Association Rules and Deep Learning Paradigms for Big Data Processing

P-ISSN: 1680-9300
E-ISSN: 2790-2129
Vol. (24), No. (4)
pp. 36-45

Mohammed F. Alsarraj

Dept. of Information Techniques Management, Administrative Technical College, Northern Technical University, Mosul, Iraq

Abstract:

This in-depth study looks at data on fuel use in two main ways to find trends and make better predictions. One way is to learn how to use machine learning and association rule mining to try to guess what will happen. It uses association rules to show how things in a set are linked and how they do a lot of different things together. We can learn more about the parts that work together to change how much fuel is used. The RNN, TCN, and LSTM machine learning models can all guess how much fuel will be used, but they can do so in different ways. The TCN plan works out the best. The results show how important it is to choose a model design that makes the dataset's features better by putting together numbers and people's ideas about what might be important. We might be able to fully understand how fuel use changes over time if we put together what machine learning and association rule mining tell us. The numbers make it clear that the collection should be used for more research. There are different sets of methods that should be used for machine learning and more general statistical methods. People who give money, make rules, and try to guess how people will use fuel in the future should think about these ideas. It was found that people will be able to make better predictions in the future if they learn more about complicated machine learning design and link rules. The study is a good way to find out how much fuel people use when they switch sources of energy. When people use different types of fuel, we can also see how much they use. We can make the most of what we have this way.

Keywords: RNN, TCN, LSTM, MSE, MAE

1. Introduction:

You can plan ahead and make better use of the energy you have if you know how much energy different people in Europe use. A lot of places, like homes, companies, and places that make energy, need more gas (Tian et al., 2023). People change how things are used when new tools and data come out. It is easy to guess how fuel use will change if you look at how it has changed in the past. We can better plan the economy and make energy sources work well and last a long time when we use technology together. People must be able to live. People need to know how much petrol will be used in order to plan for it. France will always have fuel this way. This group of businesses needs prediction model data (Meira et al., 2022) since petrol is the main fuel used for heating, making energy, and many other business chores. For them, they should be able to explain and apply the rules towards the environment, structures and energy. In this case, tools and data serve a variety of purposes. Lots of data are sent and received by IoT

Journal of Prospective Researches
Vol.(24), No.(4)
The paper was received in 12 August 2024; Accepted in 11 October 2024; and Published in 30 November 2024
Corresponding author's e-mail: mohammed_alsarraj@ntu.edu.iq

10.61704/jpr.v24i4.pp36-45

(internet of things) things. With these tools learning of many things becomes fast and simple. One typical tool that analyzed a great deal of integrated data quantized different sorts of data to illustrate how the utilization of fuel evolved through time. Those aspects of the big files that are difficult to discern, indicate something perhaps when looked at by employing machine learning. They use two different machine learning when they use the historical analysis to have a good estimate of how much fuel is consumed. Many things can make things change this include factors like the weather the state of the economy and the fashion. These models should be used when you consider such things or something like these. You can always be in the know. This way, you may always control as to how much fuel is being consumed and make choices from the most current data. Predicting modellers in particular have to be able to switch their views within a short time if forecasts fail. That is why it is becoming possible and desirable to work simultaneously with it. It becomes easy for people to understand how we can use fuels less or perhaps more depending on how those numbers are presented. Those in authority can comprehend what is presented on the computer screens. This in turn enables them make right choices on how to spend their resources. However, it proves very difficult to determine how much fuel is being consumed when it comes to data, or computers (Agand et al., 2023). Data has to be protected and the prediction tools have to be employed appropriately. It may be attempted to predict how much fuel is to be consumed but it's also necessary to incorporate new technology that won't compromise people's rights. Last but not least, how popular the green energy methods are in EU countries will depend on how well the computers and data function. About energy, resource, world protection and making decisions upon them – we need prediction models in order to decide properly. As the smart technologies advance people would need to start becoming more truthful and trusting to results. This will assist in optimizing on usage of the available energy sources in the world.

2. Literature Survey:

There are severe economic and environmental impacts of shocks including natural disaster and extreme weathers in the United States of America (USA) (Yuan et al., 2022), (Ismail et al., 2021). Evaluations of the kind of effect which the climatic alterations take on the EURO COUNTIRES and global warming effects were made several years back. An association of fossil fuel consumption and emission as the cause of changes in climate and natural disasters in the EURO COUNTIRES is recognized. Meeting these challenges requires transition to cleaner fuels, and the use of natural gas as a cleaner fuel starts to become seen as a progressive step (Günay and Tapan, 2023). However, influences such as availability of resources, negative effects and costs are still barriers to this mission (Jiao et al., 2023). The USA being endowed with Natural gas deposits has been at the forefront of reforming the Natural gas market since 1970's and has one of/ the world's longest Natural gas Pipeline networks (He et al., 2023), (Thabit and Ibraheem, 2019). Even though they are concerned with the cost and the returns with energy decisions especially with natural gas that has been both expensive and risky to invest in the EURO COUNTIRES has cautiously embarked on construction of underground natural gas storage (Wang et al., 2023). However, since information suggesting the faster exhaustion of natural gas reserves in the EURO COUNTIRES has been received, discussions about substitutions have been made (Bahashwan et al., 2024). Technological enhancements in natural gas – the shale gas, tight gas, deep gas, coal bed methane, arctic and sub-sea hydrates are now considered to be realizable natural resources that have bright future in natural gas supply assurance (Wenninger et al., 2022). Forecasting the natural gas demand in the EURO COUNTRIES is not an easy task mainly because of existence of strong indicators as well as mediate causal factors of other related variables (Duan and Luo, 2022). Some techniques such as SVM and ANN globally accepted for forecasting natural gas consumption have numerous applications, which include forecasting crude oil price, bank performance, fraud detection of financial statements, water level, air quality parameters, global solar radiation, indoor temperature, stream flow, stock price, temperature, flood, rainfall, electric load, rice yield, energy consumption, risk (Xu et al., 2023). DLST for the natural gas supply and demand: Deep learning models utilizing ST-convolution networks are applied in Germany to provide precise and efficient forecast (Kim et al., 2021). Likewise, SVM and improved artificial fish swarm algorithms have been used to predict the natural gas consumption in the China context however, nonlinear grey Bernoulli model using Hodrick-Prescott: filter has been used in the European context (Nguyen et al., 2023), (Thabet et al., 2023). Despite the fact that the USA has employed multiple regression and grey models, among other modelling techniques for this purpose, the prediction of natural gas consumption through the application of only SVM and ANN algorithms has not been done much in the literature (Chen et al., 2021). While dealing with the difficult task of natural gas consumption fore-casting, these complicated algorithms give the chance to enhance the forecasting decision, as well as to make an informed choice in the area of energy consumption. A study on the forecasting of the natural gas consumption in the EURO COUNTIRES has revealed how SVM and ANN modelling methodologies have especially helped capture non-linearity in long-term data sets (Yi et al., 2023). SVM modelling has been applied in numerous studies as it proved to be very successful, however, the results are sensitive on the selected model parameter (Chen et al., 2023). Likewise, application of ANN has benefited in enhancement of the fitting and forecasting capacity where non-linear and complex data set of natural gas use have been more successful. This work fills literature void by applying two widely used accurate prediction models, ANN and SVM, on the EURO COUNTIRES natural gas consumption data set. Comparatively, the performance academic of these machine learning methods with the traditional multiple linear regression methods is evaluated. The study's structure is set up as follows: The literature review for a number of different topics related to the use of natural gas in the United States as well as its antecedents and consequences is provided in the following section.

3. Forecasting Model:

It is argued that macroscopic perspective of the forecasting strategy is important as well as the assessment of the possible risks that may hinder efficient forecasting that is a core element for decision making. Different forecast at the same logical level can stem from different decision-making angle the outcome and different forecasting strategy can be ascribable to different categorization such as point, interval, direct and rolling forecasting and probability density forecasting. While rolling forecasting is a kind of rolling method where forecasts a given number of months or weeks ahead are created successively with prior results as inputs, direct forecasting involves building a model with a fix time horizon for direct forecast value. Notably, rolling forecasting that uses several steps can avoid the basic cumulative mistakes, while one-step straight forecasting often reveals higher precision. Whereas interval and probability density forecasting are non-deterministic in the output by providing forecast ranges of values and probability distribution respectively point forecasting provides exact values.

Many forecasting models and techniques are demonstrated in a number of studies, including (Yan et al., 2022). To forecast natural gas loads with confidence intervals, for example, a hybrid forecasting algorithm is applied, or better singular spectrum analysis (SSA) and LSTM hybrid models are used for short-term load forecasting. At (Di Sorbo, 2022), using kernel support vector quantile regression (KSVQR) and Copula theory show the significance of probability density forecasting, which has not received as much attention in the field of natural gas forecasting. In contrast, mixed-frequency dynamic factors for NGD in China are the main focus of (Thabit er al., 2023).

One approach of the flexible rolling forecasting methodologies is the multi-step rolling forecasting process that is most effective when implemented in real-time or near-real-time environment whereas the one- step direct forecasting is much more effective for the longer-term trend analysis. Various objectives are achieved with point, interval, and probability density forecasting. This paper also provides information about what strategies are more accurate, fast and computationally efficient for the given problem and which strategies are not. Forecasts for natural gas consumption (NGC) are divided into three time periods: these include; short-term, medium-term and long term. The measurement of the spatial scale is also considered; the areas are divided into global, country, regional, and site.

Consequently, there are many ways of making predictions of NGC, for example, by feature processing, data analysis, forecast models and the optimisation. It just gets better the more of the steps in removing and picking that you do. It is disassembled which means that it is disintegrated into components which are used to enhance models and outcomes. The first and the main instrument is a forecasting model. These models can be individual or interconnected, which form a network. Specific examples of these models are graphs, statistics, fuzzy logic, and grey models. These are the components that constitute the method for forecasting NGC. By making predictions by empirical data or probabilistic-models it shall serve the goal of providing useful models for predictions. This list says everything that need to be said or need to know about the NGC models and methods. The information assists the workers and students to make a wise decision in using the appropriate energy.

4. Spark Based Mining:

Dealing with large volume of data in regular computations environments will results in over load on the kernel and hence the lag on the process. There are many ways to find useful data in Spark apps' big groups. Apache Spark is free to get. This tool can handle a lot of data on computers that are shared. You can easily choose what to do when you know a lot about it. A lot of different things can be done with Spark. You can use it to quickly look at data, find patterns, and guess what will happen next. Thus, one of the suitable environments for processing a large volume of data is called as Spark. One reason is that when many computers work together, they can finish jobs quickly. Spark is based on something called "Resilient Distributed Datasets" (RDDs). The same data can be used by many things at the same time. It's easier and faster to look at big files when you parallelize them. It also makes data mining tools work better. Spark can be used with a lot of different machine learning tools. This makes it easy to make models for deep data mining. When you use Spark to mine data, you need to be careful of a lot of things. Spark is converting the data into sub folds. That is effective eases the use of large amount of data by partitioning the large volume into smallest volumes that mitigates the load on the processors.

You can use MLlib, a different set of Spark tools, to do this. One way to do standard screening is to use regression. Another way is to use classification and grouping. This method works better when it's simple to connect SQL searches with machine learning. Over time, they can improve their models by using what they learn from those tests. Spark Streaming can work with data in real time, which is a big plus. Data sets that change all the time are great for live data mining because they let people learn quickly. Spark moves data in small groups, which makes it easier to get data from the data that is coming in. We can now study in ways that are adaptable and change over time. You can work with graphs with GraphX, which is an add-on for Spark. You can see information that is linked this way. Like systems that check for scams, systems that give you ideas, and systems that check out social networks. This is very important for systems that need people to talk to each other a lot. There are graph-based ways to get a lot of different kinds of data into the Spark system.

Spark is a safe and open way to browse through large datasets and spot important trends and results. Spark has facilities for machine learning and all your searches can be performed in SQL. This makes it possible for computers within several places to interconnect. Science and college people can use this tool in order to do their work every time as many times they want and even in the place real time. Large data analysis needs to be taken care of, and Spark is exactly the tool that can help businesses make the right decision. The streamed data can be easily switched with the real time data, and both are effective.

4.1 Dataset

Using the enormous collection of daily time series found in the "European Natural Gas Demand" (ENaGaD) database (Jiao et al., 2023). It provides information on the transmission networks of the 25 European members and a few additional European nations, in addition to the national natural gas consumption of each member. The dataset is presented in energy units of measurement and covers the years 2015 through 2020. In accordance with Regulation (EC) No 715/2009, the values are primarily taken from the National Transmission System Operators' transparency platform. Where possible, the daily demand is also broken down into three categories: household consumption, industrial users' consumption, and producers of heat and electricity. The ENaGaD database can also be easily accessed at https://data.jrc.ec.europa.eu, the Joint Research Centre Data Catalogue.

4.2 Apriori Algorithm

When the large scale data such as gas consumption is analyzed, the one of the most used association rule mining technique is Apriori algorithm.m. The preprocessing stage is important for arranging the data; considering discretization for the continuous form of data; and recognizing the transactions as well as items in the data set. They can be the time of day, the weather or industrial activity and the flow can be daily,

monthly or any other time considered relevant. To use the info in that way, it has to converted into a number or code first! That is different for every kind of variable. The next is to create trust and help goals. Rules just need to be interesting and for them to be so, trust must be the baseline. All that is required for something to be of interest is support. The Apriori method allows making set of similar things. These are products which are associated because most of them are usually marketed in the same categories. From this what has been identified these groups of things are then employed to make rules for the other. It worked and I want to remind that one needs to run the script and pay attention to the coefficients. If you are looking at the data that is interesting and linked between how many gas people use you have to look at the link rules that were created. Placing the rules in lists and graphs will assist people grasp them and describe the rules more articulately.

This procedure is often cyclical, and amplification is often required by altering the encoding technique, incorporating more research variables, or altering the thresholds. The last step is to apply and analyze rules as found out in the process of the study. The insight that can be gained from the regulations may help one to get a handle on the variables that may affect the trends of usage of the gases. These outcomes may afterwards be used for the betterment of the existing methods of earlier anticipation of the gas consumption, planning the use of reserves, and other decisions is to apply and interpret the rules that have been found. The regulations' insights can lead to a better understanding of the variables affecting gas consumption trends. After being interpreted, these findings can be used to enhance forecasting models for gas consumption, optimize resource allocation, and improve decision-making processes. This feature of Apriori algorithm means it is susceptible to tweaking in order to enhance discovered rules to meet the goals set out in the research and the specialist field knowledge.

The dataset of Euro gas is been analyzed in order to understand the features that are more dominant, the analysis is illustrated in the Table 1.

Table 1. Dataset reordering/filtering base of 'AT' features items and its computation from gas

COUNTRY	GASDAY	DD	MONTH	YEAR	JULIANDAY	TOT
AT	01/01/15 0:00	1	1	2015	1	350.052
AT	01/02/15 0:00	2	1	2015	2	374.715
AT	01/03/15 0:00	3	1	2015	3	337.227
AT	01/04/15 0:00	4	1	2015	4	330.771
AT	01/05/15 0:00	5	1	2015	5	380.854

The information stated in the table is the processed data from Spark, that reflects the Austria consumption in gas. The data is partitioned into columns, like “GASDAY” and “JULIANDAY” that refers to the day of the month and the date of the record the data. For instance, the first row shows data on January 1, 2015, at midnight (00:00). The day of the month can be marked by the “DD” and the year can be presented in the “MONTH” column also. The year is “2015”, and the day of the year, also known as Julian day, is a number. Total gas usage for each instance is presented into the “TOT” column showing an overall unit amount of gas for a given date or time. Table shows Austria’s consumption structure where consumption of gas was recorded on 1st January 2015 at midnight. The data is grouped by the date, time and Julian day in which the first three slots of ‘DD’ represents the first of the month. In the "TOT" column, there is a compilation of the total amount of gas that went through each instance and in a format such as 350.052 the amount of gas that passed through for that particular date and time.

Table 2 offers more significant and precise information on the volumes of gas consumed in Austria at some time, which allows analyzing the consumption dynamics. In combination, the columns offer a clear understanding of temporal distribution of the gas consumption, and any further analysis of the datasets can be conducted with ease.

Table 2. Austria's gas consumption at various points in time

COUNTRY	GASDAY	DD	MONTH	YEAR	JULIANDAY	TOT
AT	01/01/15 0:00	1	1	2015	1	350.052
AT	01/02/15 0:00	2	1	2015	2	374.715
AT	01/03/15 0:00	3	1	2015	3	337.227
AT	01/04/15 0:00	4	1	2015	4	330.771
AT	01/05/15 0:00	5	1	2015	5	380.854
AT	12/27/2021 0:00	27	12	2021	361	359.586
AT	12/28/2021 0:00	28	12	2021	362	347.54
AT	12/29/2021 0:00	29	12	2021	363	311.527
AT	12/30/2021 0:00	30	12	2021	364	261.047
AT	12/31/2021 0:00	31	12	2021	365	208.154

Table 3. Reorders of the data base on computation amount of gas

TOT	GASDAY	DD	MONTH	YEAR	JULIANDAY
90.094	08/04/18 0:00	4	8	2018	216
91.546	7/31/2016 0:00	31	7	2016	213
91.974	8/15/2015 0:00	15	8	2015	227
92.303	6/13/2020 0:00	13	6	2020	165
92.911	7/30/2016 0:00	30	7	2016	212
580.521	1/30/2017 0:00	30	1	2017	30
588.826	1/23/2017 0:00	23	1	2017	23
596.976	03/01/18 0:00	1	3	2018	60
604.141	1/24/2017 0:00	24	1	2017	24
611.22	2/28/2018 0:00	28	2	2018	59

Information concerning the amount of gas consumption is shown at the table below, and each row contains relevant details concerning the observed events. The total of gas for each record – denoted as 'TOT', is derived from the quantity of gas that BHP needs to consume in one unit of the reference year to achieve the contractual obligations. For instance, the value "90.094" in the first row is the amount of gas used at the time and date of the measuring (Table 3).

Each recorded data point's date and time are indicated in the "GASDAY" column using the format "MM/DD/YY H:mm." This column in the first row shows that data on gas consumption was gathered on August 4, 2018, at midnight (00:00). In the "DD" column, the day of the month of the occurred instance is shown, accordingly. The last numeral "4"in the column headed by the abbreviation "DD" in the first row means that the data on the gas consumption is for the fourth day of the month. Each of the occurrences noted on this table has a respective month described in the "MONTH" column; August – number 8 in the first row. Whereas in the first row '2018' which represents the particular year for which data have been collected, the column called 'YEAR' will represent the year.

The number of days in a given year counted sequentially, Julian day numbers are added in the "JuLIANDAY" column. The first row gives more time context to the gas consumption data and it is a representation of the 216th day of the year of the year is represented by the value "216" in the first row, which gives the gas consumption data more temporal context. For example, in the first cell on the gl account of 'Gas Consumption', there was '90.094' units that was used on the 4th of August in the midnight the previous year.8. Additional information concerning the temporal aspect is given by the values in the "DD", "MONTH", and "YEAR" columns, while the "JuLIANDAY" quantifies the day of the year. Table 3, gives more detailed information on the amount of gas used, which should therefore facilitate a quantitative analysis of trends in the gas usage patterns. The enhanced knowledge of the dataset is achieved through integration of the columns pertaining to the insightful information regarding the temporal distribution of the gas consumption.

Table 4. Association rules calactin for AT data.

10.61704/jpr.v24i4.pp36-45

Consequents	Antecedent	Support	Consequent	Support
(MONTH_2)	(YEAR_2016)	0.077434	0.143136	0.011341
(MONTH_2)	(YEAR_2020)	0.077434	0.143136	0.011341
(MONTH_4)	(YEAR_2015)	0.082127	0.142745	0.011732
(MONTH_4)	(YEAR_2017)	0.082127	0.142745	0.011732
(MONTH_4)	(YEAR_2018)	0.082127	0.142745	0.011732

The Table 4 displays data mining-derived association rules that illustrate the relationships between antecedent and consequent attributes and the support values that correspond to them. In the first rule, there is a support of roughly 7.74% when the attribute "YEAR_2016" is observed (antecedent), indicating a 7.74% likelihood of observing "MONTH_2" as a consequent. Based on the data, the consequent support of 0.143136 means that 14.31% of transactions independently contain "MONTH_2," and the antecedent support of 0.011341 means that 1.13% of transactions involve "YEAR_2016." In a similar vein, the second rule has "MONTH_2" as the consequent and "YEAR_2020" as the antecedent, with support values that correspond to the first rule's interpretation.

In the following rules (Rows 3-5), variations are deployed, with "MONTH_4" being the consequent, and "YEAR_2015," "YEAR_2017," and "YEAR_2018" as the antecedents. These rules talk about how likely it is, to see "MONTH_4" IF certain years are presence. The support values obtained by the counting frequency of these patterns reveal the existence of complicated and valuable dependencies between various months and years in the dataset. In other words, the table gives a higher-level understanding of the frequency analysis of the close occurrences in the dataset and shed light on the temporal aspects that are encapsulated by the association rules.

Table 5. Lift, Confedence, Leverage, conviction and zhangs metrics of the AT data

Confidence	Lift	Leverage	Conviction	Zhangs_Metric
0.146465	1.023252	0.000258	1.003899	0.024631
0.146465	1.023252	0.000258	1.003899	0.024631
0.142857	1.000783	0.000009	1.00013	0.000852
0.142857	1.000783	0.000009	1.00013	0.000852
0.142857	1.000783	0.000009	1.00013	0.000852

The table presented in this paper as the Table 5 indicates measures used in association rules mined from a dataset. In each row, the "confidence" value represents the probability of obtaining the consequent given the antecedent; the first row being 14.65%. The "lift" value determined by the perceived association strength relating to the first row, stands at 1.023 252 formalating a positive though feeble relationship. The term "Leverage" shows a little deviation from the mid-point value of 0; the first of the row gives the value, 0.000258. The "conviction" in the first line equals 1,003 899, this means that there is a positive asymmetry to the fact that the antecedent may occur without the consequent. Lastly, tests ,is basically good at 0.024631 at the first row reflecting that it indeed has moderate interestingness of the linked rule. It should be also pointed out that in case of subsequent rows, values are the same or bear great semblance to each other – it means that subsequent rules are similar or even completely identical. In combination, these metrics form the basis for evaluating the degree of quality, relevance, and directionality of association rules to inform the subsequent analysis and interpretation of the mined concrete patterns in the dataset. The model employs four types of recurrent neural networks to predict how much gas is going to be used. The three of these are the Temporal Convolutional Network or TCN, the Long Short-Term Memory or LSTM, and the Simple Recurrent Neural Network or RNN. The one employed in this paper is known as the LSTM model which has two layers. Dense_8 has 65 numbers and an output form is in the form of [none, 1]. From the LSTM layer (Lstm_5) it can be seen that the probability distribution approximates zeros and 0.64. Since they can quickly learn the long-term dependency in linear data, LSTMs are ideal for time series since they make a good prediction.

Layer (type)	Output shape	Param
Lstm_5 (LSTM)	(none,64)	16896
Dense_8 (Desne)	(none, 1)	65

The TCN model consists of total three layers of Conv1D. It produces (none, 4, 64), (none, 4, 32) and (none, 4, 16) outputs. When Flatten layer is added the changes in the output to the (none, 64) occurs. Then 65 numbers proceed to the Dense layer and a form which can be 1 or none is also sent. For the purpose of identifying patterns in time TCN design employs expanded convolutions. This makes it good for jobs that have to be done in sequence such as predicting time series.

Layer (type)	Output shape	Param
Conv1D	(none,4,64)	256
Conv1D	(none, 4, 32)	6176
Conv1D	(none, 4, 16)	1552
Flatten	(none, 64)	0
Dense	(none,1)	65

Easy RNN is the base of the RNN model. It comes back with the form (none, 64), which has 4,224 numbers. The form that Dense_1 gives you is [none, 1], and it has 65 numbers in it. Simple RNNs can tell you what short-term connections will be in straight data. This shows the model how patterns shift over time.

Layer (type)	Output shape	Param
SimpleRNN	(none,64)	4224
Dense_1 (Desne)	(none, 1)	65

The model which thinks about how much gas will be utilized contains all three types of models RNN, TCN, and LSTM. It can now do lots of things. It also shows that the best link for training LSTMs are the long term links, for TCNs the best time trends and for Simple RNNs the short term links. In this case, once these plans are made the model will be in a position to predict how much gas will be used better. This number for each layer is equal to shows how well the model can predict what will take place given the available information.

5. Results and Discussion:

Such approach allows the RNN, TCN, and LSTM models to show us how good their guess is on how much gas will be used based on their ability to identify time trends in the dataset. When the RNN model attempted to predict how much gas would be consumed, it did a pretty good job of it, with an accuracy of 0.0336 absolute error. As can be seen the model has an R2 value of 0.8492, which means that about 85% of the spread is well explained. The TCN model can guess more often and better, thanks to it, the TCN model performs better than the RNN model. The performance of TCN was higher as compared to both RNN and LSTM since its MAE was 0.0220 while MSE was 0.0014. The chosen model is rather accurate and it can account for 93% of the variation. The R2 number at 0.9323 is very high. It doesn't do so badly; the convolutions might be better made larger, as the TCN can pick up on time trends. It is good to estimate how much petrol is going to be used because of this.

The LSTM model was the highest with an MAE of 0.0553 and MSE of 0.0081. In other words, it was less accurate in its predictions, and got more things wrong. From an R2 number of 0.6041, it seems that the LSTM model can predict only 60% of variance on the given data about petrol use. This means that while the model is good at finding long-term relationships since it is an LSTMs, then it is not very effective at capturing the actual trends in this instances.

They performed better than the others as was seen above. This proves that grosser convolutions can capture differences in how people use gas over time. The model has a high R2 number which indicates that the model fits the real data arrangement as appropriate. In most cases the RNN model is correct. However, the TCN model has a better result in MAE and R2. This may mean that the RNN does not grasp the magnitude of difficulty required in tracking the trends in use of gas such as that achieved by the TCN. Even if there are many mistakes, LSTM will not be able to perceive long-term associations of the gas use data and if the R2 number is low. In other words, projects are less accurate. That all depends on the model used because, unlike simple arithmetic, it has a wide range of likely outcomes. As mentioned earlier that TCN model contains temporal convolutions hence it looks more appropriate to guess how much gas will be used. As regards to the value in performing numerical calculations and ability to predict competitors' outcomes, M is more accurate and has greater potential to guess than the RNN and LSTM models.

6. Conclusion:

One way this in-depth study looks at gas use data is through link rule mining. The other is through prediction models based on machine learning. These tips will help you see trends and guess right. It's interesting that association rule mining can look at data and find complicated ties and ways that things work together. The rules for relationships have taught us more about things that happen when two people get together. We can figure out what makes people use petrol this way. RNN, TCN, and LSTM models are used to try to guess how much gas a car will use. When it comes to how well and honestly they do their jobs, the types don't all seem to do them the same way. Every time, TCN did the same thing. It's clear from these results that we need to choose a model method that works well with we have. There is a full picture of how gas use changes over time when you use both association rule mining and machine learning estimates together. Trends from the past help computers figure out what will happen next. On the other hand, association rules help us fully understand how the data is connected and make it clear which factors may be important. You should use both advanced machine learning techniques and simple statistical methods to learn more about large datasets. These are the study's important points. It can also help people who work in energy make rules, figure out how to best use resources, and guess what people will do with gas in the future. We might be able to guess better when we look into more complex machine learning systems and link rules. The whole study helps us decide what to do, make good use of our resources, and see how our petrol use changes over time in a world where energy is changing quickly.

References

Agand, P., Kennedy, A., Harris, T., Bae, C., Chen, M., and Park, E. (2023), Fuel Consumption Prediction for a Passenger Ferry Using Machine Learning and In-Service Data: A Comparative Study, Ocean Engineering, 284, 115271.

Bahashwan, A., Ibrahim, R., Omar, M., and Amosa, T. (2024), Supervised Learning-Based Multi-Site Lean Blowout Prediction for Dry Low Emission Gas Turbine, Expert Systems with Applications, 244, 123035.

Chen, Z., Deng, Q., Zhao, Z., Sun, B., Peng, T., and Yang, C. (2021), Energy Consumption Prediction of Cold Source System Based on Graphsage, IFAC-PapersOnLine, 54(11), pp. 37-42.

Chen, Z., Zhao, Z., Deng, Q., Tang, P., Yang, C., Li, X., and Gui, W. (2023), A Knowledge Embedded Graph Neural Network-Based Cooling Load Prediction Method Using Dynamic Data Association, Energy and Buildings, 278, 112635.

Di Sorbo, A., Laudanna, S., Vacca, A., Visaggio, C., and Canfora, G. (2022), Profiling Gas Consumption in Solidity Smart Contracts, Journal of Systems and Software, 186, 111193.

Duan, H., and Luo, X. (2022), A Novel Multivariable Grey Prediction Model and Its Application in Forecasting Coal Consumption, ISA Transactions, 120, pp. 110-127.

Günay, M., and Tapan, N. (2023), Analysis of PEM and AEM Electrolysis by Neural Network Pattern Recognition, Association Rule Mining and LIME, Energy and AI, 13, 100254.

He, Y., Gong, Q., Zhou, Z., and Chen, H. (2023), Development of a Hybrid VRF System Energy Consumption Prediction Model Based on Data Partitioning and Swarm Intelligence Algorithm, Journal of Building Engineering, 74, 106868.

Ismail, S., Hussain, Z., Thabet, H., and Thabit, T. (2021). Using PI Controller Unit for Controlling the Water Temperature in Oil Fired Heaters by PLC Techniques, Przegląd Elektrotechniczny, 3, pp. 157-161.

Jiao, H., Song, W., Cao, P., and Jiao, D. (2023), Prediction Method of Coal Mine Gas Occurrence Law Based on Multi-Source Data Fusion, Heliyon, 9(6), pp. 1-9.

Jiao, Y., Tan, Z., Zhang, D., and Zheng, Q. (2023), Short-Term Building Energy Consumption Prediction Strategy Based on Modal Decomposition and Reconstruction Algorithm, Energy and Buildings, 290, 113074.

Kim, C., Kim, M., and Song, Y. (2021), Sequence-to-sequence Deep Learning Model for Building Energy Consumption Prediction with Dynamic Simulation Modeling, Journal of Building Engineering, 43, 102577.

Meira, E., Oliveira, F., and de Menezes, L. (2022), Forecasting Natural Gas Consumption Using Bagging and Modified Regularization Techniques, Energy Economics, 106, 105760.

Nguyen, S., Fu, X., Ogawa, D., and Zheng, Q. (2023), An Application-Oriented Testing Regime and Multi-Ship Predictive Modeling for Vessel Fuel Consumption Prediction, Transportation Research Part E: Logistics and Transportation Review, 177, 103261.

Thabet, H., Al-Hilali, A., Al-safi, M., Thabit, T., and Mahmood, F. (2023), Design and Simulation of an Affordable Vehicle Speed Detection System, International Conference on Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA) (pp. 260-265).

Thabit, T., Aleiwi, J., Azeez, O., Safi, H., and Raewf, M. (2023). The Mediating Effect of Accounting Practices in the Relationship Between GreenTech and Carbon Emissions, 2023 International Conference on Smart-Green Technology in Electrical and Information Systems (ICSGTEIS) (pp. 98-103).

Thabit, T., and Ibraheem, L. (2019), Implementation of Environmental Management Accounting for Enhancing the Sustainable Development in Iraqi Oil Refining Companies, 3rd Scientific

Conference of Administration and Economic College, University of Anbar, Anbar, Iraq.

Tian, N., Shao, B., Bian, G., Zeng, H., Li, X., and Zhao, W. (2023), Application of Forecasting Strategies and Techniques to Natural Gas Consumption: A Comprehensive Review and Comparative Study, Engineering Applications of Artificial Intelligence, 129, 107644.

Wang, F., Gu, W., Bai, Y., and Bian, J. (2023), A Method for Assisting the Accident Consequence Prediction and Cause Investigation in Petrochemical Industries Based on Natural Language Processing Technology, Journal of Loss Prevention in the Process Industries, 83, 105028.

Wenninger, S., Kaymakci, C., and Wiethe, C. (2022), Explainable Long-Term Building Energy Consumption Prediction Using QLattice, Applied Energy, 308, 118300.

Xu, L., Wang, Y., Mo, L., Tang, Y., Wang, F., and Li, C. (2023), The Research Progress and Prospect of Data Mining Methods on Corrosion Prediction of Oil and Gas Pipelines, Engineering Failure Analysis, 144, 106951.

Yan, Z., Zhong, S., Lin, L., Cui, Z., and Zhao, M. (2022), A Step Parameters Prediction Model Based on Transfer Process Neural Network for Exhaust Gas Temperature Estimation after Washing Aero-Engines, Chinese Journal of Aeronautics, 35(3), pp. 98-111.

Yi, J., Chen, K., Liu, H., Liang, K., Mi, H., and Zhou, W. (2023), A Hybrid Association Analysis Framework of Accident Reports for Offshore Drilling Platforms, Journal of Loss Prevention in the Process Industries, 85, 105161.

Yuan, J., Wang, L., Li, Y., Wang, Y., Ma, T., and Luo, X. (2022), Set Pair Prediction for Chinese Natural Gas Energy Security Based on Higher-Order Markov Chain with Risk Attitude, Resources Policy, 77, 102741.

10.61704/jpr.v24i4.pp36-45