



Automation of Dangerous Goods Classifications through Machine Learning

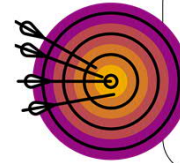
Nathaly Murillo, EHS Regulatory Documentation – SAP Content As A Service

INTRODUCTION

In many industries, product classifications for the transport of Dangerous Goods (DG) is a crucial step in SDS authoring which must be delegated to DG certified experts, whom navigate a complex classification structure with countless exceptions. Often, this is the rate limiting step for how quickly SDSs can be produced within SAP's EHS Regulatory Documentation Service (ERD).

Currently, we manage the turn around time for our DG classifications by using complex Rules-Based Algorithms. However, Machine Learning (ML) may allow us to scale automated classifications rapidly by training neural networks using ERD's expansive data sets.

Our project included teaching the model user specific patterns, i.e. specific interpretation of the regulations by different DG experts, and company specific preferences.



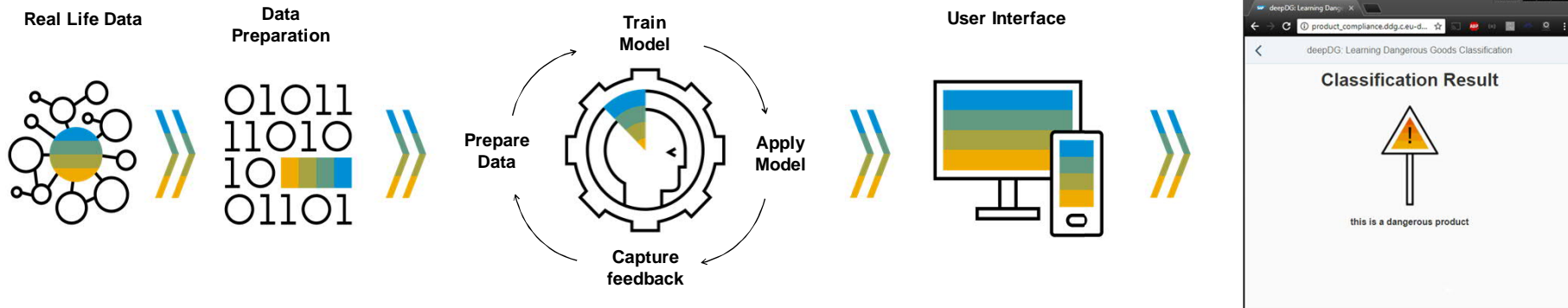
PROJECT GOALS

- Proof of concept that the reliability of DG classifications can be significantly increased using a machine learning approach.
- Can machine learning replace a rule based system?
- Can we teach user specific patterns, i.e. the personal preferences of the DG Safety Advisor, specific for each customer?
- Can machine learning enhance an existing rule based system?

EVALUATED DATA SETS

22,000 Products with a DG classification

- 16,000 are classified as non-DG
- 11 DG classifications used by 3,900 products (>100 each), spread over 9 UN numbers
- 60 DG classifications used by 10-99 products
- 300 DG classifications used by less than 10 products



MACHINE LEARNING

Advantages:

- Scalability & low operational cost
- Can predict a result without being taught explicitly (generalization)
- Can learn user specific patterns

Disadvantages:

- Huge amount of training data required
- No or only minimal explanation possible
- Changes require a new training with training data

RULES BASED ALGORITHMS

Advantages:

- Hard facts can be easily modeled (e.g. flash point ranges)
- Predictable results & allows explanation of results
- Relatively easy to adopt well defined changes

Disadvantages:

- Complex classifications (like dangerous goods) can be very time-consuming or even impossible to be fully realized
- Requires a lot of expert resources in tedious specifications and testing
- Cannot easily handle user specific classifications

CONCLUSIONS AND NEXT STEPS

- DG classifications can be automated using Machine Learning, given enough data.
- Achieved over 99% accuracy with the ML model for non-DG products.
- ML model was able to learn user specific patterns.
- ML cannot currently replace the existing rule based system. More training data needed for DG regulated products.
- Co-Innovation with partners from various industries, willing to provide real product data with SAP is needed in order to train and develop ML neural networks.