

The Company

OMV is an integrated, international oil & gas company with activities in the upstream (exploration and production) and downstream (refineries and marketing, incl. petrochemie, gas and power) area. With an annual revenue of EUR 36 billion, 25.501 employees and a market capital of EUR 7 billion by the end of 2014, OMV Aktiengesellschaft is one of the biggest industrial companies in Austria listed on the stock market.

Learning the Lesson

OMV Aktiengesellschaft (OMV) is currently using a HSSE performance reporting tool to track findings, hazards and incidents. The data contained is constantly analyzed and reviewed by OMV management.

During data entry users are manually classifying information for later reuse. Statistical reports and KPI reporting's are leveraged to create positive learn effects, increase safety standards and drive a behavioral change on day to day jobs, with the overall goal to reduce incidents.

Raising the Bar

OMV puts high efforts in incident reporting and follow up on lessons learned in order to prevent recurrence. This strong leadership focus resulted in a reduction of Lost-Time Injury Rate year over year. With an already low Lost-Time Injury Rate of 0.52 in 2014, going down 56% since 2008, OMV is searching for further ways to identify and eliminate any potential root cause for harm to people, environment or assets.

Data in the system is reported in various languages and reporting is based on the classification the user provides. Within a pre-project workshop MIOsoft data scientist in conjunction with OMV domain experts analyzed data contained in the system using techniques for semantic text analysis like rule-based analysis and machine learning algorithms. The goal was to verify the possible options of using a machine based classification system vs. the existing human classification approach to identify potential improvement areas.

Results from the workshop were very promising and indicating that the machine based classification approach will raise the bar for reported data quality.

Find Truth in Data

Working in close relationship with OMV, MIOsoft conducted a deep dive analysis on existing data using advanced data quality analysis techniques like



Company

OMV Aktiengesellschaft
(OMV)

Industry

Oil & Gas

Objectives

- Gain new insights on existing data
- Use machine learning techniques to assess data quality
- Define new categorization scheme to make data entry more efficient and increase data accuracy and consistency

MIOsoft Advantage

- MIOsoft's tool set for advanced data analytics
- MIOsoft's Data First approach to contextualize structured and unstructured data
- Out of the box thinking

Customer Benefits

- Deep insight's in current reporting behavior
- Improve data quality and report trustworthiness
- Baseline for new KPI's measuring safety leading performance

Comprehensive language recognition

OMV incident data is reported in different languages. Automated language recognition technique was used to identify the main language used on every text fragment. Data was analyzed then specifically by the detected language.

Data normalization

Before any automated classification routine could be used, data needed to be normalized, spelling errors detected and fragments using a porter stemming algorithm were created.

Visual data analysis (iterative approach)

Domain experts from OMV and MIOsoft data scientist took a deep dive into the data by using techniques like visual data inspection, tokenizing, statistical word counts, word combination / pattern detection, deep dive drill down, fuzzy search, common expressions and contextual data analysis.

Extract entity definitions and relationships

Using contextual data clustering, data objects were categorized and matching lists were created.

Rule based semantic analysis

In order to normalize the existing classification a rule based algorithm to cluster incident and finding records based on the free text fields within the defined classification matrix was used. The results gave a good indication for new classification schemes. For iterative adjustments a token based classification module with backtracking algorithm was used.

Clustering based on machine learning algorithm

To cluster the data in different topic areas, several machine learning algorithms were evaluated, finally the TF-IDF & k-Medoids machine learning algorithm was applied to cluster data.

Customer Benefit

As a result of the data analysis on OMV incident and findings data MIOsoft was able to identify areas with a high classification quality - and areas, where data was not consistently classified along the process definition.

Using the semantic text analysis approach OMV did leverage MIOsoft's data first approach to identify classification patterns in unstructured text. By this, the classification schemes could be streamlined to the user's needs, thereby data quality is aimed to be increased and time users spent on data entry should be reduced. Machine based classification was used to reclassify existing data and to normalize the existing classification.

Through new insights in potential root causes for incidents, OMV will be able to define new KPI sets which will help to further improve OMV's already high safety standards.

“MIOsoft’s unique approach to data analysis enabled us to better understand our incident data. This will help us improving our reporting process.”

Robert Lackner, OMV
Head of Safety
GTR-S Corporate Safety

