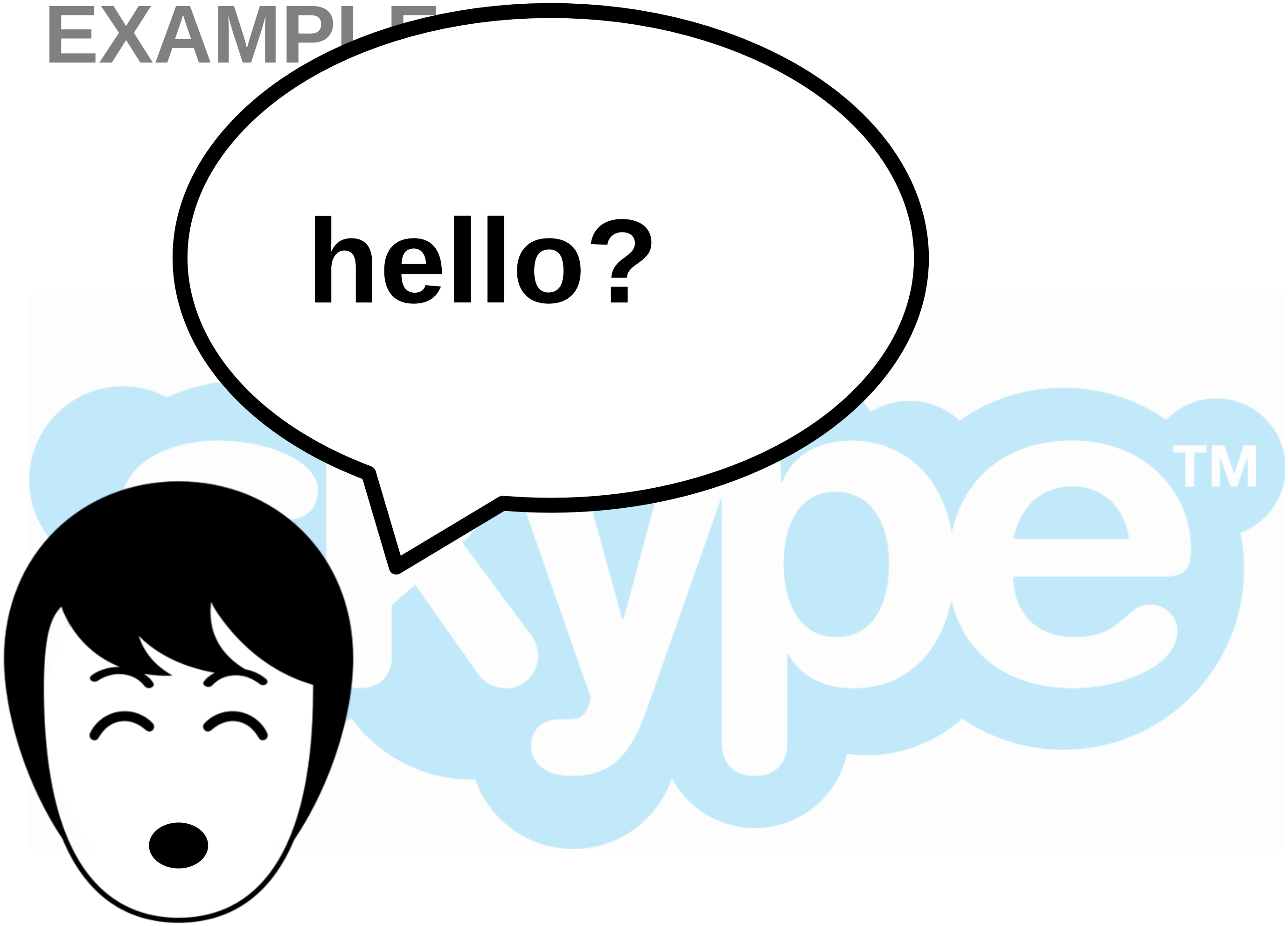
Quality of Experience

vs.

Network Performance

EXAMPLE

EXAMPLE

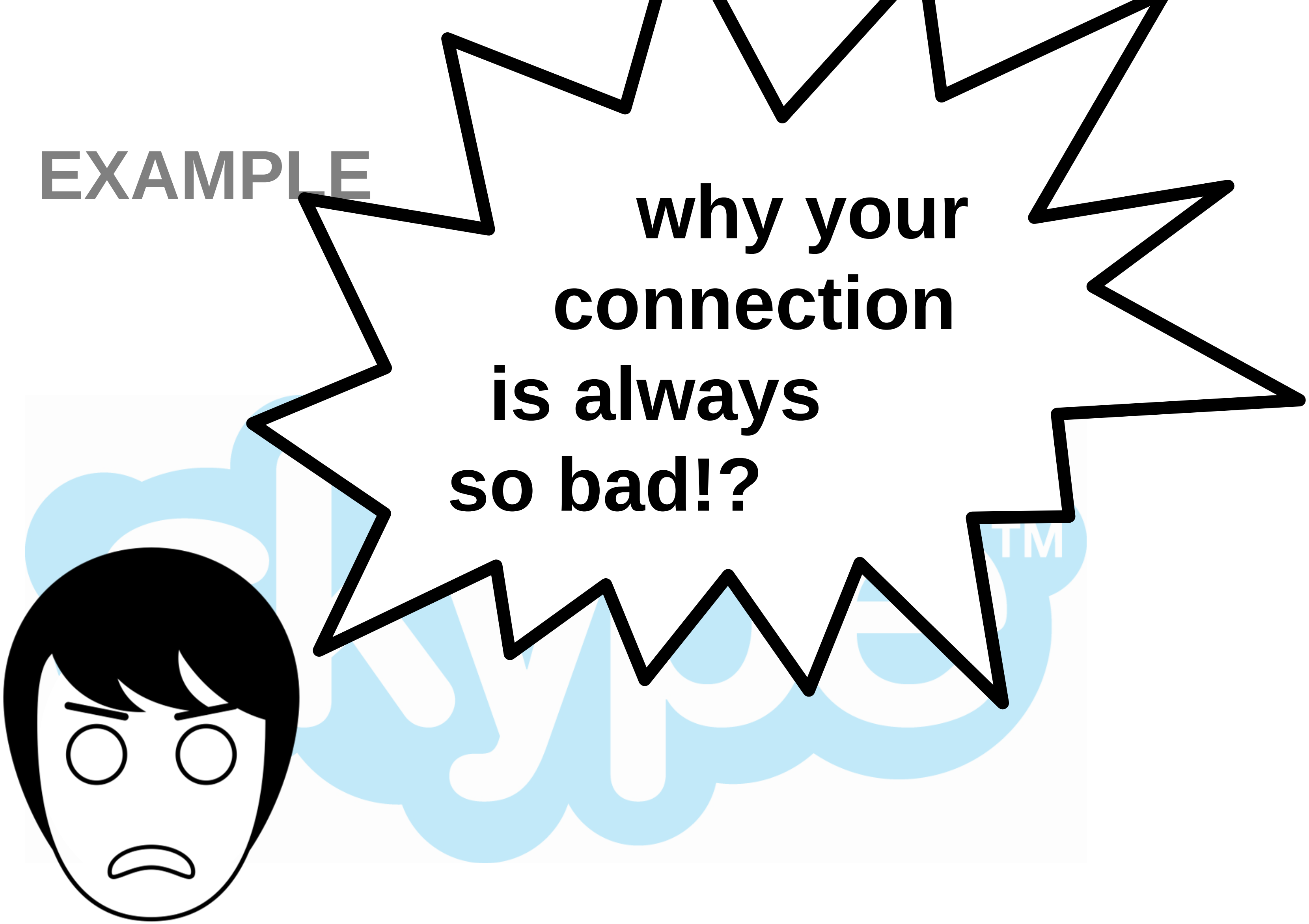


icons by: Javier Sánchez - javyliu from the Noun Project

EXAMPLE

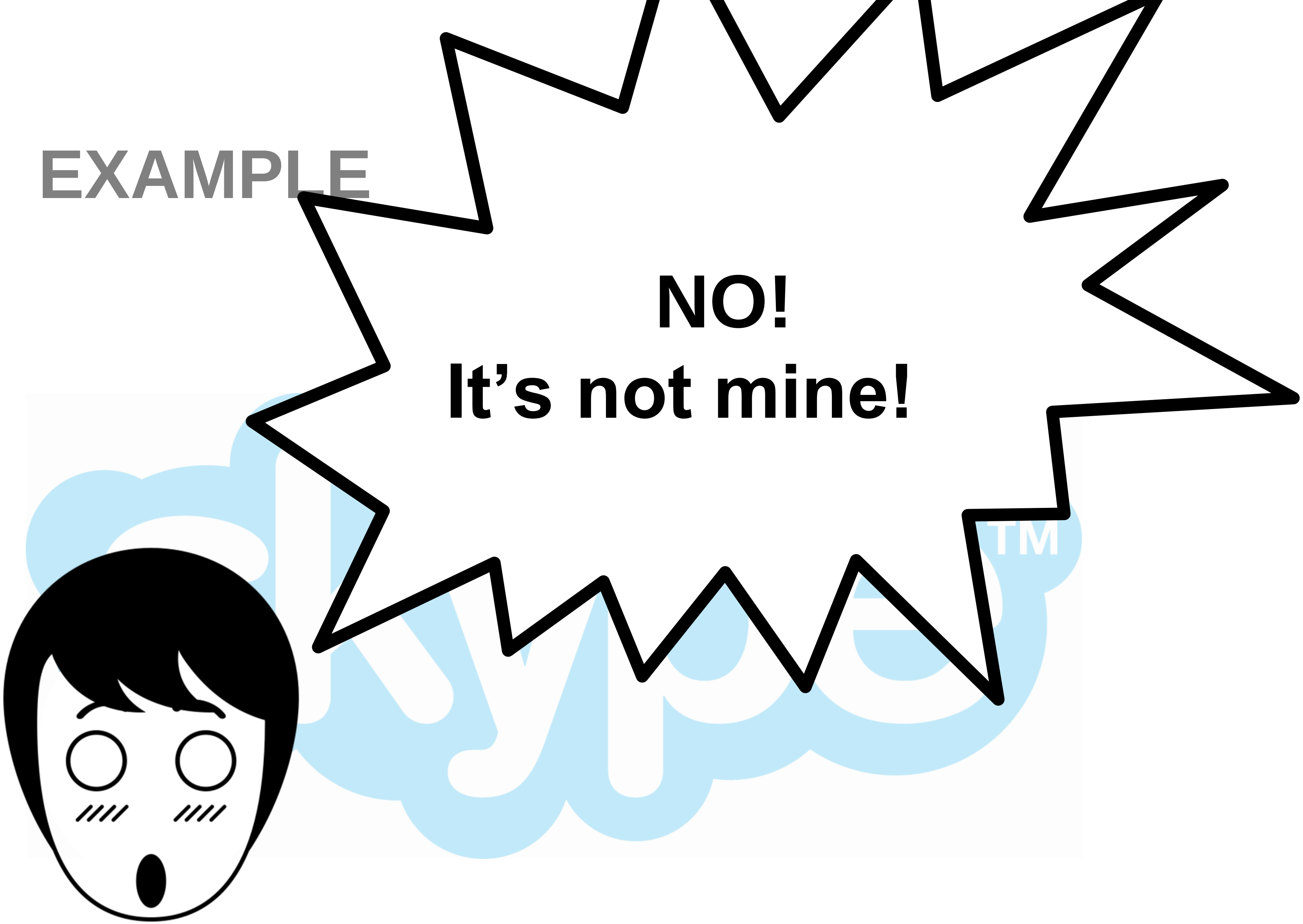
icons by: Javier Sánchez - javyliu from the Noun Project

EXAMPLE



icons by: Javier Sánchez - javyliu from the Noun Project

EXAMPLE



icons by: Javier Sánchez - javyliu from the Noun Project

GOAL

Predict applications QoE without running them!

POSSIBILITIES

- study network dimensioning
- understanding of application needs
- give user the ability to track its connectivity

STATE OF THE ART

Several contributions in the QoE domain

STATE OF THE ART

mostly

- limited scope / usage
- target single applications
- having application data
- passive measurements

STATE OF THE ART

nobody
targets
prediction
of QoE

STATE OF THE ART

nobody
aims to **relate**
QoS of connections
to the
QoE of multiple services

OUR PROJECT

ACQUA

https://team.inria.fr/diana/acqua/

OUR PROJECT

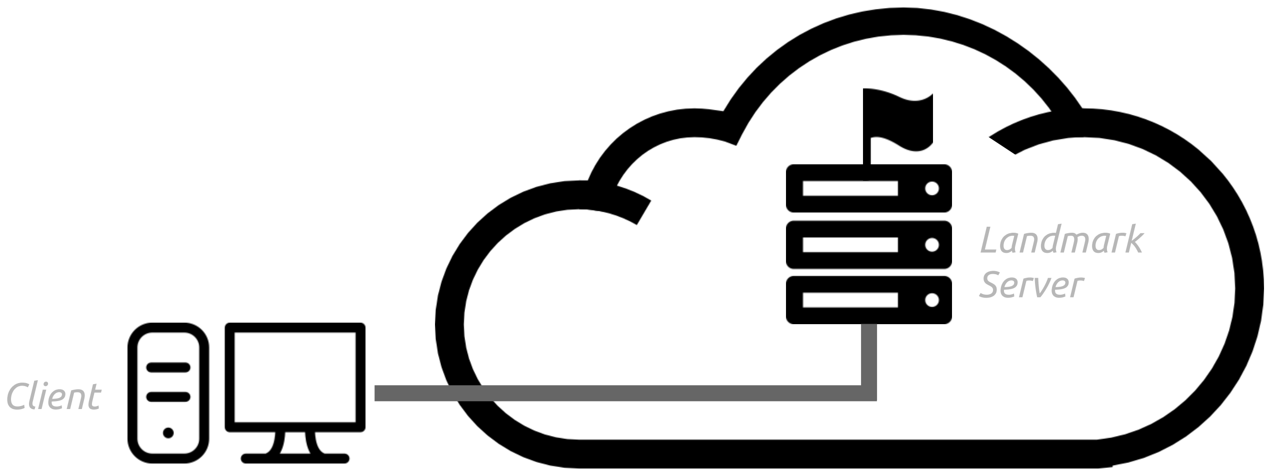
ACQUA

https://team.inria.fr/diana/acqua/

Monitoring platform for tracking QoE of Internet Services

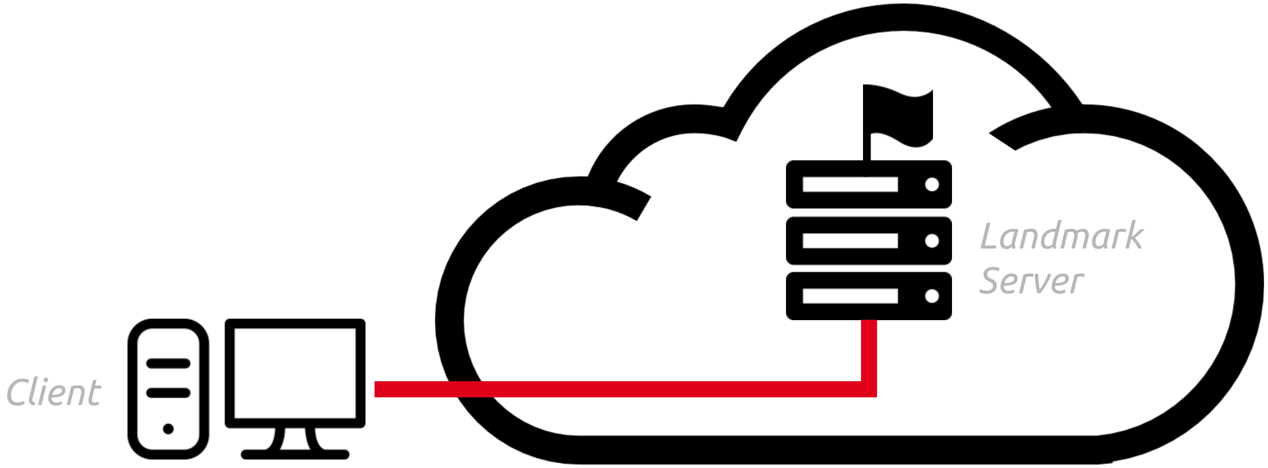
OUR PROJECT

starts with **active measurements** for detecting network conditions at the edges of the Internet



OUR PROJECT

starts with **active measurements** for detecting network conditions at the edges of the Internet

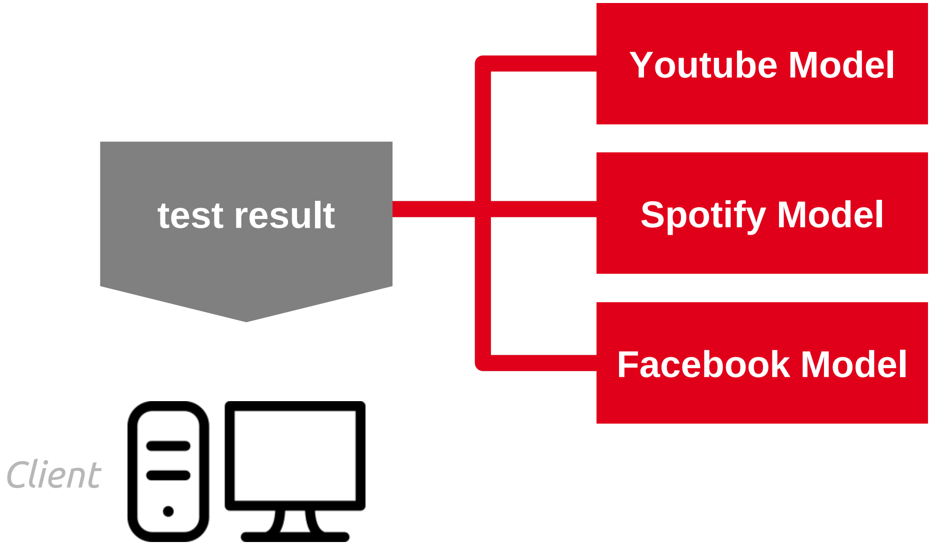


OUR PROJECT

predicts multiple applications' QoE
with only one test

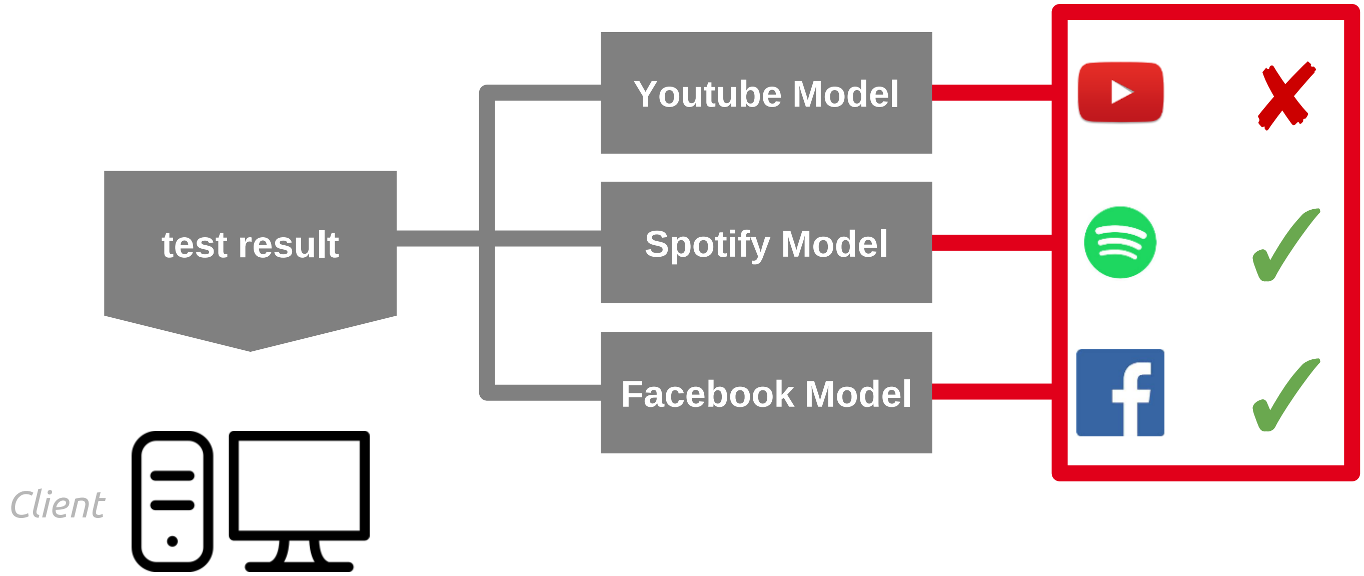
OUR PROJECT

predicts multiple applications' QoE
with only one test



OUR PROJECT

predicts multiple applications' QoE
with only one test



CHALLENGES

- find the right **metrics** for prediction
- define a robust **methodology to** collect data and **create models** (has to be applied on several services)
- **deploy** the framework

METHODOLOGY

Data Analytics

METHODOLOGY

- **create** and **analyze** a data set
- see **how to collect** such data in real world scenarios
- **compare performances** of different **classifiers** derived from such data set

THIS CONTRIBUTION

focus on

skype™

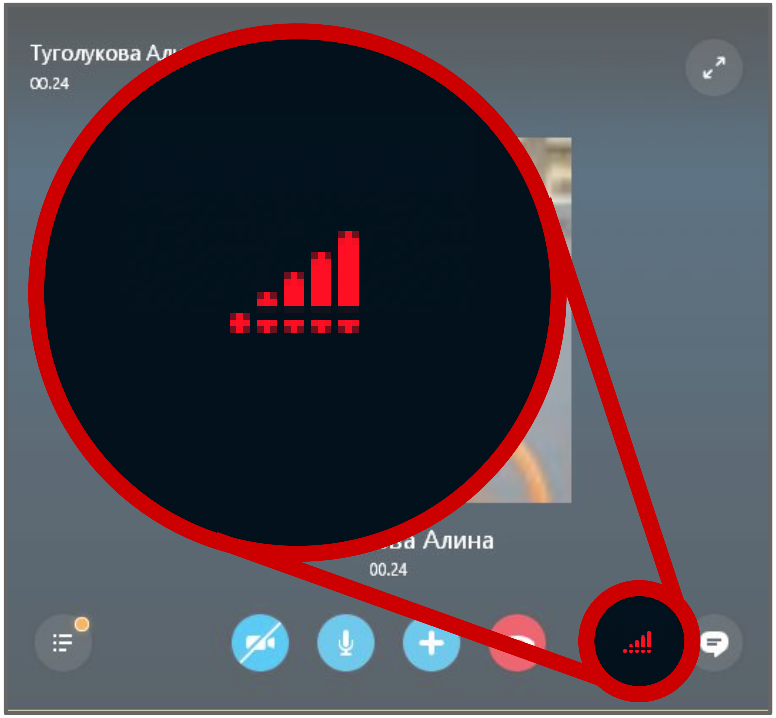
audio calls

WHY SKYPE?

Skype gives us a feedback about the QoE during each call

WHY SKYPE?

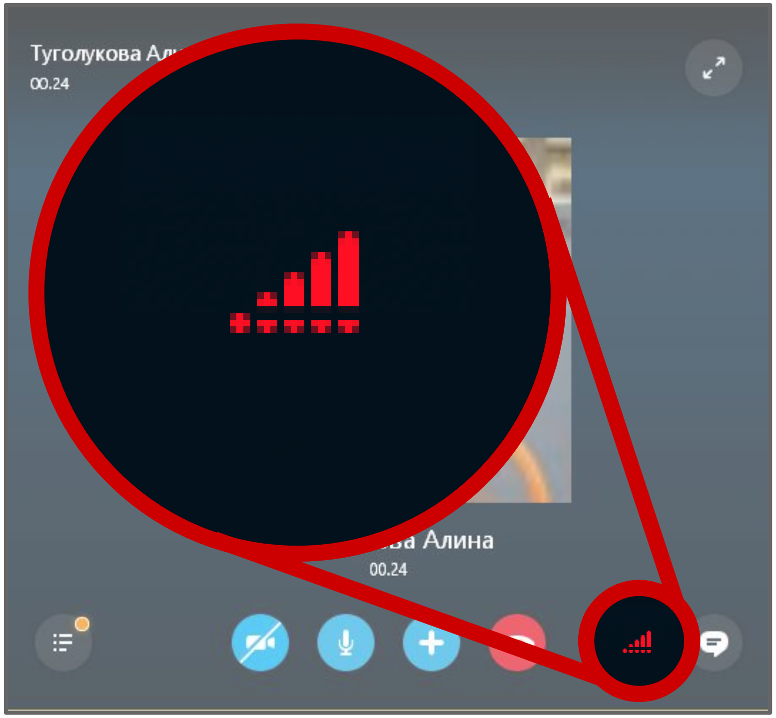
Skype gives us a feedback about the QoE during each call



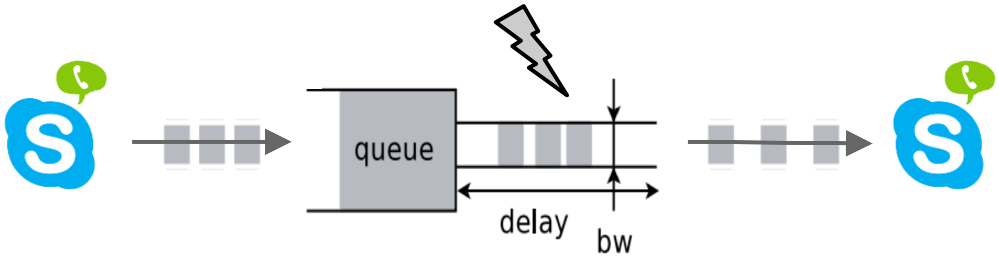
WHY SKYPE?

Skype gives us a feedback about the QoE during each call

Basically we will "reverse engineer" Skype QoE model



DATA SET

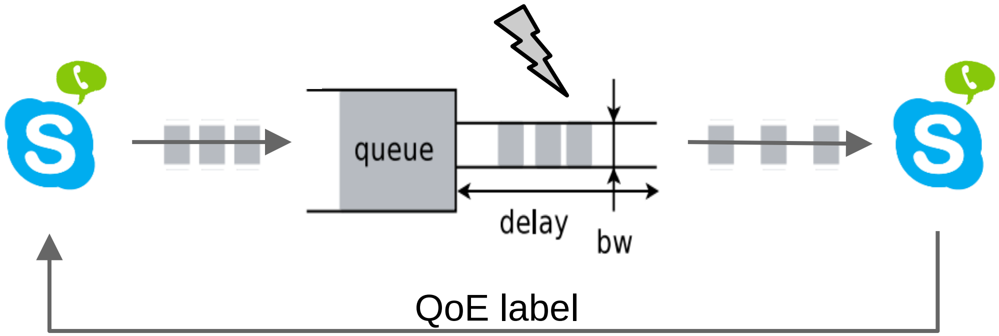


Measures obtained in a **controlled environment** collecting Skype's application feedback given

- latency
- throughput
- packet loss

Image from: Marta Carbone, Luigi Rizzo, An emulation tool for PlanetLab, Computer Communications, Elsevier, Oct.2011, doi:10.1016/j.comcom.2011.06.004

DATA SET

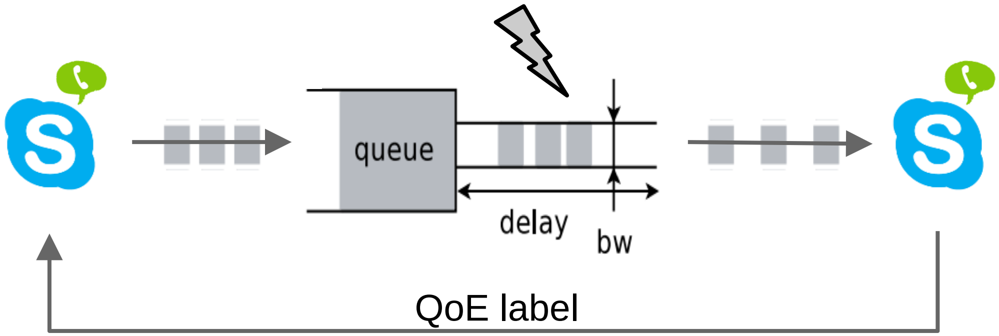


Measures obtained in a **controlled environment** collecting Skype's application feedback given

- latency
- throughput
- packet loss

Image from: Marta Carbone, Luigi Rizzo, An emulation tool for PlanetLab, Computer Communications, Elsevier, Oct.2011, doi:10.1016/j.comcom.2011.06.004

DATA SET



Measures obtained in a **controlled environment** collecting Skype's application feedback given

- latency
- throughput
- packet loss

Static Conditions

Image from: Marta Carbone, Luigi Rizzo, An emulation tool for PlanetLab, Computer Communications, Elsevier, Oct.2011, doi:10.1016/j.comcom.2011.06.004

DATA SET

Composed by 5 input metrics +
1 output QoE label

- Round Trip Time (RTT)
- passing throughput
- packet loss rate

DATA SET

Composed by 5 input metrics +
1 output QoE label

- Round Trip Time (RTT)
- passing throughput
- packet loss rate

$\Big\}$ upload and download

DATA SET

Composed by 5 input metrics +
1 output QoE label

- Round Trip Time (RTT)
- passing throughput
- packet loss rate

} upload and download

- QoE label ∈ `{NoCall, Poor, Medium, Good}`

DATA SET

FAST sampling

(**F**ourier **A**mplitude **S**ensitivity **T**esting)

to cover the configuration space uniformly

DATA SET

538 entries (training)

DATA SET

538 entries (training)

+

100 entries (validation)

MODEL GENERATION

Data-Driven Approach

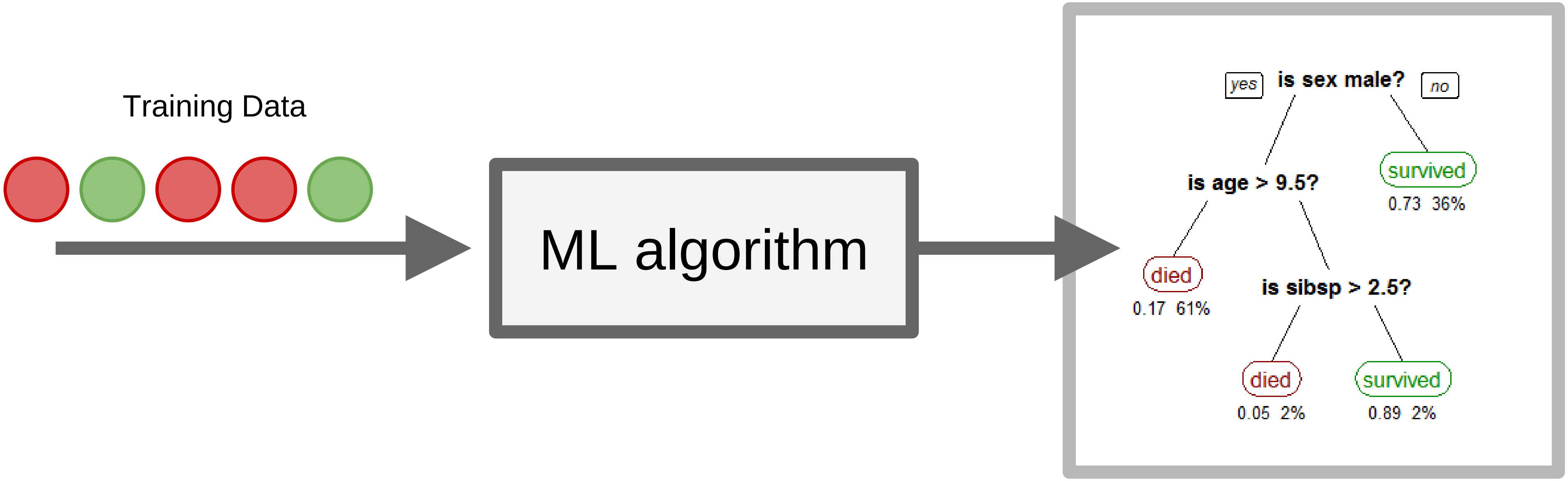
- **compare** different **ML techniques**
  using the **Weka** toolkit
- **check** classifier
  - **performances** using **independent test** set
  - **stability** using **Cross Validation**

HOW A CLASSIFIERS WORKS?

ML algorithm

HOW A CLASSIFIERS WORKS?

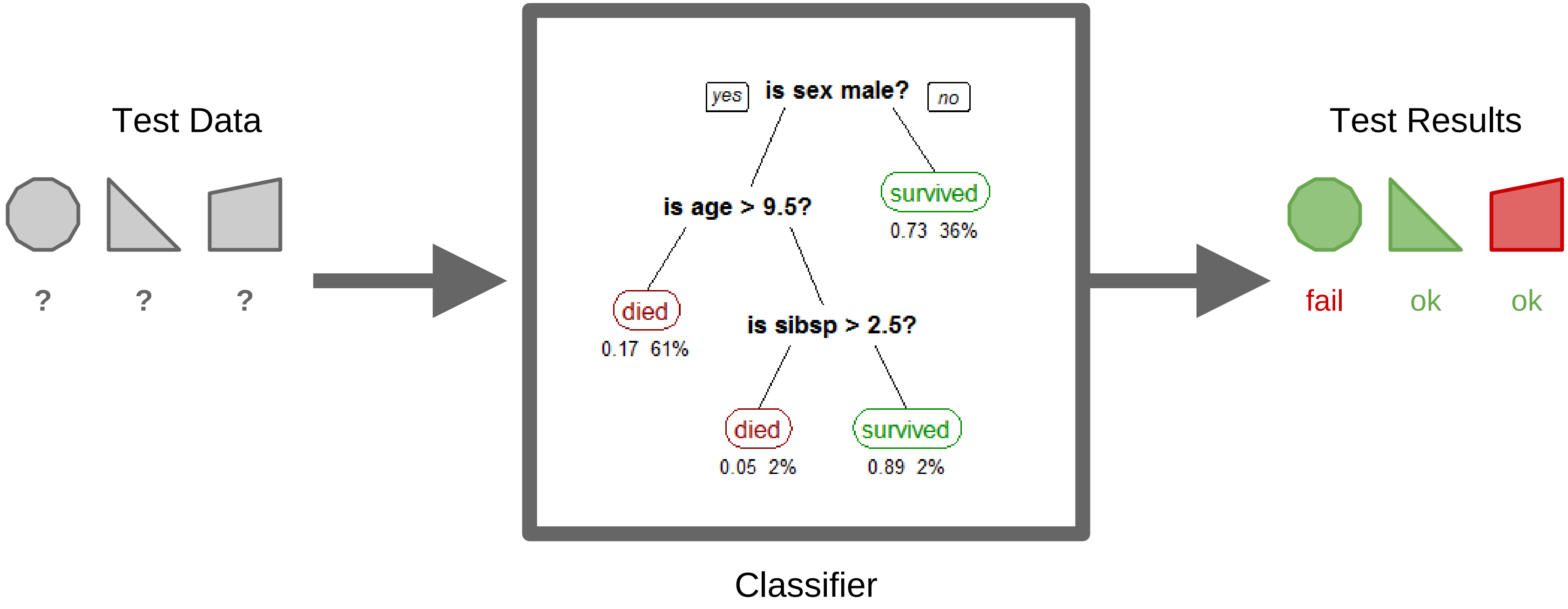
Step 1: calibration



Classifier

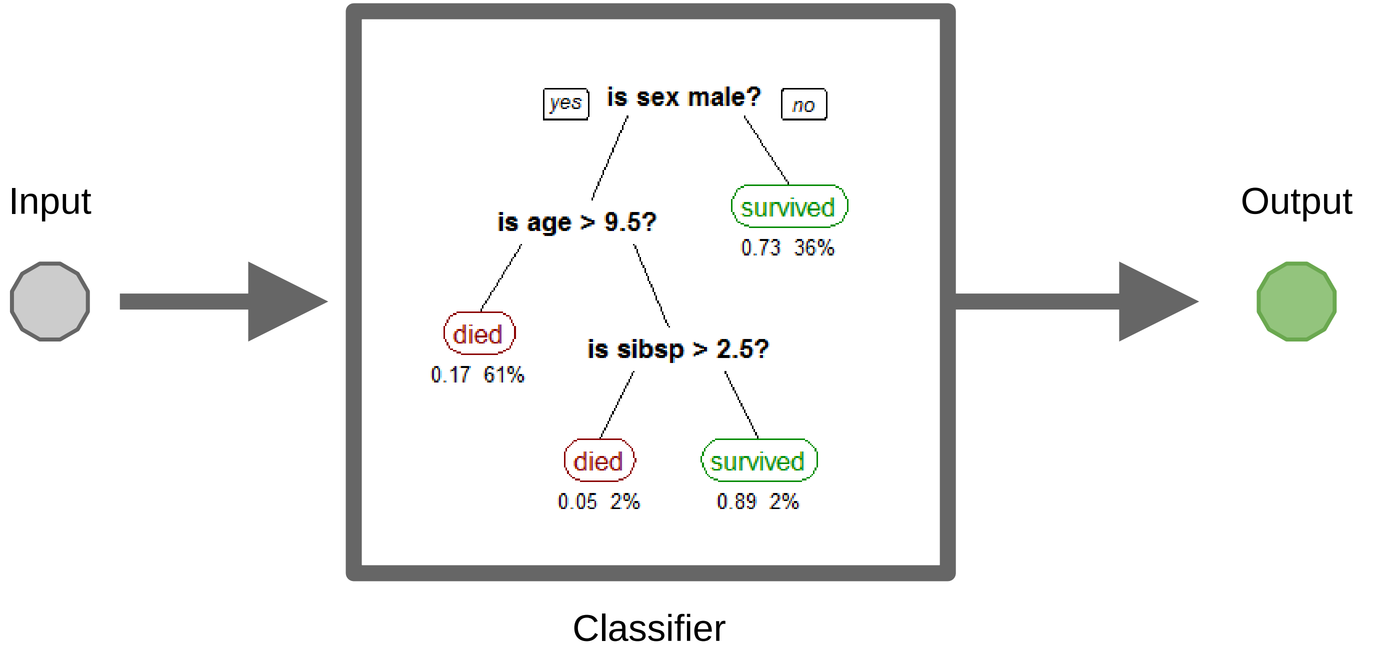
HOW A CLASSIFIERS WORKS?

Step 2: validation



HOW A CLASSIFIERS WORKS?

Step 3: deploy



CONSIDERED TECHNIQUES

a classifier for each different family

- **Decision Trees** (C4.5)
  and **Rule Inference** (JRip, FURIA)
- **Lazy Learners** (kNN)
- **Probability based** (Naive Bayes, Neural Networks)
- **Support Vector Machines** (SVM)
- **Random Forests**
- **Meta Techniques** (Boosting, Bagging)

PERFORMANCE METRICS

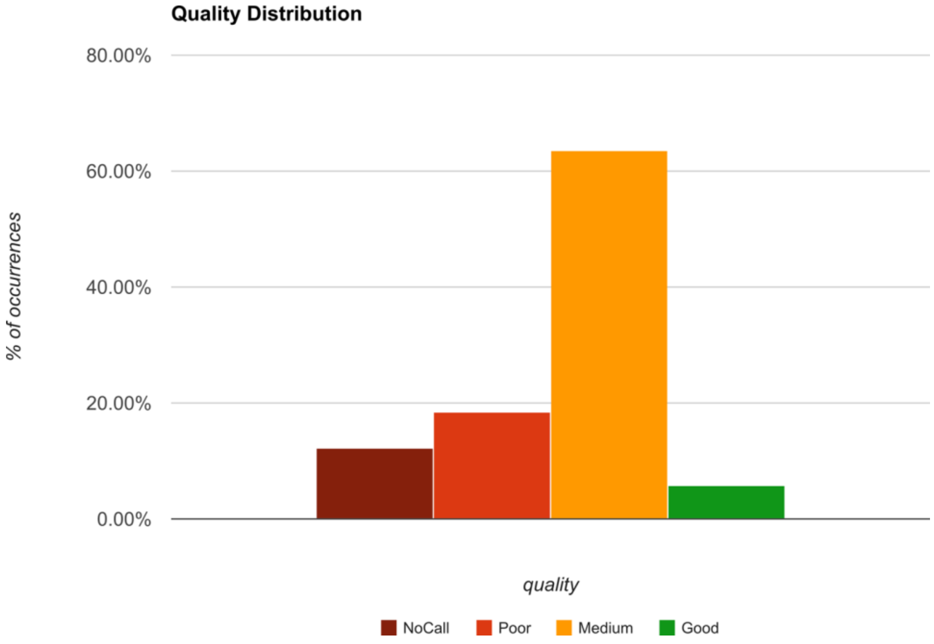
How to consider the effectiveness of a classifier?

PERFORMANCE METRICS

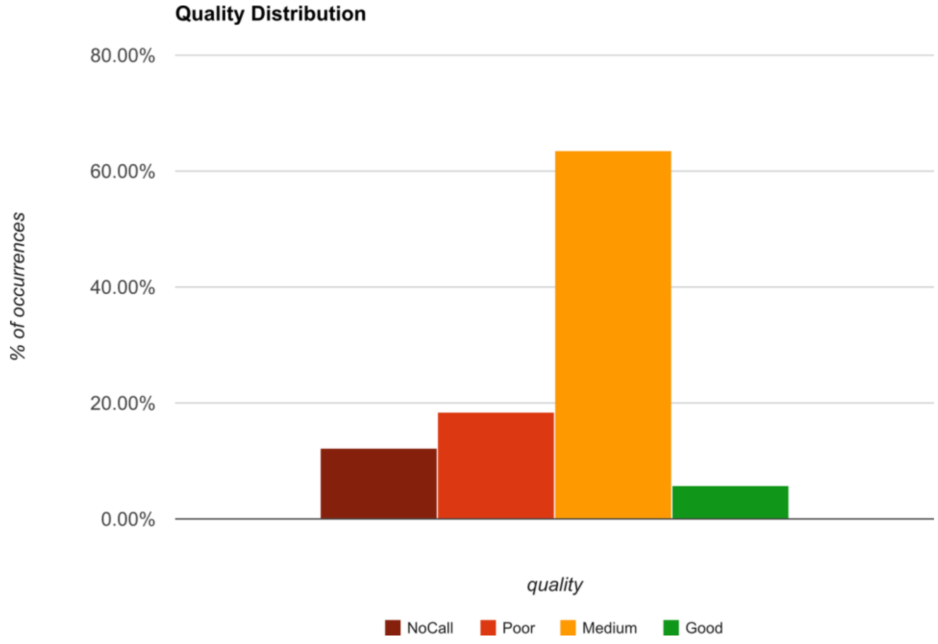
How to consider the effectiveness of a classifier?

$$\text{ACCURACY} = \frac{\text{\#Correctly Classified Instances}}{\text{\#Instances}}$$

PERFORMANCE METRICS

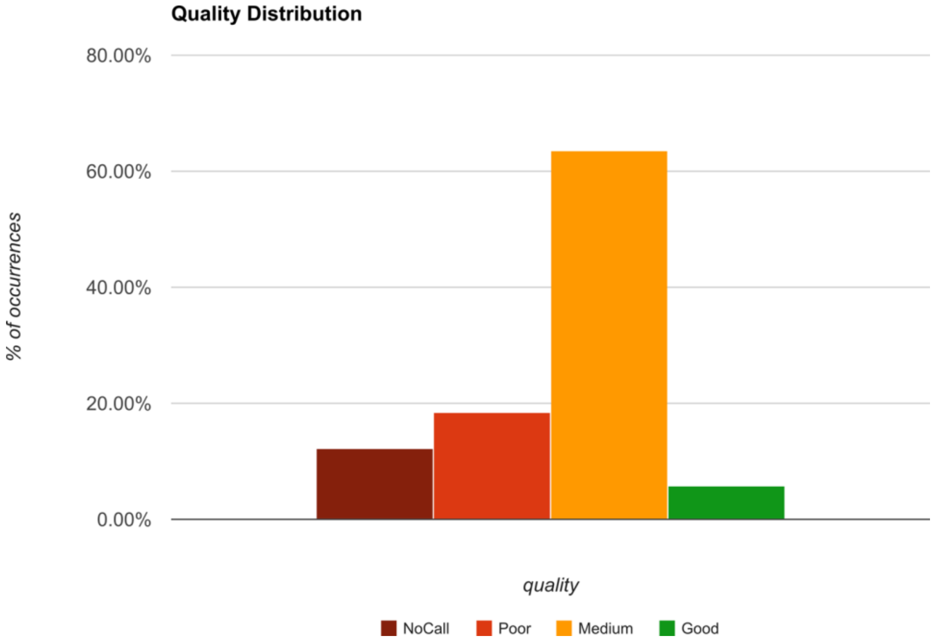


PERFORMANCE METRICS



Classifying all the instances as **medium**

PERFORMANCE METRICS

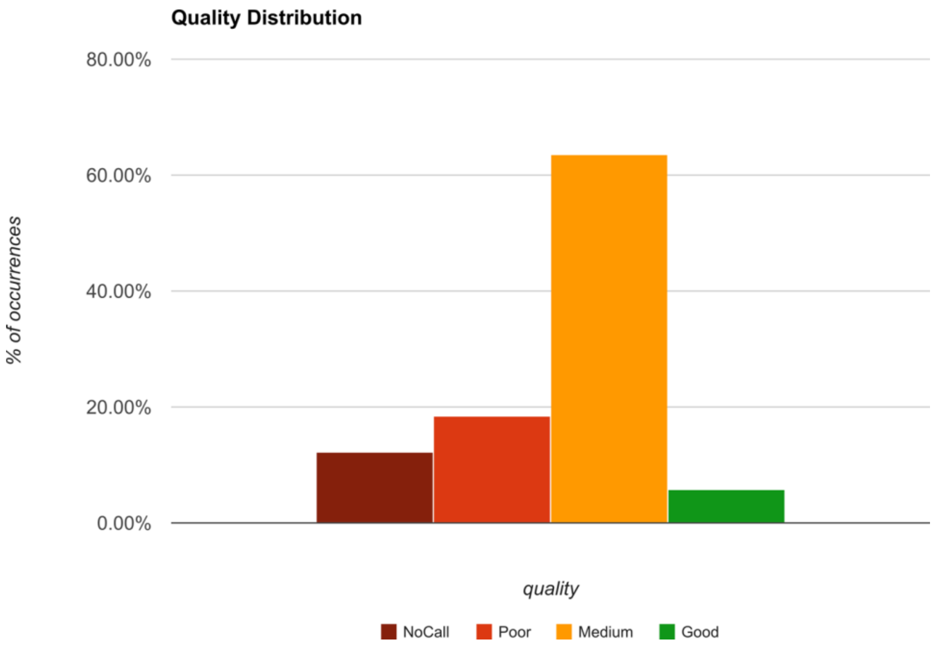


Classifying all the instances as **medium**

over

60% accuracy

PERFORMANCE METRICS



Classifying all the instances as **medium**

over

60% accuracy

Not enough to have a global overview

PERFORMANCE METRICS

How to consider the effectiveness of a classifier?

- Accuracy

PERFORMANCE METRICS

How to consider the effectiveness of a classifier?

- ~~Accuracy~~

PERFORMANCE METRICS

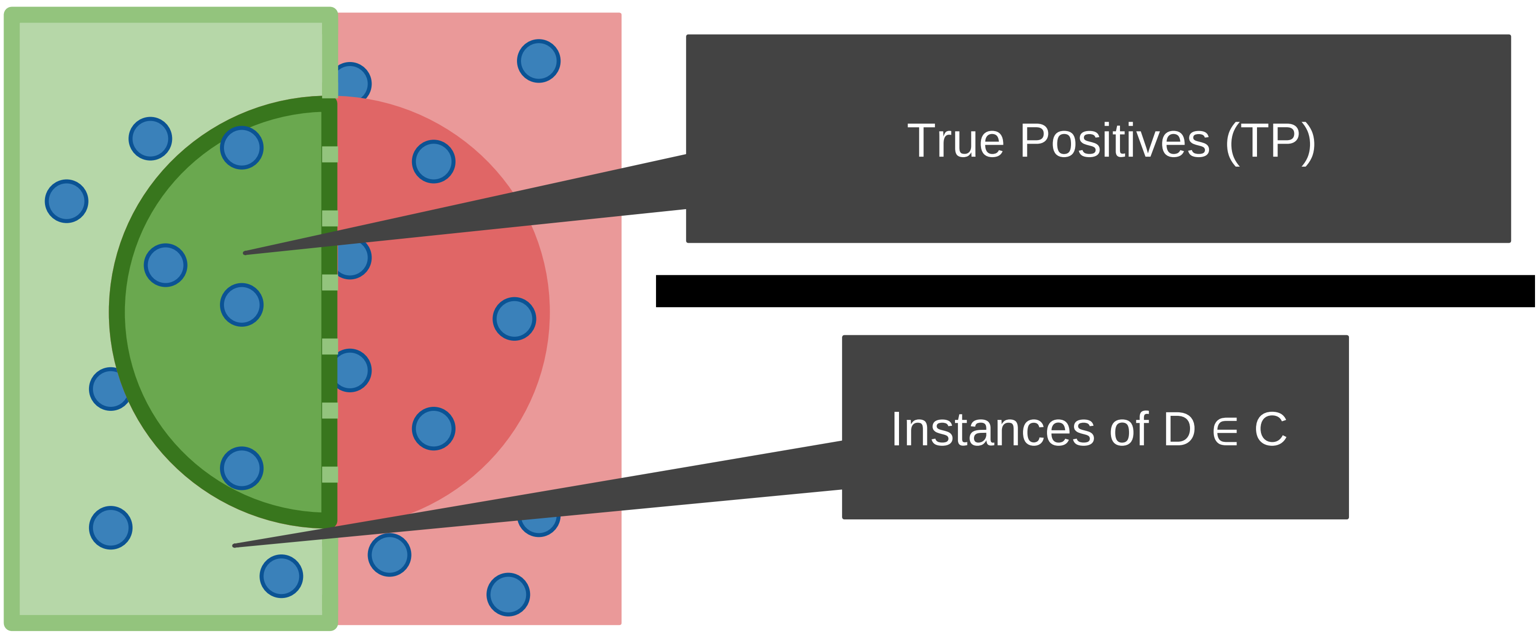
How to consider the effectiveness of a classifier?

- ~~Accuracy~~
- Precision / Recall

  for each class

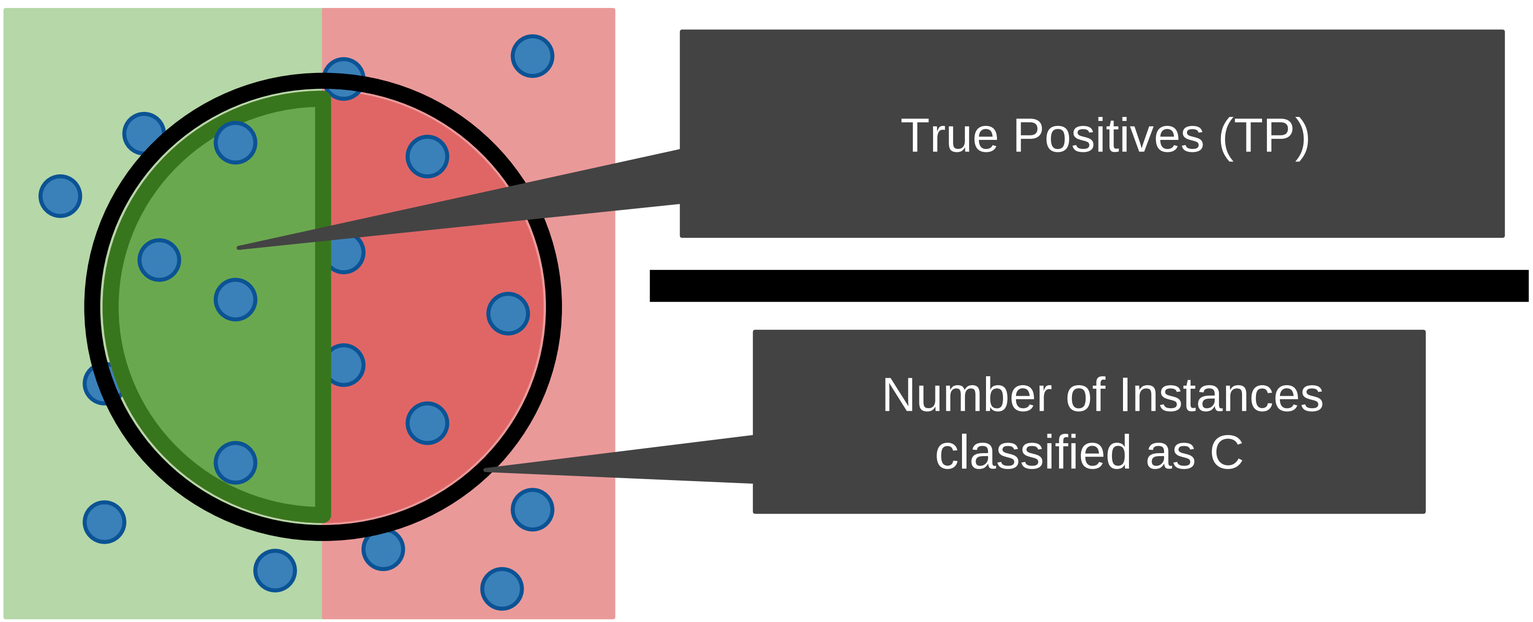
PERFORMANCE METRICS

RECALL (completeness)

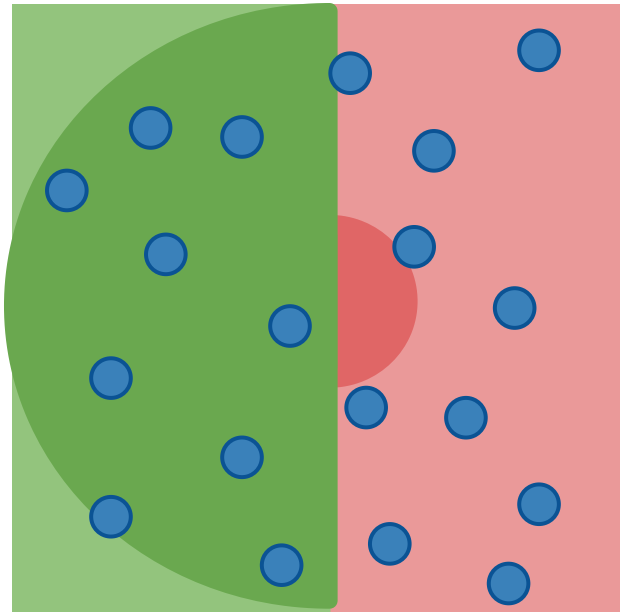


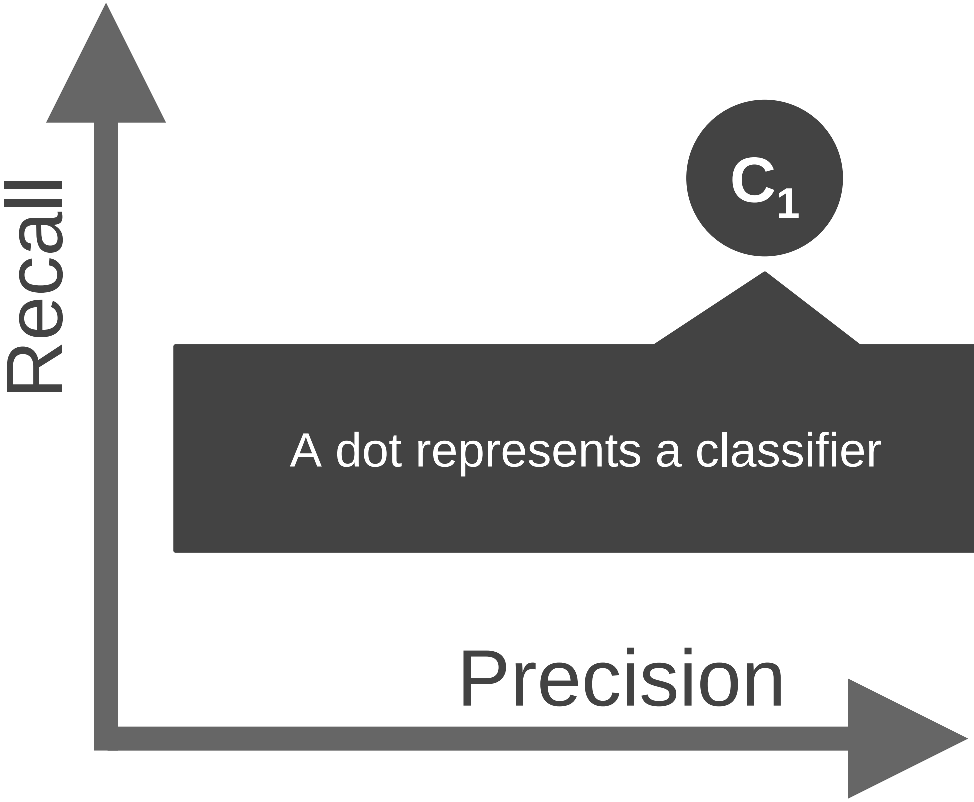
PERFORMANCE METRICS

PRECISION (quality)

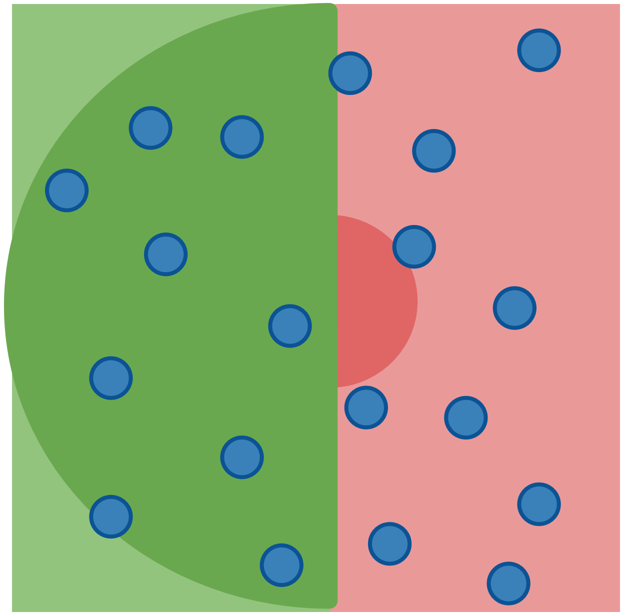


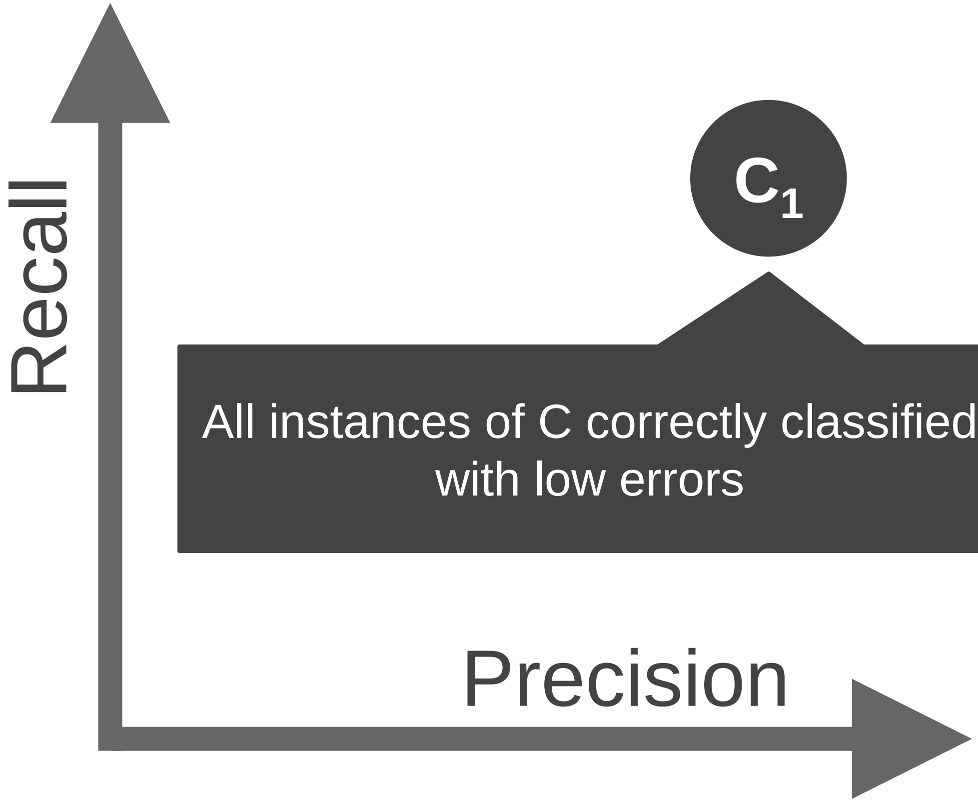
INTERPRETATION



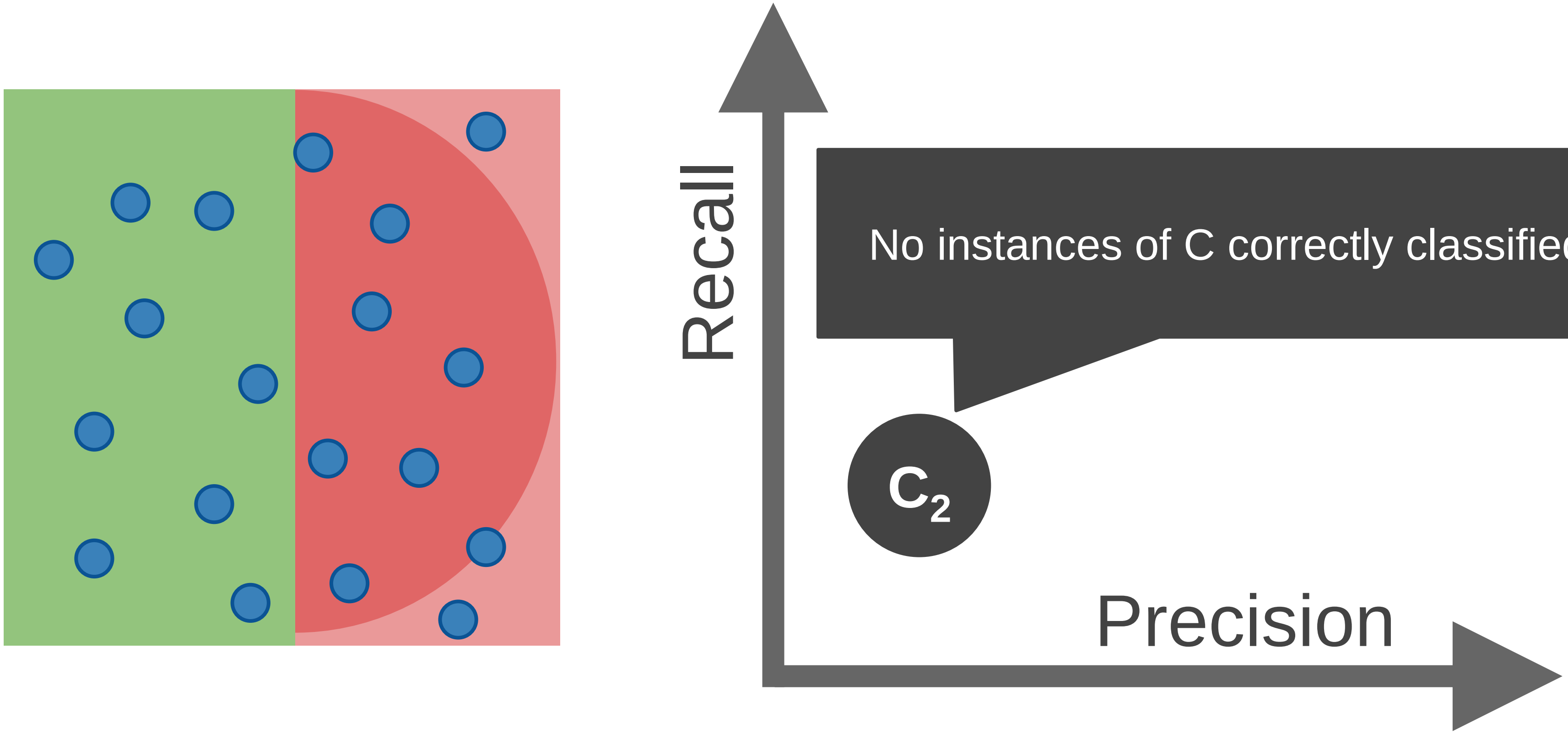


INTERPRETATION

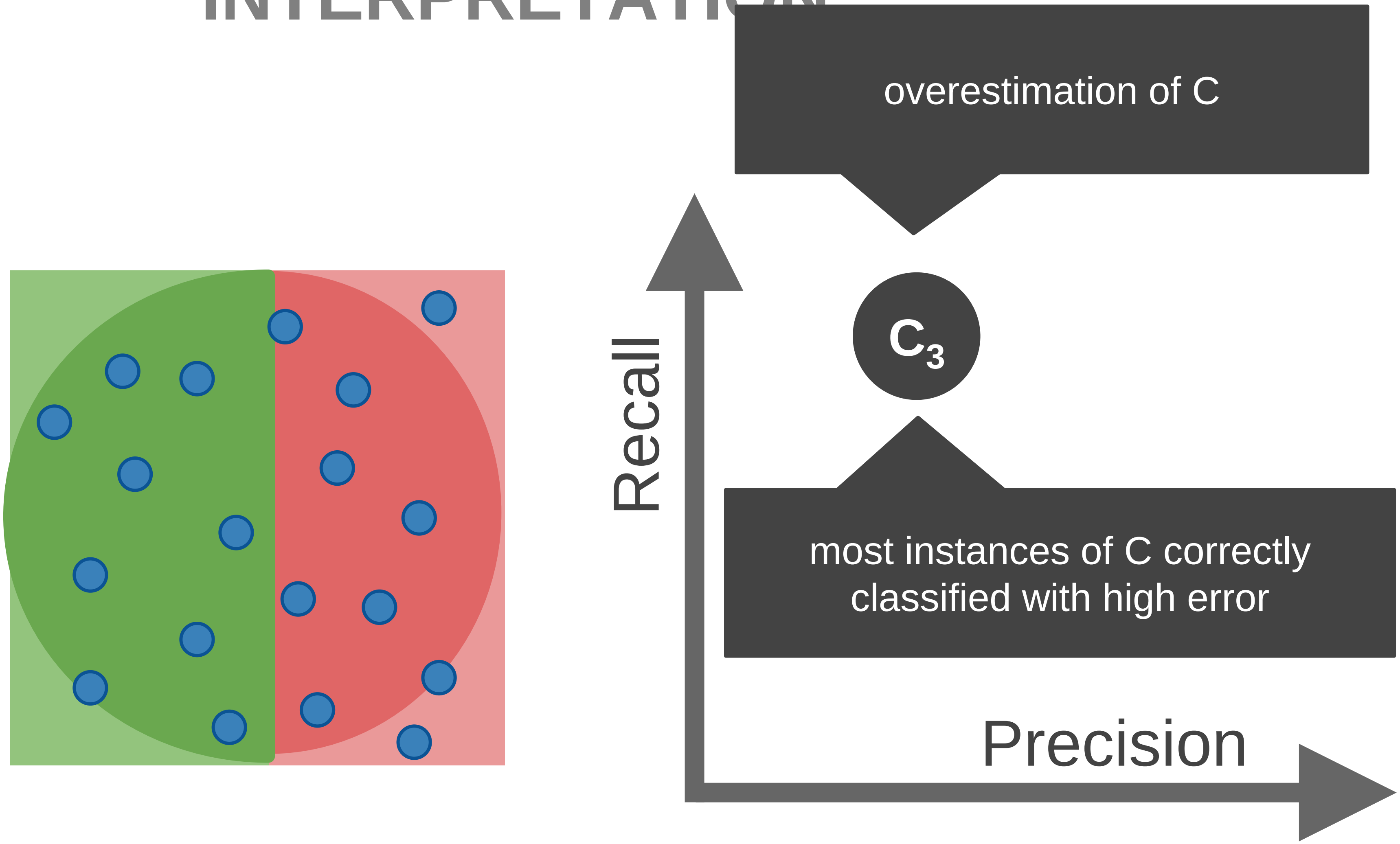




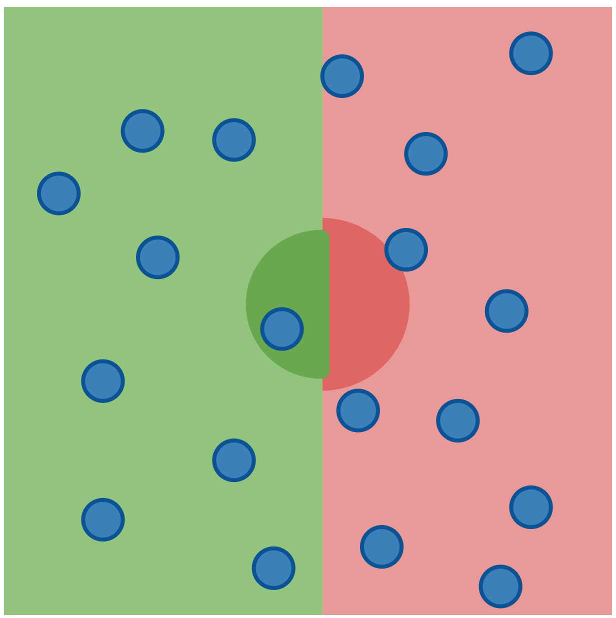
INTERPRETATION

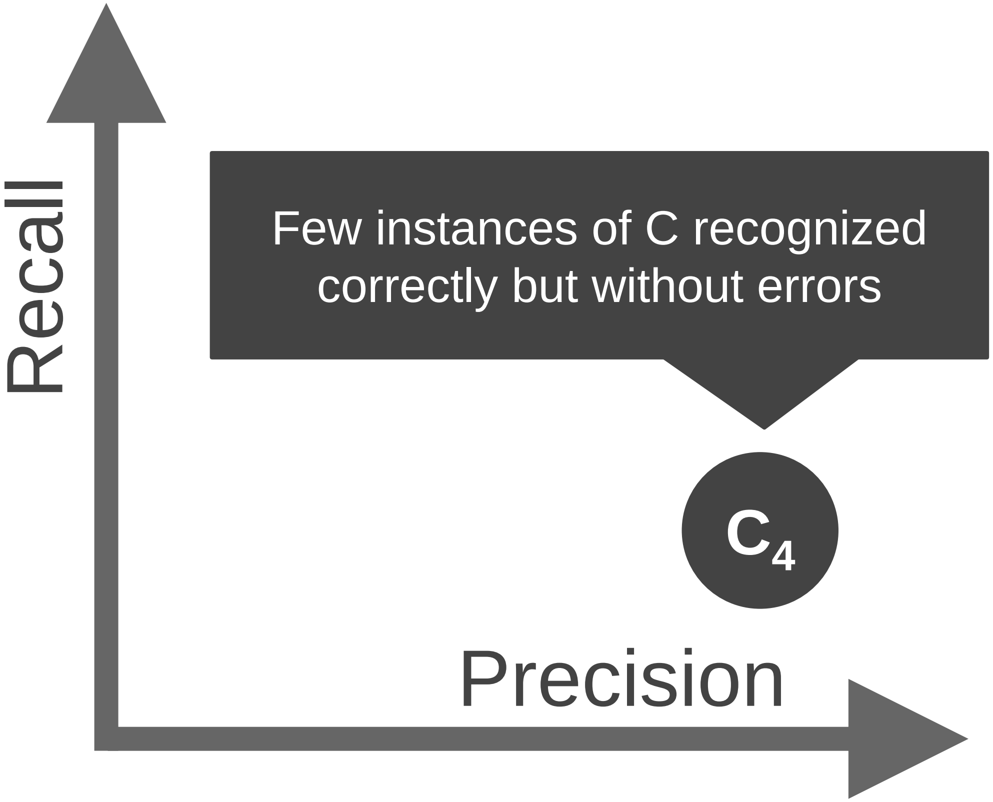


INTERPRETATION

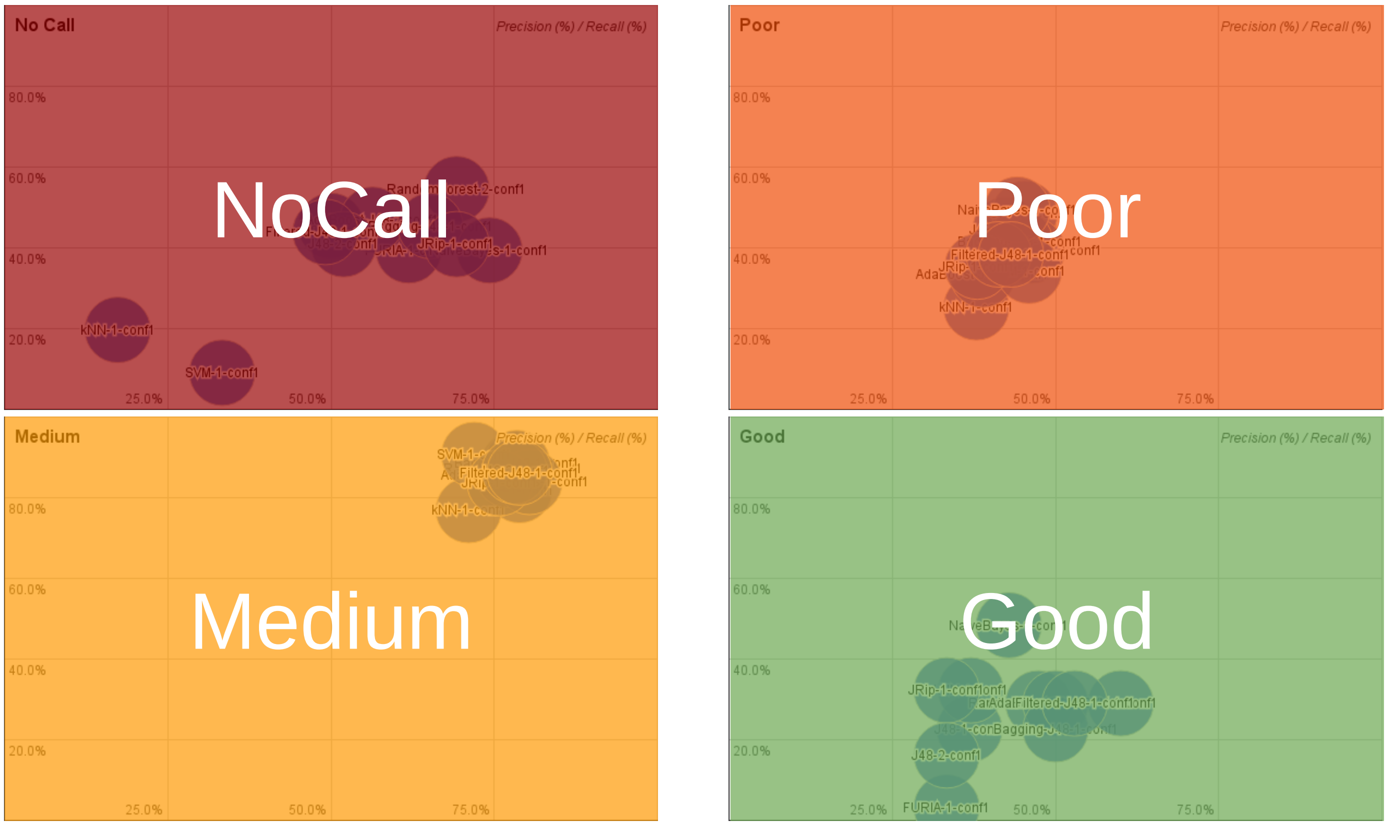


INTERPRETATION

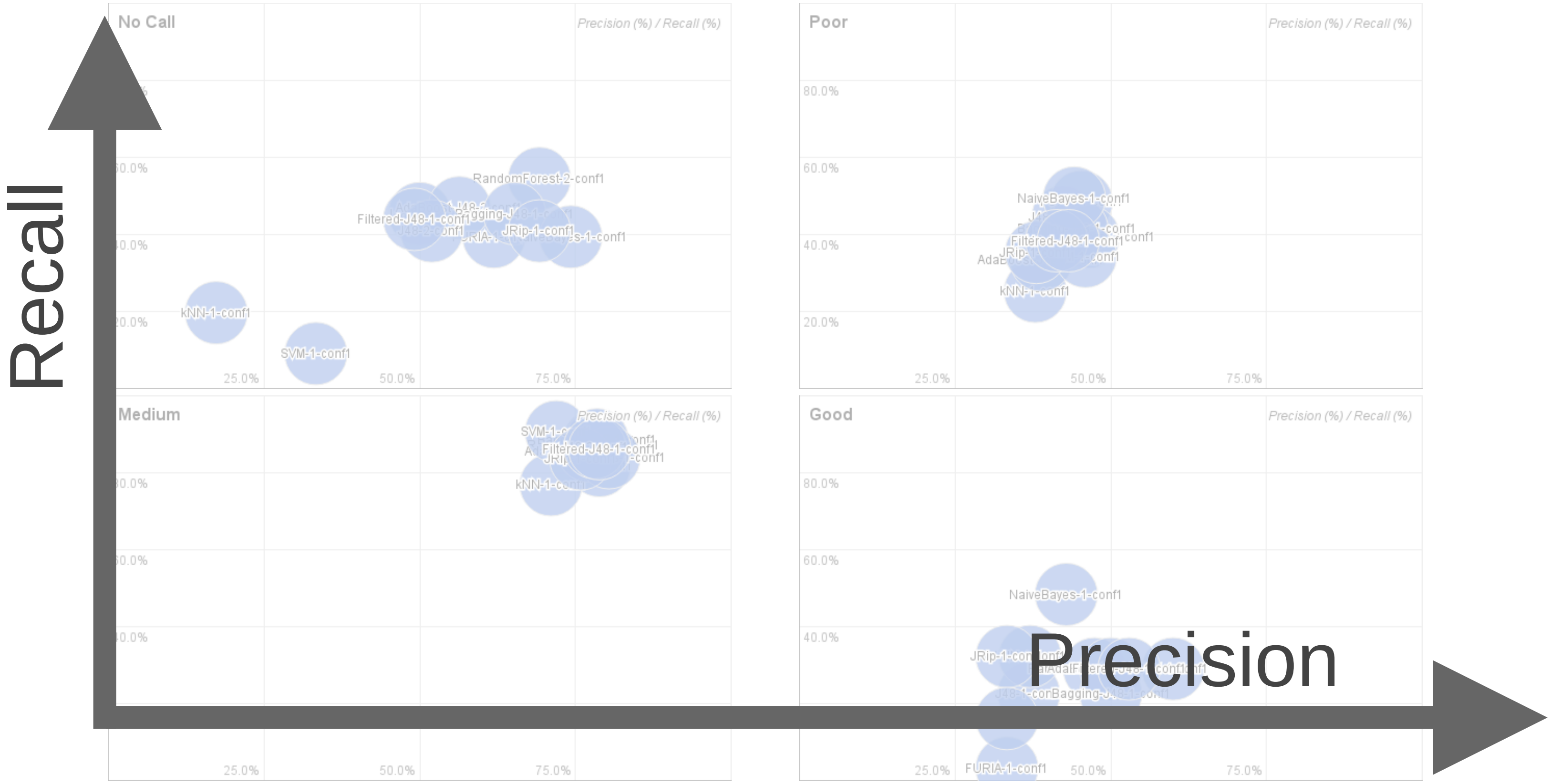




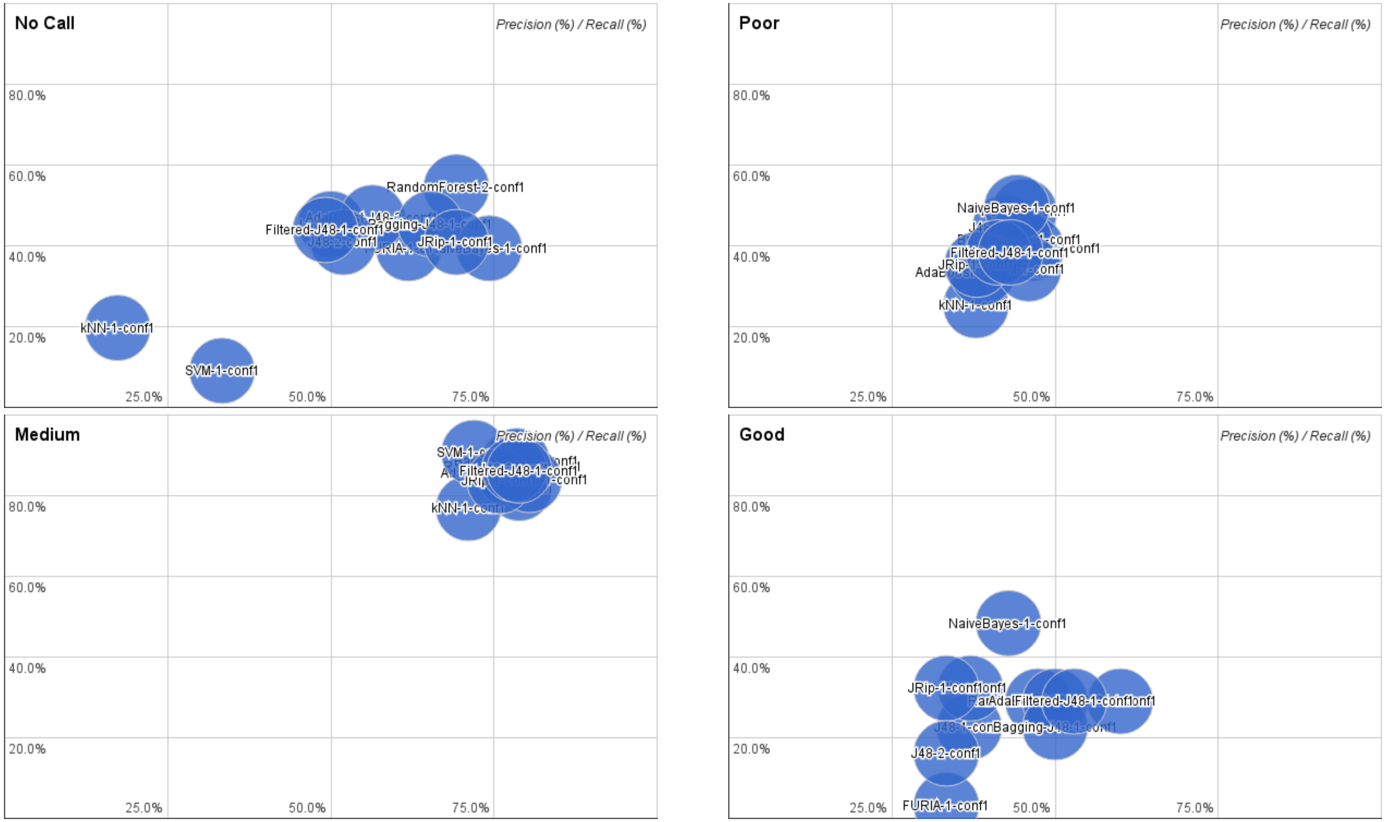
RESULTS - Cross Validation



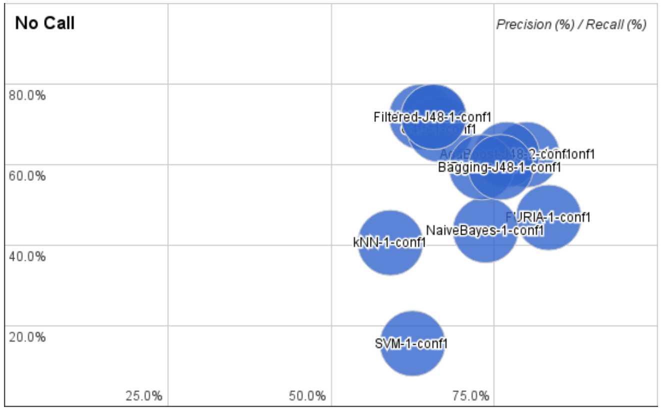
RESULTS - Cross Validation

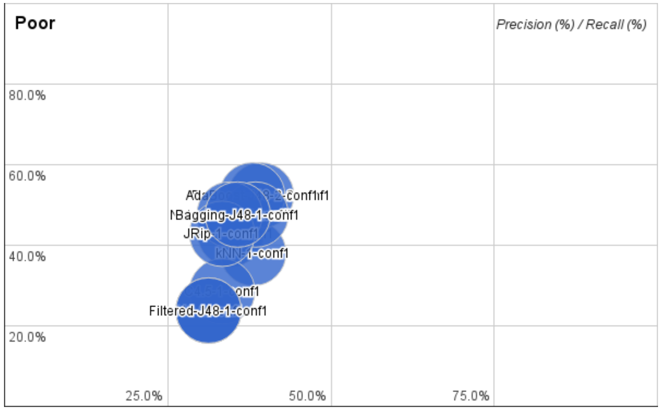


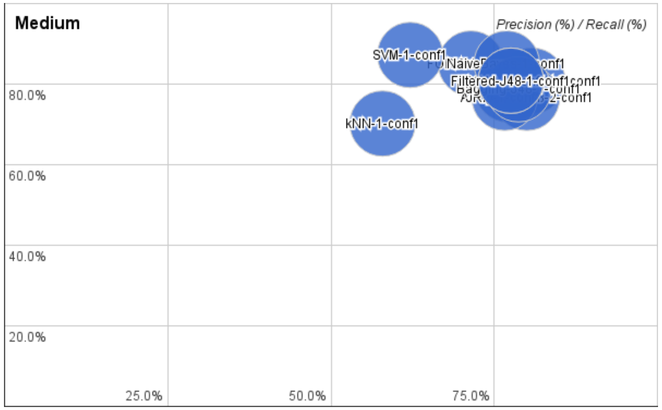
RESULTS - Cross Validation



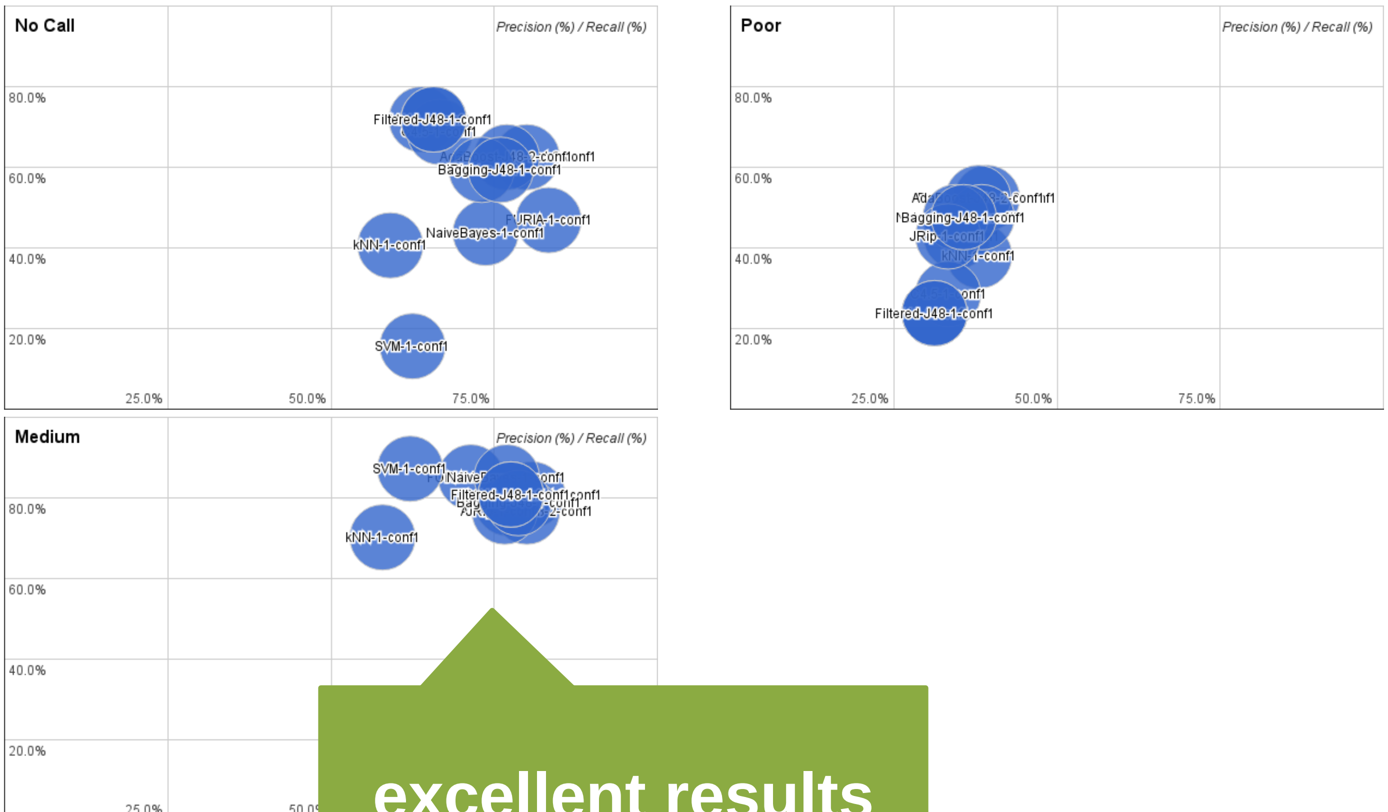
RESULTS - Validation





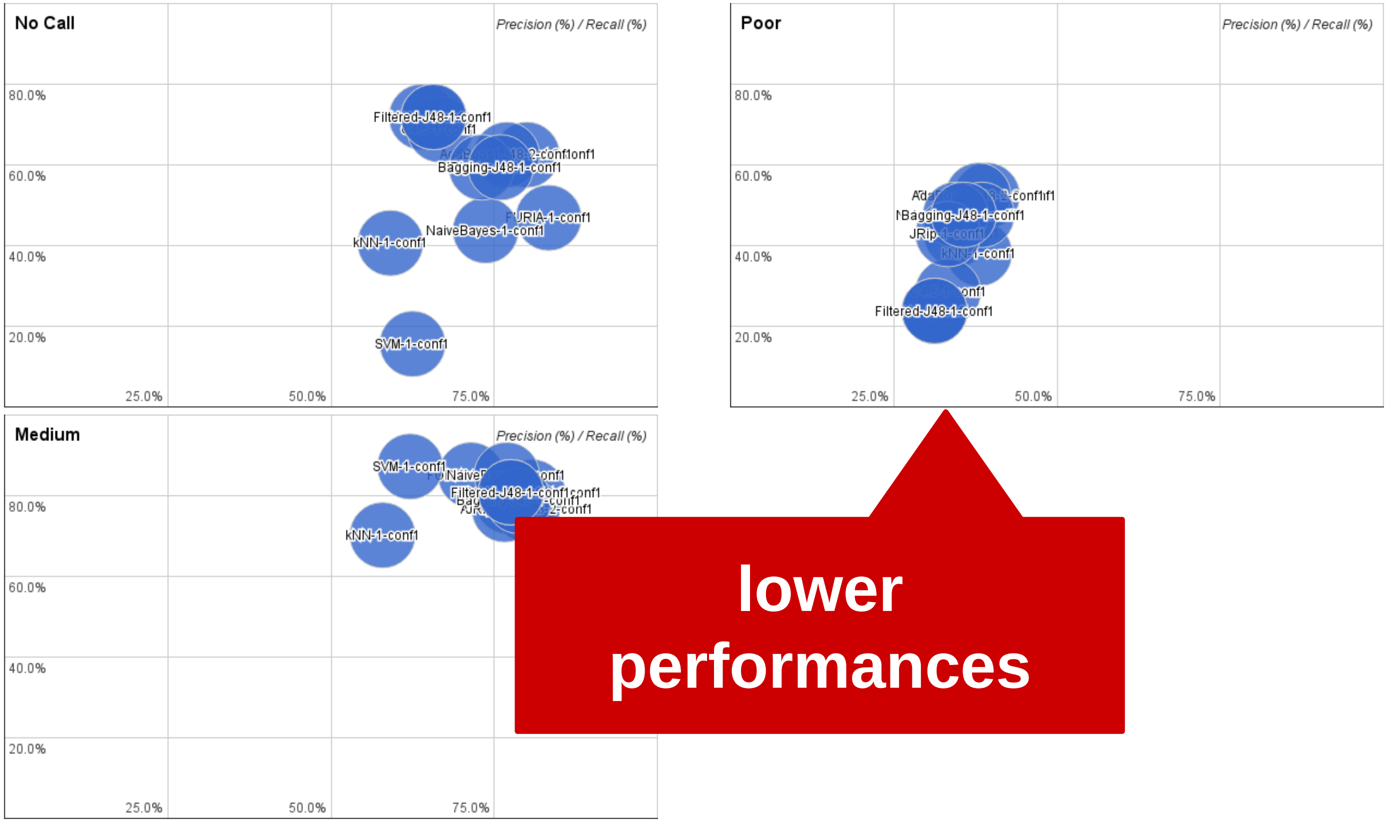


RESULTS - Validation



excellent results

RESULTS - Validation



lower performances

CAN WE GET MORE?

CAN WE GET MORE?

Unbalanced training set could affect classification

CAN WE GET MORE?

Unbalanced training set could affect classification

⇒ **inflate** data to reduce bias

CAN WE GET MORE?

Unbalanced training set could affect classification

⇒ **inflate** data to reduce bias

⇒ **shrink** data to reduce bias

CAN WE GET MORE?

Unbalanced training set could affect classification

⇒ **inflate** data to reduce bias

⇒ **shrink** data to reduce bias

WORSE RESULTS

CAN WE GET MORE?

Data Set
could be
too small

CAN WE GET MORE?

Data Set could be too small

⇒ **inflate** data

CAN WE GET MORE?

Data Set could be too small

⇒ **inflate** data

WORSE RESULTS

FINAL ACCURACY

about **70%**
of accuracy
in general

CLUSTERED RESULTS

no outstanding
performances of
a single classifier

LIMITED ERROR SPREADING

93% of instances **classified correctly or** in the **adjacent classes**

(C4.5 classifier)

ANALYSIS

12kbps
minimum
bandwidth
to start a call

ANALYSIS

up to 800ms
of RTT
do not
affect quality

FUTURE WORKS

Enrich our data set with
crowd based feedbacks
(Android application under development)

Currently applying an
extended methodology to
the video streaming domain

FUTURE WORKS

Deploy ACQUA framework

THANK YOU

Title

From network-level measurements to expected Quality of Experience: the Skype use case

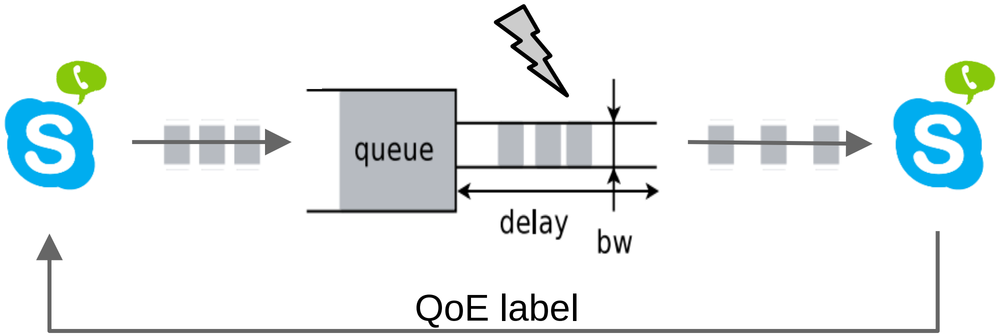
Authors

Thierry Spetebroot (INRIA Sophia Antipolis, France)
Salim Afra (University of Calgary, Canada)
Nicolás Aguilera (NIC Chile Research Labs, Chile)
Damien Saucez (INRIA Sophia Antipolis, France)
Chadi Barakat (INRIA Sophia Antipolis, France)

informatics mathematics

BACKUP SLIDES

DATA SET



Measures obtained in a **controlled environment** collecting Skype's application feedback given

- latency
- bandwidth
- packet loss

Image from: Marta Carbone, Luigi Rizzo, An emulation tool for PlanetLab, Computer Communications, Elsevier, Oct.2011, doi:10.1016/j.comcom.2011.06.004

DATA SET

Composed by 6 input metrics +
1 output QoE label

- latency
- bandwidth
- packet loss rate

$\}$ upload and download

- QoE label $\in$ `{NoCall, Poor, Medium, Good}`

MEASUREMENT FEASIBILITY

In real world is difficult to measure

- One Way Delays
  clock synchronization is required
- Link Capacity
  due to link level losses

DATA SET PREPARATION

Adaptation of

- One Way Delays → Round Trip Time (RTT)
  merging both delays
- Link Capacity → Passing Throughput
  convolution of bandwidth with packet loss rate

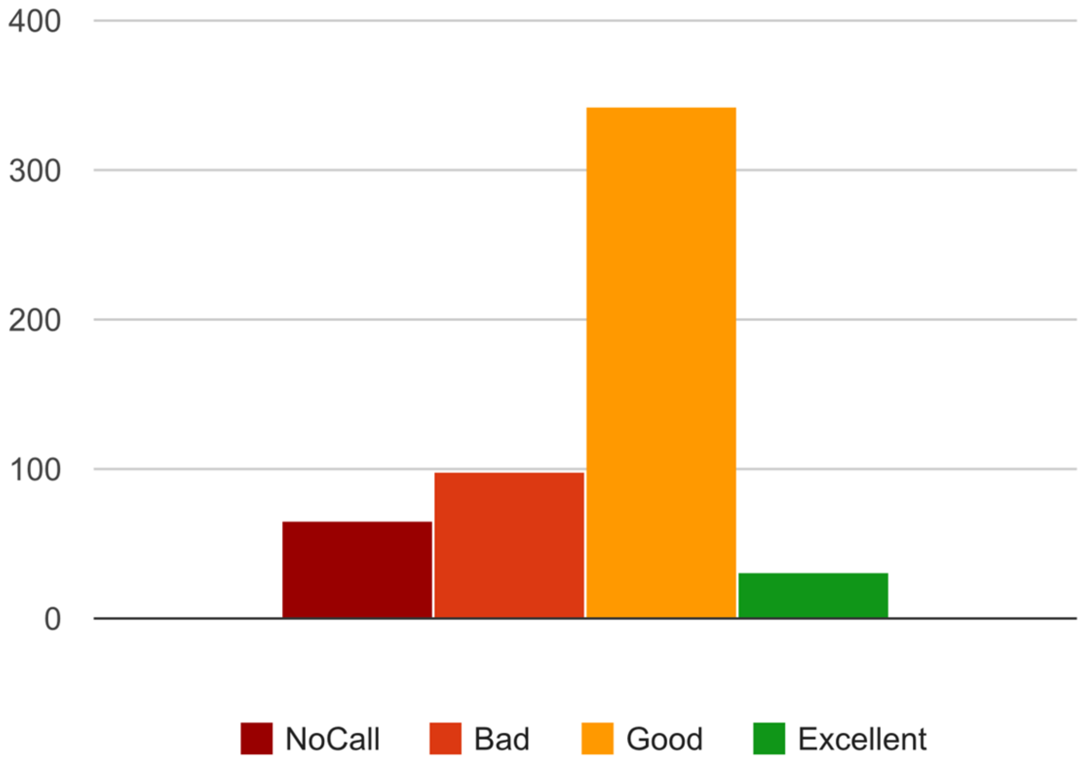
FINAL DATA SET

Composed by 5 input metrics +
1 output QoE label

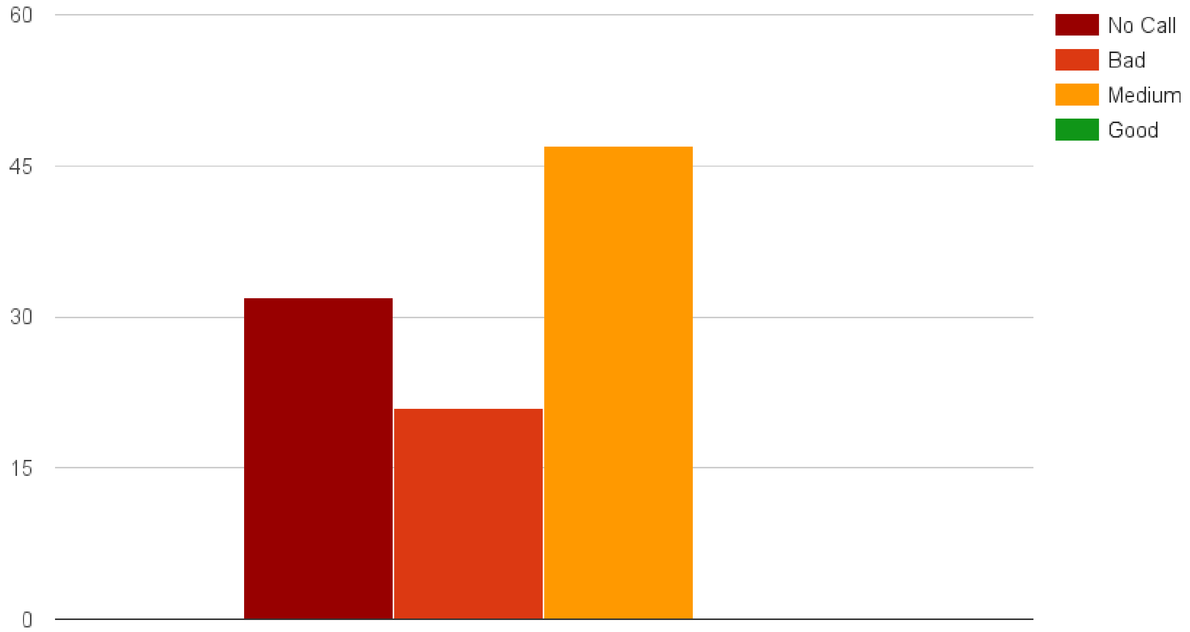
- Round Trip Time (RTT)
- passing throughput
- packet loss rate

- QoE label

TRAINING SET



TEST SET



100 random instances

RESULTS - Cross Validation

Recall 60%

Poor results in 3 cases...

NoCall

Poor

Good

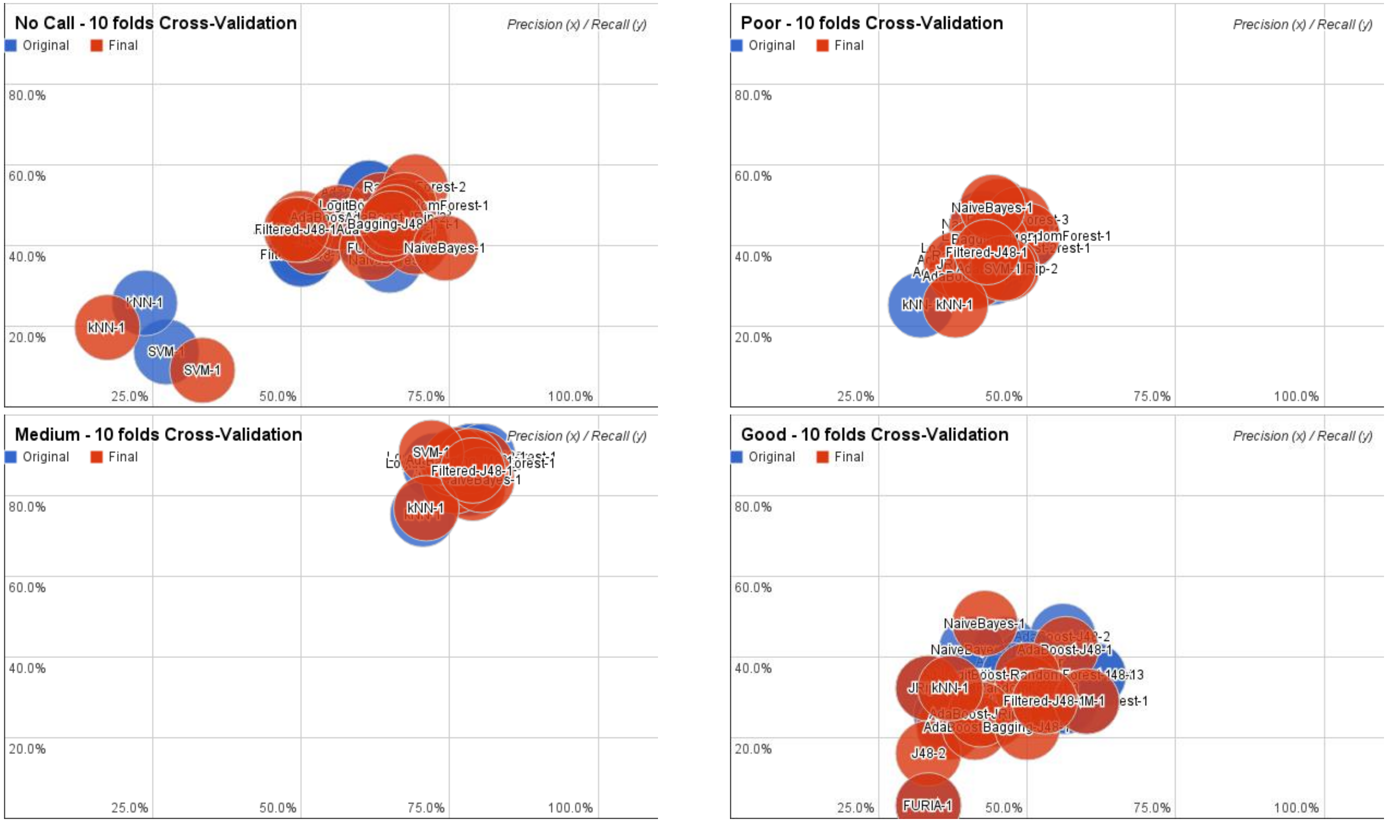
RESULTS - Cross Validation

WHY?

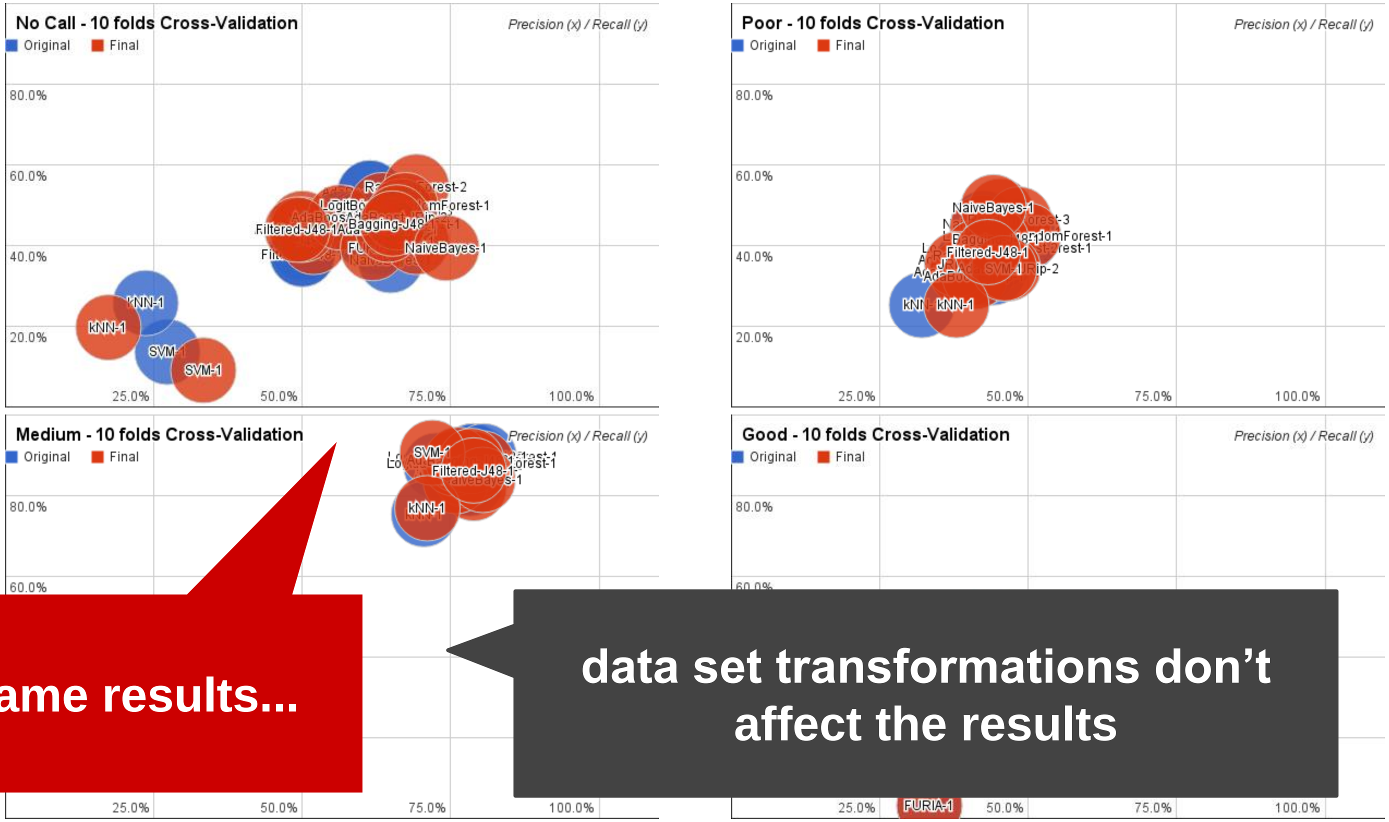
It could be due to the Data Set preparation process...

do we have information loss?

INFORMATION LOSS



INFORMATION LOSS



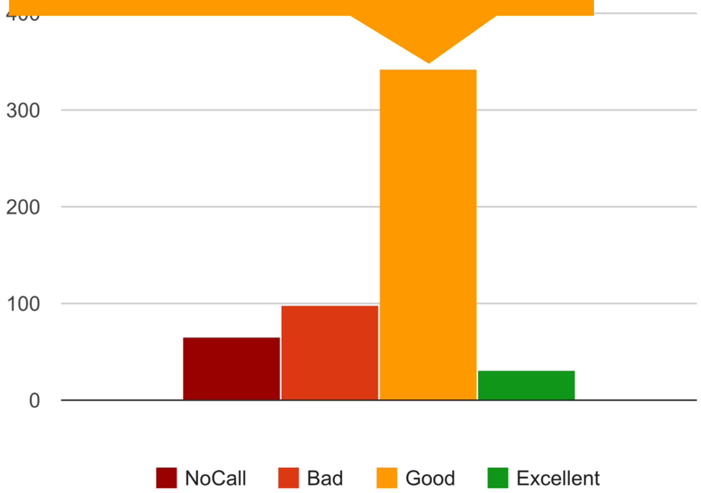
Same results...

data set transformations don't affect the results

WHY?

It could be a **Class Imbalance Problem**

Where we have better results we have considerably more data

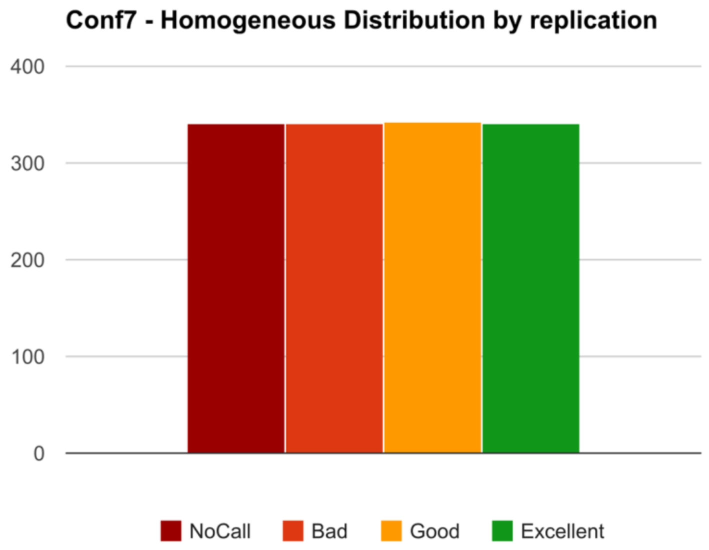


CLASS IMBALANCE PROBLEM

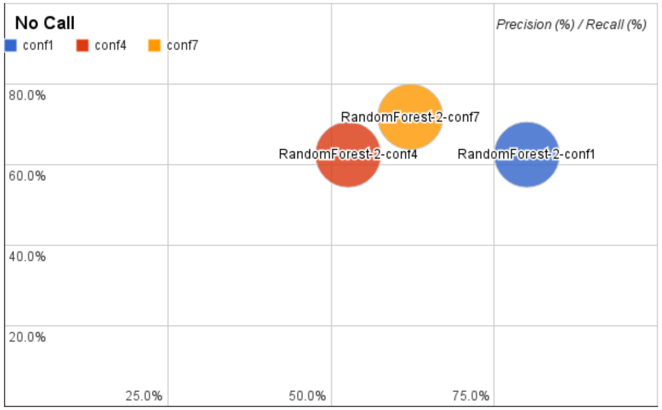
Unbalanced training set could affect classification

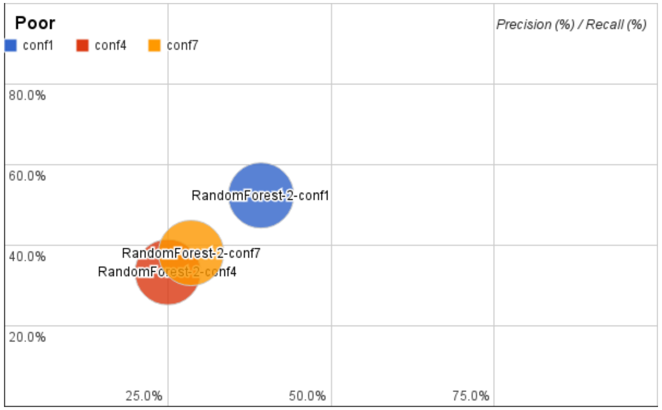
⇒ **inflate** data to reduce bias

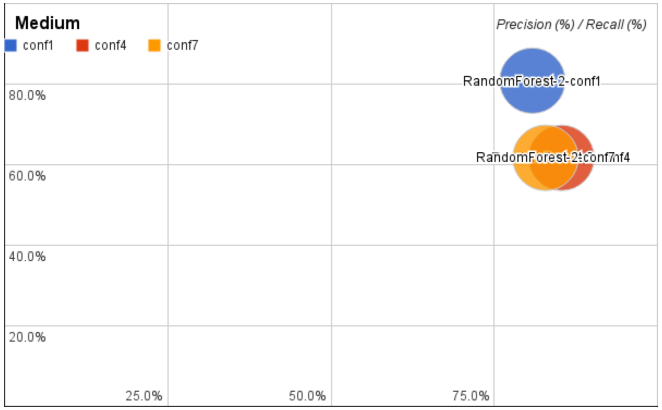
- replication
- interpolation



CLASS IMBALANCE PROBLEM



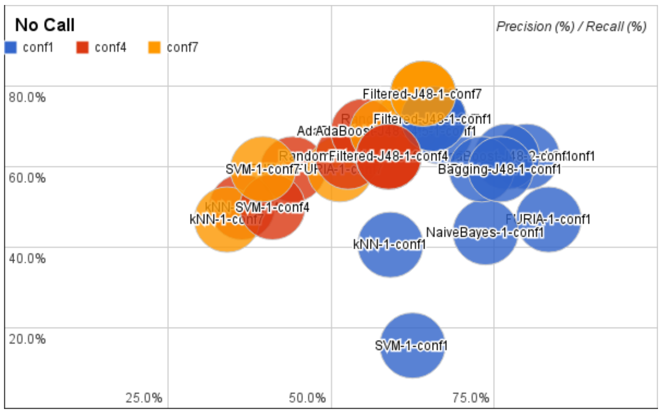


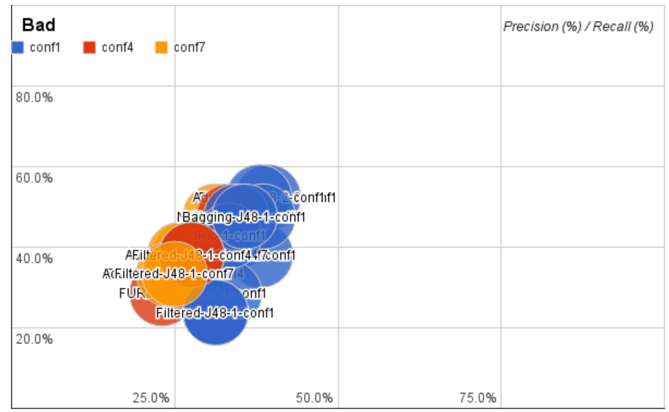


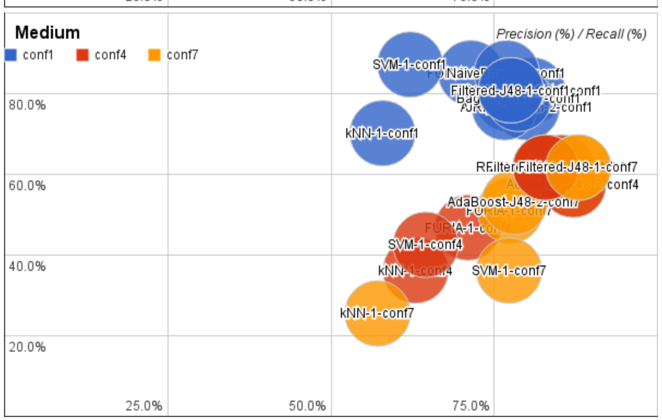
Worse results...

see all algorithms

CLASS IMBALANCE PROBLEM







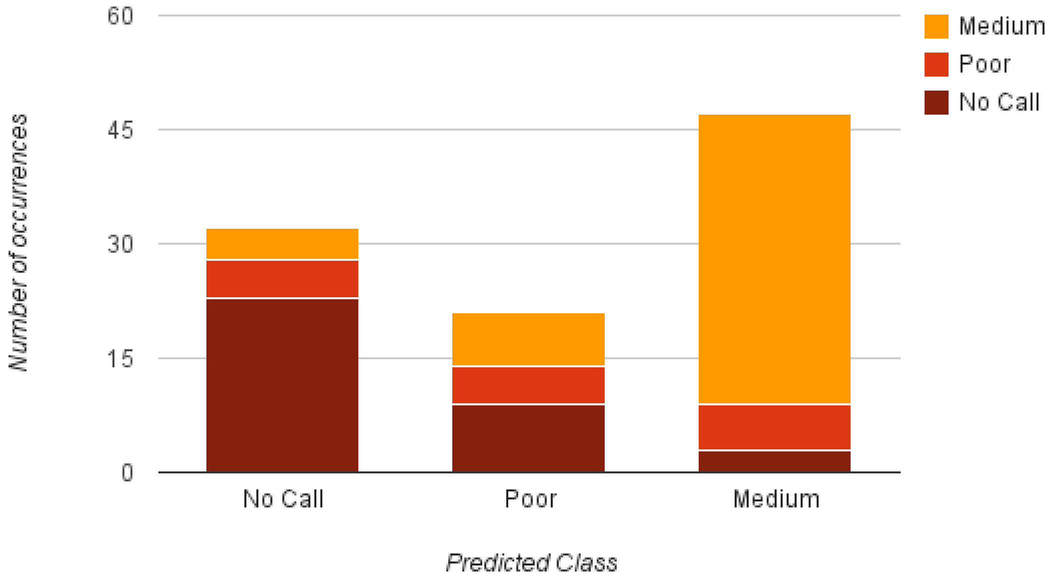
Worse results...

back

CONCLUSIONS

Limited Error Spreading

93% of instances classified correctly or in the adjacent class



(C4.5 classifier)

X

DRAFTS