Article

Predictive Model for LNG Ship Routing applying Machine Learning

Ramón Fernández-Reyes Doallo [1], Alberto Camarero Orive [2,*] and Nicoletta González-Cancelas [3]

[1] MSc Civil Engineer, Polytechnic University of Madrid, Madrid, Spain
[2] Professor in Polytechnic University of Madrid, Madrid, Spain alberto.camarero@upm.es
[3] Professor in Polytechnic University of Madrid, Madrid, Spain nicoleta.gcancelas@upm.es
* Correspondence: alberto.camarero@upm.es

Abstract: The purpose of this paper is to develop a theoretical predictive model for LNG shipping routes selection process. Strategic decisions about shipping costs could be improved if a deeper knowledge about products economic value is provided. Developments made on the extraction and industrial processes related to this fossil fuel are driving the natural gas sector towards a unique globalised market. Moreover, data analytics applications as well as machine learning are topics presented as perfect catalysers for achieving an unprecedented natural gas globalised market. Additionally, this paper aims at showing the state of the art of new techniques used in transportation engineering that might have synergies with other industries (eg. commodities cost reduction, energy supply…). Finally, this paper aims to provide foundation for further research and development using more sophisticated data and algorithms that will help to close the gap between theoretical and practical scope of this techniques.

Keywords: LNG; shipping optimization; machine learning; predictive model.

1. Introduction

Natural gas is considered the cleanest among the rest of fossil fuels (coal, oil, kerosene, propane…) and its accounts for the greatest expected growth until 2030 compared to other energy sources. During the last decade [1], natural gas achieved a 23% share in primary demand, which proves this fuel is taking the lead on global energy supply [2]. Natural gas consumption entails environmental benefits in terms of air quality and greenhouse gas emissions compared to other fossil fuels. These benefits are even more pronounced if we look at the energy efficiency level achieved during its production process [3]. Since 2010, the transition from coal to natural gas in the energy sector has saved 500 million tons of CO2. This milestone will help this fossil fuel to outperform compared to coal and oil on Sustainable Development Goals achievement [4].

During the last couple of years, naturals gas industry has experienced a globalization process due to the price reduction of this commodity. Cheaper natural gas prices have permitted to globally trade this fuel, democratizing the usage of natural gas for all kinds of industries and populations [5]. Shipping natural gas from one place of the world where it is produced to another where customers demand this product, is now possible at a reasonable price [6]. Several reason might have prompted the globalization of this market. However, technical enhancements deployed on production, storage and shipping process of natural gas have helped in a massive way [7].

Technical improvements in how natural gas is extracted eventually founded the yet known "shale revolution" in the United States in 2008 [8]. The "shale revolution" permitted the United States to evolve from exporter country to become a relevant natural gas producer [9]. Thus, one of the greatest LNG exporters in the world. Associated with this phenomenon, the main American natural gas index (Henry Hub) plummeted from an average of 13$/mcf in the 2003-2008 period to 3.69$/mcf in 2016 [10]. As a reader may assume, the United States' appearance as an exporter on the LNG landscape has eased the consecution of a global natural gas market [11]. By increasing the annual international trade, allowing access to new market participants, and outstanding this industry competitiveness [12].

From the beginning of XX century, natural gas operational costs have experienced a remarkable decrease due to advanced liquefaction plants that optimized natural gas transformation [13], storage, and shipping [14]. Liquefied Natural Gas (LNG) is natural gas that has been cooled down below -163°C to become liquid [15]. Shipping costs considerably influences the LNG competitiveness [16]. Among them are listed some specific costs such as ship's fuel cost and others more general such as loading and unloading port fares [17]. Nevertheless, the purpose of this work is to prove that a remarkable factor in LNG profitability compared to other natural gas transportation methods [18] (pipeline, for instance) is the distance between origin and destination. However, other factors such as geopolitical circumstances and socio-economic situation needs to be taken into consideration to understand LNG market. The war between Russia and Ukraine during 2022 has clearly shown the relevant role that geopolitical situation plays on gas industry [19].

Pipeline construction cost function usually increases linearly, becoming less competitive for long distances. On the other hand, LNG shipping presents higher fixed costs due to the cargo transformation from gas to the liquid state but lower variable costs per distance unit [20]. Considering that, this work proposes a methodology for identifying the economic influence that an optimum selection choice has on the commodity price [21].

After technical developments undertaken in the mining and chemical industry, transportation engineering shows as the subsequent driver in the consecution of a global natural gas market [22]. Discovering new solutions to enhance shipping competitiveness is a major challenge this industry is facing nowadays [23]. Big data availability, storage capacity, and innovative analysis methods such as Machine Learning or Artificial Intelligence are questions not yet responded to in Transportation Engineering [24]. Predictive analysis could be counted as one of the most used applications of Machine Learning nowadays. Numerous analyses are under development in futures commodity prices analysis [25]. Aiming at going a bit beyond the analysis already completed, we decided to study this market from a more innovative angle. This paper tries to bring shipping distance as a new variable that should be considered as well as others like weather, imports or volume stored to analyse natural gas price. It is important to notice that this study was done before the 2021 Suez Canal Obstruction which collapsed the logistic industry worldwide and prompted an increase in raw materials prices, including LNG. Therefore, the main intention is to start considering this variable as important others in this market which was proven two years later. Some of the features considered are weather, imports, volume stored, or the number of extraction days [26]. LNG price forecasting could be extremely useful for improving decision-making, risk reduction and supply and demand balance as well as optimising the efficient use of limited resources based on accurate predictions.

Similarly proceeding as the analyses mentioned above, this paper presents a methodology for LNG price prediction based on different features (distance, contract type, predicting year) [27]. Considering the natural gas industry's trend approaches a global market with similar index prices around the world, this tool could be valuable for shipping route's choice decision-making [28]. Leaning on this method, LNG importers could decide under what conditions would be desirable to choose one specific route aiming reduce the cost of operating ships. The paper was intended to provide a theoretical model that will evolve towards more sophisticated predictive and practical models. Being this an academic research, data was limited and difficult to extract from public sources. Potentially, more features would be added to the model in case the practical scope of the work is not achieved. For the sake of simplicity, it is assumed vessels with similar fuel consumption. Vessel fuel efficiency could be a new feature to be added in future papers.

For the sake of this predictive model, the United States has been considered the preferred LNG exporter. More precisely, the Sabine Pass liquefaction plant was chosen as the order place from which the cargo departs [29]. Apart from the United States being one of

 doi:10.20944/preprints202212.0252.v1

the most promising LNG traders in the coming years, Sabine Pass-bounded routes are especially dense in terms of shipped cargo traded with the Asian market [30]. Besides, the Panama Canal Expansion opening in 2016 widened Sabine Pass' route choices and unveiled new possibilities for optimising transport costs [31]. Conclusions obtained in this study are expected to serve as new improvements avenues in the branch of transportation engineering intended to advance towards a global market for natural gas.

2. Methodology

In order to apply Machine Learning for developing predictive models, it is appropriate to exercise precise attention to data gathering, cleaning, and handling. By doing so, the learning algorithm will have as much representative data as possible for achieving the purpose at hand.

This paper's objective is to predict LNG export price for a group of routes named "Target Routes". Thus, it is necessary to define those routes and, even more important, the thought process behind their selection, outlined in Figure 1. A great effort of this study has been exercised at this stage to avoid downstream inconsistencies or incoherent results. "Target Routes" will meet two conditions that may affect the LNG export price: a considerable annual LNG shipped volume and at least two route choices.

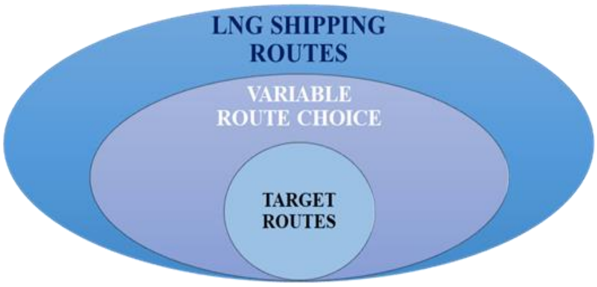

Figure 1. Target Routes Procedure Funnel Depiction. Source: own elaboration.

Because the origins and destination variability are considerably broad, a procedure for narrowing the alternatives while choosing those most valuable will be developed. Starting from identifying the biggest trade nodes in the LNG market, those origins and destinations with more impactful alternatives will be chosen. In this case, routes that will potentially cross the Panama Canal and Suez Canal. Finally, those routes with relatively small annual shipped volume will be discarded to avoid model bias.

Once the Target Routes identification procedure has been performed, different machine learning algorithms will be considered to create a function capable of predicting LNG export price. Because this study considers two scenarios (Canals' existence and No Canals' existence), insights about the economic impact of these infrastructures are expected to emerge.

2.1 Target route selection procedure

2.1.1. Main global LNG importers

The first stage of the analysis consists of determining the potential destination points. One effective way to narrow the alternatives' diversity is to identify the principal global LNG importers. By analysing the LNG volume traded in 2018 [32] and focusing on the demanding points has been identified as the main global importers. Subsequently, a similar analysis will be performed for the supplying points to finally identify the most crowded LNG shipping routes.

As shown in Figure 2, the Asian-Pacific region presents the greatest annual growth and imported volume. Japan, China, and South Korea are the countries with the highest demand. Additionally, the European Region witnesses lower imported volume and steady annual growth. It appears relevant to highlight that Europe has a sophisticated pipeline network meaning the natural gas transportation is segmented in LNG and PNG.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 December 2022 doi:10.20944/preprints202212.0252.v1

LNG IMPORTERS MARKET							
2017				2018			
#	Country	Volume (MT)	Annual Growth	#	Country	Volume	Annual
1	Japan	84.5	1.1%	1	Japan	83.2	-0.6%
2	China	39.5	12.7%	2	China	54.8	15.8%
3	Shout Korea	38.6	4.9%	3	Shout Korea	44.5	6.4%
4	India	20.7	1.5%	4	India	23.3	4.0%
5	Taiwan	16.8	1.8%	5	Taiwan	17.1	0.3%
6	Spain	12.2	2.3%	6	Spain	10.8	-1.4%
7	Turkey	7.8	2%	7	Turkey	8.5	0.6%
8	France	7.6	2%	8	France	8.4	0.9%
9	Egypt	6.2	1.1%	9	Pakistan	7.1	2.4%
10	Italy	6	2%	10	Italy	6.3	0%

Figure 2. Global LNG Imports Market Overlook (2018). Source: International Gas Union.

2.1.2. Main global LNG exporters

According to the LNG Industry Report in 2018, Qatar, Australia, and the United States were the biggest export countries in terms of annual LNG traded volume and growth over the year (Figure 3). Even though Australia shows a more prominent growth trend than other countries (Figure 4), this model will not consider it as an exporter country. The main reason is the geographical location of this country. All the routes departing from Australia do not use the Panama Canal nor Suez Canal. This fact exacerbates the possibility of route choices that may impact the LNG price. On the other hand, Qatar and the United States' location permits considering different route choices with a significant variance in their trip distance thanks to presence of canals.

LNG EXPORTERS MARKET							
2017				2018			
#	Country	Volume	Share	#	Country	Volume	Share
1	Qatar	81	27.6%	1	Catar	78.7	24.9%
2	Australia	56.2	19.2%	2	Australia	68.6	21.7%
3	Malaysia	26.4	9.0%	3	Malaysia	24.5	7.7%
4	Nigeria	21.3	7.3%	4	USA	21.1	6.7%
5	Indonesi	16.2	5.5%	5	Nigeria	20.5	6.5%
6	USA	13.1	4.5%	6	Russia	18.9	6.0%
Others	Others	86	26.9%	Others	Others	82.7	26.5%

Figure 3. LNG Exports Market Distribution by annual LNG volume share (2018). Source: International Gas Union.

LNG EXPORTERS MARKET							
2017				2018			
#	Country	Volume	Growth	#	Country	Volume	Growth
1	Qatar	81	3.7%	1	Catar	78.7	2.0%
2	Australia	56.2	11.9%	2	Australia	68.6	12.2%
3	Malaysia	26.4	1.5%	3	Malaysia	24.5	-2.0%
4	Nigeria	21.3	2.8%	4	USA	21.1	10.2%
5	Indonesi	16.2	-0.4%	5	Nigeria	20.5	2.8%
6	USA	13.1	10.2%	6	Russia	18.9	3.0%
Others	Others	86	-	Others	Others	82.7	-

Figure 4. Global LNG Exports Market Overlook (2018). Source: International Gas Union.

2.1.3. Origin-Destination matrices

Results from Figure 3 shows Qatar as the biggest LNG exporter in the world. Also, the presence of Suez Canal allows a huge distance reduction on its shipping routes to Europe. Therefore, volume traded, and potential distance reduction makes Qatar as one of the best candidates to benefit from the results of this model. Considering Qatar and United States as the origin points for further analysis, every potential destination may be considered. Sailing distance (provided by https://sea-distances.org/) are the main characteristics to be focused on. Nevertheless, the target routes will be eventually found out by filtering the ones likely to use the Panama and Suez Canal. This analysis will conduct to generate an origin-destination matrix (Figure 5).

POTENTIAL ROUTES WITH ORIGIN USA			
Volume (MT)	USA	PAN/SUZ	ALT
KOR	4.74	1	0
MEX	3.59	0	1
JPN	2.48	1	0
CHN	2.26	1	0
IND	1.04	0	1
GBR	0.88	0	1
JOR	0.83	0	1
CHL	0.82	1	0
BRA	0.74	0	1
ARG	0.51	0	1
ITA	0.34	0	1
FRA	0.31	0	1
PRT	0.26	0	1
TUR	0.26	0	1
PAK	0.25	0	1
TWN	0.25	1	0
NLD	0.24	0	1
ESP	0.2	0	1
DOM	0.16	0	1
KWT	0.16	0	1
EGY	0.13	0	1
COL	0.09	0	1
PAN	0.09	0	1
GRC	0.07	0	1
MLT	0.07	0	1
POL	0.07	0	1
ARE	0.07	0	1
JAM	0.06	0	1
ISR	0.06	0	1

POTENTIAL ROUTES WITH ORIGIN QATAR			
Volume (MT)	QAT	SUEZ	ALT
KOR	14.45	0	1
IND	11.61	0	1
JPN	9.98	0	1
CHN	9.19	0	1
TWN	5.03	0	1
ITA	4.71	1	0
PAK	4.59	0	1
ESP	2.48	1	0
TUR	2.15	1	0
GBR	2.11	1	0
THA	2.02	0	1
BEL	1.89	1	0
POL	1.68	1	0
KWT	1.43	0	1
ARG	1.05	0	1
EGY	1.02	0	1
FRA	0.86	1	0
PRT	0.73	1	0
BGD	0.7	0	1
SGP	0.43	0	1
NLD	0.28	1	0
JOR	0.19	0	1
BRA	0.06	0	1
GRC	0.06	1	0

Figure 5. Origin-Destination Matrix for USA and Qatar. Source: own elaboration.

Figure 5 shows two matrices one for each origin point. Each row contains every destination of LNG cargo in 2018.

- Colum 1: Annual LNG volume traded in 2018.
- Colum 2: values 0/1 if Panama Canal of Suez Canal usage implies an increase/reduction in distance trip.
- Column 3: values 0/1 if Panama Canal of Suez Canal NO usage implies an increase/reduction in distance trip.

Potential Routes for each destination are highlighted in green. These routes are more likely to use Panama Canal or Suez Canal. Discarded routes are highlighted in red as less likely to use Panama Canal or Suez Canal, thus, do not have routes alternatives.

In order to avoid considering destinations with low volumes that may increase model bias, the following condition will apply:

$$\frac{\sum(Target\ Routes)_i}{Potential\ Routes} \geq 80\%$$

This condition will lead identify as Target Routes those shown on Figure 6.

Target Routes with Doha Origin (QAT)			
Destination Terminal	Volume (MT)	Suez (miles)	Alternative (miles)
Livorno (ITA)	4.71	4,415	10,868
Barcelona (ESP)	2.48	4,657	10,520
Marmaris (TUR)	2.15	3,460	11,656
Grain (GBR)	2.11	6,302	11,035
Zeebrugge (BEL)	1.89	6,277	11,010
Swinoujscie (POL)	1.68	7,020	11,753

Target Routes with Sabine Origin (USA)			
Destination Terminal	Volume (MT)	Panama (miles)	Alternative (miles)
Inchon (KOR)	4.74	9,998	15,416
Senboku (JPN)	2.48	9,481	15,552
Guangdong (CHN)	2.26	10,789	14,399
Cochi (IND)	1.04	9,840	11,761

Figure 6. Target Routes with annual volume traded and distance by route choice. Source: International Gas Union.

For analyzing the distance between origin and destination places has been considered each country's biggest regasification plants.

2.2 LNG price prediction using machine learning

Once target routes have been defined, the next step will be determining their economic value. An effective way to do so could be by predicting the annual volume traded and the price at which the cargo is sold. As explained previously, the target routes count with two main choices: the Panama or Suez Canal usage and the other one is the shortest trip without crossing through these infrastructures. The proposed model aims at predicting a specific LNG price for each alternative and eventually determine the most attractive *sailing route option*. Apart from cost and risk reduction, the results of this study are expected to help for assessing the economic impact of transportation infrastructures on the natural gas industry.

After having clearly defined the output variable of the predictive model, a brief description and state of the art of Machine Learning Methodology will be provided. First, two insightful descriptions given by some of the most prestigious scientist in this domain could help to better understand what it is. "Machine Learning (ML) is the field of study that gives computers the ability to learn without being explicitly programmed" (Arthur Samuel, 1959) [33]. "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E" (Mitchell, 1997) [33].

This methodology presents a vast number of applications and one of the most useful is predictive analysis that belongs to what is called Supervised Learning. For this purpose, different learning algorithms are used depending upon the data nature or characteristics. Other significant application of Machine Learning could be clustering or market segmentation that uses other algorithms specifically designed for this objective, and typically belongs to what is called Unsupervised Learning.

For the shake of predicting LNG price, it will be more fruitful to use Supervised Learning in order to create functions that approximate the future price behaviour. In this project was selected a Multiple Linear Regression Algorithm that considers different variables linearly related that define the function. Nonetheless, other algorithms similarly functioning will be explained to serve as a reference for further analysis. Thus, in the future would be possible to reach the same conclusions in different ways which will help to gain the purpose intended.

Due to the complexity of the market and lack of data in all the potential variables, we decided to start this analysis with simple and measurable variables. Hence, we started the research using "Multiple Linear Regression" which is indicated for correlated and continuous variables (shipping distance, time and vessels volume). Furthermore, this paper was intended to be the foundation for further research and development using more sophisticated data and approximation functions.

For the sake of predicting LNG price, it will be more fruitful to use Supervised Learning in order to create functions that approximate the future price behaviour. In this project was selected a Multiple Linear Regression Algorithm that considers different variables linearly related that define the function. Nonetheless, other algorithms similarly functioning will be explained to serve as a reference for further analysis. Thus, in the future it would be possible to reach the same conclusions in different ways which will help to gain the purpose intended.

2.2.1. Data base creation

Every Machine Learning project requires one database to be divided into at least two subsets. The first one is usually called "Training Set" and serves for developing the variables of the function. The second one is called "Test Set" and serves for determining the model accuracy.

The first step towards creating a database for an ML project is defining, cleaning, and classifying the data needed for predicting LNG price. Considering the current industry trend, an accurate database is expected to reflect a price unification with declining values.

The United States is one of the greatest leverages towards a global unified natural gas market and its location relative to the Panama Canal makes the route choices to China more attractive. For these reasons, data used on the training set will be extracted from annual reports of the US Department of Energy: Fossil Soil Energy – Office of Oil & Natural Gas. In these reports, every LNG vessel delivered from American terminals are registered. Among others, here are listed some features provided for each delivery: Departure Date, Contract Type (spot, short-term, long-term, and optional contracts), Destination Country, Vessel Name, Export Terminal, Volume Shipped, Price. The date range used for this project is from 2015 to 2019, for collecting examples *on June 26, 2016*, the opening of the Panama Canal Expansion.

In Figure 7 different deliveries examples are shown.

LNG Export Price in USA (2015-2019)						
Year	Contract Type	Country	Vessel Name	Terminal Name	Volume (mcf)	Price $/MMBtu
2019	Short-Term	India	Golar Maria	Sabine Pass, Louisiana	3,108,482	3.62
2018	Long-Term	South Korea	Woodside Rees Withers	Sabine Pass, Louisiana	3,701,775	4.18
2017	Short-Term	South Korea	Golar Celsius	Sabine Pass LNG Terminal	3,422,112	5.25
2017	Long-Term	Japan	Woodside Chaney	Sabine Pass LNG Terminal	3,693,568	6.53
2017	Short-Term	Japan	Creole Spirit	Sabine Pass LNG Terminal	3,668,152	5.88

Figure 7. Database used for developing machine learning model. Source: US Department of Energy - Fossil Energy. Monthly report.

2.2.2. Data handling and classification

The database was classified following different criteria. Aiming at finding the most accurate features for training the model. This procedure will allow identifying the most representative values, data distribution, and accuracy.

- **Contract Type:**

The database shows a trend on the price of short-term contracts or "Spot" towards unification. For the case of long-term contracts, there is not a clear trend in price evolution. This high bias may trigger further anomalies in the learning algorithm, and then only Short-Term contracts will be considered for the training set. Apart from selecting the training set properly, this analysis confirms the price reduction since 2015 prompted for both the "Shale Revolution" and New Panama Canal opening. Figure 8 may reflect the increase of competitiveness of this sector is led by the spot market with prices ranging from 8$/MMBtu in 2015 to near 3.5$/MMBtu in 2019.

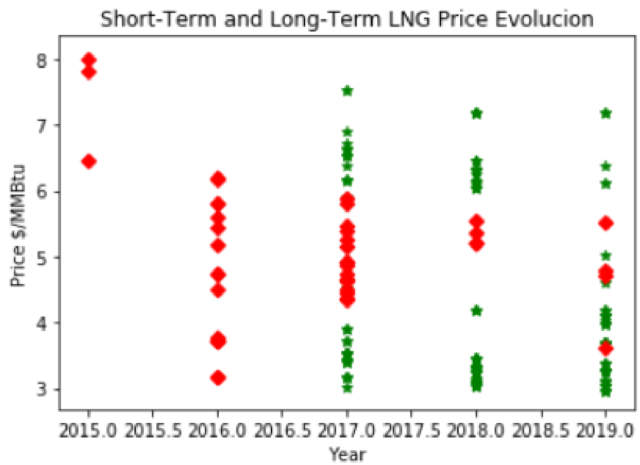

Figure 8. Short-Term and Long-Term LNG Price Evolution ($/MMBtu). Source: own elaboration.

- **Distance Trip:**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 December 2022 doi:10.20944/preprints202212.0252.v1

The first step is assessing short, medium, and long trips representativeness on the database. Those will be categorized depending on their distance between origin and destination. As explained before, this is a fundamental variable for natural gas pricing, so it is important to make sure the database is not biased. Each training example will be labelled based as Short Trip, Long Trip, Medium Trip.

- Short Trip 201
- Long Trip 102
- Medium Trip 0

Name: Trip_Category, dtype: int64 (variable name in phyton scrypt)

Based on the results, the ratio between short and long trips is approximately 2/1 that is deemed not biased. Distance Trip is likely to be an accurate feature for subsequent analysis.

- **Volume Shipped:**

The LNG volume shipped appears to be a potential feature for determining the unit price for each delivery. In order to identify the influence of this variable in the dataset, every training example will be classified based on their volume. Similarly, as for the distance trip, there will be 3 different labels: Large Vessel, Medium Vessel, and Small Vessel.

- Large Vessel 296
- Medium Vessel 4
- Small Vessel 3
- Small Vessel 2 0

Name: Vessel_Category, dtype: int64 (variable name in python scrypt)

Based on the results, it would be fair to consider every vessel as large. That means the volume shipped will not be a promising variable for determining LNG price.

Aiming at better visualizing the identified trends, variables Trip Distance and Type Contract has been compound. As shown in Figure 9, the price reduction is even more significant for the case of long-distance trips

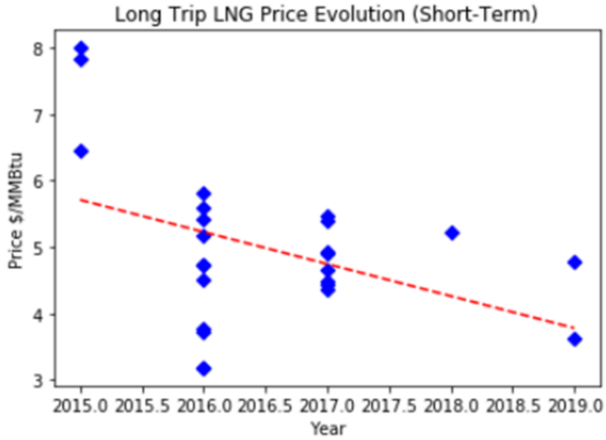

Figure 9. Short-Term Contract LNG Price Evolution for Long-Distance Trips. Source: own elaboration.

Once the database has been handled and selected the variables that will primarily affect the model, the Training Set is defined based on data availability and industry understanding. As the name infers, this subset of data sought to train the learning algorithm of the predictive model. After the model is trained, introducing more data would be expected to lead to more accurate results. Moreover, this analysis served for recognizing time and distance as the most influential features in the database. The next steps will consist of determining what learning algorithm will be used to create a function capable of predicting the price.

2.2.3. Random Forest

As one reader may assume by the name of this algorithm consists of a great number of decision trees working simultaneously. Each of those decision trees reaches a particular solution, and the outcome of the hole model will be the most frequently repeated. This model is founded on the idea that a considerable large number of not related models reach better results than each of them working independently. Based on this fact, the larger the number of non-related the more likely to obtain accurate predictions on this model. As the database available for this project is not extremely large, this model does not seem to be quite promising [34].

2.2.4. Support Vector Regression

This model is effective for function parameter estimation. Same as most of the unsupervised learning methods, SVR helps to estimate function parameters that will serve for predictions. This model foundation consists of training a symmetric function that penalizes low and high estimations. For this purpose, a flexible tube of minimum radium is created around the estimated function. Then, the model improves using predictions beneath a given threshold. SVR solves the training problem as an optimisation problem. The main goal is to find the smaller threshold possible around the function, while the prediction error is minimized. In other words, the distance between predicted and expected outcomes is reduced.

2.2.5. Multiple Linear Regression

The purpose of this model is to create a multivariate linear equation capable of predicting desired outcomes. Depending on the number of features of the database, this function will present more or fewer variables. Different optimisation procedures may be used in order to train the variables. Some considerations worth mentioning for applying this model are:

1. Relationship between features (i.e., distance or year) and predicted value (LNG price).
2. Models perform better if residual values are normally distributed.
3. Features or variables must not be multicollinear.

These considerations prompt this model to be especially useful when fundamental variables that predict a specific product are known. As demonstrated in the data classification step of this paper, some relevant characteristics that drive LNG prices are considered.

3. Selected model: multiple linear regression

Multiple Linear Regression is an extremely useful model when various parameters deemed as fundamental for the predicted outcome are known. As presented in this paper, a significant amount of work has been invested in finding promising parameters that drive LNG prices. Thus, the form of the function developed by the training algorithm will be:

$$h_\theta(X) = \theta_0 + \theta_1 * x_1 + \theta_2 * x_2 + \theta_3 * x_3 \ldots + \theta_i * x_i$$

Where θ_0 is the intercept term added in order to improve the optimisation problem performance and θ_i are the parameters obtained after the training process. Finally, the independent variables of the model will be the trip year (X1) and the distance between origin and destination (X2). $h_\theta(X)$ is the predicted LNG price in \$/MMBtu. Once the parameters are trained, the following function results:

$$Price(X, Y) = 474.2948 - 0.2324967 * Year - 3.2540e^{-5} * Distance$$

Figure 10 shows target route's predicted price in 2020 considering two major trip choices:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 December 2022 doi:10.20944/preprints202212.0252.v1

Predicted LNG Export Price by Target Route (Origin: Qatar)				
Destination	**Route**	**Year**	**Distance (miles)**	**Price $/MMBtu**
ITA	SUEZ	2020	4,415	4.513
	ALT	2020	10,868	4.303
ESP	SUEZ	2020	4,657	4.505
	ALT	2020	10,520	4.315
TUR	SUEZ	2020	3,460	4.544
	ALT	2020	11,656	4.278
GBR	SUEZ	2020	6,302	4.452
	ALT	2020	11,035	4.298
BEL	SUEZ	2020	6,277	4.453
	ALT	2020	11,010	4.299
POL	SUEZ	2020	7,020	4.428
	ALT	2020	11,753	4.275
Predicted LNG Export Price by Target Route (Origin: USA)				
Destination	**Route**	**Year**	**Distance (miles)**	**Price $/MMBtu**
KOR	PAN	2020	9,998	4.332
	ALT	2020	15,416	4.196
JPN	PAN	2020	9,481	4.348
	ALT	2020	15,552	4.191
CHN	PAN	2020	10,789	4.306
	ALT	2020	14,399	4.229
IND	SUEZ	2020	9,840	4.337
	ALT	2020	11,761	4.298

Figure 10. Predicted LNG Price across target routes applying Machine Learning. Source: own elaboration.

4. Results

The results provided by the predictive model serve for estimating the price at which the target routes' LNG transported volume will be sold annually. Considering 80% of the LNG fleet counts with a capacity greater than 180.000 m^3, this figure has been used for estimating the number of ships annually traveling across target routes. Finally, the annual number of ships and the cargo price permit calculating the unit price of an LNG ship traveling across the target routes. As explained previously, the analysis has been divided into two scenarios: No Canals Usage (Scenario A) and Canals Usage (Scenario B). Figure 7 and Figure 8 show the value of each target for both scenarios in order to easily visualise the consequences of choosing one route or another.

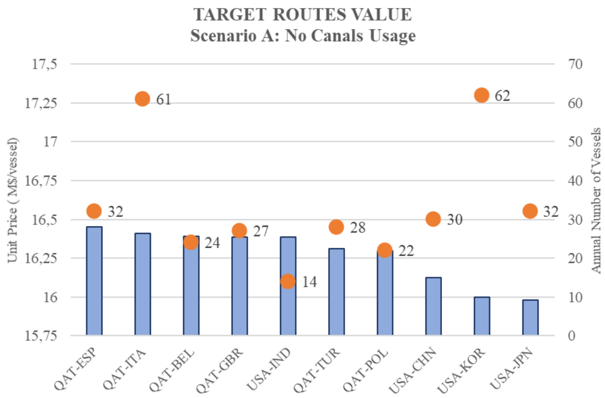

Figure 11. Ranking of Target Routes Value assuming No Canals Usage. Source: own elaboration.

Scenario A. No Canals Usage: trip chosen implies no using Panama Canal nor Suez Canal when traveling from target route's origin to destination. Large distances and a market moving toward global unification would trigger a reduced scattering between each route unit price. In accordance, with the predictive model, the mean unit price for target routes is 16.27 M$/ship and a typical deviation of 0.174.

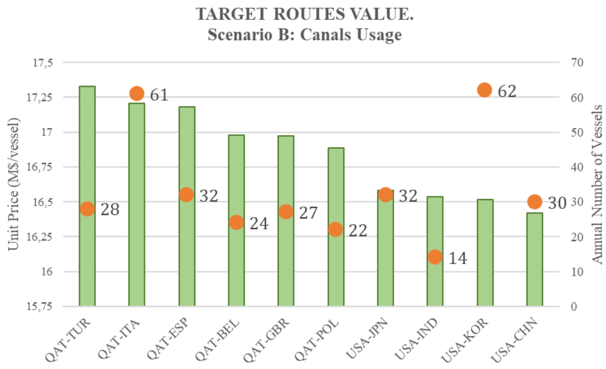

Figure 12. Ranking of Target Routes Value assuming Canals Usage. Source: own elaboration.

Scenario B: Canals Usage: The trip chosen implies using the Panama Canal or Suez Canal when traveling from the target route's origin to destination. For every target route, the value of LNG ship increases partly motivated by the distance reduction. The mean unit price for target routes reaches 16.86 M$/ship and a typical deviation of 0.327. Results' scattering demonstrates a variation between each target route benefit from canals usage. For the case of target routes with Qatar origin, QAT_TUR would benefit the most. Moving from the fifth to the first position in target routes' economic value ranking. For the case of target routes with USA origin, USA_JPN would benefit the most. Moving from the fourth to the first position in target routes' economic value ranking.

As the LNG market is moving to a global unification, the decision-making behind choosing one specific route or another could be a worth considering concern for traders. This model is intended to serve as tool for strategic decisions aimed at cost reduction. As expected, distance between origin and destination is a relevant factor when it comes to analysing LNG spot price. The results showcase the great consequences that this variable has on target routes' economic value. Nonetheless, taking into consideration the transported volume throughout each rote, QAT_ITA spots the first position in terms of annual LNG ships (62) in contrast with QAT_TUR that is lastly positioned

Finally, on Table 1 are shown the percentage increase unit price on LNG target routes by using Canals.

Table 1. Target route value increase using canals. Source: own elaboration.

Destination	Unit price increase (%)
QAT-TUR	6.23
QAT-ITA	4.88
QAT-ESP	4.42
USA-JPN	3.75
QAT-POL	3.60
QAT-GBR	3.58
QAT-BEL	3.58
USA-KOR	3.23
USA-CHN	1.82
USA-IND	0.90

5. Conclusions

The predictive model results reveal the influence of an accurate decision in the process of choosing LNG shipping routes could be critical. A direct impact on LNG Price is presumed and consequently in the natural gas industry. As suspected, the greater the distance between origin and destination, the lower the price of a product. This fact is mainly prompted by a transport cost increase. Thus, for target routes' economic valuation should be considered those with the greater export price. By implementing the predictive model shown in this paper, target routes can be shorted by their economic value.

Apart from that, the impact transportation infrastructures have on the liquified natural gas industry can be evaluated. Analysing those routes with various trip choices yield fruitful insights. For Scenario A where the Panama and Suez Canals remain unused, distances travelled are larger. Consequently, the mean LNG ship price is 16.27 M$/unit. For Scenario B, the usage of these infrastructures may trigger a surge in the mean LNG ship price up to 16.68 M$/ship. Analyses like this pretend to serve as a reference for shipping route choice evaluations in other cargo segments. This tool can also help for new shipping routes' evaluation, for instance, the Artic Route that may totally change the trade between Russia and China. Thus, this paper and the predictive model presented on it are expected to help for future infrastructure investment decisions, especially in transportation engineering.

6. Limitations and biases

Lack of data was a major problem during this project. As part of academic research, it was difficult to obtain real data that was sensitive for trading companies. For that reason, the model could have been insufficiently trained to achieve a great performance. In the future it is hoped to use real data and bigger computational power to transform this theoretical model into more realistic.

By the time this analysis was done, the training data was selected considering the natural gas industry landscape at that moment (2018 LNG Exports Market from International Gas Union). Besides, other assumptions out of the scope of the article like vessels capacity, fuel prices, geopolitical landscape might differ in the present. Therefore, not only the training model but the whole analysis should be updated at a certain frequency in the future to prove its accuracy. However, the main purpose of this first version is to show the technique and validate it with future publications using more recent data.

References

1. Shaikh, F., & Ji, Q. (2016). Forecasting natural gas demand in China: Logistic modelling analysis. International Journal of Electrical Power & Energy Systems, 77, 25-32
2. Asif, M., & Muneer, T. (2007). Energy supply, its demand and security issues for developed and emerging economies. Renewable and sustainable energy reviews, 11(7), 1388-1413
3. Ecofys, MRI Research Associates, Japan (2018). International comparison of fossil power efficiency and CO2 intensity.
4. International Energy Agency, IEA (2019). World Energy Outlook 2019.
5. Höök, M., & Tang, X. (2013). Depletion of fossil fuels and anthropogenic climate change—A review. Energy policy, 52, 797-809.
6. Krane, J. (2017). Climate change and fossil fuel: An examination of risks for the energy industry and producer states. MRS Energy & Sustainability, 4.
7. Hibbard, P. J., & Schatzki, T. (2012). The interdependence of electricity and natural gas: current factors and future prospects. The Electricity Journal, 25(4), 6-17.
8. Aguilera, R. F., & Radetzki, M. (2014). The shale revolution: Global gas and oil markets under transformation. Mineral Economics, 26(3), 75-84.
9. Medlock, K. B., Jaffe, A. M., & O'Sullivan, M. (2014). The global gas market, LNG exports and the shifting US geopolitical presence. Energy Strategy Reviews, 5, 14-25.
10. Kenneth B. Medlock III (2016). The shale revolution and its implications for the World Energy Market.
11. Cummins, M., Murphy, B., Roncoroni, A., & Fusai, G. (2015). Natural gas markets and products (pp. 135-180). West Sussex: John Wiley & Sons Ltd.
12. Schernikau, L. (2010). Economics of the International Coal Trade. London: Springer.
13. Molina Serrano, B., González-Cancelas, N., & Soler-Flores, F. (2018, December). Sustainable short sea shipping: social component analysis through decision networks. In Proceedings of the Institution of Civil Engineers-Maritime Engineering (Vol. 171, No. 4, pp. 135-144). Thomas Telford Ltd.Author 1, A.; Author 2, B. Title of the chapter. In *Book Title*, 2nd ed.; Editor 1, A., Editor 2, B., Eds.; Publisher: Publisher Location, Country, 2007; Volume 3, pp. 154–196.
14. Ríos-Mercado, R. Z., & Borraz-Sánchez, C. (2015). Optimization problems in natural gas transportation systems: A state-of-the-art review. Applied Energy, 147, 536-555.
15. Pajączek, K., Kostowski, W., & Stanek, W. (2020). Natural gas liquefaction using the high-pressure potential in the gas transmission system. Energy, 117726.

 doi:10.20944/preprints202212.0252.v1

16. González-Cancelas, N., Serrano, B. M., & Soler-Flores, F. (2019). Seaport sustainable: Use of artificial intelligence to evaluate liquid natural gas utilization in short sea shipping. Transportation Journal, 58(3), 197-221.
17. Psaraftis, H. N., & Kontovas, C. A. (2013). Speed models for energy-efficient maritime transportation: A taxonomy and survey. Transportation Research Part C: Emerging Technologies, 26, 331-351.
18. Serrano, B. M., González-Cancelas, N., & Soler-Flores, F. (2020). Analysis of the port sustainability parameters through Bayesian networks. Environmental and Sustainability Indicators, 100030Author 1, A.; Author 2, B. *Book Title*, 3rd ed.; Publisher: Publisher Location, Country, 2008; pp. 154–196.
19. Carral, L., Tarrío-Saavedra, J., Crespo-Pereira, D., Fernández-Campoamor, M., & Sabonge, R. (2019). Competitive position of the expanded Panama Canal in the shipment of liquefied natural gas. Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment, 233(4), 1154-1176.
20. Ritz, R. A. “A Strategic Perspective on Competition between Pipeline Gas and LNG.” The Energy Journal 2019 (n.d.).
21. Sakarya, S., Eckman, M., & Hyllegard, K. H. (2007). Market selection for international expansion: Assessing opportunities in emerging markets. International Marketing Review, 24(2), 208-238.
22. Bhutto, A. W., Bazmi, A. A., & Zahedi, G. (2013). Underground coal gasification: From fundamentals to applications. Progress in Energy and Combustion Science, 39(1), 189-214.
23. Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information sciences, 275, 314-347.
24. L’heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. IEEE Access, 5, 7776-7797.
25. Su, Moting, Zongyi Zhang, Ye Zhu, Zha Donglan, and Wenying Wen. “Data Driven Natural Gas Spot Price Prediction Models Using Machine Learning Methods.” Energies 12 (May 3, 2019).
26. Mishra, P. “Forecasting Natural Gas Price - Time Series and Nonparametric Approach,” 2012.
27. Cardin, M. A., Ranjbar-Bourani, M., & De Neufville, R. (2015). Improving the lifecycle performance of engineering projects with flexible strategies: example of on-shore LNG production design. Systems Engineering, 18(3), 253-268.
28. Jensen, J. T. (2004). The development of a global LNG market: is it likely? If so, when?. Oxford Institute for Energy Studies
29. Al-Kubaisi, A. (2008, January). An Insight into the World's Largest LNG Ships. In International Petroleum Technology Conference. International Petroleum Technology Conference.
30. Inchauspe, J., Aguilera, R., Packey, D., Burns, K., & Trueck, S. (2018). GAS AND OIL PROSPECTS.
31. Quijano, Jorge L. “THE EXPANDED PANAMA CANAL AND ITS IMPACT ON TRADE,” 2017
32. International Gas Union. (2020). 2019 World LNG Report (Pag.20)
33. Géron, A. (2017). Hands-on Machine Learning with Scikit-Learn & TensorFlow. (N. Tache, Ed.) (First Edition).
34. Yiu, T. (2019). Understanding Random Forest: How the Algorithm Works and Why It Is So Effective.