Categoría: Marque con una "X"

- Artículo Técnico [x]
- Tesis Pregrado [ ]
- Tesis Posgrado [ ]

ACIPET

Supply Chain Towards Smart Management in Oil&Gas

Edgar A Camargo C, 3SAiTech Energy, 3saitech@gmail.com
Angel E Lacret S, Cobuildlab, alacret@gmail.com,+1(786)991-3467

Abstract— In this article, an Autonomous Cycle of Data Analysis Tasks is presented, to evaluate the behavior of the supply chain in Oil&Gas. The model considers, among other things, the market situation, commercial events that affect production, productive behavior of the countries, etc. The purpose of this paper is to investigate the role of supply-chain management in the oil and gas industry. So, Data Mining Techniques are applied to each task of the autonomic cycle. For example, to predict Monitoring events, Identifying Patterns in Oil Prices and Prediction of Behavior Prices Oil. Also, to establish the price behavior of the product, several predictive models were developed, based on Learning Machines (for example, Support Vector Machines, Random Forest, Gradient Boosting, and Decision Trees). On the other hand, for the development of the autonomic cycle that allows analyzing the behavior of the price of the product, in this work, the MIDANO methodology ("Methodology for the development of Data Mining applications based on organizational analysis") is used in its three phases.

Keywords—Supply Chain Management Oil&Gas, Autonomous Cycle, Smart Supervisor, Autonomous Computing, Machine Learning.

Introduction

The manufacturing industry has faced growing challenges in recent years. Technologies such as Cloud computing, Cyber-physical systems, The Internet of Things, Blockchain, Edge computing, etc. So, manufacturers have benefitted from data‑driven innovations for demand planning and logistics management of their supply chains but Tracking production across entire processes and managing the supply chain as an integrated platform is now an urgent need. Hence, there is a demand to further explore the futuristic technologies for intelligent manufacturing and supply chain management since Supply Chain is one of the trickiest parts of a business and as the chances of error are quite high, the need for proper synchronization and analytics is very important. Something simple getting out of hand can make the business lose a lot of money as it can create a break or kink in the supply chain. When it comes to strategic and operational parts of the Supply Chain Management, the need for resolving issues on the go to keep the chain continuous is very important or every step in the continuation can become a pain point that will block the resources in the wrong places. Given the many steps involved in energy processing, transportation and logistics, the oil and gas supply chain is understandably complex. The oil and gas industry operates through a supply chain(SCMO&G) that includes domestic and international transport, trading, shipping, ordering, and inventory visibility and control. Other supply chain elements include material handling, import/export facilities, and the distribution of refined energy products from points of origin to market. So, typically, the supply chain is divided into three segments. The upstream segment finds and produces crude oil and natural gas. The midstream segment handles the processing, storing and transporting of energy commodities. And the downstream segment encompasses oil refineries, retail outlets and natural gas distribution companies. So, Yet with global competition on the rise, industry professionals find that supply chain improvements are both necessary and cost effective. This is the reason why Artificial intelligence (AI) and Machine Learning [1] now play a leading role in enhancing the quality of the manufacturing process. With the use of machine learning in the supply chain, not only is the operation becoming easier but the services also get improved. So, Artificial intelligence and Machine learning [1], the importance of analytics has been generally perceived and played a critical role in Supply Chain Management (SCM).

In everyday terms, there are several types of assets an association is based on, such as innovative assets, specialized, administrative capabilities, and IT-based assets. To competently deal with these assets, organizations should have the option to adequately deal with their supply chains, keeping in mind the unique competitive environment in current world-class business environments. This underscores the need for effective reconciliation and joint effort among supply chain accomplices; this can be implemented with the latest improvements in innovation and results. These frameworks are, in general, with Supply Chain (SC) and attached data and financial flows. Generally [2], SCM tries to half-control or link the creation, shipping, and allocation of an item. By dealing with the supply chain, organizations can reduce excess redundancy costs and deliver products to the customer faster. This is complemented by tighter control of internal inventories, inside creation, circulation, deals, and inventories of organization vendors. SCM depends on the probability that almost every item that comes to display is due to the efforts of different associations that make up a supply chain. Although supply chains have been around for a very long time, most organizations have focused on them recently to add value to their tasks. The supply chain is the organization of different types, firms, entities, exercises, and technologies dealing with the assembly and offering of effective management. A supply chain begins with raw materials moving from a supplier to a manufacturer and closes with the finished product or administration moving to the end customer [2]. SCM oversees every touchpoint of an organization's element or management, from the start of creation to the final deal. With so many points along the supply chain that can gain reputation through efficiency or lose credibility due to extended costs, proper SCM can increase revenues, lower expenses, and affect an organization's core concern. SCM is often depicted as having five key components: arrangement, supply of raw materials, assembly, transportation, and returns. So, Today, there are more opportunities for coordinating activities across a supply-chain even in such complex operations as oil and gas, because of improving information systems and communication technologies. Integrating operations management with other functions of the operation allows all functions to be involved in the supply-chain management decisions

Some examples of works of this type are the following: In [3], an inductive methodology of meeting type was applied to obtain new observational data. The obtained observational data were then analyzed using the topical examination technique with the help of the previous literature and case studies. Moreover, the discoveries offer new methodologies and perspectives that the latest analytics apply at both basic and operational management levels that shape supply chain management plans. So, [4] This paper explores the effect of data-driven supply chain capabilities on transportation (train-based). To illustrate the shortest path calculation, London Underground Transportation's open- source data have been analyzed by implementing different data mining tools. The findings indicate that a data-driven supply chain has a significant time-efficient effect on logistics support. Coordination, using available data, and supply chain responsiveness are positively and significantly related to time and cost-efficient performance. [5] Based on the analysis of the existing traditional supply chain financial model, this article proposes an edge intelligence-enabled supply chain financial model based on Business-to-Business (B2B) platforms, combining the operation mechanism of the model and the quantitative analysis thinking of the traditional supply chain financing. This article uses the model to construct and evaluates the cost–benefit model of dealers, manufacturers, and B2B e-business platforms from the perspective of the supply chain finance for B2B platforms. [6] Results point to the central role of supply chain adaptability in capturing the benefits of supplier technological intelligence for enhanced product innovation capability, new product launch success, and firm financial performance. In contrast, product innovation capability serves as the generative means by which customer and competitor intelligence is translated into more successful new product launches, which, in turn, produce superior firm financial performance. Overall, these findings contribute to a better understanding of factors that can explain why certain product launches are more successful than others, and offer practical insights for appropriate investments in the development of related knowledge resources. [7] In This work focused on various research avenues in supply chain networks, particularly vendor selection, transportation, inventory routing, agent-based modeling, reverse logistics. In recent days researchers have triggered very high interest in Artificial Intelligence (AI) tools in the supply chain network problems. [8] Currently, Industry 4.0-based intelligent supply chain has attracted a lot of attention from academia; however, studies on performance measurement indicators of an intelligent supply chain are still lacking. To fill this gap, this paper first introduces the impact of Industry 4.0 on supply chain management. After analyzing the key characteristics of an intelligent supply chain, the paper proposes a performance measurement indicator framework consisting of seven indicators. This indicator framework enriches the theoretical knowledge of supply chain performance evaluation and provides an efficient way to improve the operational performance of intelligent supply chain management.

Now, this paper takes a different approach. First of all, it is part of the autonomous macro cycle for intelligent supervision, which integrates different data analysis techniques in an autonomous cycle called SAiSC (Intelligent Autonomous System for Supply Chain), to adapt the said process to world events. The concept of an autonomic cycle (AC) of data analysis tasks has been developed for the first time in [9], based on the idea of defining cycles of MAPE-K (Monitor-Analyze-Plan-Execute plus Knowledge) distributed in each one of the services of SOA applications (Service-oriented Architecture) [9-10], loosely coupled, to make them tolerant to faults. Other autonomic cycles have been developed in [10] for Learning Analysis tasks in an intelligent classroom (ACOLAT, Autonomic Cycle of Learning Analysis Tasks), where a set of learning analytics tasks are defined that allow improving the learning process of the students in the intelligent classroom, based on the analysis and interpretation of what happens in the environment. Within SAiSC, there are data analysis tasks dedicated to supervising the production process, tasks dedicated to optimizing it, and in particular, tasks that allow monitoring the behavior of the world market of the process under study. Each group of tasks makes up regional micro cycles with specific objectives, which are integrated into SAiSC, allowing the supply process to be optimized. So, the approach proposed in this article consists of an autonomous microcycle of data analysis tasks within SAiSC in charge of analyzing the behavior of product price (AB2P), considering the endogenous variables (for example, world supply of a product, world demand for product, the current price of a product and product inventories). In this sense, the autonomic micro-cycle of data analysis tasks considers a mix of data analytics (AdD). In this way, the cycle integrates data analysis techniques (such as machine learning and optimization lineal). In particular, the autonomic micro-cycle (AB2P) is composed of three (3) tasks; one for monitoring events that affect the endogenous variables that characterize the product market, another for identifying patterns in product, and finally, one for predicting the behavior of a product. In addition, this work uses the MIDANO methodology [11] for the design of the autonomic micro-cycle, as well as for the specification of data analysis tasks and data preparation.

The structure of the article is as follows: section 2 presents the theoretical framework used as a basis for the development of the autonomic micro-cycle. Section 3 describes the design of the autonomic micro-cycle, while section 4 deals with the experiments. Finally, section 4 deals with the conclusions of the work.

Theoretical framework

Supply chain management oil&gas

The supply chain of the oil and gas (SCMO&G) industry is divided into two different, yet closely related, major segments: the upstream and the downstream supply chains. [12]. The upstream supply chain involves the acquisition of crude oil, which is the specialty of the oil companies. The upstream process includes exploration, forecasting, production, and logistics management of delivering crude oil from remotely located oil wells to refineries. The downstream supply chain starts at the refinery, where the crude oil is manufactured into the consumable products that are the specialty of refineries and petrochemical companies. The downstream supply chain involves the process of forecasting, production, and the logistics management of delivering the crude oil derivatives to customers around the globe. Despite many enhancements in the processes and technologies introduced over the years, downstream distribution remains the main contributor to supply chain operation costs and process inflexibility [12]. So, For example, exploration includes seismic, geophysical and geological operations, while production operations include drilling, reservoir, production, and facilities engineering. Refining is a complex operation and its output is the input to marketing. Marketing includes the retail sale of gasoline, engine oil and other refined products. Very few industries can benefit from maximizing supply-chain efficiencies more than the oil and gas companies. In this industry, the types of shipments made vary widely from gloves to pipes, valves, cranes, chemicals, cement, steel, and drilling rigs, just to mention a few. In addition, very few industries require this immense array of supplies to be moved daily and frequently in large quantities domestically, globally, onshore and offshore. In the oil and gas industry, almost all significant and important operations are planned in advance. Thus, the whole process can be massaged and fine-tuned into a high-performance money-making machine. The goal of supply chain management is to provide maximum customer service at the lowest possible cost. In the industry supply-chain, exploration operations create value through seismic analysis and identifying prospects. Production operations become the customers that use the output of exploration. The refining is the customer of production while marketing is the customer of refining and the consumer of refined products such as gasoline is the ultimate customer. There is a need to ensure that each company or operator along the supply-chain can respond quickly to the exact material needs of its customers, protect itself from problems with suppliers and buffer its operations from the demand and supply uncertainty it faces. One of the weaknesses of a supply-chain is that each company is likely to act in its best interests to optimize its profit. The goal of satisfying the ultimate customer is easily lost and opportunities that could arise from some coordination of decisions across stages of the supply-chain could also be lost. If suppliers could be made more reliable, there would be less need for inventories of raw materials, quality inspection systems, rework, and other non-value adding activities, resulting in lean production. Tubes and tubular goods are among the important goods supplied to the oil and gas industry on a daily basis. The supply-chain in tubular goods is the process through which oilfield tubular goods such as pipes, tubing, and casing are ordered, manufactured, transported, stored, prepared, and then delivered to the website for installation into a well. Managing this part of the supply-chain can be both an operational and logistics nightmare for most oil and gas companies. Delays in the arrival of pipes, casing, tubing, and other accessories, can result in extensive rig downtime and consequently high operating costs.

Smart supervisor

Industrial processes are those in which complex operations are carried out and are monitored using sensor equipment. Generally, various parts of the equipment that act and carry out the operations carry their own automatic controllers, and the supervision work is in the hands of operators or high-level computer systems. Currently, with the advances in technology, high levels of industrial supervision can be achieved, based on the integration of data, information and knowledge, using supervision systems capable of learning, adapting to operating contexts, among other things. Particularly, the Intelligent Supervision (SI) technology is an evolution of the artificial intelligence area [11], to achieve autonomous computational systems in supervision tasks, with adaptive capacities, to allow self-diagnosis dynamics, self-organization, among other things, in the processes. Thus, an SI system must have the capacity to monitor a process and act on it, according to the anomalies detected. It is made up of components that "interpret" the behavior of the supervised process, and propose and execute actions to maintain normal operating conditions, even in the presence of faults [11,12]. In general, SI must perform tasks of monitoring, fault detection, and diagnosis, control, optimization, planning, among others. These tasks should exploit the data of the supervised process, for which the use of intelligent techniques to perform their tasks becomes a very useful tool [12].

Autonomous computing

Autonomic computing is a self-management model inspired by the human nervous system, which was proposed by IBM [13]. It is based on the idea of incorporating sensors and effectors to observe and act on the environment, based on analysis and planning processes. Thus, to model an autonomous system, a reference model composed of [13] has been defined: Managed elements: Any resource (hardware or software) embedded in the system, which has attributes that can be self-managed. Sensors: Collection of mechanisms for capturing information about the elements handled. Autonomous administrator: Implements intelligent control loops that automate the self-regulation tasks of the applications). It is composed of four modules, which characterize the autonomous control loop: Monitoring, Analysis, Planning and Execution (MAPE) [14]. The monitoring module collects the events/data from the sensors. The analysis module identifies and examines situations of interest. The planning module decides and organizes the tasks to be carried out, based on particular states of the system and all the internal knowledge of the system (K, "Knowledge"). The execution module allows sending the results obtained to the effectors and Actuators or Effectors: It carries out the changes in the elements handled. Thus, autonomic computing [13-14] promotes the development of self-managed software. That is, its goal is to allow software systems to be able to manage themselves, minimizing the need for human interaction. To make this possible, the autonomic process has a set of objectives or goals that serve as guides, which it is responsible for interpreting to self-manage to ensure that they are met.

Data analytics

AdD solutions enable an organization to use data from its processes to answer complex business questions. AdD implies a cognitive process, that is, a learning process from data, such that it is necessary to study mechanisms that are involved in the creation of knowledge. It should be noted that the AdD allows for building knowledge models. The typical knowledge models to build with the AD tasks are very varied, and some of them are:

Descriptive Models. These models use historical data to define how things are being done. It is one of the most used methods, and its objective is to show a photograph of the processes. From this, it is possible to analyze the current situation, to make decisions with a high degree of success. Tasks such as clustering, association rules, and sequence discovery are examples of such models. These methods seek to answer questions such as: ¿What are the characteristics of the process understudy? What's going on?

Prediction Models. They are models that allow predicting what is going to happen in advance. They are models that look to the future, to understand how our environment can evolve. The goal of the model is to make forecasts. It seeks to extract knowledge in the form of patterns or trends, which help to guess future situations. Classification and regression tasks are examples of these methods. They seek to answer questions such as: What will occur? Why will it occur?

Prescriptive Models. They are models that define a set of actions to perform and the order between them. These models can be based on descriptive and predictive models, to establish the actions to follow. An example of a prescriptive system collects business information, predicts based on that information, what impact the different policies or actions to be taken will have, and chooses and orders the set of actions that provide the greatest benefit according to an objective. or given a goal, through an optimization process. Answer questions like: ¿What should be done? Why should it be done? What happens if you try that?

Optimization Model. Optimization models have an objective function and a set of constraints, in the form of equations or inequalities. Optimization serves to find the solutions that provide the best result to achieve the defined objective. In general terms, optimization models look for what values the variables should have, so that a mathematical expression has the highest possible numerical value (maximization), and/or the lowest possible numerical value (minimization).

To carry out AdD, techniques linked to the machine learning paradigm, statistics, among others, are mainly used [1]. In our work, we will use the AdD paradigm for supply chain management, which will be identified as supply chain data analytics (AdDSC). It will aim to create knowledge from the data of the supply chain value chain is a network of retailers, distributors, transporters, storage facilities, and suppliers who take part in the production, delivery, and sale of a product that convert and move the goods from raw materials to end-users, it describes the processes and organizations involved in converting and conveying the goods from manufacturers to consumers, based on AdD tasks.

Midano

Methodology for the Development of Data Mining (MD) based on Organizational Analysis (MIDANO) [11] for the analysis, understanding, cleaning, and modeling of the data, as well as for the development of the autonomous cycle of AdD tasks. It is made up of three major phases. So, Phase 1: Identification of sources for extracting knowledge in an organization: This phase aims to carry out a knowledge engineering process, oriented to the problem and its organizational/business context. The main objective of this phase is to get to know the organization, its processes, its experts, among other aspects, and thus define the scope of the AdD application. In our case, it will determine the scope of the autonomous cycle of AdDSC in the organization. In addition, it performs an initial design of said tasks. Phase 2: Preparation and treatment of the data: In this phase, the preparation of the data is carried out, developing two stages: a) determination and extraction from the sources, b) engineering of data characteristics. Phase 3 Development of AdDSC tasks: In this phase, the AdDSC tasks are developed. In this stage, the computational tools to be used for the development of the AdDSC tasks are evaluated and the autonomic cycles are implemented. For this purpose, the CRISP-DM methodology is used, it is one of the most used methodologies [11] in data mining. The methodology is structured in six methods:

1. Understanding the business: Understand and detail the processes on which you want to intervene.
2. Understanding the data: The data provided by the supply chain process will be explored through descriptive statistics, to understand its nature, relevance and potential applicability in solving problems.
3. Data preparation: The data will be subjected to pre-processing, where it is sought to clean it and then extract statistically good information through the use of robust statistical techniques.
4. Modeling: Different statistical techniques of "machine learning" are evaluated to generate operational and organizational knowledge models.
5. Evaluation: Validation of the proposed models through cross-validation techniques, both in the sample and in production, measuring the efficiency of the models with relevant metrics depending on the technique used.
6. Deployment: After the model has been built and evaluated, a system containing the proposed models for application in the energy industry will be developed.

In this project, CRISP-DM will be used in its phases 4, 5, and 6, in conjunction with phase 3 of MIDANO, which foresees stages of development of the AdD tasks, and validation of the knowledge models built, as well as of the product at a functional level; in this case, of the deployment platform of the autonomous cycle of intelligent supervision of the Supply Chain System, a specific case of the European community.

Intelligent autonomous system for supply chain management (SAiCS*)*

The autonomic cycle is made up of a series of AdD tasks, which are distributed in the different components of the MAPE model. Thus, in the Monitoring phase, 3 data tasks will be developed:

(1) Monitoring of events affecting the global supply chain market.
(2) Production cost identification.
(3) Monitoring of trends of operating variables.

While the first task of SAiCS presents events that modify the price of asset sales, the calculation of production costs represents expenses, which also significantly mark the economic destiny of a company. Also, while income, particularly from sales, is associated with the company's marketing sector, the cost of production is closely related to the technology sector.

Consequently, it is essential that such costs, such as initial investment, labor costs, maintenance, energy, contracted services, taxes, asset depreciation, factory asset defect, be considered in this SAiCS task. In the Analysis phase there are two tasks to interpret the information provided in the monitoring phase:

(4) Product Price Prediction.

(5) Evaluation of economic profitability.

The prediction of the prices of the product is fed directly from the results of phase 1, for which forecasting techniques can be applied, being able to obtain a model that is capable of predicting the prices of the product. For the evaluation of profitability, the Return on Investment (ROI) formula will be used, which indicates a relationship between the investment and the benefits generated. In the Planning phase, 2 tasks are performed to achieve the final SAiCS goal of optimizing the supply chain at the lowest possible cost:

(6) Optimization of operational performance.

(7) Maximization of production at the lowest cost.

Tasks 6 and 7 are the heart of SAiCS, since through them it is possible to optimize the supply chain autonomously, applying multi-objective optimization theories. Finally, in the Execution phase, there is only one task, whose purpose is to take the necessary actions based on the results in the planning phase: (8) Output variable adjustment. In this way, in the Execution phase, supervisory techniques based on Optimal Fuzzy Rules will be designed, to propose actions and achieve the final objective of the regional system, which is multi-objective optimization applied to the supply chain system.

I. Specification of the Tasks Analysis of the Behavior of the Product Prices

Experimentation

In this section, the design of the regional micro-cycle for the Analysis of the Behavior of Oil Prices (AC2P) is presented, which is framed within another regional cycle called SAiCS. AC2P is made up of a series of tasks as illustrated in (Figure 1).

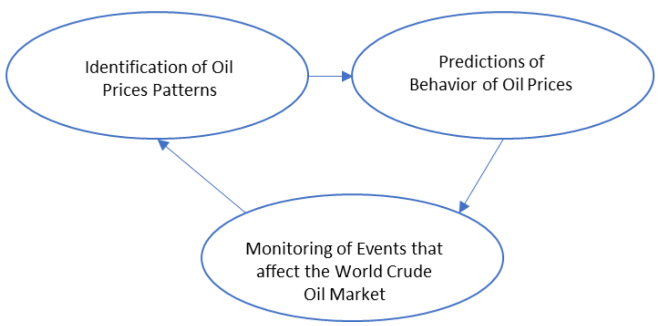

Fig.1. Micro-Cycle AC2P

In the monitoring or observation phase, there is the task of Social Data Analytics, which automatically captures information on different events in the world crude oil market from different Web portals.

The analysis phase aims to interpret, decipher, and find out the meaning or importance of a news event reported by the Web portal. In this task, dynamic knowledge models are built in order to determine behavior patterns.

The third task is the prediction of the price of oil, which helps decision-making, since it generates the necessary knowledge to predict the trend in the price of crude oil. The result of this task is sent to the SAiCS crude oil extraction activity calculation task.

Task Specifications

Task 1: Monitoring of events to calculate the exogenous and endogenous variables that characterize the world crude oil market. In this task, the events that affect the world crude oil market are automatically captured, such as refinery maintenance, hurricanes, prices of other world crude oils such as WTI (West Texas Intermediate), BRENT, URALS, among others, amounts of crude oil placed on the markets, asphalt prices, war problems in countries in conflict that affect oil installations or operations, statements by energy authorities of countries or world organizations and any other information from oil Web portals, such as Argus and Platts .

Task 2: Identify patterns in oil prices. With the information from the first task, the aim is to identify or determine the relevance or importance of a word, or collection of words, for the news event. The main objective is to determine patterns that identify news events of rising oil prices, low oil prices, or neutral, based on the frequency of the words or set of words in the text under study (Figure 2).

Fig.2. Relations with Task 2

Task 3: Prediction of the behavior of oil prices. In this task, a prediction of the impact of an event or set of events on the price of oil is made, such that if the impact is positive, then an increase in the current price is predicted (S = Rise); or if the effect is negative, then a decrease in the current price is predicted (B = Low); or if it has no impact then your prediction is neutral and the forecast is for the price to hold (N = Neutral). For this task learning machines were used (Figure 3).

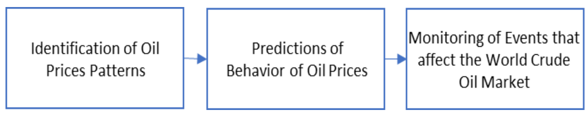

Fig. 3 Relation with Task 3

AC2P Experimentation: Case study

To define the case study, a total of 1217 news events were taken, from December 2015 to March 2017. These events were subjected to a process of extraction, treatment, loading and classification. On the other hand, the Venezuelan Boscan crude oil was taken as the reference price, in the markets of the Gulf of Mexico Coast of the United States (CGME), Northwest Europe (NOE) and Asia. On the other hand, the statistical software R version 4.3.4, in its version RStudio Inc version 1.1.383, was used as a computational tool. In particular, the R Text Mining libraries were used. On the other hand, several classification techniques (Support Vector Machines, Random Forest, One Ripper) were used for the prediction task, to select the best of them as the AC2P prediction algorithm. To develop the predictive model, the R Caret, e1017 and Random Forest libraries were used.

Monitoring of Events that affect the World Crude Oil Market

The monitoring of news events in the Platt and Argus Web portals was done through an API (Application Programming Interface) under the streaming model. In addition, by means of a script, the information is stored in a Postgres database, under the structure indicated in Table 5. A total of 1276 instances were monitored, which will serve as input for the tasks of discovering patterns and predicting the price behavior of the market. Petroleum.

Discovery of variable patterns that affect the price of oil

In order to determine these variables (and the patterns behind them), a data treatment process was first carried out, and a classification of the collected news. For the discovery of patterns, all the resulting words were used after carrying out the treatment process (cleaning) and the frequency of each one of them was determined. For this, the relative order of frequent words was defined and word cloud and bigram graphs were constructed, which determined the relationship between frequent words.

Word cloud: Figure 4 shows this cloud, which allows visually determining the set of words with the highest frequency and indicating to which class it belongs (trend down, up or neutral, with respect to the price of oil). For example, notice in (Figure 4) that words that appear very frequently are “South Korean” and “supplier”, and they are in the neutral price behavior class, and one linked to the rising price of crude oil is the word “displaced”.

Fig. 4 Word Cloud

Bigrams: (Figure 5) shows a bigram, a visual exploration tool to discover relationships between two words, and their frequency within the text. A minimum frequency of 30% was established as a filter for the words. With this bigram, associative rules can be established as a result of the combinations of said words. For example, crude oil is related to precious, asphalted, and season; or gasoline with margin and refining. This establishes, in some way, dependency relations between these variables, which allows the construction of association rules between them (causal relations).

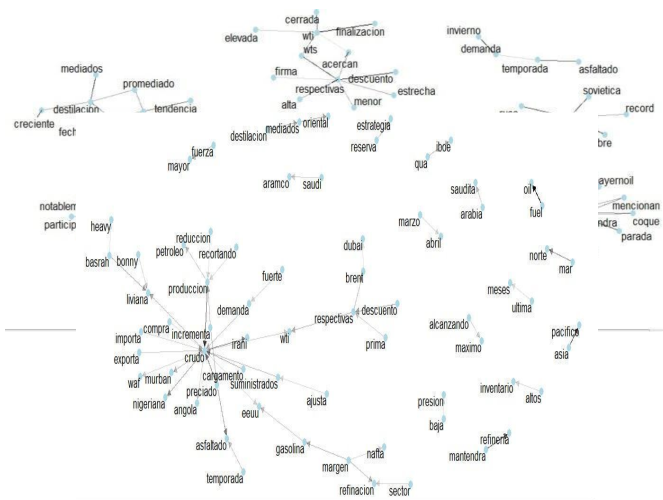

Fig. 5 Monitored Dataset Bigrams

Analysis of Relationships: (Figure 6) shows the relationships between the words with a downward trend, for the specific case of Boscan crude. For this graph, a frequency threshold of 17% and a minimum correlation of 30% were established. The same can be done, to analyze the relationships between the frequent words in the news that speak of an upward and neutral trend in the price of oil. So, In Figure 6 the downward trend is related to the fact that there is a “stable demand” or there is a “healthy stable increase in demand”. Thus, each of the news classes (high, low, neutral) is created a matrix of frequent terms (words, bigrams and relationships) that characterize them. To determine the frequency of words, the metric TfxIdf (Term frequency – Inverse document frequency) is used, which establishes how relevant an attribute is within a group of instances [26, 27]. Next, the most important attributes of each class are selected, which reduced from 5028 attributes to 398 attributes (the most relevant terms of each of the classes under study (S,N,B)). This array is the minable view used in the A2CP prediction task.

Fig. 6 Relationship between words of the supervised group with a downward trend

Prediction of the Behavior of Oil Prices

To choose the prediction model, an accuracy greater than 80% was required. Of the total of 1217 instances, 70% was used for machine training and the remaining 30% for model testing. As said before, the classification techniques used were Support Vector Machines, Random Forest, One Ripper and JRipper. The quality metrics defined in section 2 were determined for each of them. The data sample used was as follows. For the training data, the instances used by classes were (B = Instances Trending Down, N = Instances Trending Neutral, S = Instances Trending Up): B=236, N=384 and S=222 and the dates for test: B=106, N=176 and S=93. So, this Table 1 shows the comparisons of the prediction results obtained between the Support Vector Machines (SVM), the Random Forest Algorithm, the One Ripper Algorithm, and the JRipper Algorithm. Table 6 shows that the Random Forest algorithm had better accuracy, with 83.7%, followed by MSV with 80.27%, which represents a good prediction value for both models. One Ripper and JRipper are outside the stated purposes of this work. Finally, the following algorithm makes it possible to predict the impact of a news event on the price of a barrel of oil, whether it affects it positively, negatively or neutrally, and indicates the accuracy of the prediction. Also, the Random Forest algorithm had a higher percentage in the classification tasks of true positives, for classes B and S, compared to the MSV algorithm, which had the best result classifying class N (the least relevant, in the case of this article). On the other hand, its sensitivity is also better than the rest of the algorithms.

Table 1. Results obtained

	MSV			Random Forest			One Riiper			JRipper		
	B	N	S	B	N	S	B	N	S	B	N	S
Exactitud (%)	80,27			83,74			57,07			73,07		
Verdaderos Positivos (%)	75	86	79	99	84	95	51	61	44	69	75	75
Sensibilidad (%)	89	72	87	86	97	80	46	87	13	71	84	56

The algorithm first determines the type or class of each news event (B, N, A). Based on the results obtained in the type of each event, the type of events (B, N, A) that has the most amounts is selected. The final newsgroup trend will be the class that has the highest average and passes a minimum average threshold (in this case, 50%). In case none of the classes is above this threshold, the algorithm indicates that no prediction can be issued. The test group was subjected to the developed algorithm, with the result of "Neutral Trend" with a precision percentage of 83.49%. The algorithm managed to classify 91 news with a downward trend, 171 with a neutral trend, and 75 with an upward trend. The results are shown in (Figure 7).

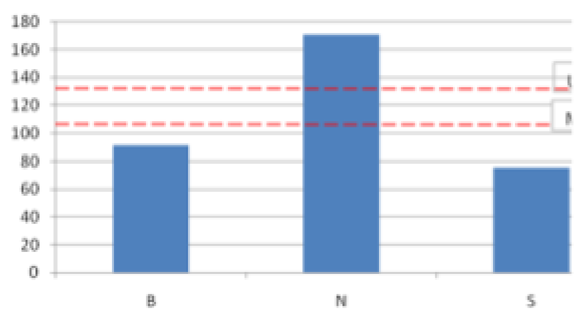

Fig. 7 AC2P Prediction

Discussion of Results

AC2P is an innovative proposal for predicting the behavior of crude oil prices, since its data entry is not based on structured data, but on unstructured data from Web portals in natural language. Similarly, by having your data input from these portals, all variables, both endogenous and exogenous, are treated with the same relevance. Table 2 shows the most relevant differences between AC2P and other investigations. The model developed in this work is the only one based on AdD, and it proposes a CA that mixes several techniques in order to always be monitoring oil prices, and thus update its trend models (it is an oil price supervision scheme in the market). In addition, it mixes exogenous and endogenous variables without any problem, which allows you to scale the number of variables to use in the future.

Table 2. AC2P and other research

Criterios	Modelos	Técnicas usadas	Variables en estudio
AC2P	AdDS	Clasificación y Minería de texto	Exógenas y Endógenas
[15]	Econométricos	Regresión	Endógenas
[16]	Econométricos	Regresión	Endógenas
[17]	Econométricos	Regresión	Exógenas y Endógenas

Conclusion

Techniques for extracting knowledge through the AdD provide a powerful tool for the generation of knowledge, such as the prediction of the behavior of oil prices, for the case study of Boscan oil. In addition, they allow the analysis of unstructured information and natural language, to extend the sources of information to be used. AC2P has great advantages over previous works, since it takes into account the endogenous and exogenous variables that affect the oil market, and can predict the impact of a set of events as soon as they occur, which greatly helps in decision making. AC2P is a computer expert who can predict the behavior of the price of a given type of oil. AC2P is capable of providing important information to SAiCS, in its intelligent supervision tasks, for the method of production of oil wells by Electro-Submersible Pumping (ESP), whose main objective is the optimization of the production system, adapting to the economic environment, and it is there where AC2P provides important information for decision making. Other AdD tasks will be developed in SAiCS, both in the operational and financial environment, which will grant SAiCS the necessary autonomy in making its decisions. Also, future research will incorporate the analysis of other types of oil.

ACKNOWLEDGMENTS

Sincere thanks to Dr Jose Aguilar and Hector Lozada their contributions to this research.

References

[1] E. Babaee, S. Sadeghi, F. Manson, Haidi R, "Application of Machine Learning in Supply Chain Management" Mathematical Problems in Engineering, vol. 1, pp. 1–10, April 2021.

[2] D. Campionp, Caza, B., "Multiple jobholding: An integrative systematic revire and future research" Journal of Management, vol. 65, pp.3123–332, May 2020.

[3] N. Tuptuk, S. Hailes, "Security of smaryt manufacturing systems," Journal of Manufacturing Systems, vol. 47, pp. 93–106, April 2018.

[4] H. Sebastia, P. Godinho, "Forecasting and Trading Cryptocurrencies with Machine Learning under changing market conditions" SpringerOpen, vol. 7, pp. 93–106, April 2021.

[5] Li Da Xu, "Industry 4.0 Frontiers of fourth industrial revolution" Systems Research Behavioral Science, vol. 37, pp. 529–783, July 2020.

[6] P. Jaquart, D. Dann, C. Weinhardt, "Short term bitcoin market prediction via machine learning" The Journal of Finance and Data Science, vol. 7, pp. 45–66, Nov 2017.

[7] Z. Jiang, D. Xu, "A deep reinforcement learning framewrok for the financial portfolio management problem" ArXiv, vol. 7, pp. 5–11, Nov 2017.

[8] D. Nazarov, A. Klarin, "Taxonomy of Industry 4.0 research Mapping schoarship and industry insights" Special Issue Industry 4.0, vol. 37, pp. 535–556, July 2020.

[9] Vizcarrondo J., Aguilar J., Exposito E., Subias A. "MAPE-K as a Serviceoriented Architecture". Revista IEEE Latinoamerica Transactions, Vol. 15, No. 6 (Pp. 1163-1175), 2017..

[10] J. Vizcarrondo, J. Aguilar, "ARMISCOM: Self-healing Service Composition. Service Oriented Computing and Applications" Service Oriented Computing and Applications, vol. 11, pp. 345-365, July 2017.

[11] H. Lozada, E. Camargo, J. Aguilar, "Analysis of the Behavior of Petroleum Prices" Revista Ingeneria al Dia, vol. 3, pp. 20–45, July 2017

[12] E. O'Dwyer, I. Pan, S. Acha,N.Shart, "Smart energy systems for sustainable smart cities: Current developments, trends and future directions", *Applied Energy*. Elsevier.vol. 237,pp.581-587, March 2019.https://doi.org/10.1016/j.apenergy.2019.01.024.

[13] IBM Autonomic Computing Architecture Team, "An architecturalblueprint for autonomic computing", Technical report, IBM Corporation,Hawthorne, NY, Fourth Edition, Junio 2006.

[14] J. Vizcarrondo, J.Aguilar, E. Exposito, "MAPE-K as a Service- oriented Architecture",*Revista IEEE Latinoamerica Transactions*, vol. 15, No. 6, pp.1163-1175, 2017.

[15] Zamani M. (2004). An Econometrics Forecasting Model of Short Term Oil Spot Price. 6th IAEE European Conference, Zurich, Switzerland.

[16] Mirmirani S., Li H.C. (2004). A Comparison of VAR and Neural Network with Genetic Algorithm in Forecasting Price of Oil, in Applications of Artificial Intelligence in Finance and Economics (Advances in Econometrics, Vol. 19), Binner J.M., Kendall G., Chen S.H., Emerald Group Publishing Limited (Pp. 203-223).

[17] Kilian L., Murphy D.P. (2013). The Rol of Inventories and Speculative Tranding in the Global Market for Crude Oil. Journal of Applied Econometrics, Vol. 29, No. 3 (Pp. 454- 478)