Predict DGA data using machine learning
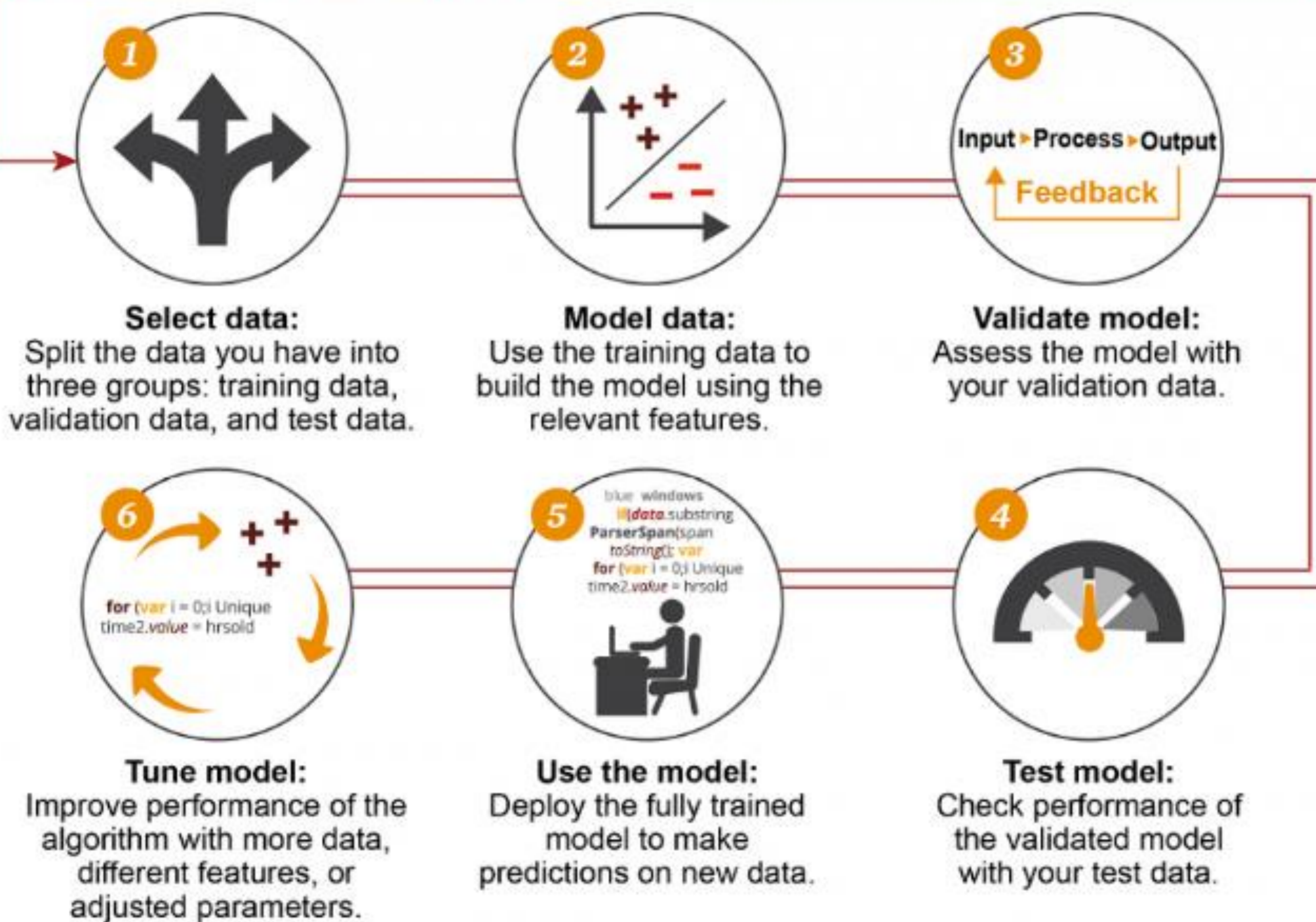
Freek Doornkamp	Marcel Hooijmans	Tom Huybregts	Jos Slangen	Roger de Jong	Bob Rombach
Ksandr/Stedin	DGA/Stedin	Enexis	TenneT	VU Amsterdam	VU Amsterdam

How does machine learning relate to artificial intelligence?

Pattern recognition
Statistics
Neurocomputing
Deep learning
Data mining
Machine learning
AI
Databases
Knowledge discovery
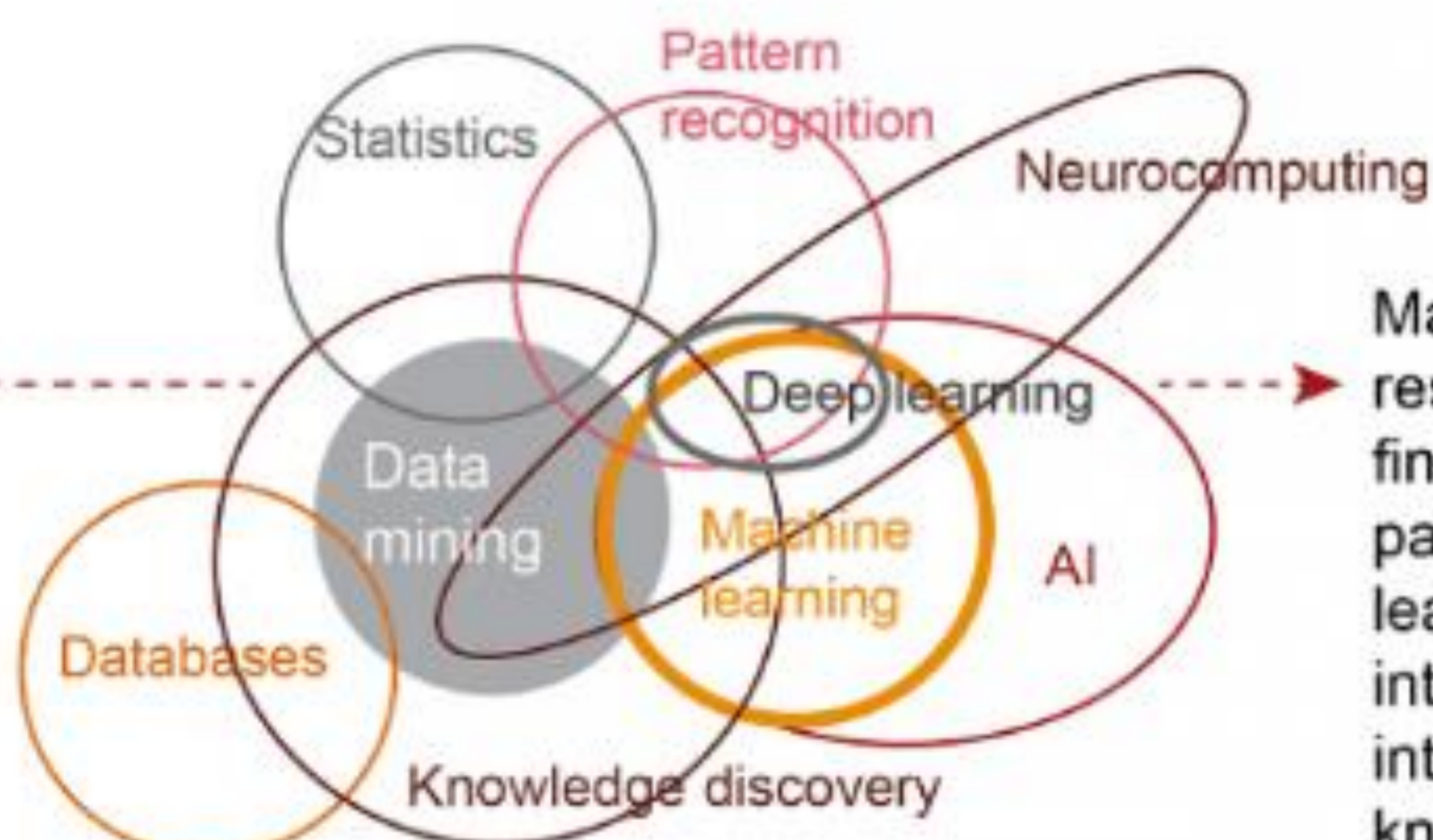
Machine learning is a category of research and algorithms focused on finding patterns in data and using those patterns to make predictions. Machine learning falls within the artificial intelligence (AI) umbrella, which in turn intersects with the broader field of knowledge discovery and data mining.

How machine learning works

1 Select data:
Split the data you have into three groups: training data, validation data, and test data.

2 Model data:
Use the training data to build the model using the relevant features.

3 Validate model:
Assess the model with your validation data.
Input ▸ Process ▸ Output
Feedback

6 Tune model:
Improve performance of the algorithm with more data, different features, or adjusted parameters.
for (var i = 0;i Unique time2.value = hrsold

5 Use the model:
Deploy the fully trained model to make predictions on new data.
blue windows
data.substring
ParserSpan(span
toString(: var
for (var i = 0;i Unique
time2.value = hrsold

4 Test model:
Check performance of the validated model with your test data.

DGA

DISSOLVED GAS ANALYSIS

(DGA) is the single most comprehensive asset condition assessment and management tool for oil-filled transformers. DGA offers advanced detection of incipient fault conditions leading to almost all of the failure modes listed below.

Pie chart values: 2%, 3%, 6%, 6%, 11%, 13%, 22%, 12%, 25%

- Lightning
- Through Faults
- Insulation Deterioration
- Inadequate Maintenance
- Moisture
- Loose Connections
- Workmanship
- Overloading
- All Others

Chart Source: William H. Bartley, P.E.
The Hartford Steam Boiler Inspection and Insurance Co.

METHODS

OIL COLLECTION

Manual Collection - A small volume of oil is collected for laboratory analysis and transferred into a gas-tight container from a dedicated fitting and then transported to the laboratory. ASTM method D3613 details procedures for oil sample handling.

On-Line Collection - In the case of On-Line Transfomer Monitors a small volume of oil is continuously circulated through the monitor and then returned to the transformer. The circulating oil is sampled and analyzed for gas content. On-Line Monitors offer a closed-loop repeatable oil collection process.

SHAP plot labels: C2H4_lag 0.254, C2H6_lag 0.165, C2H2_lag 0.123, H2_lag 0.054, C2H2_lag2 0.026, CH4_lag 0.023, C2H4_lag3 0.018, C2H4_lag2 0.014, V1_Elin 0.013, V1_Smit 0.009
SHAP value (impact on model output)
Feature value Low — High

Pentagon 1
H2
C2H6
C2H2
CH4
C2H4
D2
T3

Triangle 1
100% 0%
CH4
C2H2
C2H4
D2 DT T3

Transformer diagram labels:
HV and LV Insulator Bushings
Pressure Relief Diaphragm
Conservator
Buchholz Relay
Cooling Radiator Pipes
HV Winding (LV winding inside)
Oil Filling
Earthing Lug

SHAP (SHapley Additive exPlanations) is a game-theoretic approach to explain the output of any machine learning model. It connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions

As we can see from the picture above, the plot represents:
- The most important feature of the model on the y-axis in a descending order (at the top the most important one).
- The SHAP value on the x-axis.
- The feature value with colors. A high value is represented with purple, while a low value with yellow.
- Each point represent a prediction result.

We have chosen a summary plot as output (SHAP value). It is striking that two brands are included in the top 10. Based on the data, the Smit transformers are reliable and the Elin transformers are more likely to fail
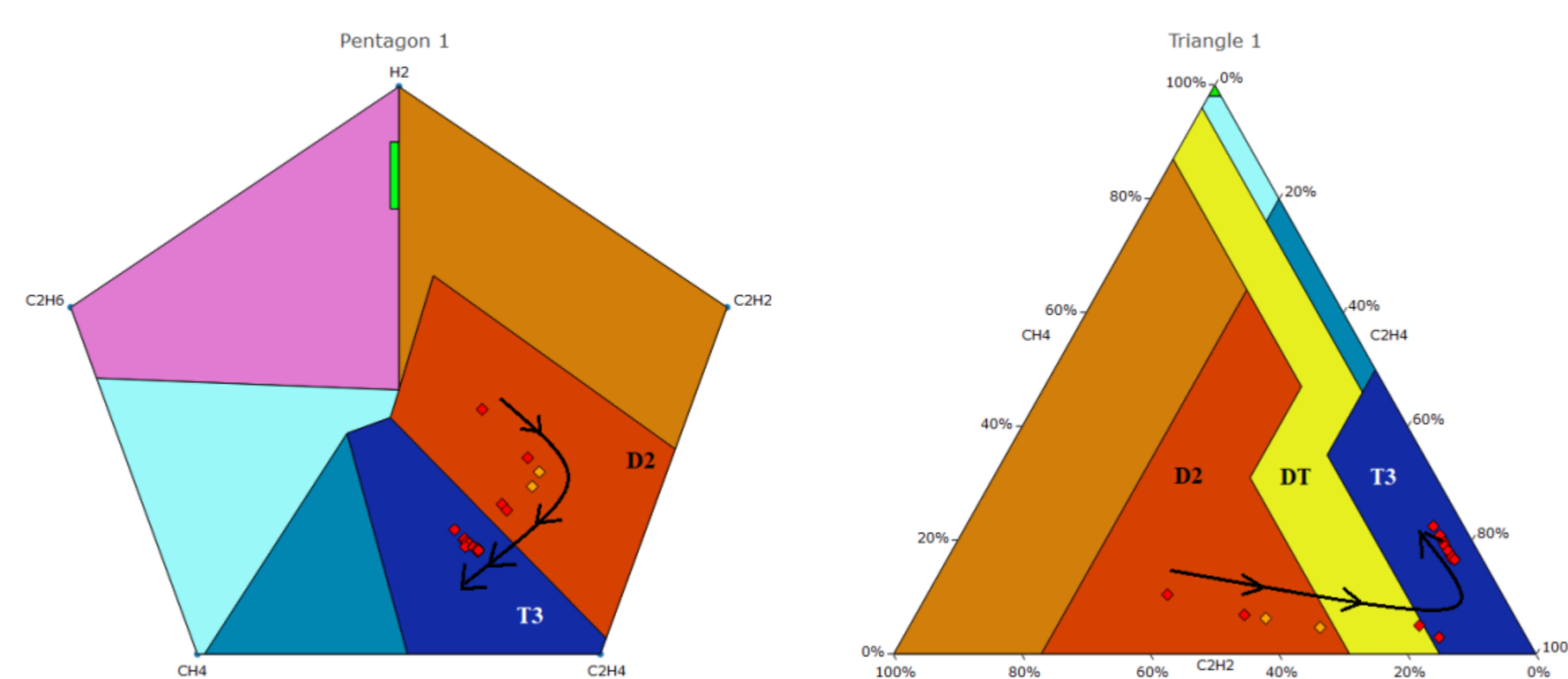
by making data visual, we are able to analyze the data better and to provide insight into the trend together with the outcome of the predictive model (machine learning)