# Towards the Forecasting of Delays in Supply of Offshore Platforms

**Maricruz A. F. Cepeda**
Federal University of Rio de Janeiro, Brazil.
maricruzcepeda@oceanica.ufrj.br

**Rafael Basilio da Silva**
Federal University of Rio de Janeiro, Brazil.
rafaelbasilio@poli.ufrj.br

**Jean-David Caprace**
Federal University of Rio de Janeiro, Brazil.
jdcaprace@oceanica.ufrj.br

**Título: Pronóstico de Retrasos en el Suministro de Plataformas Offshore.**

**Resumen:**
Actualmente, la industria Offshore de Petróleo pretende mejorar la logística relacionada con la producción petrolera. El suministro de plataformas marinas requiere un alto nivel de servicio para utilizar el mínimo de recursos. Las condiciones meteorológicas y buques cesados son las principales variables que inducen retrasos en la planificación de operaciones, lo cual afecta la producción, almacenamiento y descarga de petróleo en unidades marinas (FPSO). El indicador clave de rendimiento (KPI) real muestra un potencial de mejora. El objetivo de este trabajo, es pronosticar los retrasos producidos durante la entrega de los suministros para plataformas, costa afuera operadas en Brasil. La base de datos analizada presenta 2,851 viajes de una-vía de barcos de suministro de plataformas (PSV). El objetivo es desarrollar una metodología basada en minería de datos, utilizar el algoritmo A priori, árbol de decisiones y Perceptrón multicapa (MLP). El tipo de carga (carga seca y líquida al granel o contenedores), la prioridad de carga y el tipo de operación (carga, descarga o transbordo) son algunos de los parámetros de entrada considerados en el modelo. Los resultados proporcionan una nueva manera de abordar la eficiencia y el rendimiento de suministro de la logística a las plataformas, incluso, si el modelo requiere mejoras futuras. Conocer de antemano cuáles serán los retrasos más susceptibles que podrían ocurrir en la cadena de suministro, ayuda a los planificadores a anticipar sus estrategias y rutas de entrega.

***Palabras Claves:*** *Costa afuera, cadena de suministro, logística, buque de suministro mar adentro, plataforma offshore*

**Title: Towards the Forecasting of Delays in Supply of Offshore Platforms.**

**Abstract:** Nowadays, the oil Offshore industry seeks to improve the logistic related to oil production. Supply of Offshore platforms requires an elevated level of service using minimal resources. Weather conditions and vessel off-hire are the main variables that induce delays in operations planning, which may affect the oil production of floating production storage and offloading (FPSO) marine units. Actual key performance indicator (KPI) related to logistic shows a potential for improvement. The objective of this paper is to forecast the delays that are occurring during the delivery of the supplies for offshore platforms operated in Brazil. The database analyzed in this study present 2851 platform supply vessels (PSV) one-way travels. To achieve this goal, a methodology based on data mining is developed using A priori algorithm, decision tree and multi-layer perceptron (MLP). The type of cargo (dry bulk, liquid bulk or container), the cargo priority, and type of operation (load, backload or transshipment) are some of the input parameters that has been considered in the model. The findings provide a new way to address efficiency and performance supply logistics of Offshore platforms even if some future model improvements are required. Knowing in advance were the delays are more susceptible to occurs in the supply chain allows the planners to anticipate their strategies and delivery routes.

**Key Words:** Offshore, Supply Chain, Logistic, Platform Supply Vessel, Offshore Platform.

**1. Introducción:**

Nowadays, around a third of the oil & gas extracted worldwide comes from offshore production, (Ocean, Institute, & Mare, 2014). In addition, the distance between offshore units is constantly increasing. This explain why the offshore logistics uses constantly increasing amount of infrastructure to maintain and develop operations of marine units, composed by airports, ports, hubs, warehouses, specialized vessels, among other, (F. Cepeda, da Silva, & Caprace, 2015). Therefore, the study of logistics between a land warehouse and the offshore installations can be a complex problem to solve (Nordbo, 2013).

Key challenges must be addressed in order to find a suitable level of operations planning to use a minimum amount of resources with high service level. Logistics is fundamental in the petroleum offshore industry and optimizing the transport logistics network thus becomes an economically critical issue. The presence of stochastic elements in cargo transportation to and from offshore units affects the offshore supply operations planning justifies the implementation of an analysis and prediction approach. Today, there are few researches that consider planning and logistics problems in the offshore supply chain and even less that take into account the uncertainty and possible disruptions that may arise in the logistic chain.

A Maritime Cargo Fulfillment Indicator (MCFI) has been developed in this study. This indicator measures the percentage of cargo supplied before deadline in relation to total amount of cargo supplied to offshore units for a period of one month. For logistics operations, the major challenge is to keep the indicator above 76%. Various issues affect this indicator such as weather conditions, vessels off-hire (near 10%), availability of deck area and number of vacant berths, unexpected failures, multiple goods and priority of some cargoes.

This paper is focusing the analysis in the logistic of campos brazilian basin. Distance to offshore units is one of the main factor that affects the logistic operational performance as Brazil oil production is concentrated more and more in deep seas.

The logistics process has the following stages: transport request creating, picking list generating, cargo picking start, cargo picking and cargo consolidation end, transport request releasing, transport request fulfillment creating, cargo delivered to port, and cargo delivered to offshore unit. Materials for the units can be stored either in own warehouses or supplier warehouses. The requests orders are used to plan the available supply vessels after analysis and consolidation of the data. Some information as cargo weight and dimensions are important because it limits the amount of cargo to be loaded on the ships. Furthermore, it is very important to taking into account the delivery deadline. The earliest date and latest date are associated to each request. The quality of the service will be affected whether the cargo is delivered on time or not. (Ferreira Filho, 2015).

The principal objective of this paper is to analyze the request order database and recovers hidden information, that can be found through data mining.

Three different algorithms have been applied. First, association rules have been detected amongst available transactions on the database such as cargo weight versus delay status, cargo weight versus type of operations, etc. Then, decision trees and multilayer perceptron neural network models were used to carry out predictions on cargo delays.

**2. Methodology and Data Model:**

Data Mining (DM) is the analysis of datasets to find patterns, associations or unsuspected relationships and to summarize the data in understandable and useful ways to the data owner (Hand, Mannila, & Smyth, 2001). DM uses various analytical techniques and involves the creation of a model, so that the concluded result will become useful information or knowledge, (Ting, Tse, Ho, Chung, & Pang, 2014). DM also uses different algorithms to predict information about entire project. Finally, DM is a multi-disciplinary field that is at the intersection of statistics, machine learning, DB management, and data visualization (Feelders, Danielsa, & Holsheimer, 2000).

The analyzed DB consists of a list of cargoes orders, which are supposed to be delivered to offshore units by PSVs. Each one of 239 921 rows of the DB represent a transport request item and have key information about cargo which will be supplied to offshore units.

Extreme outlier's values were excluded using Equation 1 and 2. Outliers values represent 24 116 records on 239 921 which mean 10% of the DB. Where p are data points, Q1 is the lower quartile, Q3 is the upper quartile, and IQR is the distance between Q1 and Q3.

$$p < Q_1 - (3 \cdot IQR) \quad (1)$$
$$p > Q_3 + (3 \cdot IQR) \quad (2)$$

In this study, the MCFI is measured to know the logistic operations performance and in the first semester of 2014 the goal was not reached (MCFI>76%). The average was 61% (minimum value 52% and maximum value 67%). The delay measured in days is the difference between delivery date and deadline. The higher frequency is presented between 1 to 3 days of delay. A classification of the delays has been defined as follow: 0 to 1 days (40%) [Much Lower - 37375 records], 2 to 3 days (~35%) [Lower - 30060 records], 4 to 5 days (~14%) [About the same - 12602 records], 6 to 7 days (~4%) [Higher - 4234 records], and more than 7 days (~6%) [Much Higher - 6311 records].

In addition to the fact that logistics service has not performed above the desired goal (MCFI > 76%), the percentage of cargoes delayed plus one day is around 40% of the total amount of delays, which can be considered very high.
The proportion of normal operations are about 84% and emergency priorities operations are 16%, the proportion of each type of operations are Load 70%, Backload 28% and Transshipment 2% over the analyzed period.

Cargoes movements can be considered as one of the most important operations, because it could affect oil production. Indeed, some cargoes have a huge impact on the development of the production and the habitability (foods and water) of the offshore units. We can also conclude that the amount of transport requests items delivered after deadline is higher for load than backload and transshipment operations. The proportion of delayed items is respectively 48%, 18% and 23% for load, backload and transshipment operations. The load frequency per weekday over the covered period is in average 14% per day, the higher percentage is on Sundays (17%) and lower percentage is on Fridays (13%), that is the distribution of delays is almost constant over the week. The cargoes were delivered before the deadline (within of time) are 64% and out of time are 36%.

### 3. Results and Discussions

Modeling step has been divided in three stages explained in this section. First, association rules inference algorithm (A priori) is applied on the DB to find possible relations between transactions such as cargo weight versus delay status, cargo weight versus type of operations, etc. Then, multilayer perceptron neural network models is used to carry out predictions on cargo delays starting from voyages features such as weekdays, origin, destination, cargo class and type of operations. Finally, decision tree model is used to undertake predictions about delay status and type of operations (load, backload and transshipment). The load order request DB (239 921 rows) have been gathered in 2 851 voyages. Selected input and outputs of the different models are presented in 1.

*Table 1 – Input and output values of the models*

| Item number | Inputs name | Outputs |
|---|---|---|
| 1 | Quantity of cargo items | 1. A priori: Association rules<br>2. Multilayer Perceptron: Average of delay days<br>3. Decision Trees: Average of delay day |
| 2 | Total weight of Cargo items | |
| 3 | Average Delay days | |
| 4 | Priority: Normal (N) or Emergency (E) | |
| 5 | Cargo Classification Type: General Cargo (GC), Liquid bulk Cargo (LC), or Dry Bulk Cargo (DC). | |
| 6 | Type of shipment: Load, Backload, or Transshipment | |
| 7 | Status of delay: Within of time or Out of time | |
| 8 | Ship ID | |

### 3.1 Apriori Algorithm

A priori model is performed by recognizing the most frequent individual items in the DB and extending them to even larger item sets as long as those item sets appear sufficiently often in the DB to determine any possible association rules, which allows to identify possible trends in the DB.

Support and Confidence are major indicators in this technology. Support indicator is defined as a fraction of transactions that contain an item set, i.e, is the frequency of occurrence of an item set. A frequent item set is one whose support is equal or greater than a threshold (minus in the algorithm). Thus, confidence measures how often items in Y appear in transactions that contain X (Agrawal R., 1994). X and Y are called precedent and consequent of the rule respectively, see Table 2. For this work we used an A priori algorithm. Confidence and support values have respectively set to 90% and 80%.

A priori model have provided relations that were not known previously. Table  shows the fifteen association rules considered satisfactory from an operational perspective.

*Table 2 - A priori Results: Association Rules*

| Consequent | Precedent |
|---|---|
| Delays average | Backload, Dry bulk |
| Delays average | Emergency |
| Delays average | Transshipment, Liquid Bulk |
| Transshipment | Outdated |
| Backload | Delays average, Emergency |
| Transshipment | Outdated, Delays average |
| Delays average | Emergency, Transshipment |
| Outdated | Emergency |
| Backload | Outdated, Emergency |
| Backload | Delays average |
| Delays average | Transshipment |
| Delays average | Dry Bulk |
| Dry Bulk | Backload |
| Backload | Delays average, Outdated |
| Backload | Emergency, Delays average |

### 3.2 Multilayer Perceptron (MLP)

A MLP neural network is a feedforward artificial neural network model that maps sets of input data on a set of appropriate outputs. A MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLP utilizes a supervised learning technique, called back-propagation for training the network (Rosenblatt, 1961), (Rumelhart, 1986).

The learning set of the MLP has been set to 70% while test set used 30% of the data available. MLP has been used to predict Average of delay days for each trip. The prediction carried out for delays average on logarithm base shows good correlation as we can see in *Figure 1*, R2 is 0.991, mean absolute error is 0.007, mean squared error is 0.001, root squared error is 0.026, and mean signed difference is 0.001. This gives a good confidence in the results of the model.

### 3.3 Decision Trees (DT)

A DT can be used as a model for sequential decision problems under uncertainty. This describes graphically the decisions to be made, events that may occur, and outcomes associated with combinations of decisions and events. Probabilities are assigned to events, and values are determined for each outcome. A major goal of the analysis is to determine the best decisions (Middleton, 2015). DT are commonly used in operational research, specifically in decision analysis, to help to identify a strategy most likely to reach a goal. Another use of decision trees is as a descriptive means for calculating conditional probabilities. (Quinlan, 1987), (Y. Yuan, 1995).

The learning set of the DT has been set to 70% (1 995 records), of the data while test set used 30% (856 records), of the data available. Decision tree, have been used to predict average of delay days for each trip. *Figure 2* shows results for this model, which provided a good accuracy (99.87%). Indeed, error is 0.121% in relation to prediction delays average, and wrong classified one on 824.

*Figure 1 – Multilayer Perceptron Neural Network error of the test set, X axis: LOG average delay days; Y axis: prediction average delay days.*

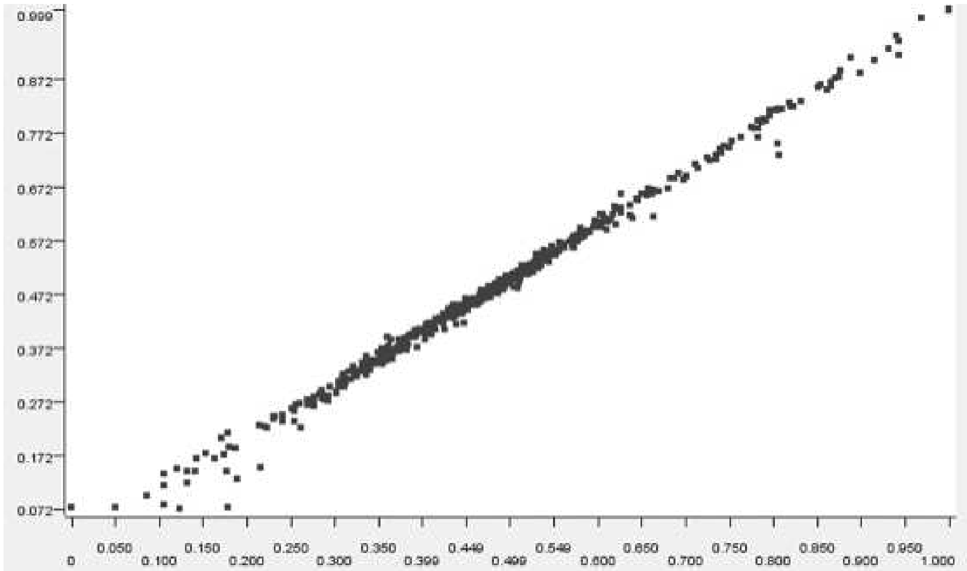
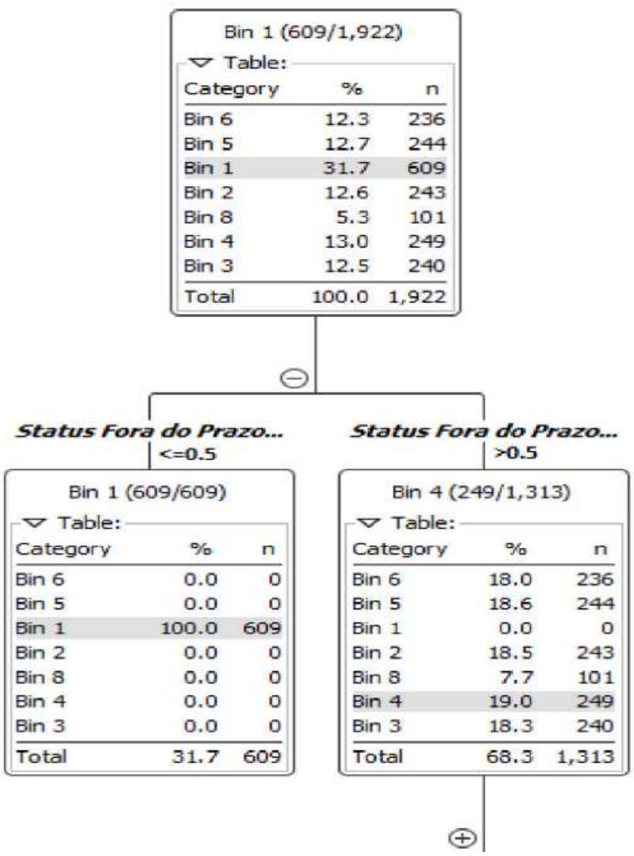



*Figure 2 – Decision Trees Results: View and Accuracy*



## 4. Conclusions and Future Work

This paper presented an application of machine learning algorithm on the analysis and prediction of offshore supply delays. Three models have been used, the A priori association rules, the multilayer perceptron neural network and the decision tree. All presented a great reliability (accuracy better than 95%) during testing, (test set have been set up to 30% of available data).

Planners can apply the model proposed to analyze the delays of the PSV fleet.

Knowing in advance were the delays are more susceptible to occurs in the supply chain allows the planners to anticipate their strategies and delivery routes. The findings provide a new way to address efficiency and performance supply logistics of offshore platforms even if some future model improvements are required.

It is interesting to highlight that the most of working time have been spent for data treatment.

A potential future work, can focus the development of the operation performance over time and included as well as more input data, e.g. one year span instead of a semester.

## 5. References

(1) Agrawal R., S. R. Fast Algorithms for Mining Association Rules, Proceedings of the 20th VLDB Conference. Santiago, 1994.

(2) Feelders, A., Danielsa, H., & Holsheimer, M.,. Methodological and practical aspects of data mining. Information & Management, 37(5), 271–281. doi:10.1016/S0378-7206(99)00051-8, 2000.

(3) Ferreira Filho, V. J.,Gestão de Operações e Logística na Produção de Petróleo. Rio de Janeiro, RJ, Brazil: ELSEVIER, 2015.

(5) Hand, D., Mannila, H., & Smyth, P. , Principles of Data Mining. Cambridge, Massachusetts: MIT Press, 2001.

(6) Leite, R., Maritime transport of deck Cargo to Petrobras fields in Campos Basin: an empirical analysis, identification and qualification of improvement points. Rio de Janeiro: PUC-RIO., 2012.

(7) Middleton, M. R., Introduction to Decision Trees. In M. R. Middleton, TreePlan Tutorial (pp. 157-168), 2015

(8) Quinlan,J. R., Simplifying decision trees. International Journal of Man-Machine Studies 27 (3): 221. doi:10.1016/S0020-7373(87)80053-6., 1987.

(9) Rosenblatt, F. x., Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Washington DC: Spartan Books, 1961

(10) Rumelhart, D. E., Learning Internal Representations by Error Propagation (Vol. Volume 1: Foundations). MIT Press, 1986.

(11) Ting, S., Tse, Y., Ho, G., Chung, S., & Pang, G. (2014). Mining logistics data to assure the quality in a sustainable food supply chain: A case in the red wine industry. International Journal of Production Economics, 200–209, 2014.

(12) Y. Yuan, S. M. (1995). Induction of fuzzy decision trees. Fuzzy Sets and Systems 69, 125–139, 1995.