

Elevate Decision-Making through Predictive Analytics

Amersfoort, 28.09.2023



Roberto Mastrangelo
EMEA Solutions Architect

Agenda

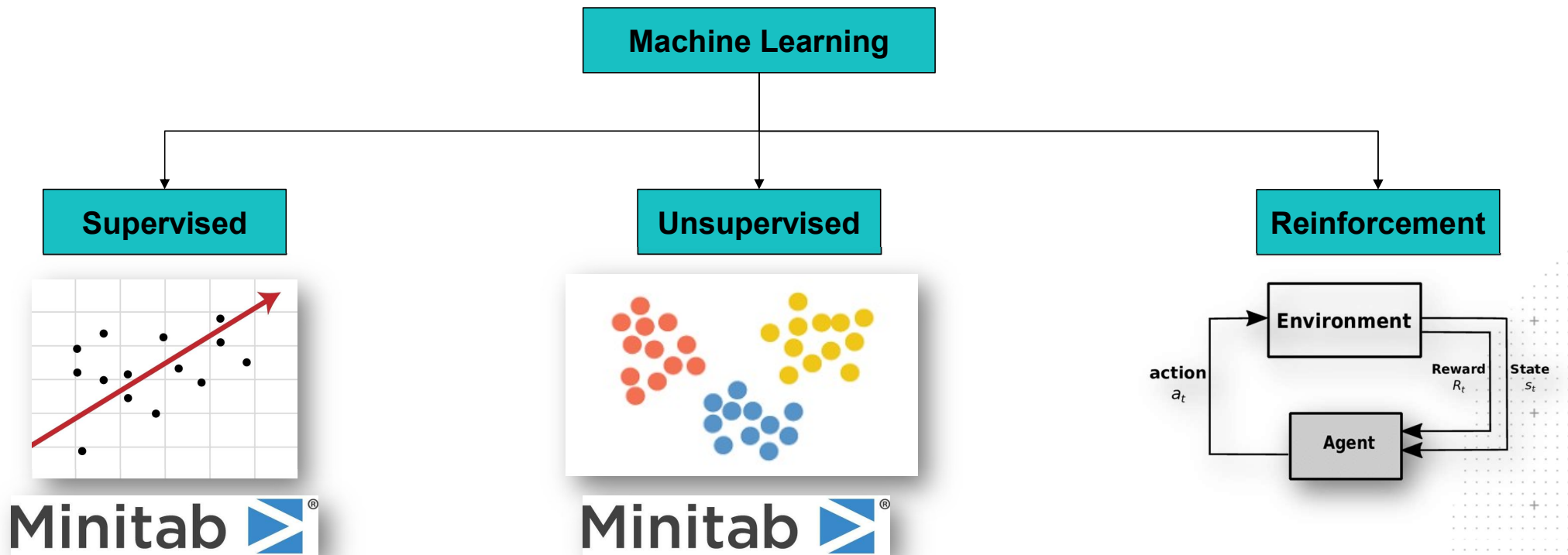
1. *Introduction to machine learning and the algorithms in Minitab*
2. *Today's use case: "Improvement of the paper brightness"*
3. *Summary of today's session*

1. Introduction to Machine Learning

Why Machine Learning from Minitab?

1. The **amount** of data available and its **complexity** keep increasing
2. There is a need for **data democratization**
 - Also **non-data scientists** should be able to gain important insights from the data available
3. Compared to classic statistical methods, tree-based models are suited for:
 - larger datasets
 - more variables
 - messy data
 - missing values
 - random outliers and
 - nonlinear relationships

General understanding of Machine Learning

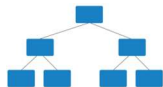


Tree-based Algorithms and MARS® in Minitab

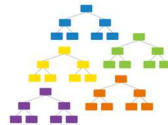
equipped to handle larger data sets with more variables, messy data, missing values, random outliers, and nonlinear relationships

CART®

Classification & Regression Trees



Random Forests®



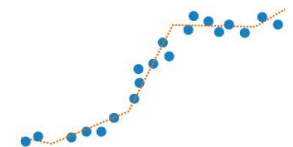
TreeNet®

(Gradient Boosting)



MARS®

Multivariate Adaptive Regression Splines



- Decision trees intuitive to interpret (no need to be a data scientist)
- Quick alternative and useful insights provided (quickly understand the key drivers of a process)
- Ranking of variable importance of individual predictors available

- Non-parametric regression technique
- Extension of linear models

Increase in accuracy / predictive model performance*

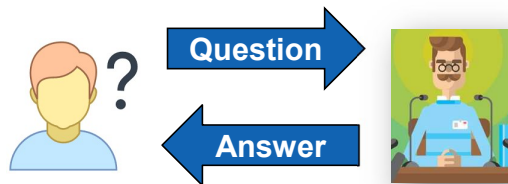
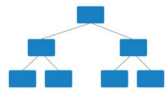
* This ordering depends on the dataset/case

Tree-based Algorithms in Minitab

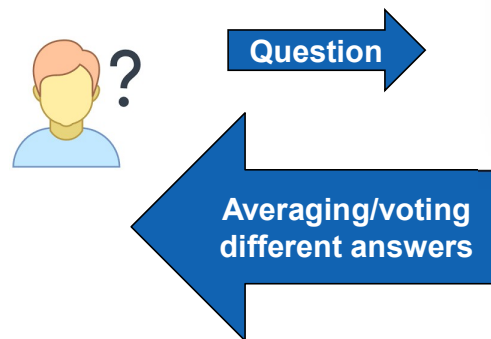
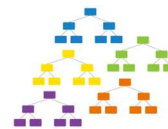
equipped to handle larger data sets with more variables, messy data, missing values, random outliers, and nonlinear relationships

CART®

Classification & Regression Trees



Random Forests®



TreeNet® (Gradient Boosting)



Increase in accuracy / predictive model performance*

* This ordering depends on the dataset/case

Typical use cases* 1/2

Use case	Description
Predictive Maintenance	Forecast when machines and equipment are likely to fail , allowing for timely maintenance and reducing downtime .
Health care	Assess the risk of disease outbreaks and patient readmissions facilitating more effective treatment plans .
Quality Control	Identify potential defects . This saves both time and resources and reduces rework or recalls.
Fraud Detection	Identify potentially fraudulent activities in real-time by analyzing trends and patterns in transaction data.



*This list is not exhaustive and contains a selection of possible applications

Typical use cases* 2/2

Use case	Description
Customer Retention	Predict which customers are likely to churn , and plan targeted actions to retain valuable customers .
Sales Forecasting	Forecast sales trends , which can then inform production, inventory management, and marketing efforts.
Supply Chain Management	Forecast demand , and optimize inventory , thus reducing costs and improving efficiency.
Energy Consumption Forecasting	Enable better energy resource planning and minimize waste to allow more sustainable energy management.



*This list is not exhaustive and contains a selection of possible applications

2. Today's use case

Today's example

„**Paper Bright**“, a manufacturer of paper products, lost a large order because the quality of their products did not meet the customer's requirements (**low paper brightness**, min spec=**75**).

Management has mandated the Operational Excellence Team to:

1. **Improve the brightness** of the paper products.
2. **Understand the key variables** and their influence on the brightness.
3. Predict and **prevent future quality issues**.

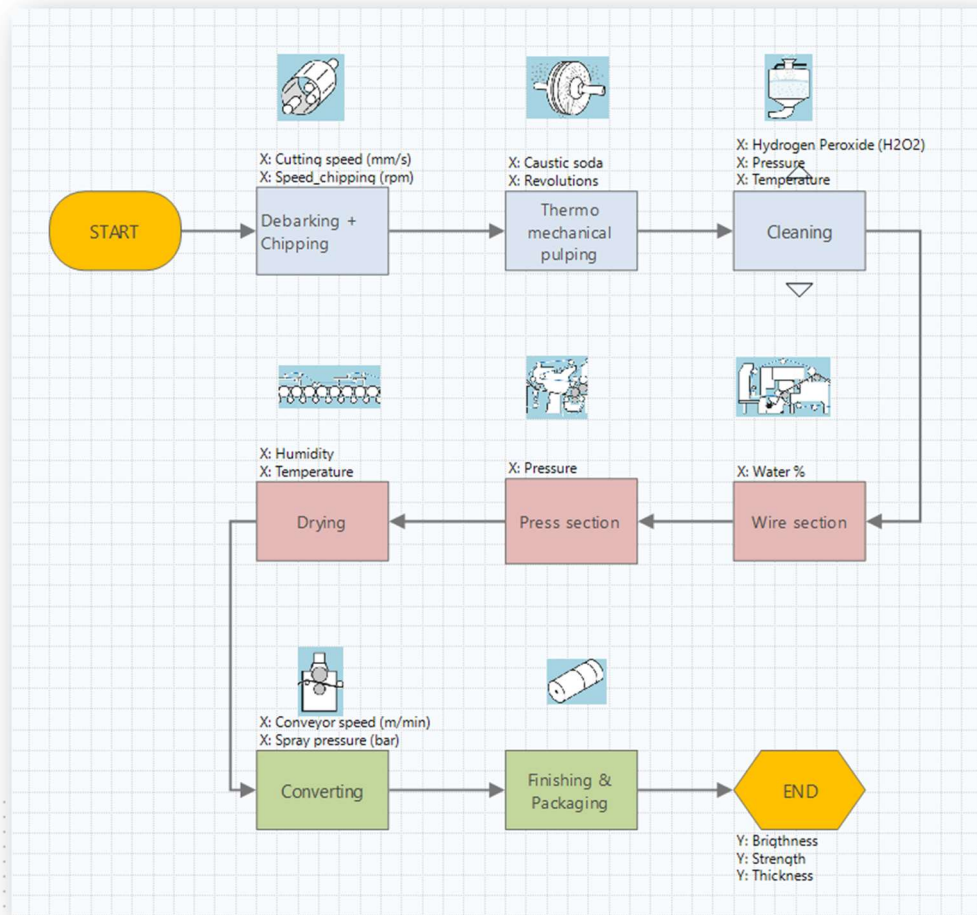


Project Team Results Preview

Thanks to its methodical approach, the project team was able to:

- **Increase the brightness** level on average by **5%** and to reduce its **standard deviation** by **31%**
- **Regain trust** with one of their main customers
- Expand the business into **premium segments**
- Generate an **additional 4% in revenue**

Process Map high level



Main processes

1. Pulping

Sampling * 2

2. Papermaking

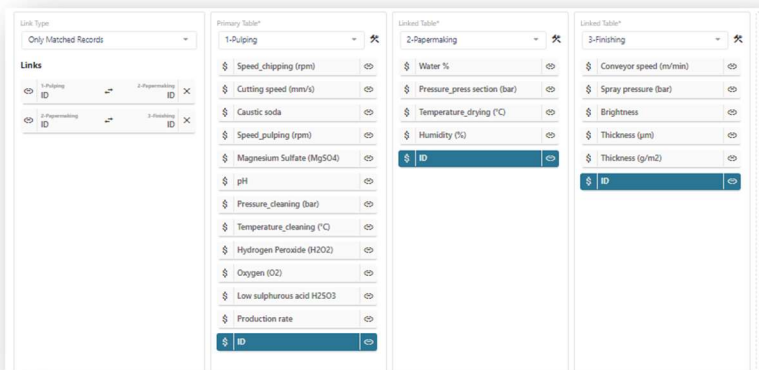
Sampling * 1

3. Finishing

Sampling * 1

Data set of today's example

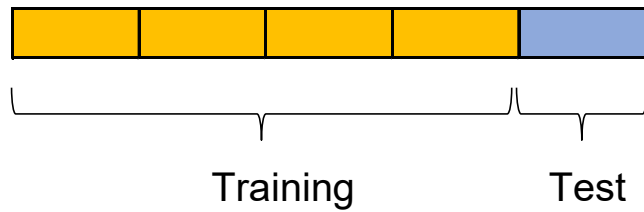
After merging the data from 3 different databases, our data set consists of 22 columns, which are available to us for the analysis.



#	Main process	Process steps	Features / Predictors	Type	Variable type
1	Pulping	Debarking + Chipping	Speed_chipping (rpm)	Predictor	Continuous
2	Pulping	Debarking + Chipping	Cutting speed (mm/s)	Predictor	Continuous
3	Pulping	Thermo mechanical pulping	Caustic soda	Predictor	Continuous
4	Pulping	Thermo mechanical pulping	Speed_pulping (rpm)	Predictor	Continuous
5	Pulping	Thermo mechanical pulping	Magnesium Sulfate (MgSO4)	Predictor	Continuous
6	Pulping	Thermo mechanical pulping	pH	Predictor	Continuous
7	Pulping	Cleaning	Pressure_cleaning (bar)	Predictor	Continuous
8	Pulping	Cleaning	Temperature_cleaning (°C)	Predictor	Continuous
9	Pulping	Cleaning	Hydrogen Peroxide (H2O2)	Predictor	Continuous
10	Pulping	Cleaning	Oxygen (O2)	Predictor	Continuous
11	Pulping	Cleaning	Low sulphurous acid H2SO3	Predictor	Categorical
12	Pulping	Cleaning	Production rate	Predictor	Continuous
13	Papermaking	Wire section	Water %	Predictor	Continuous
14	Papermaking	Press section	Pressure_press section (bar)	Predictor	Continuous
16	Papermaking	Drying	Temperature_drying (°C)	Predictor	Continuous
17	Papermaking	Drying	Humidity (%)	Predictor	Continuous
18	Finishing	Converting	Conveyor speed (m/min)	Predictor	Continuous
19	Finishing	Converting	Spray pressure (bar)	Predictor	Continuous
20	Finishing	Converting	Brightness	Output	Continuous
21	Finishing	Finishing & Packaging	Thickness (µm)	Output	Continuous
22	Finishing	Finishing & Packaging	Thickness (g/m2)	Output	Continuous

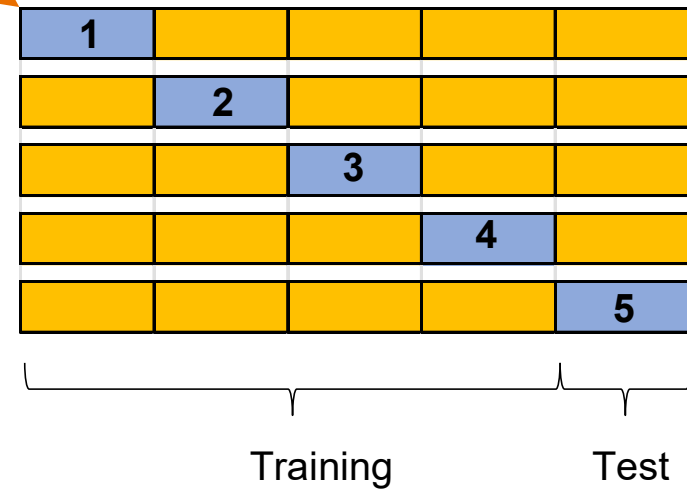
Training and Test

Single train-test split



Cross validation (5-fold)

Validation fold



DEMO

3. Summary of today's session

Technical Summary 1/2

The function “Automated Machine Learning” ranks **Random Forests®** and **TreeNet®** as the best algorithms for this dataset.

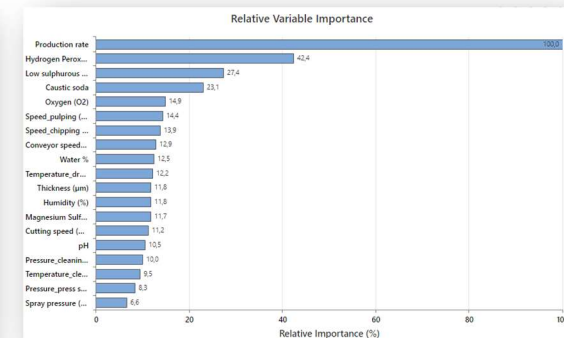
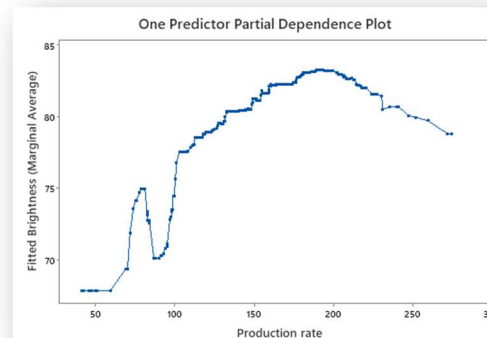
Based on the selected model from **TreeNet® Regression** the 2 most relevant predictors influencing brightness are:

1. **Production rate** – 180 to 200
2. **Hydrogen peroxide (H2O2)** – as high as possible

Since these 2 predictors belong to the “**Pulping-cleaning**” process step, the team decides to focus their efforts on this step.

Best Model within Type	R-squared (%)	Mean Absolute Deviation
Random Forests®*	88,17	1,0283
TreeNet®	81,27	1,4688
CART®	79,73	1,4104
MARS®	76,91	1,6466
Multiple Regression	61,91	2,0730

* Best model across all model types with maximum R-squared. Output for the best model follows.

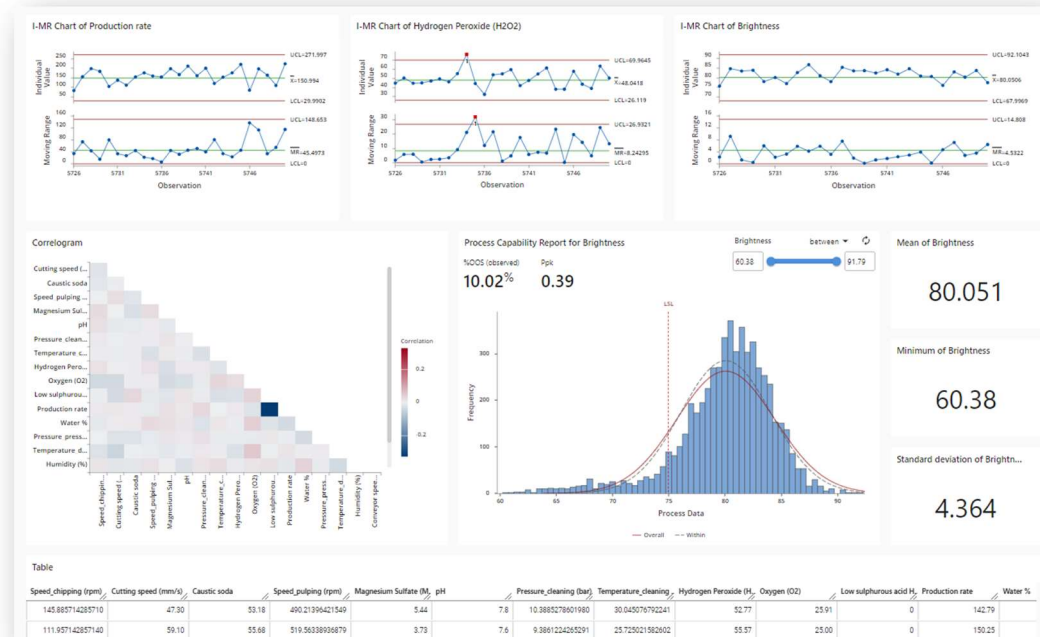


Technical Summary 2/2

The changes on the process allow for an **average increase in brightness of 5%** and a **reduction of its standard deviation by 31%**

To avoid future quality issues, the team decides to **monitor** these and other **key variables** with **dashboards and alerts**.

In addition, the team could identify **why** the brightness is optimal at a production rate **between 180 and 200**.



Financial Summary

By implementing the measures shown in the use case, the project team was able to:

- **Increase the brightness level** on average by **5%** and to reduce its **standard deviation** by **31%**
- **Regain trust** with one of their main customers
- Expand the business into **premium segments**
- Generate an **additional 4% in revenue**

General Conclusions

1. Easy-to-Use Minitab Solutions

- "No code, just click"
- Predictive analytics is also for non-data scientists
- Easy to discuss outputs with experts
- From data silos to a central repository

2. From descriptive to predictive analysis

- Recognize and understand relationships and enable a forecast (concrete recommendation for future decisions).

3. Significant financial improvements

4. How Minitab Can Help

- Software solutions, training, and consulting



Thank you!

Contact us to learn how our
Minitab solutions can help you.

Jos Huijsmans, Partner Sales EMEA

jhuijsmans@minitab.nl

Roberto Mastrangelo, Solutions Architect EMEA

rmastrangelo@minitab.de



Discover



Forecast



Achieve



Transform

Thank You!