Forest Grove
Data-Driven Decision Making

Optimising technology to unlock insights that drive better business decisions

Forest Grove Customer Conference 2023

www.forestgt.com.au | info@forestgt.com.au | 1300 559 827

Forest Grove
Data-Driven Decision Making

Optimising technology to unlock insights that drive better business decisions

[D2S7] Applying Predictive Modelling and Machine Learning

Presented by :
Jérôme Treboux, International Data Science Expert

About the session:
Several industries are increasingly adopting machine learning (ML) and predictive modelling to solve business problems and gain operational efficiencies. Many specialized platforms help implement ML and predictive modelling reducing technical constraints. For example, the KNIME platform allows the creation of visual workflows in an intuitive environment, enabling every stakeholder in the data science process to focus on what they do best. This presentation first introduces ML and predictive modelling. It then provides an overview of our applied projects in collaboration with varying sectors (e.g., marketing, customer insights, medical, and food processing). The projects use KNIME and are presented from their research and development to production deployment. It discusses the opportunities, benefits, and possible pitfalls (e.g., technical) of using ML and predictive modelling for applied projects.

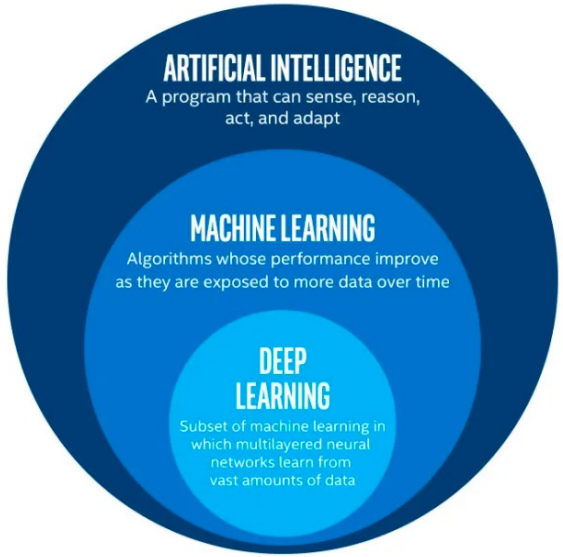
CONTENTS

Introduction

Machine Learning – Definition

“Machine learning (ML) teaches computers to do what comes naturally to humans: learn from experience” [1]

- Arthur Samuel (1959) [2]
- Algorithms “learn” information from data and improve their performance
- Training an algorithms result in a ML model
- Daily usage: deepl, chat GPT, Siri, …



[1] Peck W.G., *Machine Learning Techniques Using MATLAB*, North Charleston, SC: CreateSpace Independent Publishing Platform, 2017.
[2] SamuelA.L.,“SomestudiesinmachinelearningusingthegameofCheckers”,*IBMJournal of Research and Development*, 1959, 3(3): pp. 210–229, https://doi.org/10.1147/rd.33. 0210.
Image source: https://ai.stackexchange.com/questions/15859/is-machine-learning-required-for-deep-learning

Introduction

Predictive Modelling – Definition

“Predictive modelling is commonly used statistical technique to predict and forecast likely future behavior and outcomes using current and historical data” [1]

- Using Machine Learning
- Daily usage : email, weather forecast, finance forecast, …

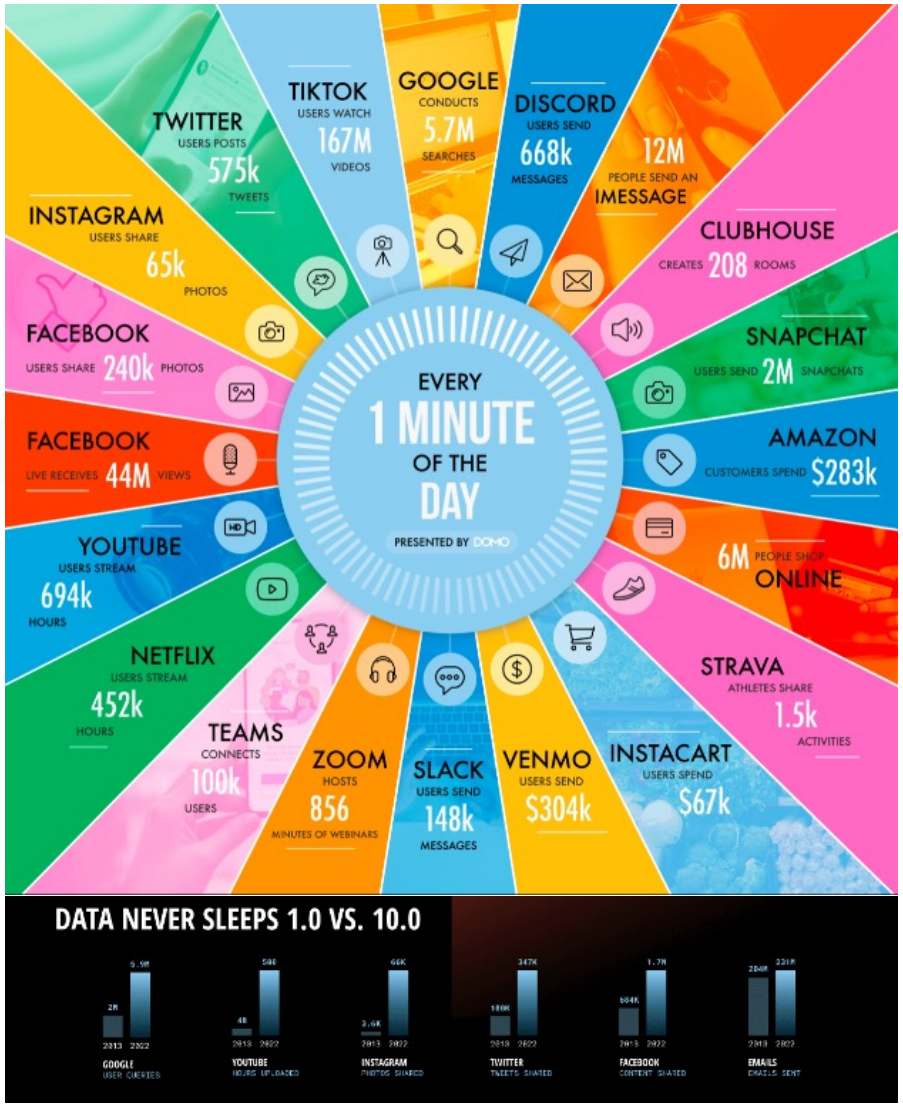
PREDICTIVE MODELING

[1] Ali R.,Predictive Modeling: Types, Benefits, and Algorithms, Oracle, September 2020, retrieved on https://www.netsuite.com/portal/resource/articles/financial-management/predictive-modeling.shtml

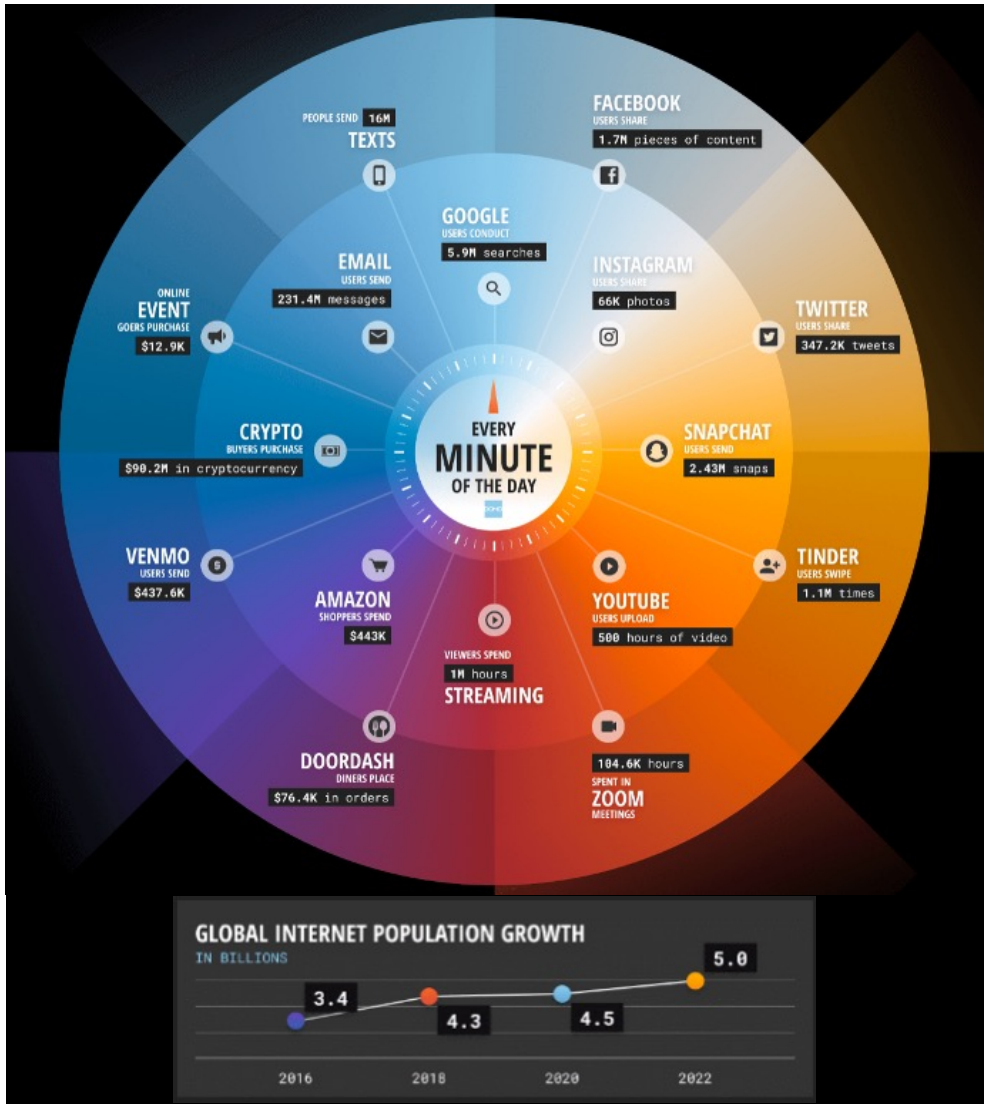
Introduction

Big Data

2021



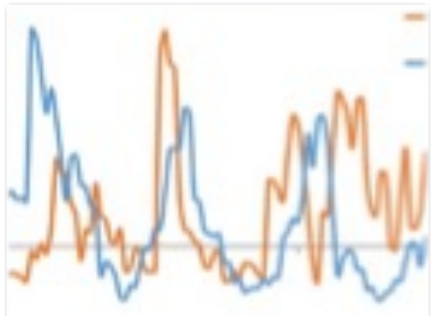
2022

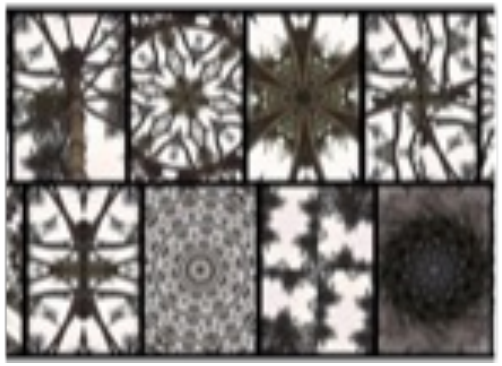


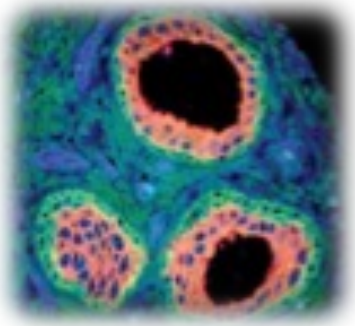
Source: https://www.domo.com/data-never-sleeps

Introduction

Machine Learning - Finding Patterns

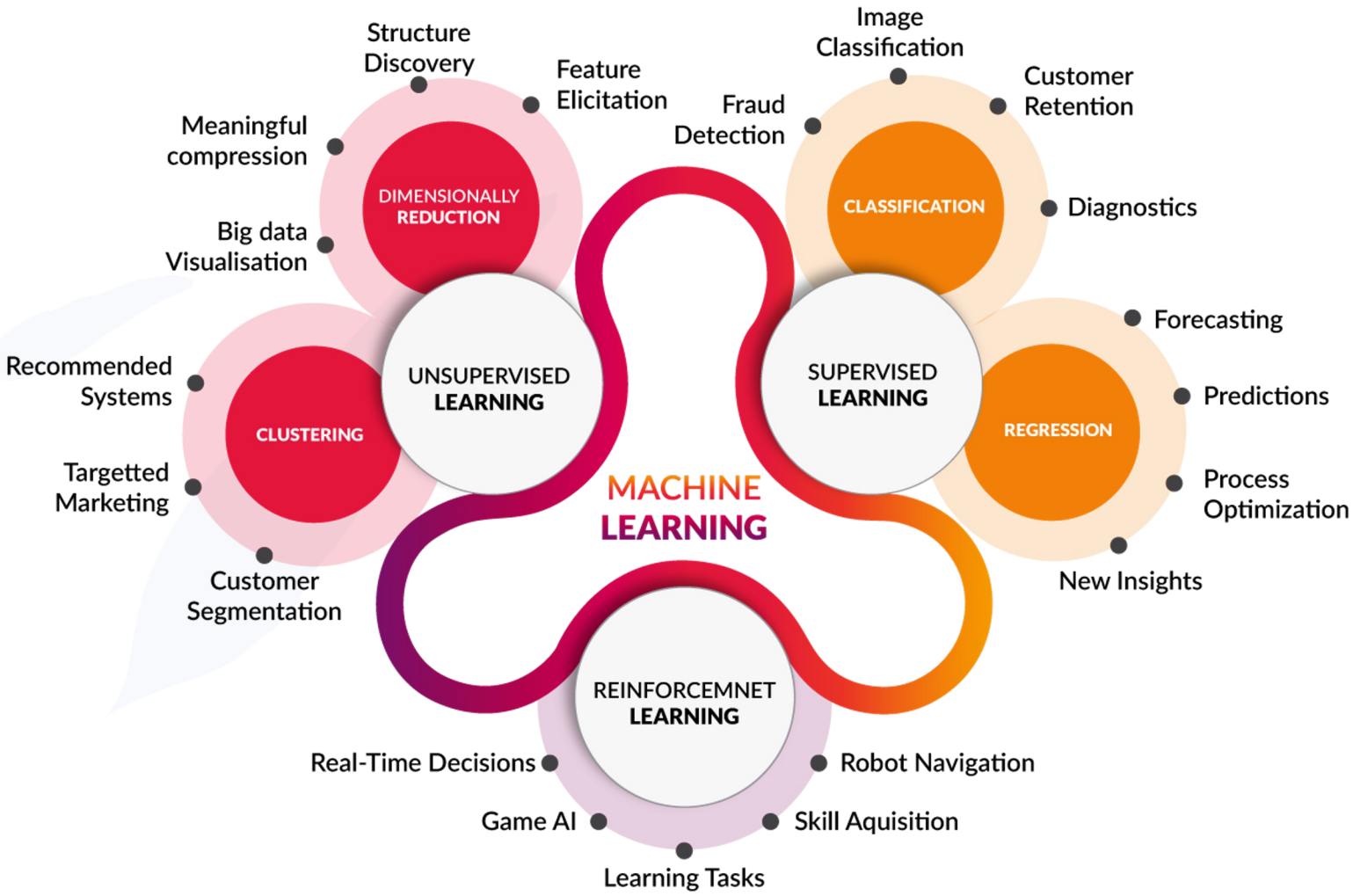






Introduction

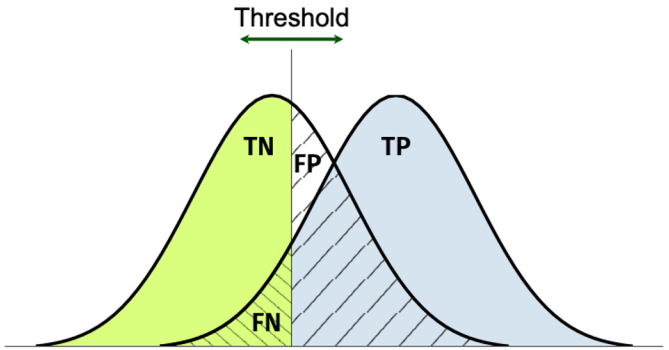
Machine Learning Approaches

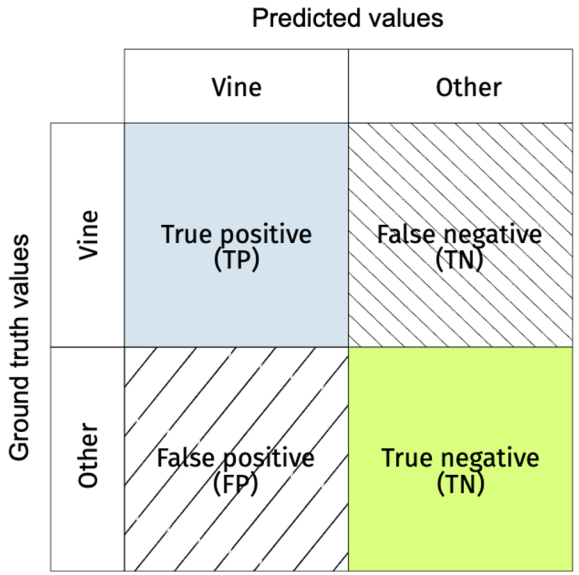


Introduction

Evaluation

- **Precision**: the proportion of objects correctly identified as relevant compared to all objects identified as relevant.
- **Recall**: determines the proportion of objects correctly identified as relevant compared to the total relevant objects of the ground truth.





$$Precision = \frac{TP}{TP + FP}$$

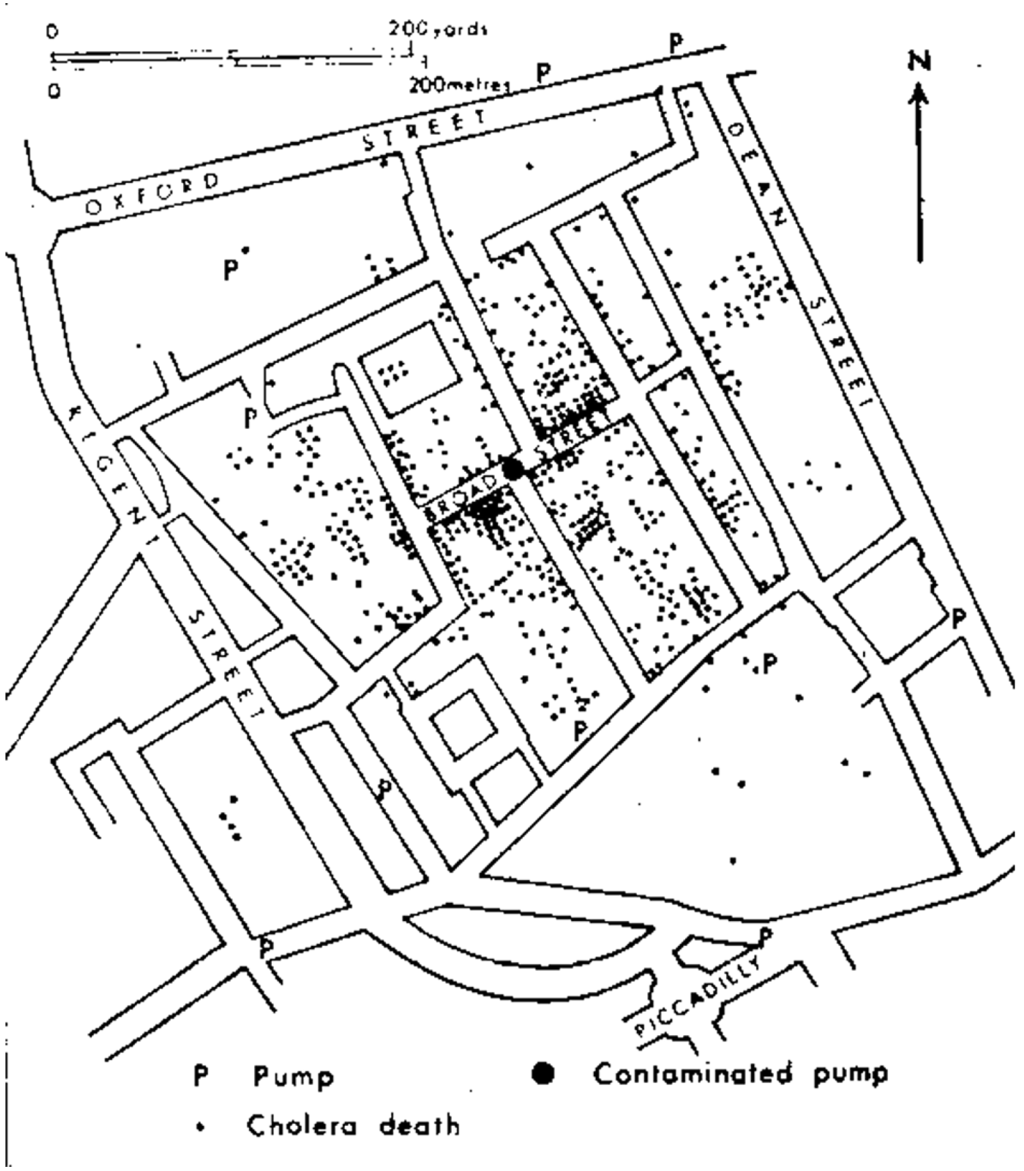
$$Recall = \frac{TP}{TP + FN}$$

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall}$$

Introduction

Visualization – Public Health

- Inform
- Analyse
- Take decision

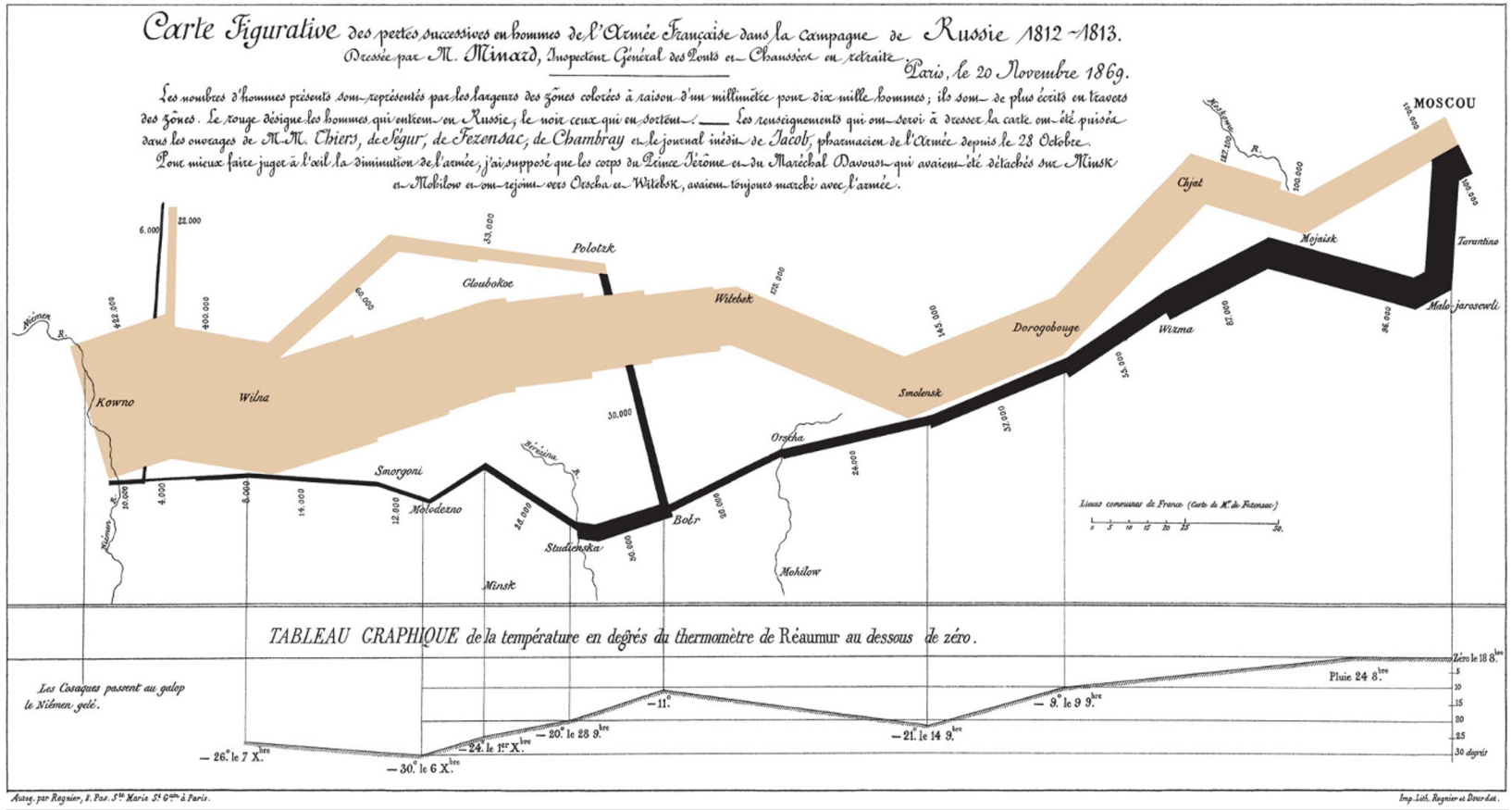


Source: https://en.wikipedia.org/wiki/John_Snow_/_1854

Introduction

Visualization – History

- Explain
- Visualize
- Tell a story understandable by everyone

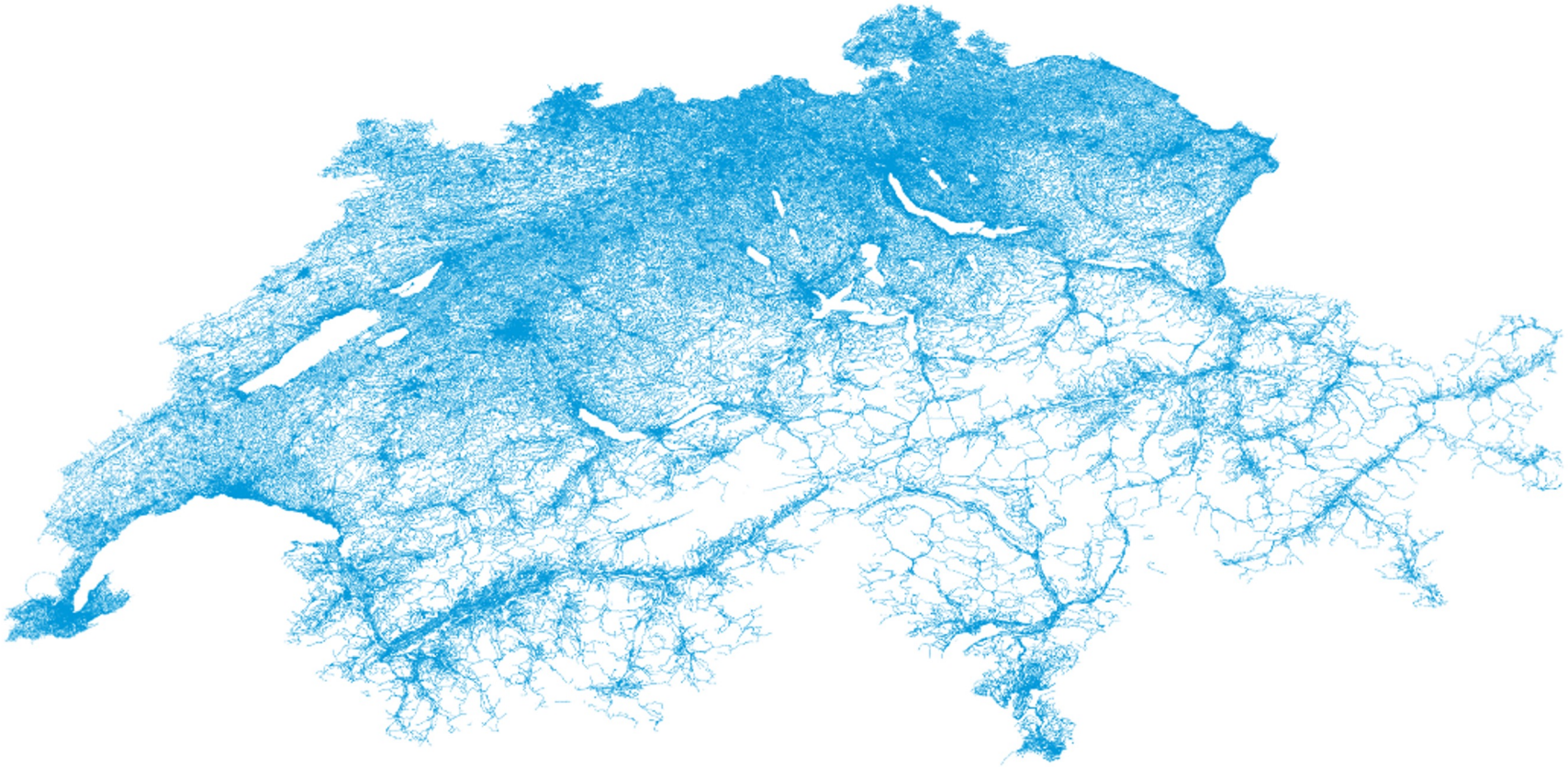


Source: https://www.edwardtufte.com/tufte/minard

Introduction

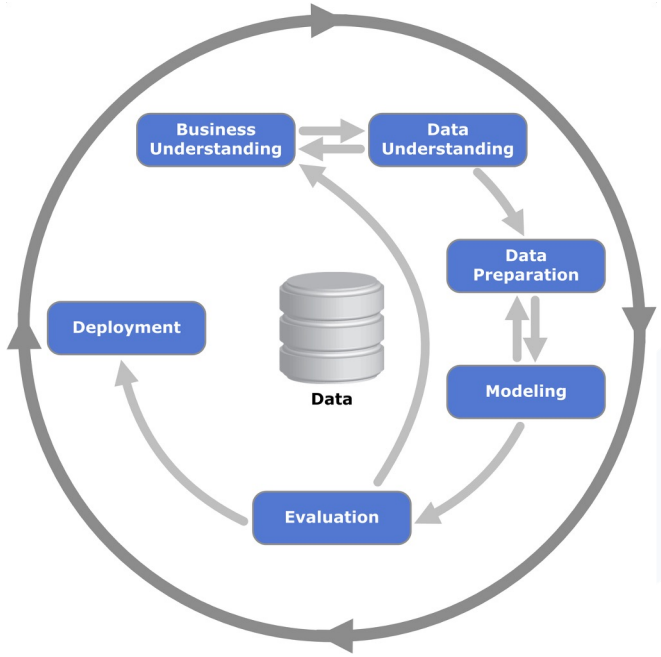
Visualization – Geography

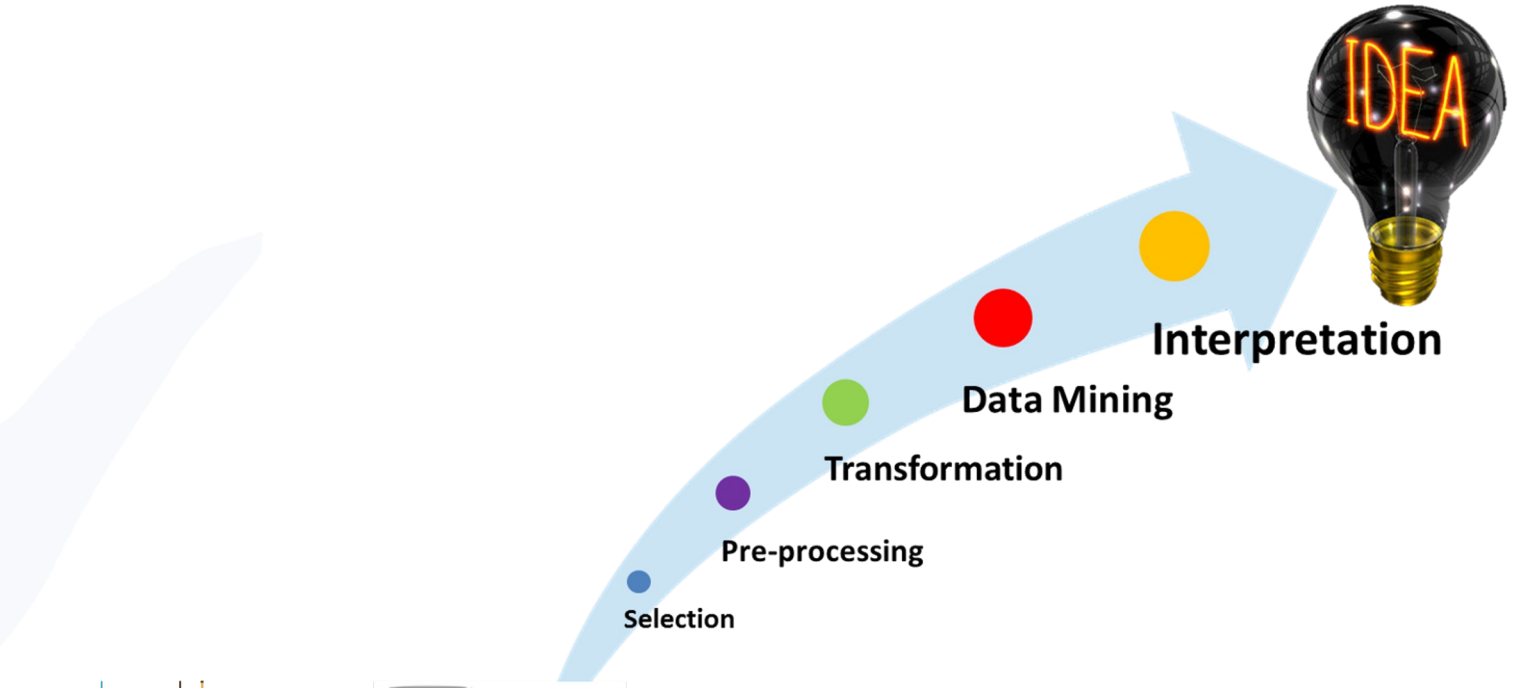
- Analyse
- Understand
- Obtain an overview

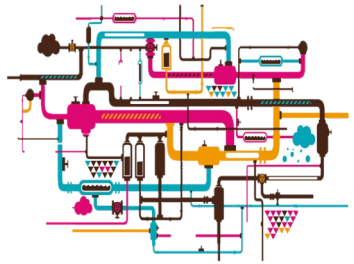


Introduction

Machine Learning Methods







Inspired from www.prolog.org

Introduction

Machine Learning Platforms (examples)

- Integrated suites
  - SAS, IBM, Dataiku, H2O
- Graphical Programming
  - Knime, RapidMiner, Sisense, Alteryx
- Libraries
  - Python, java, c++
- Big data/ cloud based
  - Azure, Amazon AWS, SAP
- Deep learning approach
  - Tensorflow (includes more than 1300 pre-trained models), Torch, Keras

Vine Line Recognition for Flight Path Automation

Food production must almost double in the next 30 years (FAO, 2018)

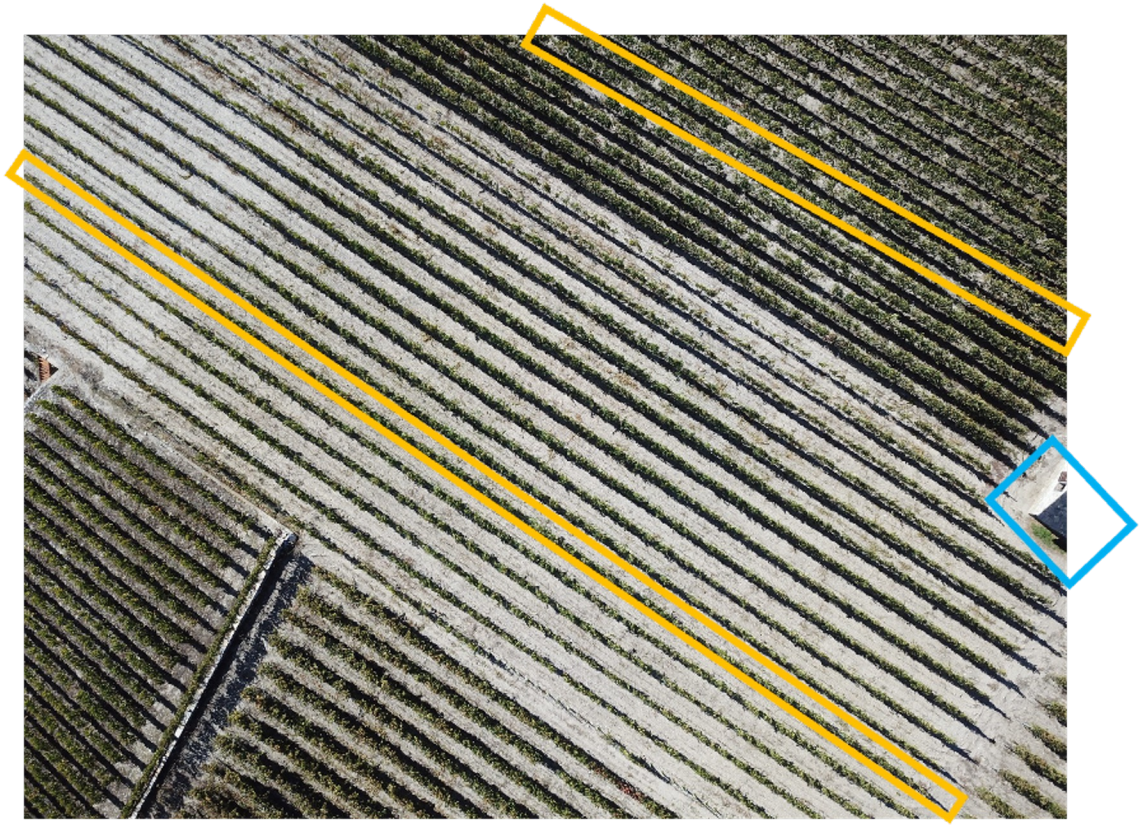
- Population growth
- Excessive demands on natural resources
- Declining productivity due to climate change
- Food wasting

One of the solution: **precision agriculture**

- Technologies integration (e.g., drones, AI)
- Determining location of crops by analyzing data and using image analysis
- Precision agriculture can be applied to all types of agriculture
- Ecological and economical impact

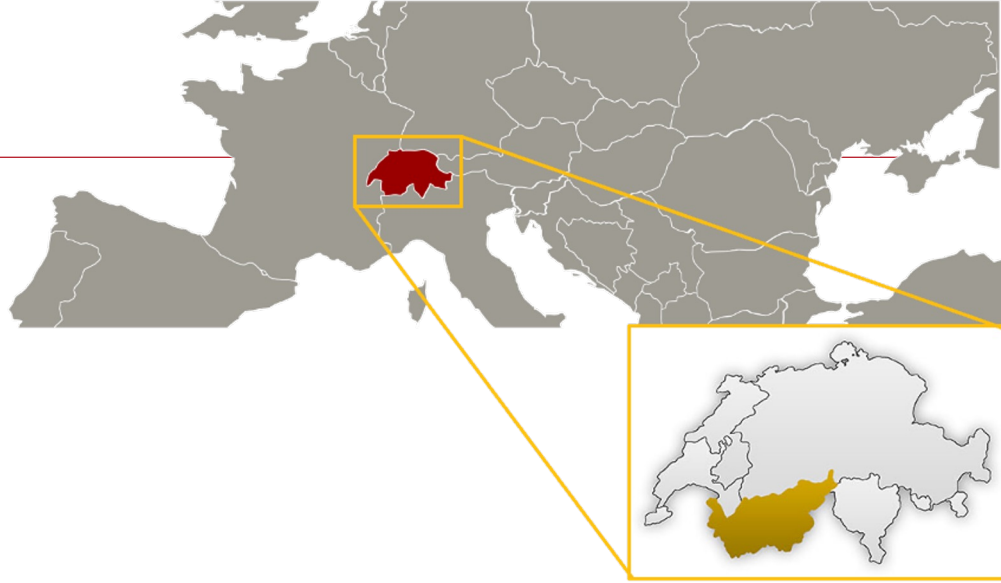
Vine Line Recognition for Flight Path Automation

Problems



- Vigne lines detection : multiple instances of the same object, crossing the image
- Agricultural objects : similar to the vines based on the shape or the color
- Other objects

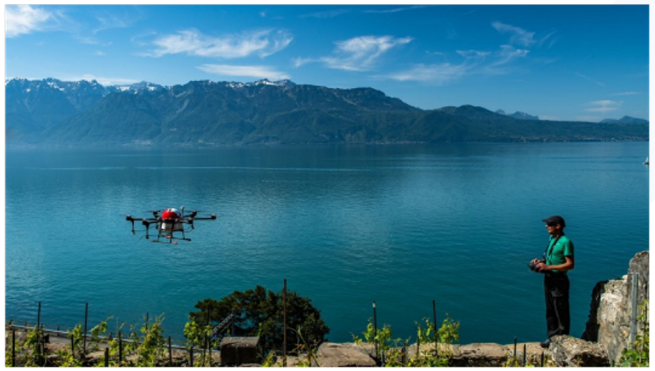
Vine Line Recognition for Flight Path Automation



Dataset

Images sources

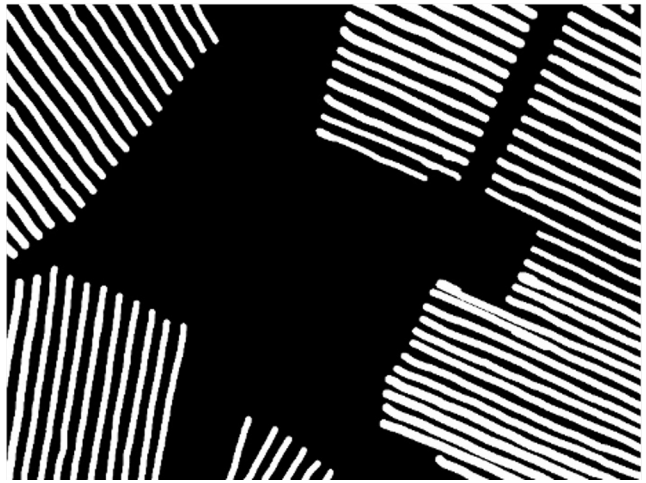
- Drone
- Swisstopo (mapping platform)



Vine Line Recognition for Flight Path Automation

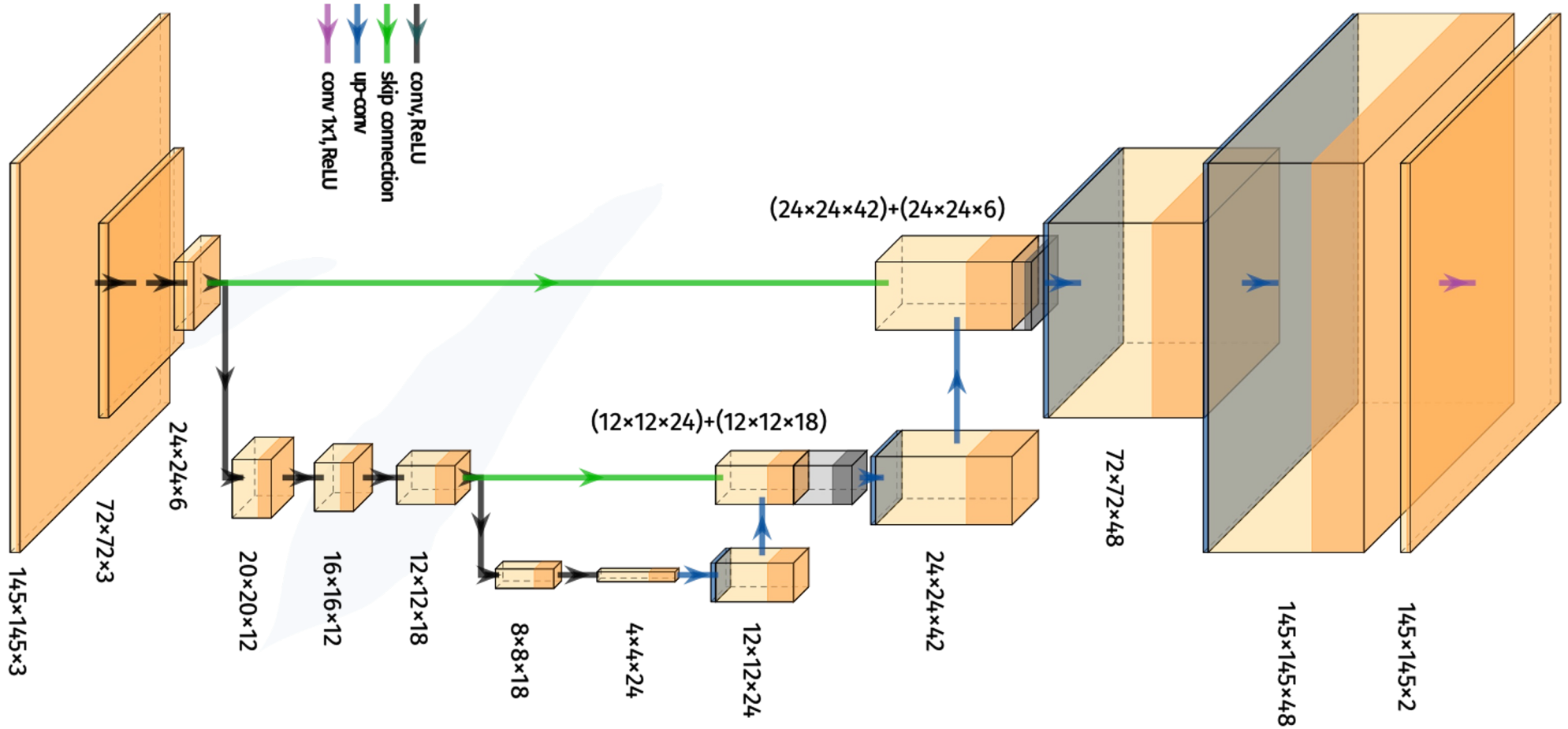
Data preparation

- Pixel-wise annotation of the vine lines
- Annotation is done manually on each image.
- White lines represent vine lines



Vine Line Recognition for Flight Path Automation

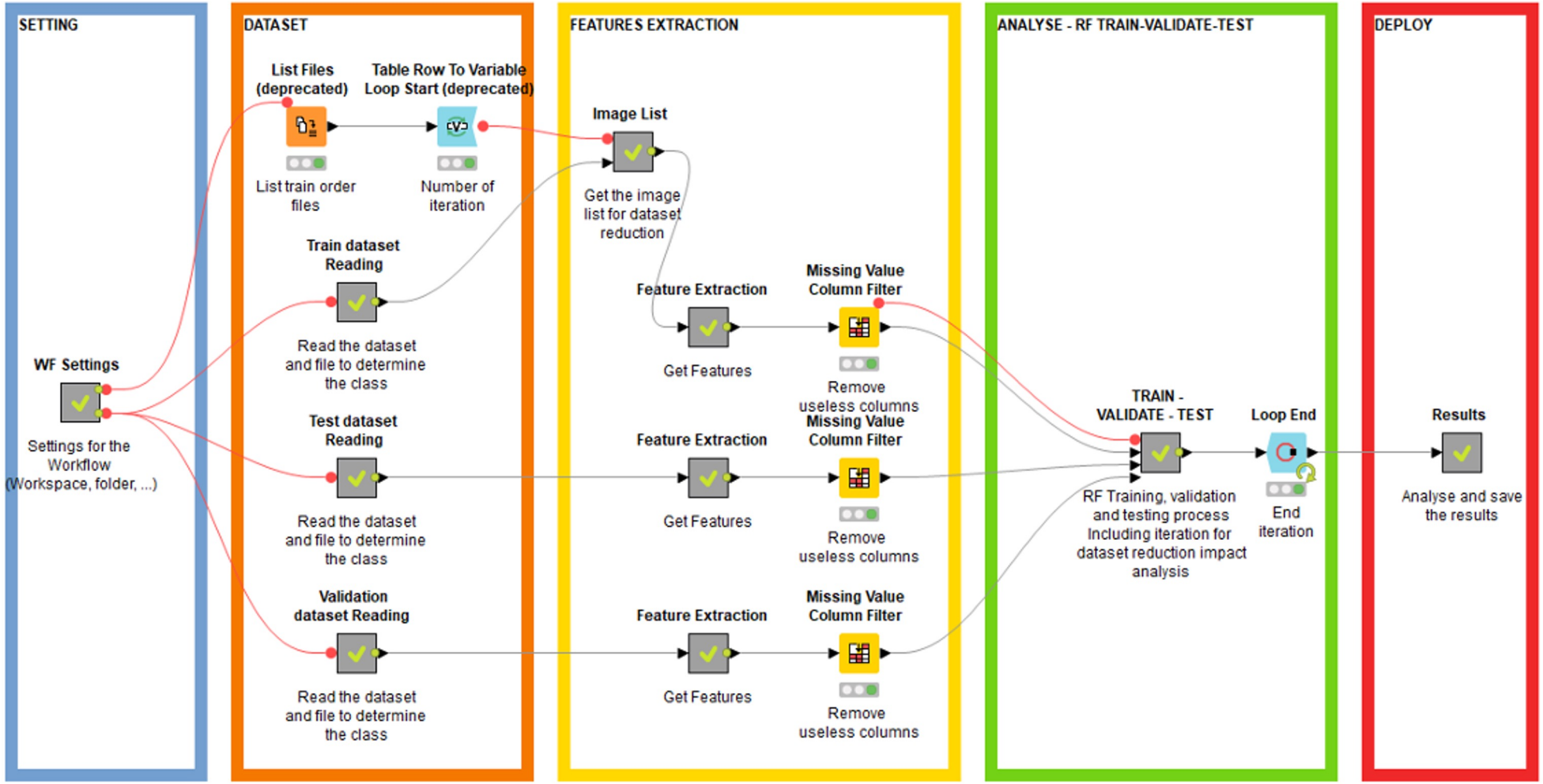
ML Model



Forest Grove
Data-Driven Decision Making

Vine Line Recognition for Flight Path Automation

KNIME Workflow

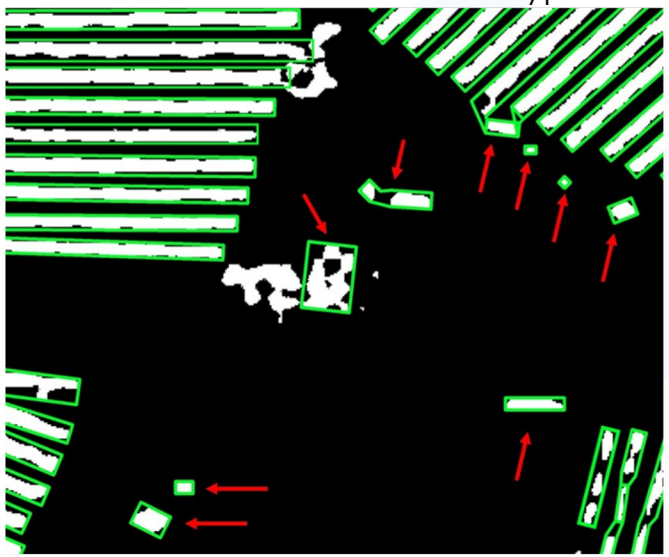


Vine Line Recognition for Flight Path Automation

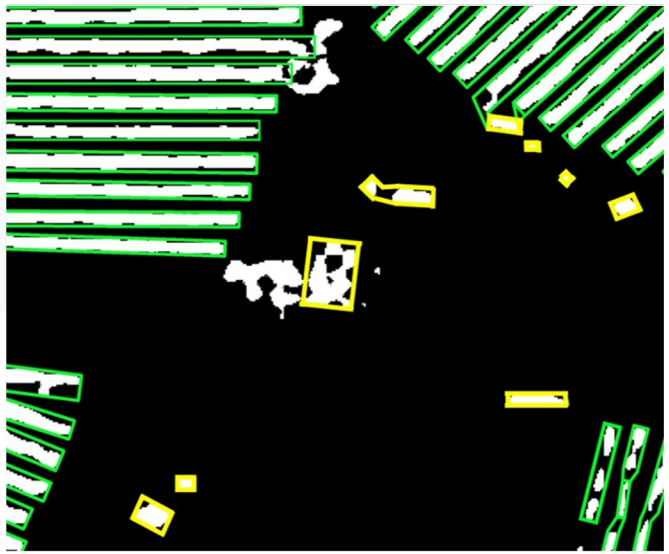
Results

- Significant improvement of the precision and f-score using a RF as vine line classifier
- The RF acts as a filter and removes False Positive (identical recall)
- Performance are reduced because of the seasonal change

- No pilot needed for a drone
- A person can visually check multiple flying drones at the same time
- Reduce the costs and the ecological impact
- Can be trained for other types of agricultures



Before the RF vine classifier



After the RF vine classifier

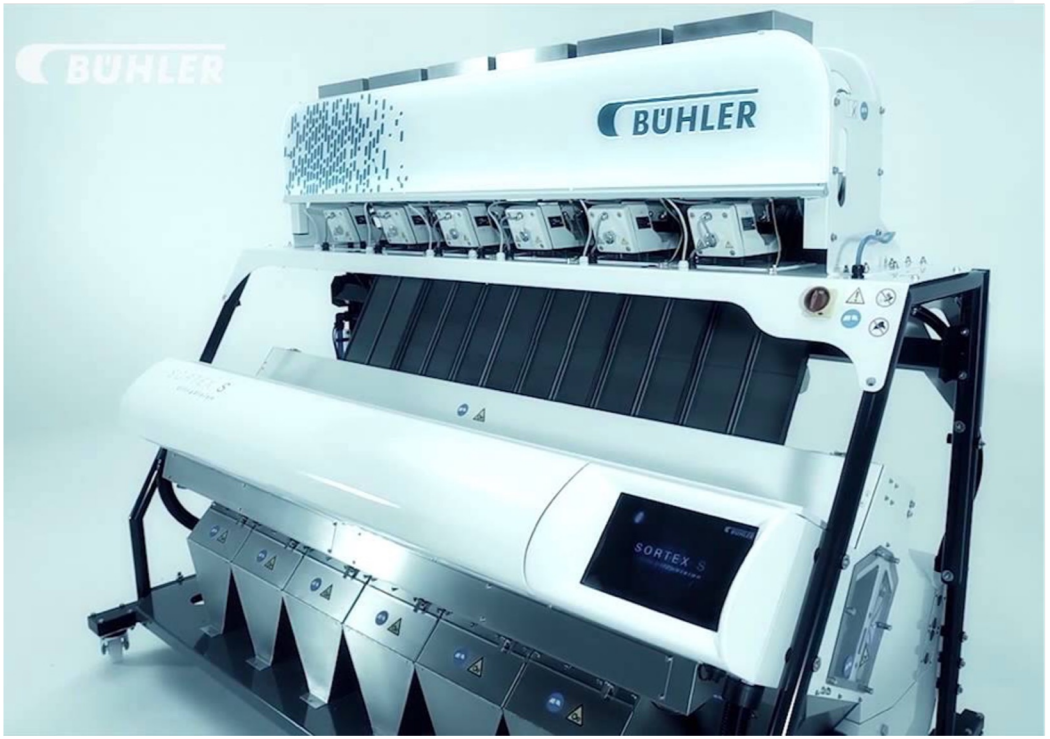
False Positives
Corrected classification with the RF
Aeras detected as vine line

	Precision	Recall	F-score
U-Net	0.30 +-0.03	0.93	0.45 +-0.03
U-Net and RF	0.41 +-0.03	0.93	0.57 +-0.03
Asymmetrical U-Net	0.42 +-0.03	0.95	0.58 +-0.04
Asymmetrical U-Net and RF	**0.53 +-0.04**	0.95	0.68 +-0.03
U-Net (drone test set)	0.82 +-0.01	0.74	0.78 +-0.02
Asymmetrical U-Net (drone test set)	0.83 +-0.01	0.78	0.80 +-0.02

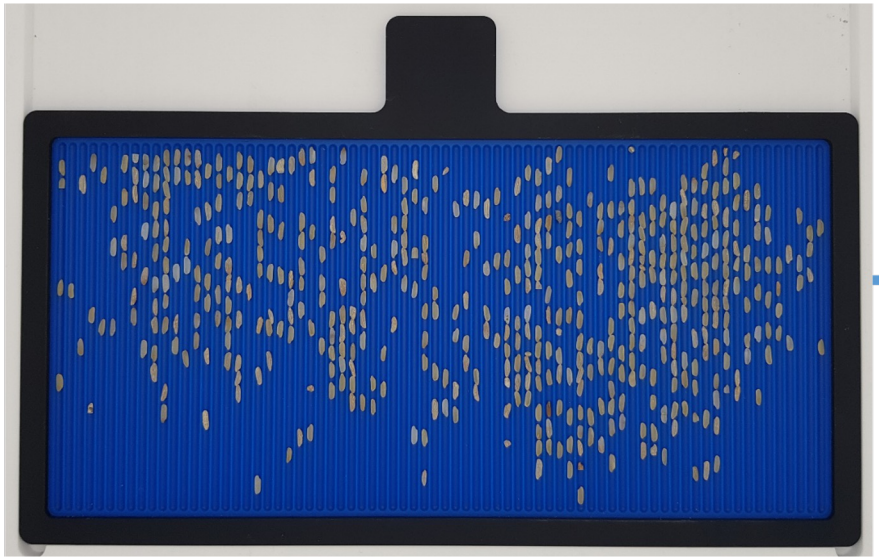
Rice Grain Recognition for Quality Control

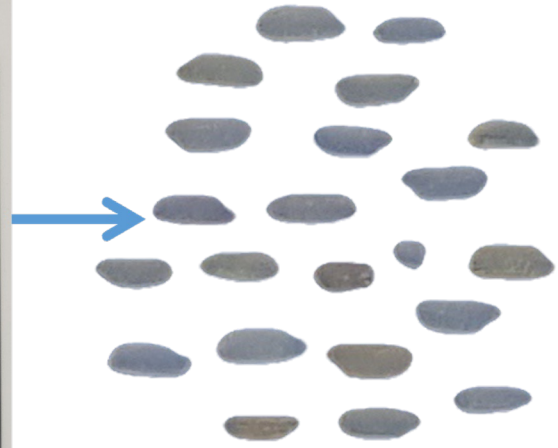
Introduction

- 120’000 varieties of rice grains
- Analysis and sorting of grains up to 20 tons per hour.
- Up to 40% of good grains are saved.
- Rice provides half of the world's food needs



Rice Grain Recognition for Quality Control





Rice Grain Recognition for Quality Control

Dataset

- Manual labelling
- 3 standards
- 3 types of cameras

International Organization for Standardization

USDA

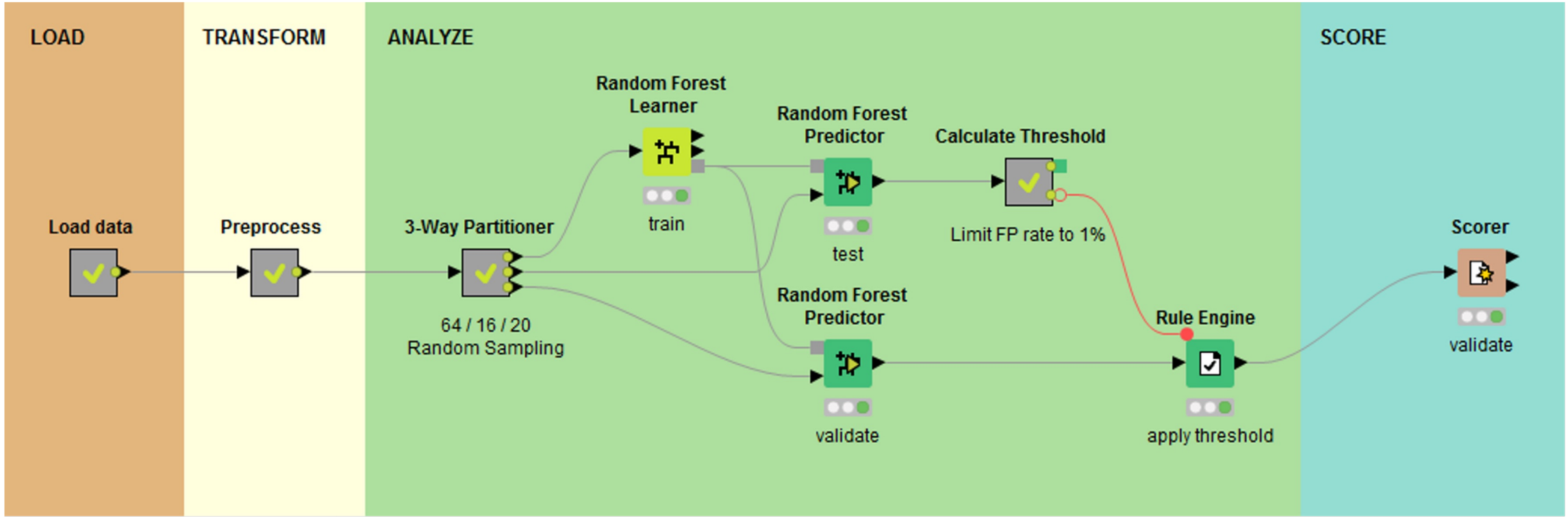
BÜHLER

```
$Sample$ = 1 => "sample"
$Sample$ = 2 => "sortex rejects"
$Sample$ = 3 => "damage (other)"
$Sample$ = 4 => "red"
$Sample$ = 5 => "damage (pecky"
$Sample$ = 6 => "damage (other)"
$Sample$ = 7 => "damage (other)"
$Sample$ = 8 => "heat damage"
$Sample$ = 9 => "damage (stain)"
$Sample$ >= 10 AND $Sample$ <= 12 => "good"
```

	samplename	class
S8_147.png	good	accept
S8_316.png	good	accept
S8_15.png	damage (stain)	reject
S8_183.png	heat damage	reject
S8_3.png	damage (other)	reject
S8_75.png	good	accept
_1.png	heat damage	reject
S8_68.png	good	accept

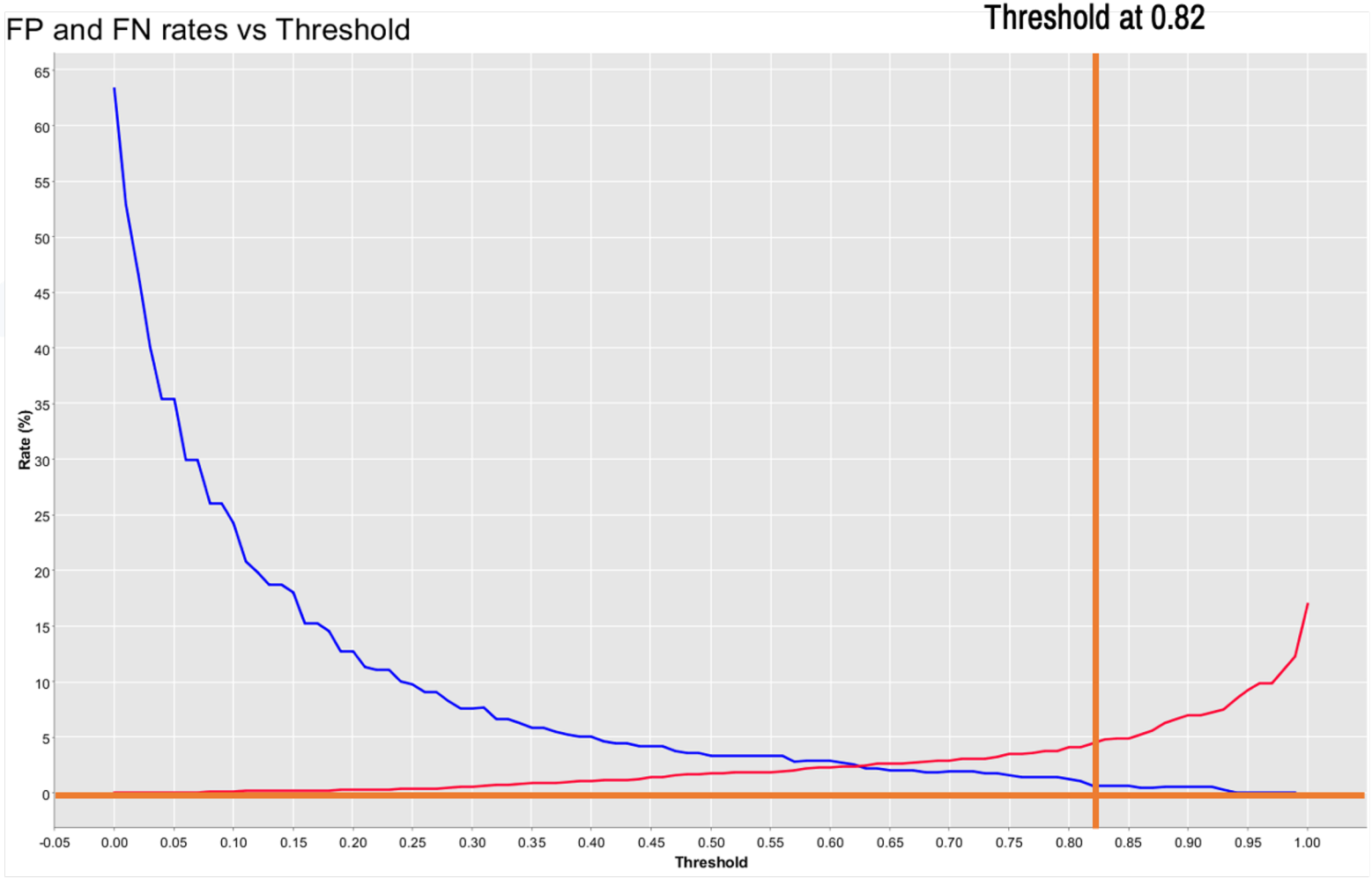
Rice Grain Recognition for Quality Control

KNIME Workflow



Rice Grain Recognition for Quality Control

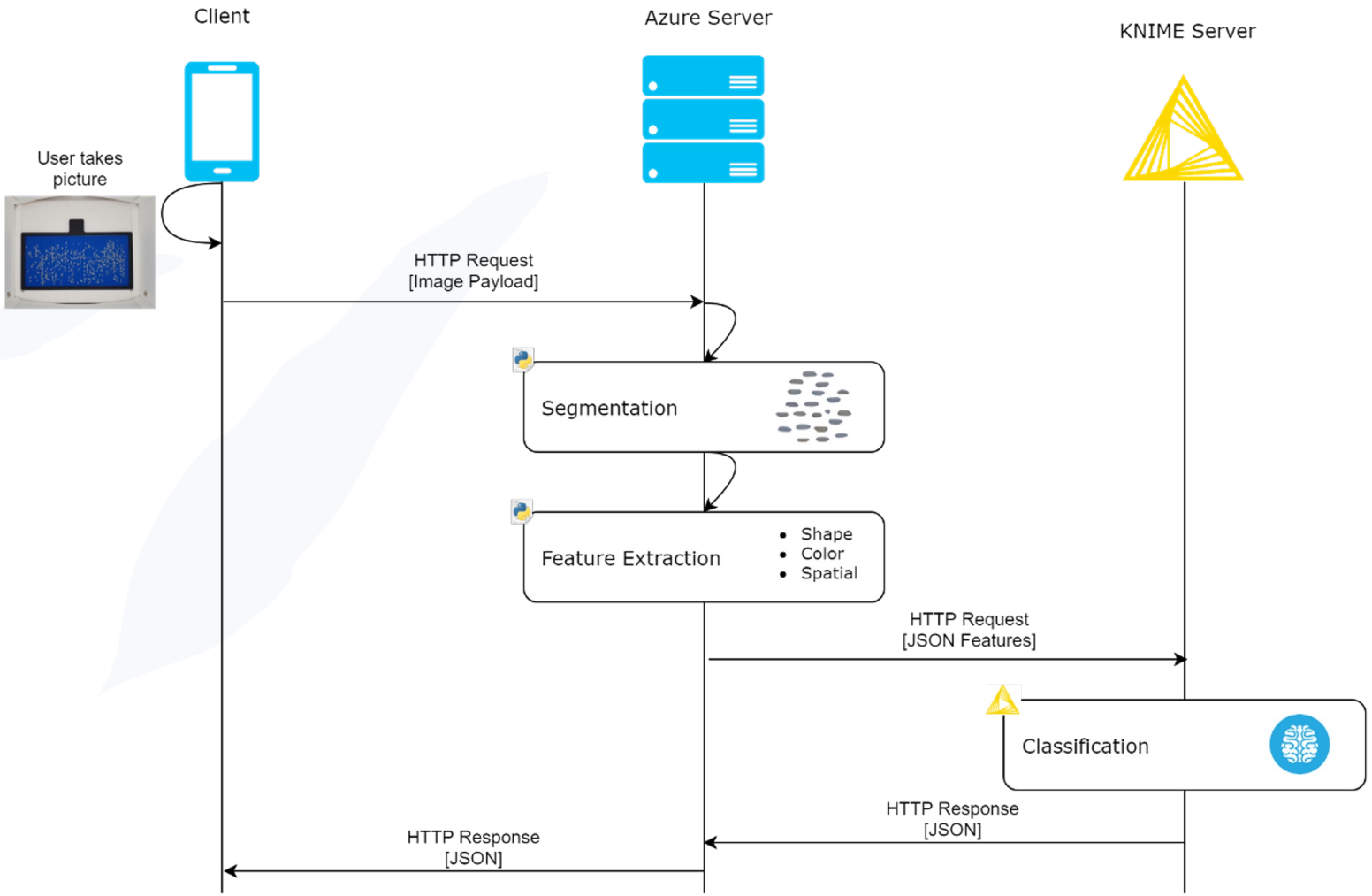
ML Model Training



FP rate at 1%

Rice Grain Recognition for Quality Control

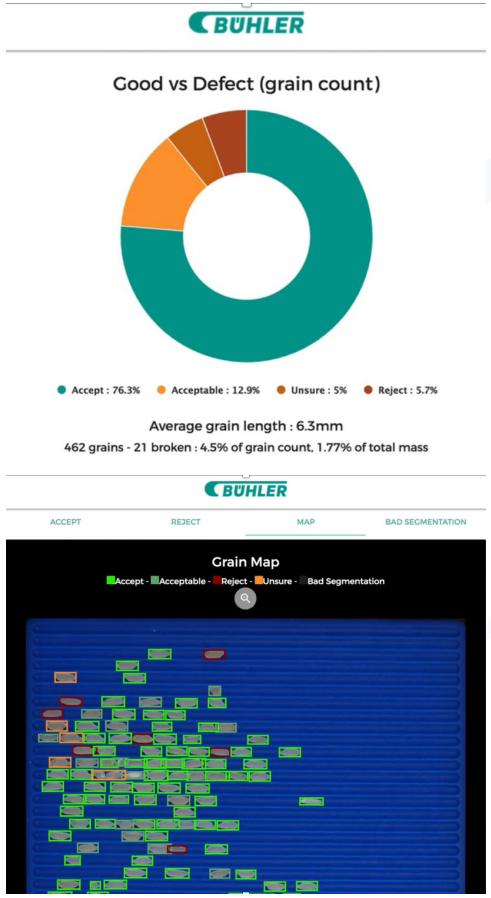
Production deployment using KNIME Server

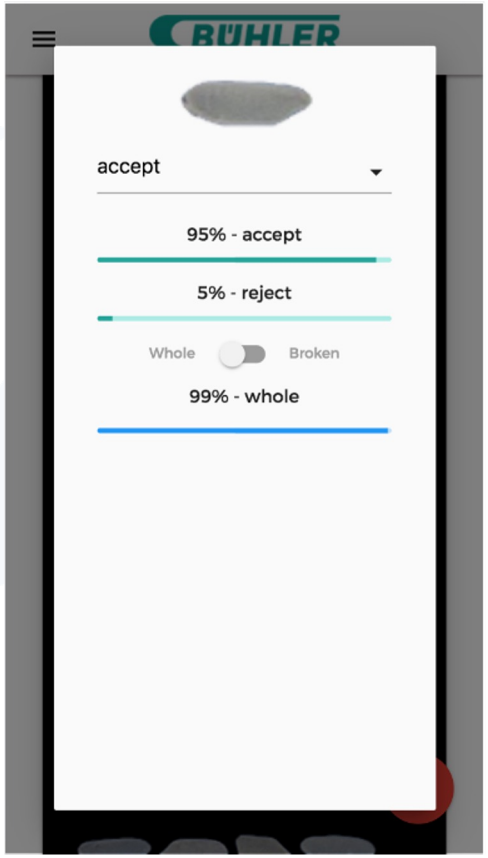


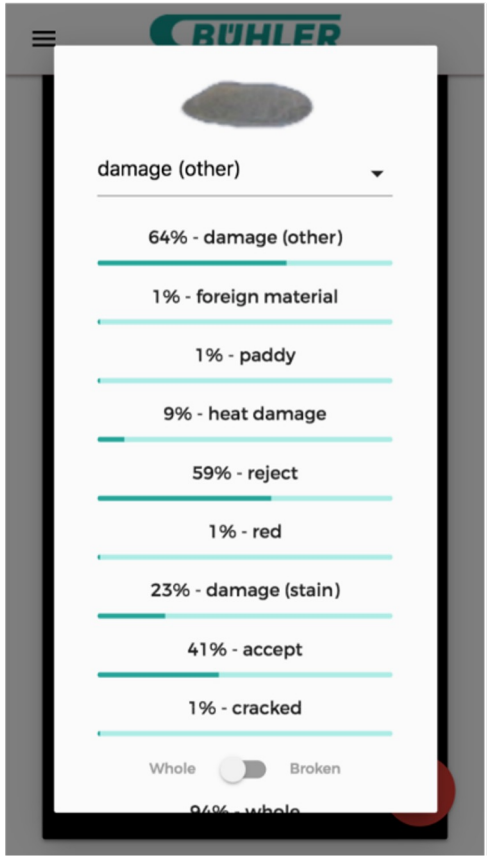
Rice Grain Recognition for Quality Control

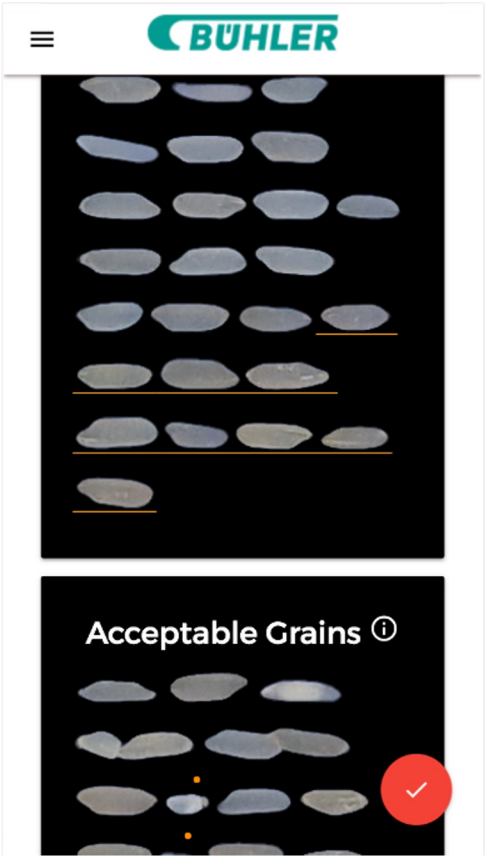
Results

- Reduce the time to control the quality
- Ensure a standard for quality control
- Retrain the ML model with new rice grain





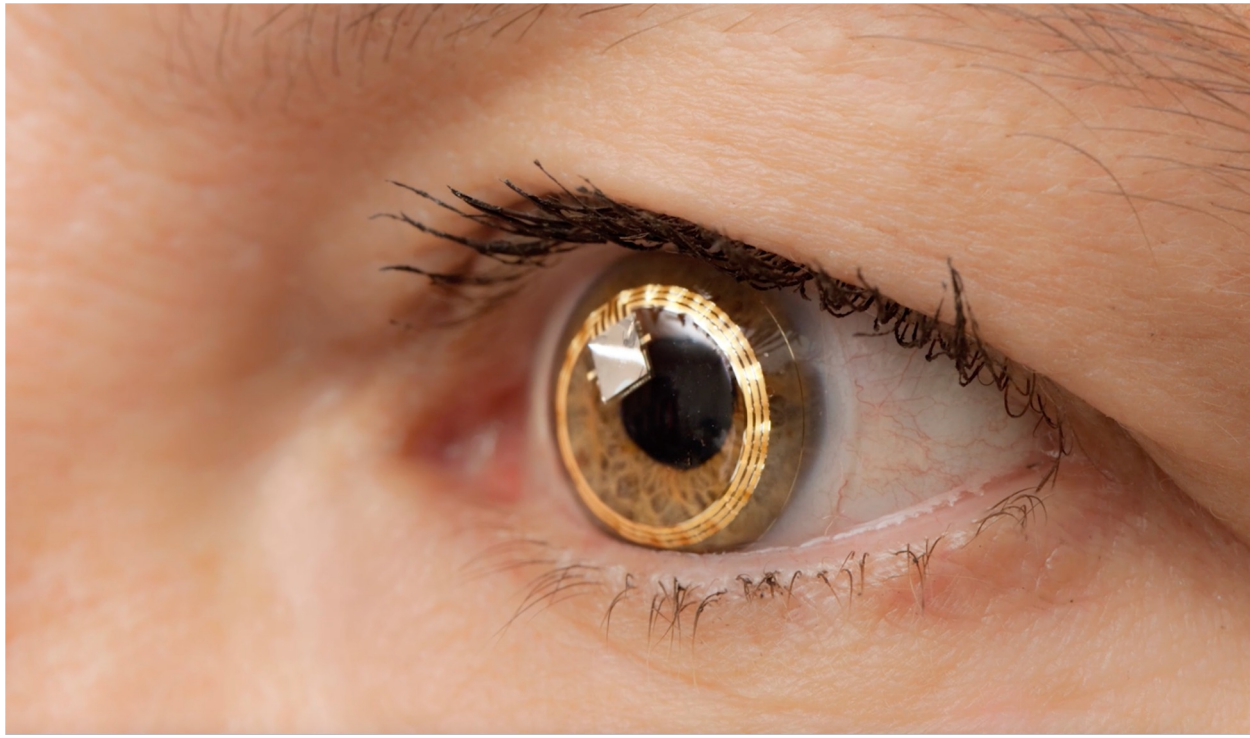




Predictive Modelling for Medical Analysis

Introduction

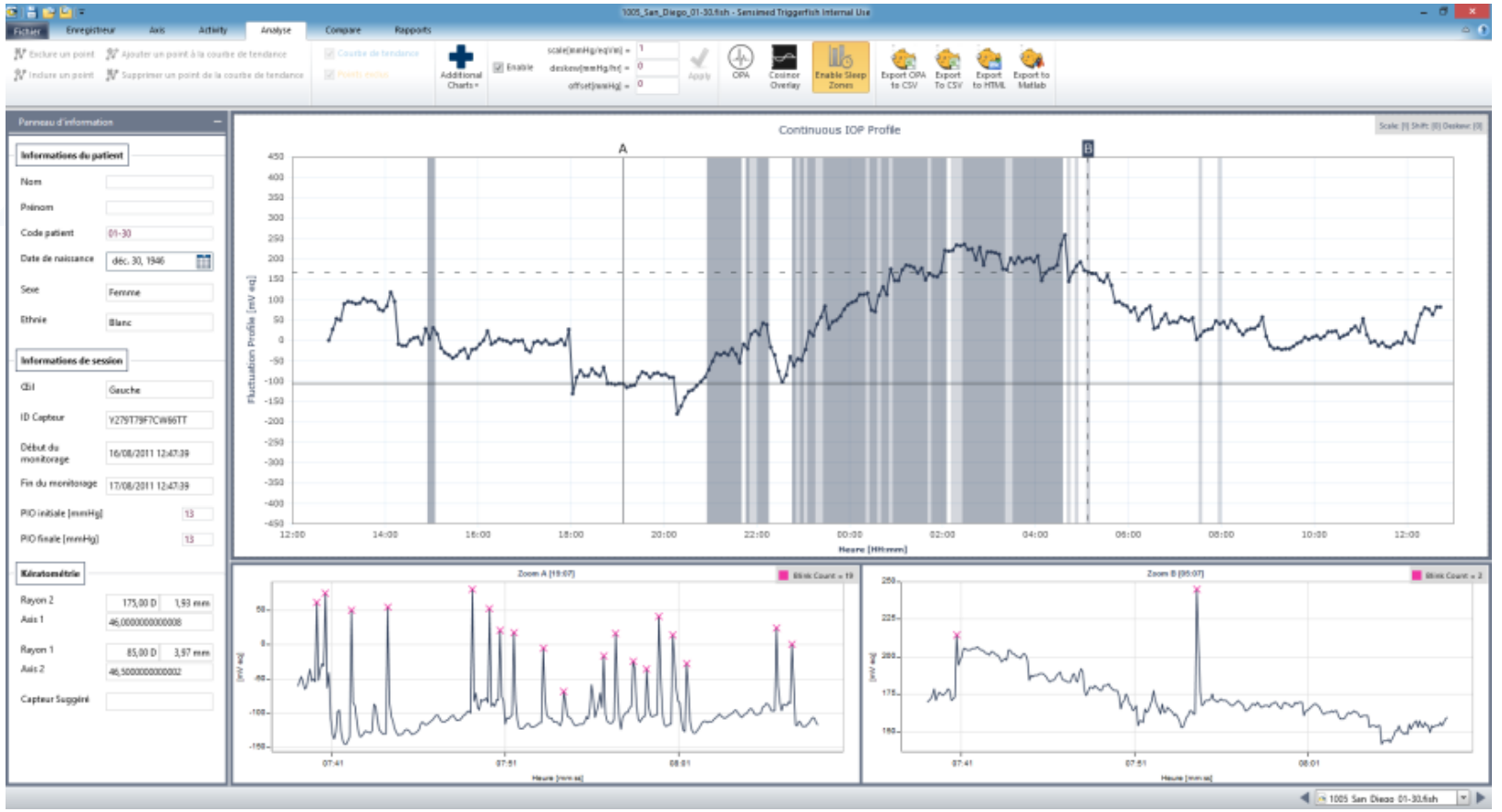
- Glaucoma detection
- Connected contact lenses
- Involved people:
  - Data scientist
  - Mathematicians
  - Medical experts
  - Medical doctors
  - Management



Predictive Modelling for Medical Analysis

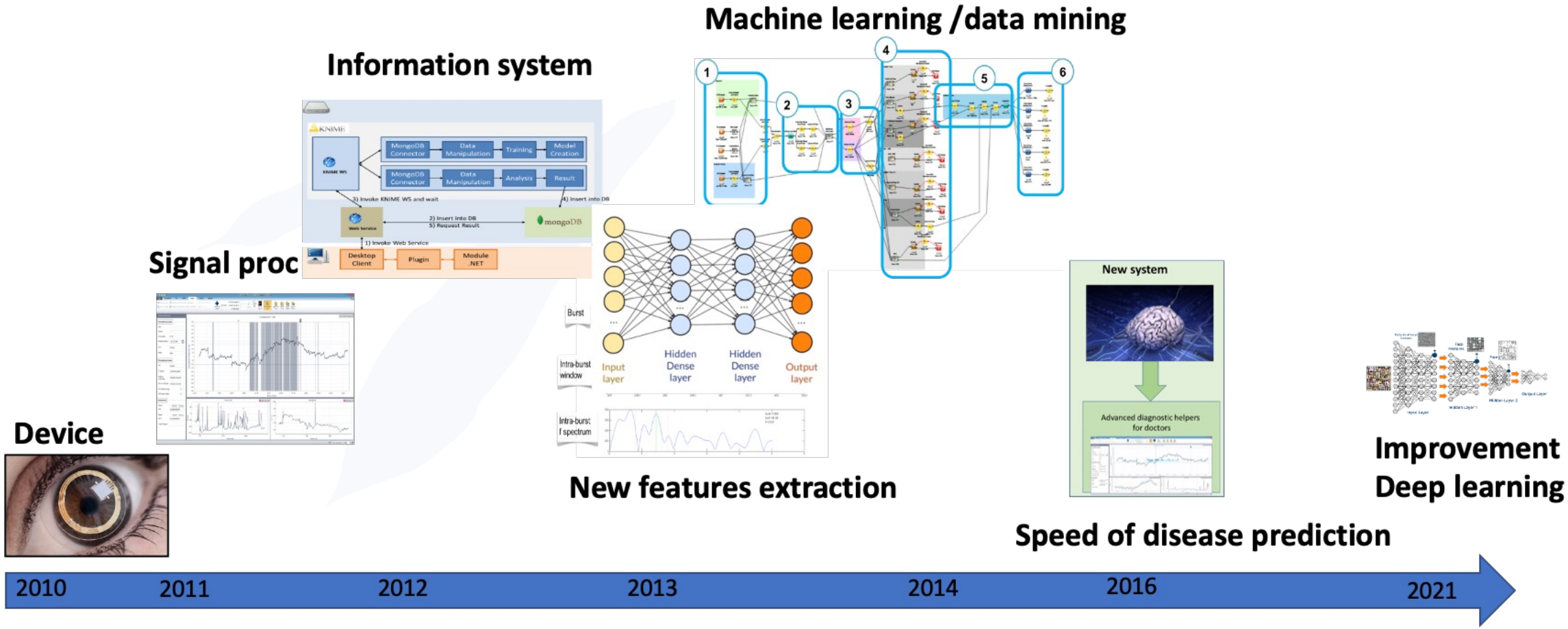
Dataset

- Collecting data from patients around the world
- 24 hours of data collection
- Daily actions/behaviors described at the same time



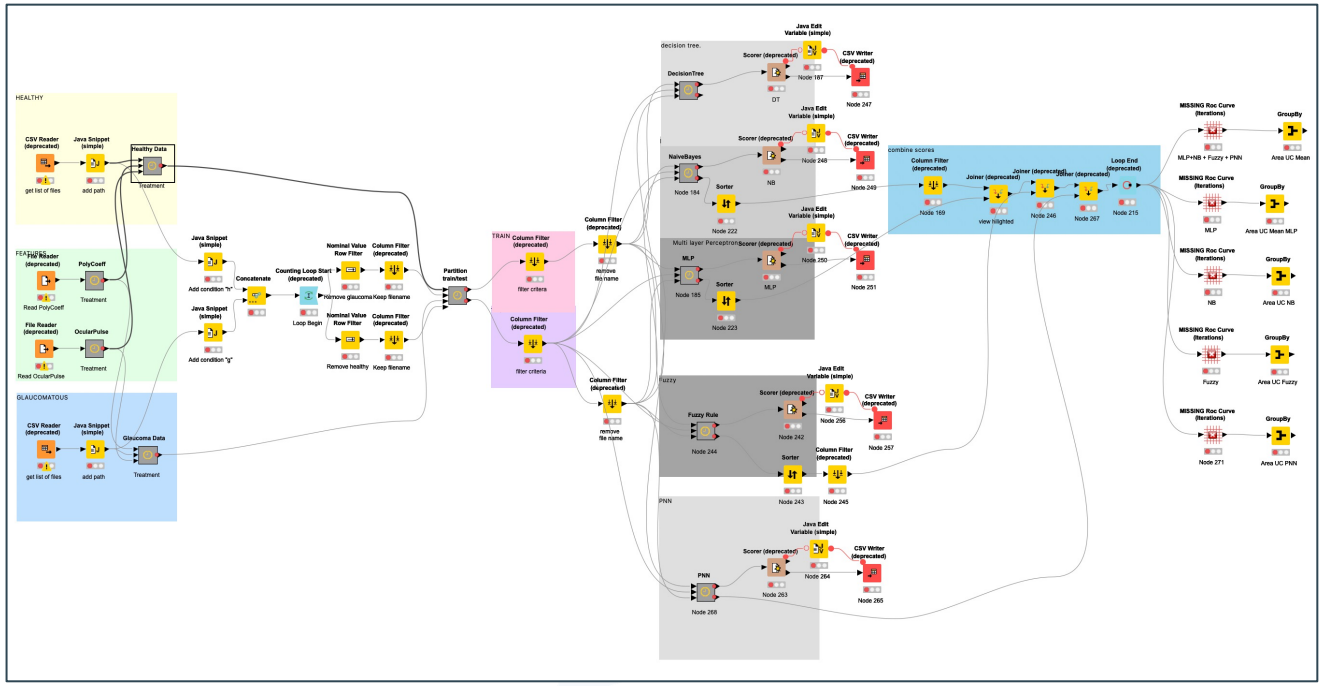
Predictive Modelling for Medical Analysis

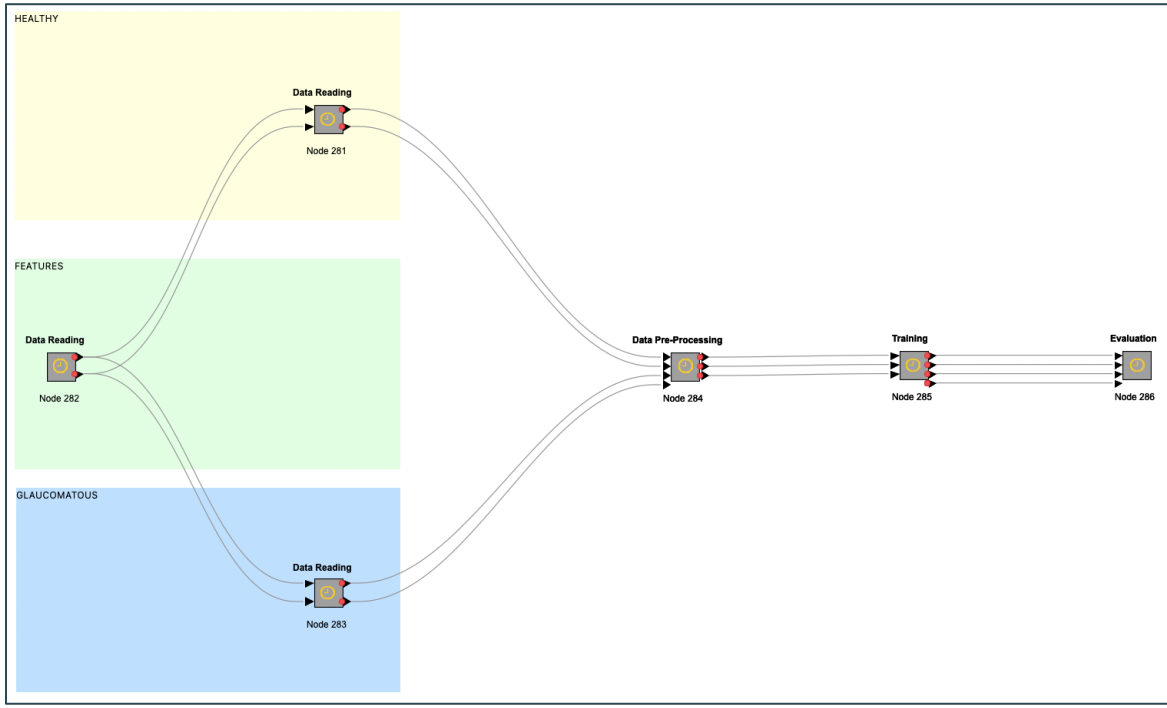
From detection to prediction



Predictive Modelling for Medical Analysis

KNIME - From a development to a deployment process





Forest Grove
Data-Driven Decision Making

Predictive Modelling for Medical Analysis

Results

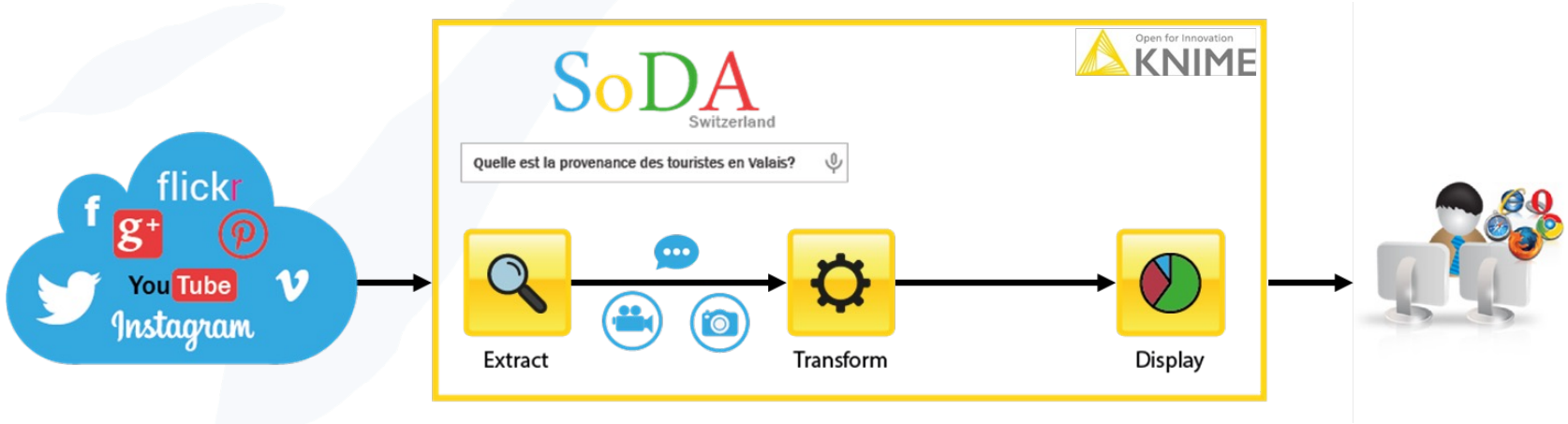
- Glaucoma disease detection
  - As precise as existing medical machines
  - More reliable than existing medical machines
  - Less intrusive

- Glaucoma disease prediction using predictive Modelling (Precision of 82%)
  - A medical machine detects, we predict
  - Used in production by laboratories, hospitals and ophthalmologists

ML and Predictive Modelling for Marketing

Introduction

- Collect data from the web (web scrapping)
- Transform, enrich, analyze and display the results
- Introduce predictive modelling to the displayed results

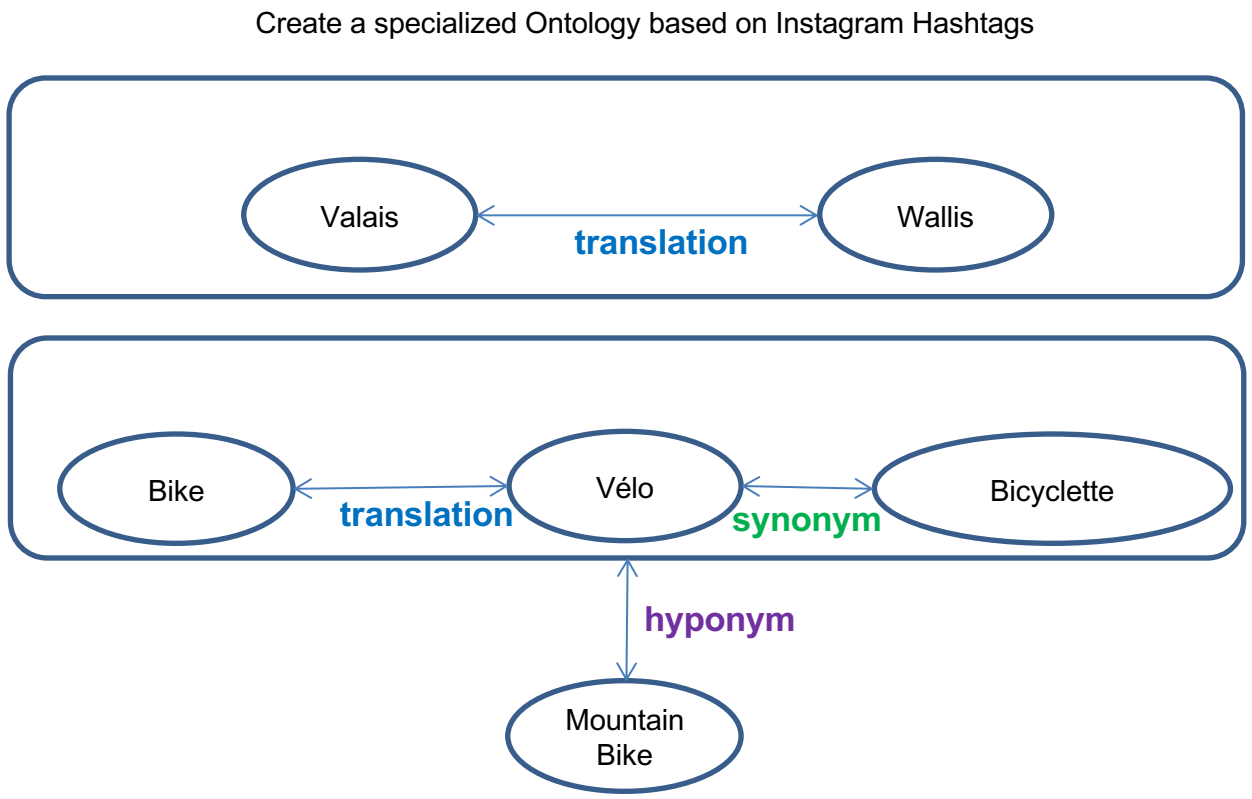


ML and Predictive Modelling for Marketing

Dataset

- Heterogenous data sources
  - Images
  - Text
  - Public profiles
- Social Medias
  - Twitter
  - Instagram
  - Facebook
  - Flickr

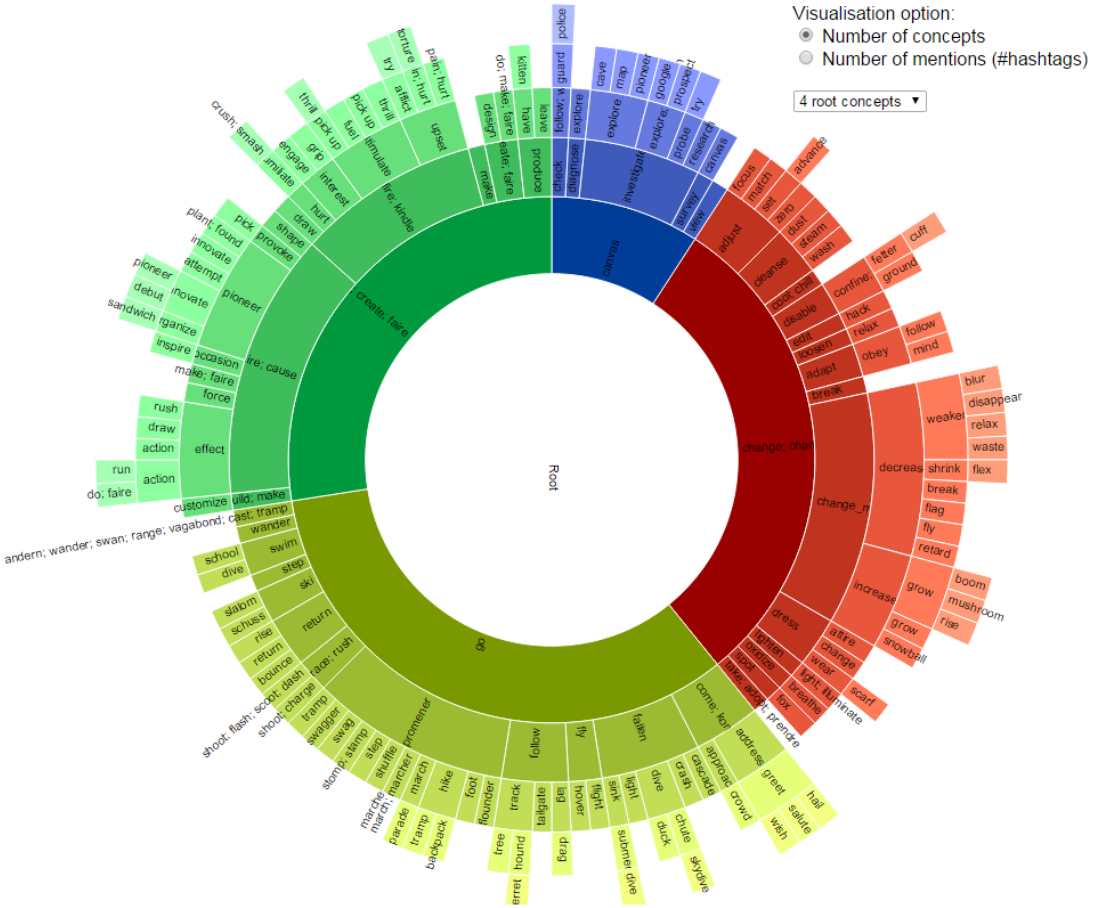
- Transformation and enrichment
  - Semantic
  - Image augmentation



ML and Predictive Modelling for Marketing

Visualization

- Analysis of the first results
- Interpretate the information

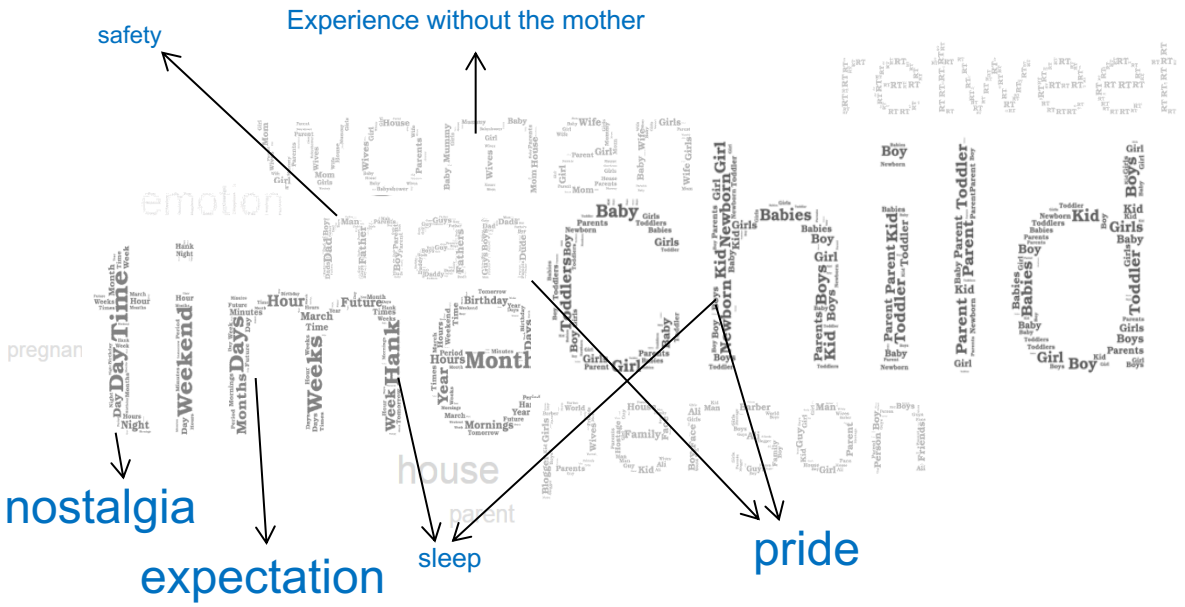


Forest Grove
Data-Driven Decision Making

ML and Predictive Modelling for Marketing

Results

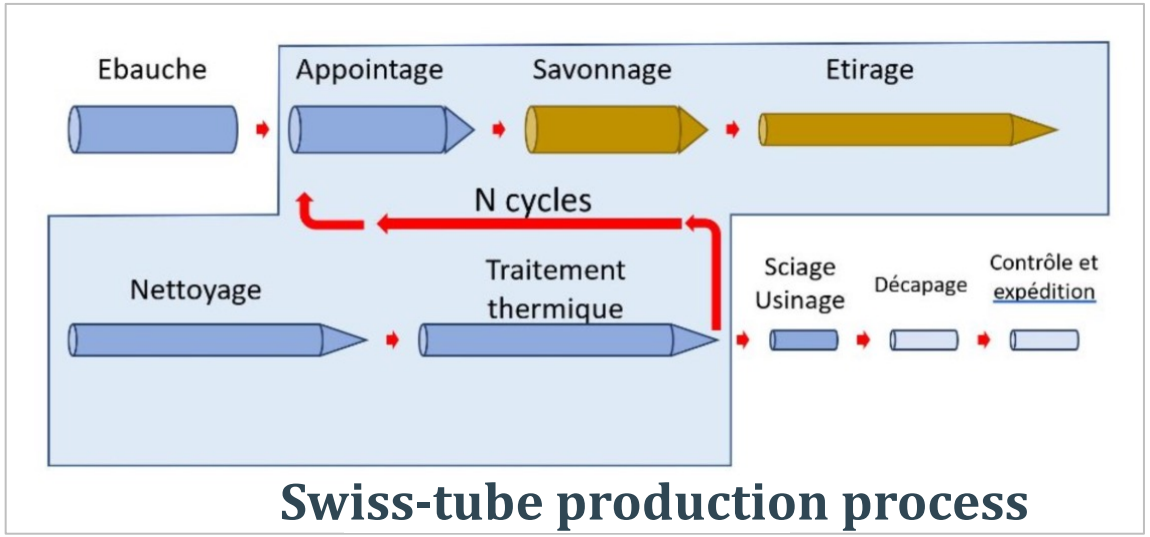
- Prediction of the future clients
  - Improved their markets share
  - Used as a daily decision tool
- Interpretation of change points



ML and Predictive Modelling for Industry

Introduction

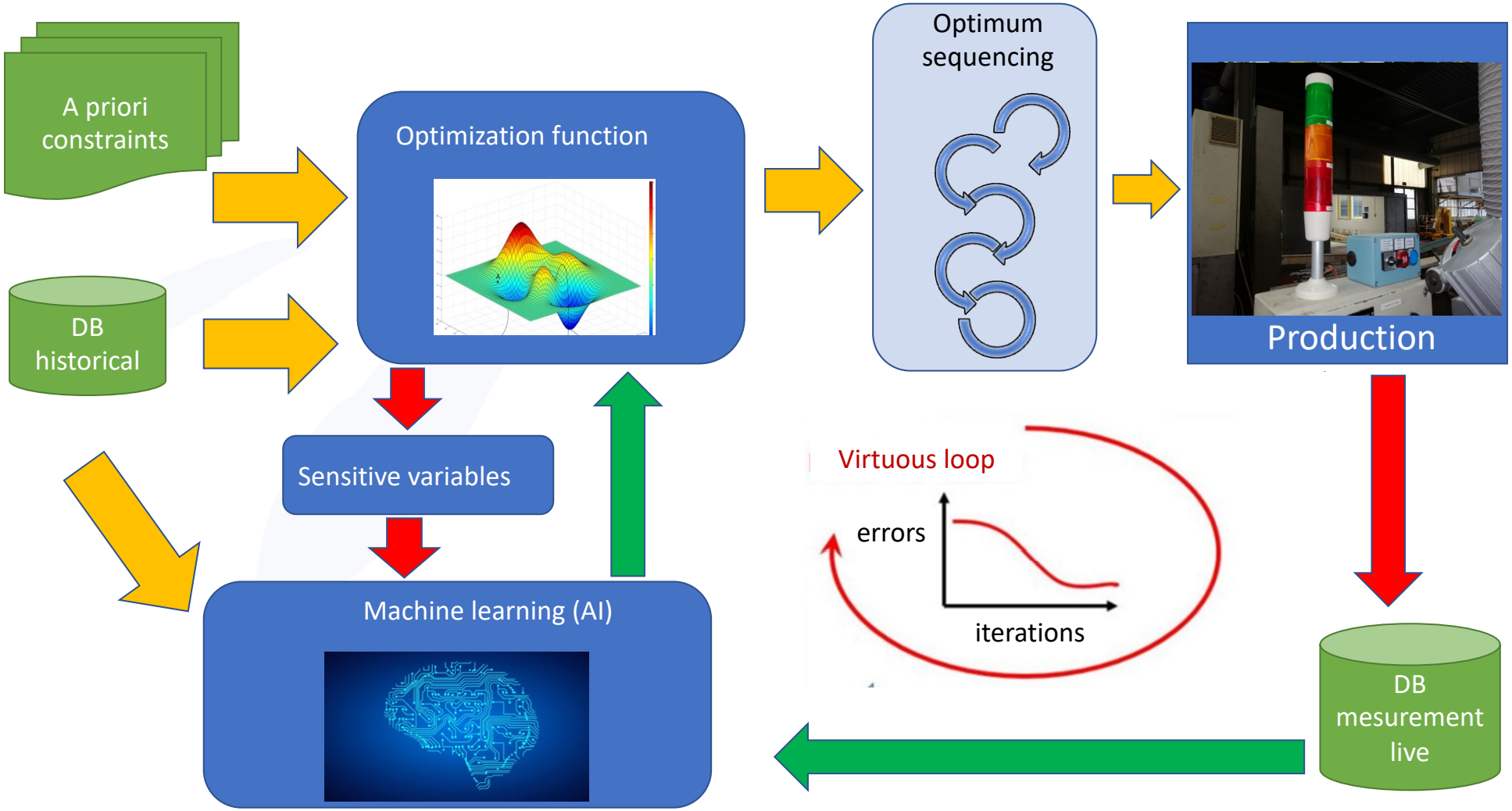
- Multi-usage tubes production
- Production optimization
  - Control and optimize the deliveries
- Quality control
  - Image analysis and classification
- Predictive modelling
  - Predictive maintenance for the machine
  - Human resources optimization



Swiss-tube production process

ML and Predictive Modelling for Industry

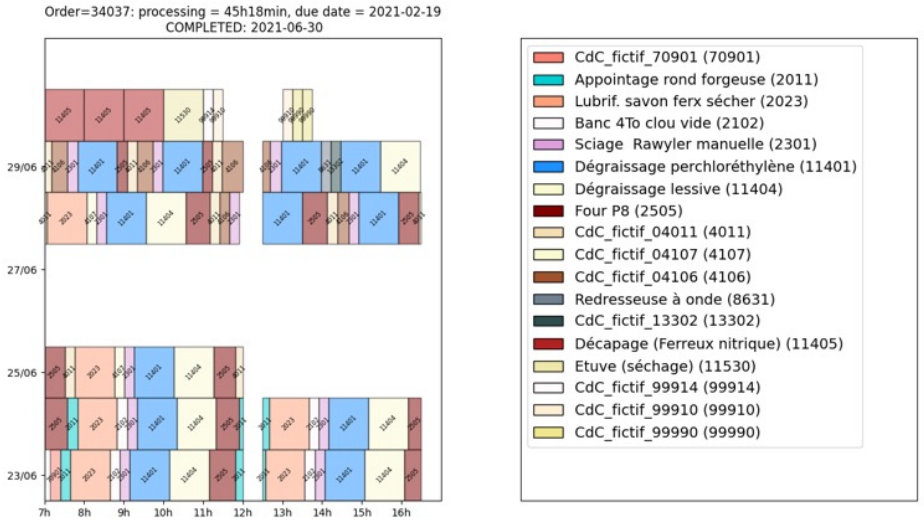
Dataset and methodology



ML and Predictive Modelling for Industry

Results

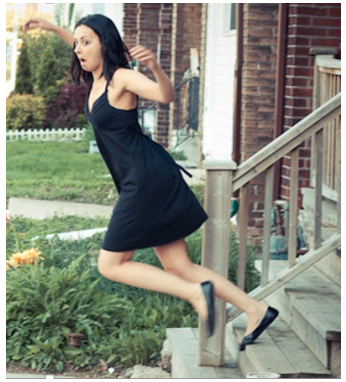
- A web application
  - to optimize the sequences
  - to optimize the human resources
  - to plan the maintenance of the machines

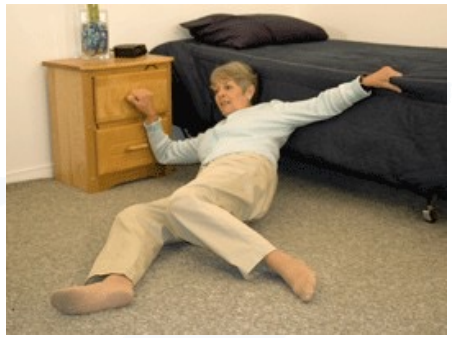


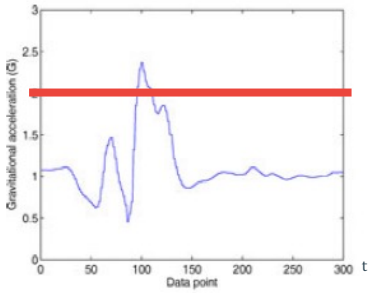
	Average days of delay	Average passage time
Real scheduling	5.7d	12.8d
Optimized scheduling	3d (-47,4%)	11.9d (-7,3%)
Optimized scheduling and ML	2.5d (-56.1%)	11.4d (-10.9%)

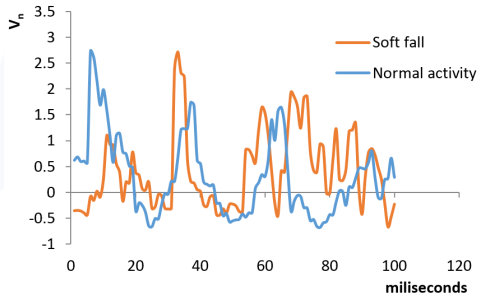
Other Examples

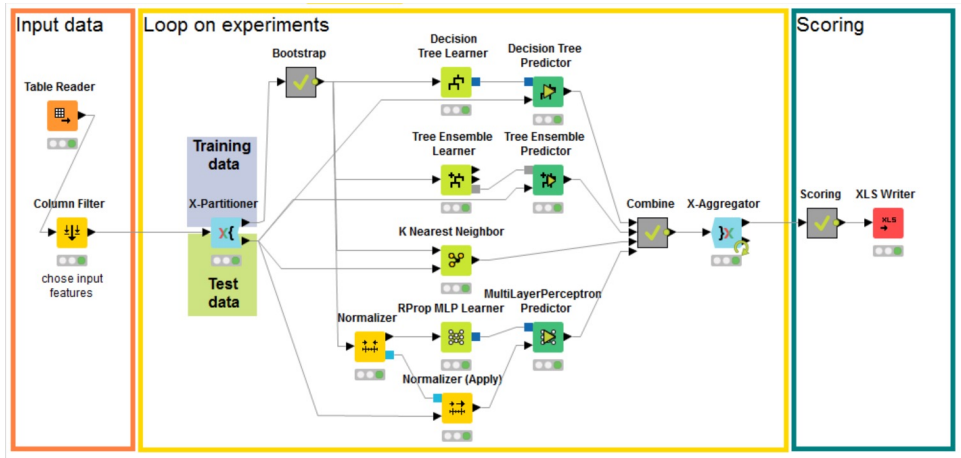
ML - Soft fall detection using an Android device

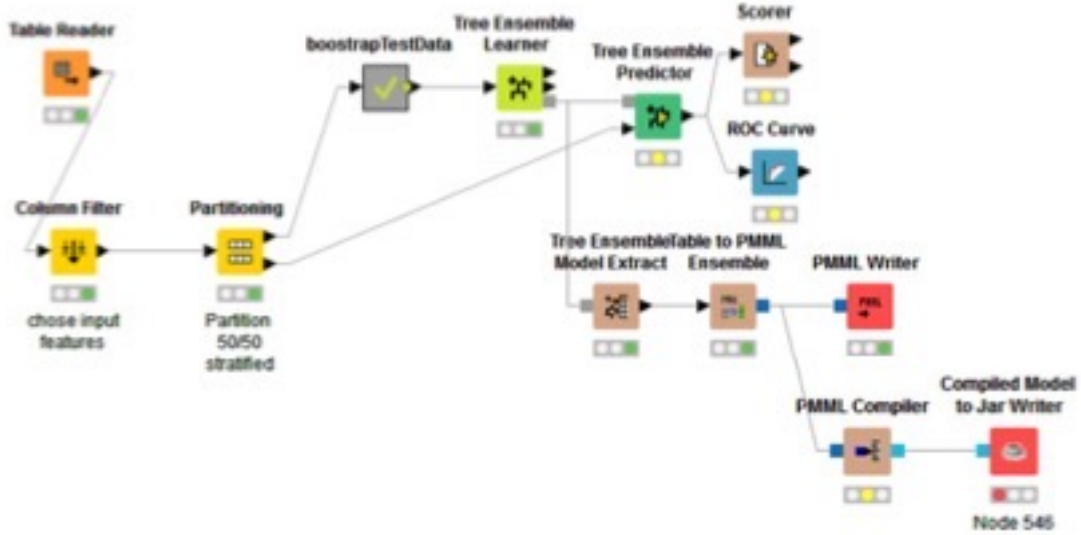












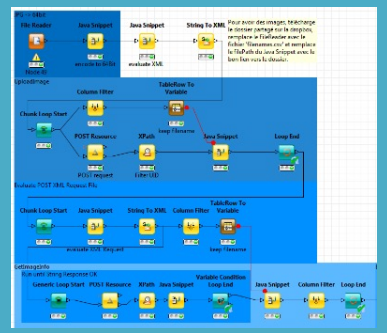
Forest Grove
Data-Driven Decision Making

Other Examples

ML – Face recognition

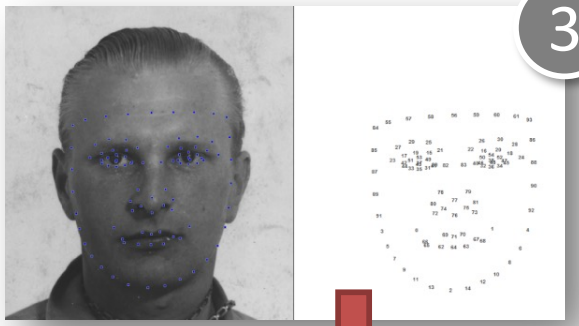
1

concentre (1)
concentre (2)
concentre (3)
concentre (4)
concentre (5)
concentre (6)
concentre (7)
concentre (8)
dilate (1)
dilate (10)
dilate (11)
dilate (12)
dilate (13)
dilate (14)
dilate (15)
dilate (16)
dilate (17)
dilate (18)
dilate (19)
dilate (2)
dilate (20)
dilate (21)
dilate (22)
dilate (23)
dilate (24)
dilate (25)
dilate (3)
dilate (4)
dilate (5)
dilate (6)
dilate (7)
dilate (9)
reagissant (1)
reagissant (10)
reagissant (11)
reagissant (12)
reagissant (2)
reagissant (3)
reagissant (4)
reagissant (5)
reagissant (6)
reagissant (7)
reagissant (8)
reagissant (9)



2

Preprocessing
KNIME



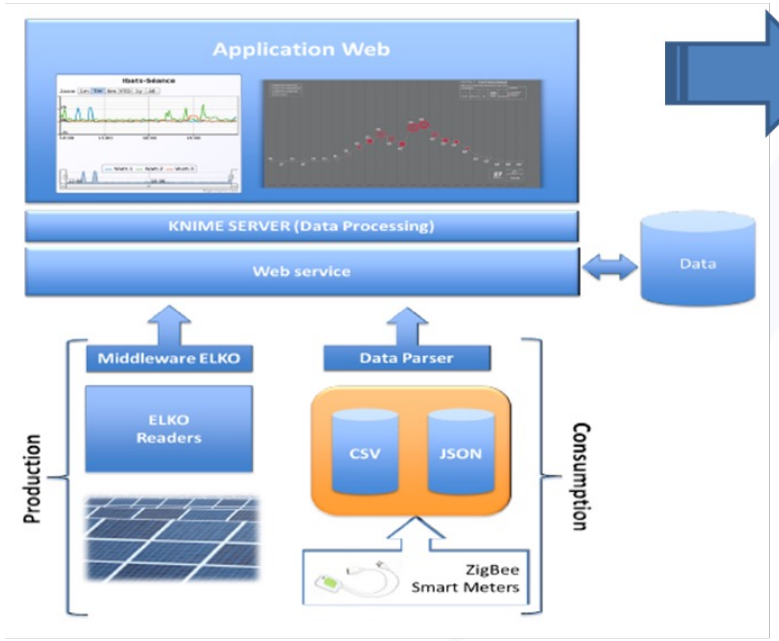
3

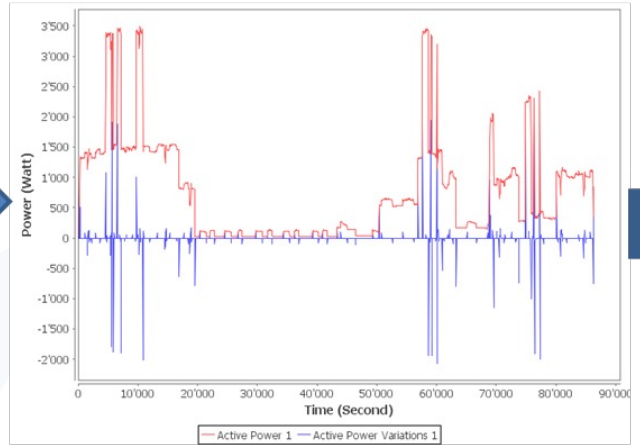
Analysis & visualization

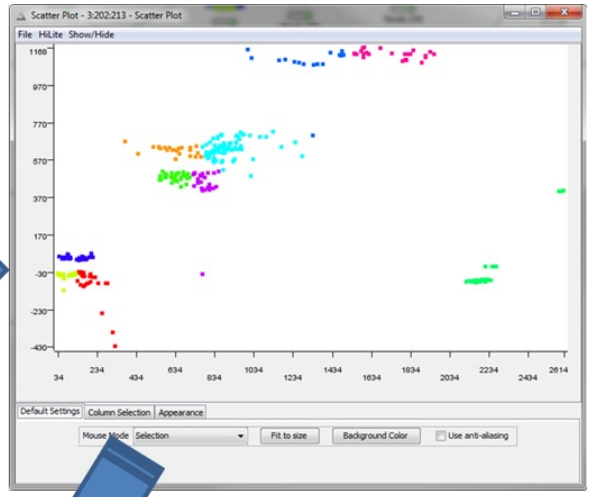
Next steps:
- **Classification**
- **Error analysis**
- **Features selection**

Other Examples

ML and Predictive Modelling - Energy prediction







Forest Grove
Data-Driven Decision Making

Chat GPT

following this customer conference in Adelaide about data analytics, data storytelling, and business intelligence, what should I do for the next one and where in Oceania should it be hold ?

For the next customer conference about data analytics, data storytelling, and business intelligence in Oceania, you could consider the following ideas:

1. Predictive Analytics: Focus on the importance of predictive analytics in business decision-making. This could include topics such as machine learning algorithms, time series analysis, and forecasting.
2. Data Privacy and Security: Discuss the latest trends and best practices in data privacy and security. Your customers may be interested in learning about how to protect sensitive data while still using it effectively in analytics.
3. Cloud Analytics: Explore the latest developments in cloud-based analytics platforms, and how they can be used to streamline data analysis and decision-making. This could include topics such as cloud data warehouses, data lakes, and analytics-as-a-service.
4. Business Intelligence Strategy: Highlight the importance of developing a robust business intelligence strategy, and how to align business objectives with data analytics initiatives. This could include topics such as goal setting, KPI development, and change management.
5. Data Ethics and Responsibility: Discuss the ethical considerations involved in data analytics and how to ensure that your business is using data ethically and responsibly. This could include topics such as privacy, bias, and the social impact of data analytics.

In terms of where to hold the conference, you could consider cities in Australia or New Zealand, such as Sydney, Melbourne, Auckland, or Wellington, which have a strong data analytics presence. You could also consider locations in other parts of Oceania, such as Singapore or Hong Kong, which are major business hubs in the region. Alternatively, you could host the conference online or in a hybrid format to accommodate customers who are unable to travel.

Revolutionize Your Business with Machine Learning and Predictive Modelling

Empowering You to Make Smarter Decisions

- Machine Learning and Predictive Modelling can help you
  - gain insights
  - make more accurate predictions
  - make better decisions for your business
- With the power of data, you can take your business to the next level.

Get in touch

Jérôme Treboux
treboux.jerome@gmail.com
http://jerometreboux.com/
https://linkedin.ch/in/jerometreboux

TR provide me the best ending slide for a presentation about machine learning and predictive modelling

Sure, here's an example of an ending slide you can use for your PowerPoint presentation about machine learning:

[Background image or solid color]

[Title]: Thank You

[Subtitle]: Harnessing the Power of Machine Learning

[Body text]: Today, we've explored the exciting possibilities of machine learning and how it can transform the way we interact with technology. By unlocking insights and making predictions, we can create a better future for ourselves and those around us.

[Closing statement]: Thank you for joining us today. Let's continue to harness the power of machine learning to drive innovation and progress.

[Your Company Logo]

[Contact information, such as email or website]

[Copyright information, if necessary]

Feel free to customize the design and content to fit the style and messaging of your presentation. Good luck!