Machine Learning in the Oil & Gas Supply Chain
AGENDA

1. MACHINE LEARNING OVERVIEW
2. CASE STUDY – DATA CLASSIFICATION
3. LEARNINGS
4. FUTURE USE CASES
5. QUESTION & ANSWERS
USE CASE – Accurate Categorization of Supply Chain Data

eCommerce-enabled companies have immense volumes of digital, yet unstructured data.

<pidx:LineItemDescription>
5/8 x 3-3/4 b-7 stud bolt w/nuts
</pidx:LineItemDescription>

x (millions of records)

How do we accurately, consistently, and efficiently classify this data to a taxonomy?

GL  
unspc  
Other Taxonomy/Spend Heirarchy

How do we accurately, consistently, and efficiently classify this data to a taxonomy?
Current Approaches to Data Classification

1. Manual Coding
   Manual taxonomy assignment of:
   - Contract/catalogue item
   - Purchase Order item
   - Invoice line item

2. Rules-based Engines
   IF ‘VLV’ THEN ‘Pipes/Valves/Fittings’

3. Offshore Post Processing Service
   Usually a combination of #1 & #2

TIME-CONSUMING
EXPENSIVE
INCONSISTENT
INACCURATE
SUB-OPTIMAL CATEGORY MANAGEMENT
A Supervised Approach: Data-Driven Hierarchy

**Model Input**
- Contents of ~2.7 million invoices from were provided
  - Vendor information
  - Description of items / services

**Model Approach**
- Used categories assigned to these invoices to supervise the machine learning process
- Trained an artificial neural network to learn the categorization logic from invoice contents to categories
  - A word embedding technique was applied in the artificial neural network to process natural language
  - Classification of invoices was based on processed natural language content and vendor information

**Model Performance**
- The artificial neural network learned semantic patterns and categorization logic from historical data successfully
- A system to filter human errors
  - ~10% historical invoices are flagged as being mislabeled or contents being too vague
- Insights from model outputs useful to guide improvement of data hierarchy
Key Findings

Artificial Neural Network (ANN) model learned categorization logic
There was a 10% difference between model categorization and human categorization of vendor invoices

The data suggests that the model better / more accurately categorizes invoices (for procurement purposes)

Given these findings, automated machine learning invoice mapping would likely help enable:

More efficient and effective analysis and management of procurement spend

A significant increase in the productivity and accuracy of invoice categorization
Questions

andy@preminor.com
+1.403.909.7677