SEDGE THE SOLUTIONS EDGE

OIL TANKER VETTING ANALYSIS

A white paper on factors influencing the success of Oil Tanker Vetting

www.sedge.ai

Vetting Analysis Project

Introduction

The project was initiated by Oil Tanker company, to analyze Tanker Vetting data for a period of 2 year. The objective of the exercise was to load the tanker vetting data in SEDGE, and to understand the fields which have a high correlation between the Inspection class and other fields, which determines whether the ship has passed its vetting inspection with G- Good, B-Bad, F-Fair or E-Excellent status. The study was appreciated by our client and the client comments as to how the value was generated is mentioned below.

Client feedback of Study

Solverminds used its innovative machine learning and predictive modeling tool SEDGE to mine Teekay Tankers' Ship Management vetting inspection data

Whilst Teekay Tankers OCIMF SIRE Vetting inspection results are amongst the best in the industry, to continue to be a leader in this business critical metric, a project was undertaken to establish the main criteria that will significantly move the needle to further improve fleet-wide inspection results

SEDGE identified key attributes and their correlation (Linear & Non Linear) that influence vetting inspection outcomes. SEDGE automatically mined the order of importance, cluster and hierarchy of each attribute. These findings are instrumental in shaping up initiatives taken to achieve 2016 vetting inspections goals. Solverminds is a great partner to work with. Entire project was executed promptly and to full satisfaction of Teekay Tankers Ship Management

1 https://en.wikipedia.org/wiki/Distance_correlation

Background

Vetting is an inspection of the ship by the charterers, terminal operators and government organization, to understand whether the ship is complying with all rules and regulations in order for the vessel to transport the cargo in a safe manner and for the assurance of ship safety standards.

The vetting inspection usually covers the following - safety management systems (SMS), regulation as published by International Maritime Organization (IMO), Vessel Checklists, compliance of Safety of Life at Sea (SOLAS), Marine Pollution (MARPOL), Oil companies International Marine Forum (OCIMF) publications as well as the international guide of tankers.

Oil tanker vetting inspection is carried out by Qualified Ship Inspection Report (SIRE) inspectors, mainly on behalf of Oil majors or Oil terminal. Inspectors would report deficiencies and provide the degree of the deficiency. These deficiencies are listed in an observations list, and the feedback is provided to the vessels at the end of each inspection. Vessels with SIRE inspection deficiencies cannot be used until all observations are closed and the vessel is found suitable for operations by a vetting superintendent.

Methodology

SVM team received data which contained the details of Fleet Overview, period-based budget data, purchase order, Unplanned Maintenance, Internal Inspection, Vetting Performance, Crew List & Period, Port Calls, and HSEQ Risk details. Each category was then further sub divided in to details, and fields which were considered relevant in the exercise were added and fields which were considered not relevant were excluded.

Each record was then loaded in SEDGE and data cleaning and data wrangling was performed and any missing data was imputed and the data was then analyzed. Certain fields were excluded from the data, as they had high cardinality.

Data Exploration

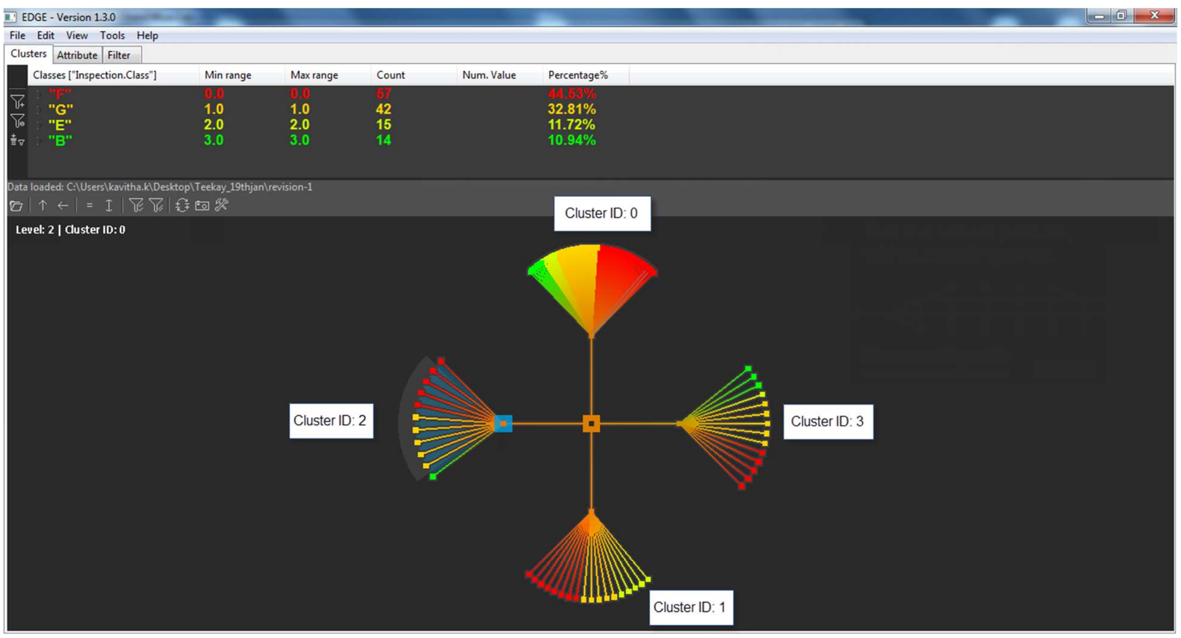
The data was loaded into the SEDGE and the system produced cluster as shown below. The field Inspection Class was used as Target variable. The target variable is the field which is of interest as this field determines as to whether the vessel passes the vetting inspection. Four categories were created based on the fields “Inspection avoidable” and “Inspection observations”. Within the Inspection class, the data “E” denotes Excellent, “G” denotes Good, “F” denotes Fair and “B” denotes Bad. These Inspection Classes based on the following formula

Observations	Avoidable	Inspection Class
0	0	Excellent
<=2	<=1	Good
<=5	<= 2	Fair
>5	>2	Bad

The SEDGE tool produced clusters of data, based on the vetting Inspection class, brown denoting Good, Yellow denoting excellent, Red denoting Fair and Green denoting Bad.

Global cluster view

Please find below the cluster which the SEDGE tool produced. Total no of clusters created were 4. Each cluster has specific characteristics, and the cluster which creates a clear segregation of Excellent, Good, Fair and Bad is the ideal cluster.

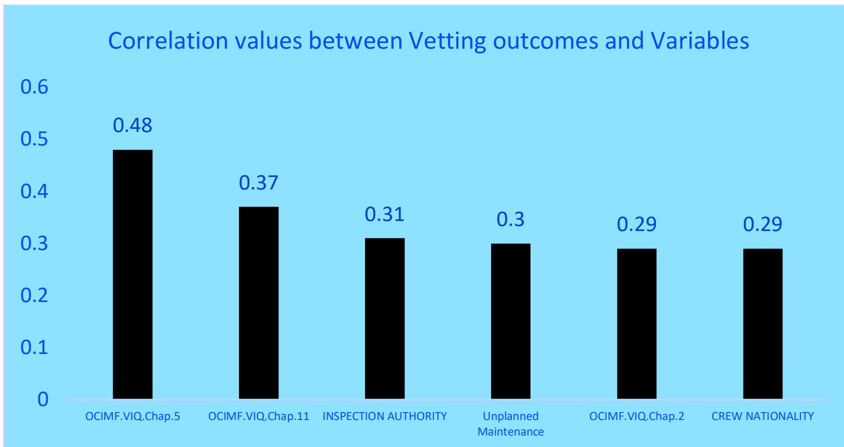


The clustering analysis produced result where features where identified which segregated the bad category from the other category. These features were further analyzed to extract the details of the specific feature influencing the vetting outcomes.

Table below defines the fields which have correlation in the Global cluster data, and these were ranked in the order of top ones have high correlation and the bottom ones, having least correlation. For calculation we have used both linear as well as non-linear.

no	Attribute	Correlation	Type of Correlation1
1	OCIMF.VIQ.Chap.5	0.48	RDC
2	OCIMF.VIQ.Chap.11	0.37	RDC
3	INSPECTION AUTHORITY	0.31	RDC
4	Unplanned Maintenance	0.30	RDC
5	OCIMF.VIQ.Chap.2	0.29	RDC
6	CREW NATIONALITY	0.29	Pearson

Table:1 *Actual data from client is protected therefore presenting simulated data to share methodology of study*



Above are examples of some of the variables which had correlation analysis, there were further fields which also generated correlation, but were excluded from this document.

Individual field Analysis - Global Cluster

To understand the impact of each field, and how the same influences, analysis was done on the individual fields which SEDGE produced and compared them against the success %age (Inspection class “E” and “G”).

OCIMF.VIQ. Chapter 5

The field OCIMF VIQ.No.5 has the highest correlation 0.48, with the target variable (Inspection class). OCIMF Chapter 5 relates to Safety Management. Below we have analyzed this field on an individual level basis, and following is the tabular analysis.

1 The type of Correlation used and the detailed explanation of each is shown in the Appendix 1

Actual data from client is protected therefore presenting simulated data to share methodology of study

Data Value	B	E	F	G	Total	G+E	F+B	Success %age ((G+E)/Total)*100%
0	5	17	28	39	89	56	33	63%
1	7	4	32	15	58	19	39	33%
2	8	1	4	3	16	4	12	25%
3	5	2	2		9	2	7	22%
Grand Total	25	24	66	57	172	81	91	

Table:2 *Actual data from client is protected therefore presenting simulated data to share methodology of study*

The OCIMF chapter 5 data was taken and the data was then split into multiple Inspection class. A field called success %age was created which was the success of data which had E and G as the data value. The field success % age was calculated ((G+E) / (Total))*100. This gave a success %age for each data value, and it was observed that the OCIMF 5 with data value 0 had a high success % age of 63%, as compared to 33% for Data value 1 and 25%, 22% for data value 2 and 3 respectively. In summary if the OCIMF 5 value is 0, there is high probability (63%) that the Inspection class will be E or G. For data value 1, there were 33% chance of success and 67% of failure. A big chunk of the data value 1 category (more than 50%) corresponds with Inspection class F, and small number of data corresponded with E or G. If the Data value is 2 or 3, then there is a high probability that the Inspection class will be B or F.

On analyzing this field with the vessel, found that vessel ABC and DEF both had an inherent issue with OCIMF chapter 5 with data values at 2 or 3. For the ship GHI, it was observed that there were 3 vettings, the first in DD/MM/YY, which the ship passed, subsequently there were 2 vetting's on DD/MM/YY and both these vettings the ship had OCIMF 5 with data value 2, which also means that the inspectors do look for defects based on the previous vetting, as a guideline or it is possible that the issues reported in previous vetting was not fixed subsequently.

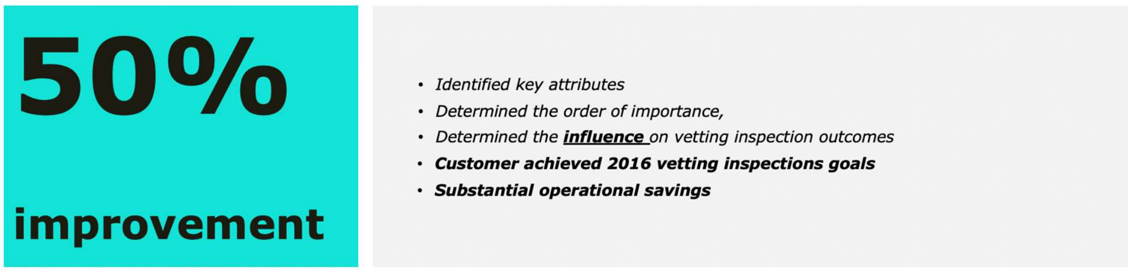
A further analysis was done for other variables such as OCIMF Chapter 11, OCIMF Chapter 2, Crew Nationality, unplanned maintenance and others. The details of these analysis are not published in this document to maintain confidentiality.

Conclusion

A detailed analysis of various fields which influence the vetting outcome was performed. It was observed that observations in OCIMF Chapters had a strong correlation to the outcome. In addition to the OCIMF chapters, other variables which had strong influencing factors were inspection authority. Inspection authorities also had an influence on the success, where the probability of Inspection class being B and F was very high. Since the inspection authority had same inspectors, analysis of inspectors and their rating of Inspection. Unplanned maintenance was an influencing factor for the vetting outcome, as compared to planned maintenance.

There were other factors such as number of crew nationalities within the ship's crew which also had an influencing fact to the number of observations. Communication between the staff was a key factor for this factor.

Business Impact



Post Study Analysis Observation

Based on the above study, the Oil tanker company identified the variables in the clusters which influenced the success of vetting. The company created policies and action plan, which were implemented to ensure that the number of observations on the key variables were minimized.

One year after the implementation of action plan, the number of observations significantly dropped which meant that more ships were accepted by oil majors, generating huge value for the Oil Tanker company. The study had a profound effect on the vetting success of the Oil Tanker company.

Recommendations

Oil tanker companies should collate similar data, and periodically perform analysis to see any variation from the above analysis. This will give oil company an edge in deeper understanding of the factors which influence the outcome of inspection and contribute to environmental sustainability, operational efficiency and commercial success.

Appendix A

Explanation on various types of correlation used for analysis-.

Correlation coefficient- A correlation coefficient is a coefficient that illustrates a quantitative measure of some type of correlation and dependence, meaning statistical relationships between two or more random variables or observed data values.

Types of correlation coefficients include:

Pearson product-moment correlation[2]

Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviations. The form of the definition involves a "product moment", that is, the mean (the first moment about the origin) of the product of the mean-adjusted random variables; hence the modifier product-moment in the name.

Spearman Correlation[3]

The Spearman correlation coefficient is defined as the Pearson correlation coefficient between the ranked variables. For a sample of size n, the n raw scores X_i, Y_i are converted to ranks x_i, y_i, and ρ is computed from:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

where $d_i = x_i - y_i$, is the difference between ranks.

Positive and negative Spearman rank correlations- The sign of the Spearman correlation indicates the direction of association between X (the independent variable) and Y (the dependent variable). If Y tends to increase when X increases, the Spearman correlation coefficient is positive. If Y tends to decrease when X increases, the Spearman correlation coefficient is negative. A Spearman correlation of zero indicates that there is no tendency for Y to either increase or decrease when X increases. The Spearman correlation increases in magnitude as X and Y become closer to being perfect monotone functions of each other. When X and Y are perfectly monotonically related, the Spearman correlation coefficient becomes 1. A perfect monotone increasing relationship implies that for any two pairs of data values Xi, Yi and Xj, Yj, that Xi − Xj and Yi − Yj always have the same sign. A perfect monotone decreasing relationship implies that these differences always have opposite signs.

---

2 https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient

3 https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient

1 https://en.wikipedia.org/wiki/Distance_correlation

The Spearman correlation coefficient is often described as being "nonparametric". This can have two meanings. First, the fact that a perfect Spearman correlation results when X and Y are related by any monotonic function can be contrasted with the Pearson correlation, which only gives a perfect value when X and Y are related by a linear function. The other sense in which the Spearman correlation is nonparametric in that its exact sampling distribution can be obtained without requiring knowledge (i.e., knowing the parameters) of the joint probability distribution of X and Y.

DCOR[4]

In statistics and in probability theory, distance correlation is a measure of statistical dependence between two random variables or two random vectors of arbitrary, not necessarily equal dimension. An important property is that this measure of dependence is zero if and only if the random variables are statistically independent. This measure is derived from a number of other quantities that are used in its specification, specifically: distance variance, distance standard deviation and distance covariance. These take the same roles as the ordinary moments with corresponding names in the specification of the Pearson product-moment correlation coefficient.

These distance-based measures can be put into an indirect relationship to the ordinary moments by an alternative formulation (described below) using ideas related to Brownian motion, and this has led to the use of names such as Brownian covariance and Brownian distance covariance.
Maximal Information Coefficient- MIC[5]

In statistics, the maximal information coefficient (MIC) is a measure of the strength of the linear or non-linear association between two variables X and Y.

The MIC belongs to the maximal information-based nonparametric exploration (MINE) class of statistics. In a simulation study, MIC outperformed some selected low power tests, however concerns have been raised regarding reduced statistical power in detecting some associations in settings with low sample size when compared to powerful methods such as distance correlation and HHG. Comparisons with these methods, in which MIC was outperformed, were made in and. It is claimed that MIC approximately satisfies a property called equitability, which is illustrated by selected simulation studies

Randomized Dependence Coefficient - RDC Correlation[6]

---

2 https://en.wikipedia.org/wiki/Distance_correlation

3 https://en.wikipedia.org/wiki/Maximal_information_coefficient

4 https://en.wikipedia.org/wiki/Distance_correlation6

5 https://en.wikipedia.org/wiki/Maximal_information_coefficient

In statistics, dependence is any statistical relationship between two random variables or two sets of data. Correlation refers to any of a broad class of statistical relationships involving dependence. Familiar examples of dependent phenomena include the correlation between the physical statures of parents and their offspring, and the correlation between the demand for a product and its price.

Correlations are useful because they can indicate a predictive relationship that can be exploited in practice. For example, an electrical utility may produce less power on a mild day based on the correlation between electricity demand and weather. In this example there is a causal relationship, because extreme weather causes people to use more electricity for heating or cooling; however, statistical dependence is not sufficient to demonstrate the presence of such a causal relationship (i.e., correlation does not imply causation).

6 http://arxiv.org/abs/1304.7717

1 https://en.wikipedia.org/wiki/Distance correlation