Decision Science Letters 13 (2024) 237–248

Contents lists available at GrowingScience

Decision Science Letters

homepage: www.GrowingScience.com/dsl

An integrated approach for modern supply chain management: Utilizing advanced machine learning models for sentiment analysis, demand forecasting, and probabilistic price prediction

Issam Amellal[a], Asmae Amellal[b*], Hamid Seghiouer[b] and Mohammed Rida Ech-Charrat[b]

[a]Ensa Berrechid, Hassan 1er University, Morocco
[b]Laboratory Of Modeling, Optimization Of Industrial And Logistics Systems MOSIL, ENSA TÉTOUAN, Abdelmalek Essaadi University, Morocco

C H R O N I C L E

Article history:
Received: July 1, 2023
Received in revised format: August 22 , 2023
Accepted: September 28, 2023
Available online: September 28, 2023

Keywords:
Supply Chain Management
Demand Forecasting
Sentiment Analysis
Price prediction
Machine Learning
Probabilistic Models

A B S T R A C T

In the contemporary business landscape, effective interpretation of customer sentiment, accurate demand forecasting, and precise price prediction are pivotal in making strategic decisions and efficiently allocating resources. Harnessing the vast array of data available from social media and online platforms, this paper presents an integrative approach employing machine learning, deep learning, and probabilistic models. Our methodology leverages the BERT transformer model for customer sentiment analysis, the Gated Recurrent Unit (GRU) model for demand forecasting, and the Bayesian Network for price prediction. These state-of-the-art techniques are adept at managing large-scale, high-dimensional data and uncovering hidden patterns, surpassing traditional statistical methods in performance. By bridging these diverse models, we aim to furnish businesses with a comprehensive understanding of their customer base and market dynamics, thus equipping them with insights to make informed decisions, optimize pricing strategies, and manage supply chain uncertainties effectively. The results demonstrate the strengths and areas for improvement of each model, ultimately presenting a robust and holistic approach to tackling the complex challenges of modern supply chain management.

1. Introduction

Understanding customer sentiment and accurately forecasting demand and price is vital in today's fiercely competitive business landscape. These insights enable organizations to make strategic decisions, efficiently allocate resources, and adeptly manage their supply chains (Pathak et al., 2020 ; Kelleher *et al*., 2019; Chopra & Meindl, 2018). The emergence of social media and online platforms has brought forth an unprecedented volume of data, useful for deriving insights and predicting customer behavior. Harnessing this data aids in effectively managing decision-making uncertainty and market volatility (Ivanov & Dolgui, 2020). Recognizing the value of these insights, businesses are now utilizing powerful analytic tools such as machine learning, deep learning, and probabilistic models. These techniques not only help analyze data but also aid in optimizing pricing strategies by considering the complex interactions within the supply chain (Feizabadi, 2022; Seuring, 2013 ; Boutselis & McNaught, 2019 ; Dou et al., 2019).

Machine learning and deep learning methodologies have proven their mettle by handling large-scale, high-dimensional data and uncovering hidden patterns beneficial for decision-making and prediction (Goodfellow *et al*., 2016). They have demonstrated extraordinary performance in diverse fields, such as sentiment analysis, demand forecasting, and price prediction, often outperforming traditional statistical methods (Feizabadi, 2022; Devlin *et al*., 2018; Tao and Yang, 2022). Deep learning models, in particular, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), excel at learning significant features from data automatically, which minimizes the need for manual feature

* Corresponding author. Tel.: +212661944309
E-mail address: asmae.amellal57@gmail.com (A. Amellal)

© 2024 by the authors; licensee Growing Science, Canada.
doi: 10.5267/dsl.2023.9.003

engineering (Amellal et al., 2023). These models are particularly equipped to handle complex tasks prevalent in supply chain management (Feizabadi, 2022).

In addition to these methodologies, probabilistic models such as Bayesian Networks offer a robust framework for managing uncertainty and modeling sophisticated dependencies among variables (Pauwels and Calders, 2020). The progressive evolution of the e-commerce landscape necessitates an intricate investigation into complex interactions among various entities like contract manufacturers, original equipment manufacturers, and platforms. To this end, game-theoretical approaches have been instrumental in exploring these multifaceted relationships (Wang *et al*., 2023). Likewise, studies in competitive markets have utilized game-theoretical models to decipher price wars and competitive behavior (Rychłowska-Musiał, 2020; Aydinonat & Köksal, 2019 ; Aamer *et al*., 2020 ; Ma *et al*., 2021). Furthermore, other modeling techniques like agent-based models (ABMs) have been employed for analyzing complex systems comprising interacting agents. In a notable study (Esmaeili *et al*., 2022), used ABMs to simulate human behavior in pandemic situations, modeling them as volunteer dilemma games. By introducing probabilistic incentives, the study highlighted an approach to mitigate individualistic behavior and encourage collective action, thus offering important insights into promoting cooperative behavior in critical societal situations. In (Fera et al., 2017), System Dynamics models have been employed to scrutinize the uncertainties within supply chains for manufacturing and services firms in the context of a globalized economy. Markov Decision Processes (MDPs) provide another example, used to formulate decision-making problems and optimize processes over time under uncertainty (Puterman, 2014). However, Bayesian Networks stand out as a compelling alternative to these approaches. They offer a comprehensive depiction of causal relationships and have the capacity to adjust their beliefs in response to new evidence (Chickering, 2002). This adaptability makes them aptly suited to handle inherent uncertainties and fluctuations within supply chain dynamics, even in the details of after-sales scenarios (Tao and Yang, 2022).

BERT, introduced by (Devlin *et al*., 2018), has revolutionized NLP by pre-training deep bidirectional transformers for language understanding. Its ability to generate high-quality language representations has advanced various NLP tasks, such as sentiment analysis and text classification. Numerous studies have built upon BERT's success, fine-tuning the model and its derivatives to enhance performance across a range of NLP tasks (Geetha and Renuka, 2021; Yarullin and Serdyukov, 2021; Durairaj and Chinnalagu, 2021). Additionally, BERT's applicability has been extended to domains such as finance and healthcare (Wu *et al*., 2022; Teo *et al*., 2020). The impact of BERT and its derivatives on NLP and other domains continues to shape the development of AI applications and our understanding of language (Brown et al., 2020).

Gated Recurrent Unit (GRU) models, widely used in demand forecasting across various industries, have demonstrated their adaptability and efficiency. Many studies have combined GRU with other techniques to enhance forecasting accuracy, while others have compared GRU with different neural network architectures. For instance, (Noh *et al*., 2020) integrated a GRU model with a genetic algorithm for product demand forecasting in supply chain management, while (Honjo *et al*., 2022 ; Huang *et al*., 2022) introduced a CNN-GRU-based deep learning model for demand forecasting in the retail industry. GRU models have proven effective in various industries, from agriculture load forecasting (Saini *et al*., 2020) to automotive spare parts demand forecasting (Ma *et al.,* 2021). However, it remains an open question whether the complexity introduced by some of these combinations is necessary for accurate demand forecasting. Simpler, well-tuned GRU models may yield comparable, if not better, results with greater computational efficiency (Pearl, 1985).

Bayesian networks, with their ability to handle complex relationships and uncertainty, have seen successful implementation in various prediction tasks across diverse domains. (Pauwels and Calders ,2020) emphasized the effectiveness of Bayesian networks in predicting business processes, while (Balta *et al*., 2021) demonstrated their value as decision support tools in complex construction projects. Moreover, in the field of water resources, (Avilés *et al*., 2016) and (Noorbeh *et al*., 2020) illustrated the potential of Bayesian networks for probabilistic forecasting. Despite these promising results, more comprehensive comparisons with other prediction techniques and broader performance evaluations across various conditions would improve the generalizability of these findings.

Our study employs Bayesian networks for price prediction, leveraging their ability to model complex dependencies, manage uncertainty, and adapt to new information, qualities particularly valuable in the inherently uncertain and complex financial markets (Fama, 1970). Bayesian networks also allow for the incorporation of both quantitative data-driven and expert knowledge into the model, a feature of significant importance in the financial domain (Buntine, 1996).

This paper presents a novel, multifaceted framework combining machine learning, deep learning, and probabilistic models to interpret customer sentiment, forecast demand, and predict prices. Our approach differs from previous work in that it integrates these advanced methods into a unified model, providing a comprehensive understanding of supply chain management processes.

2. Material and methods

In this section, we initially present the proposed framework as described in Fig. 1

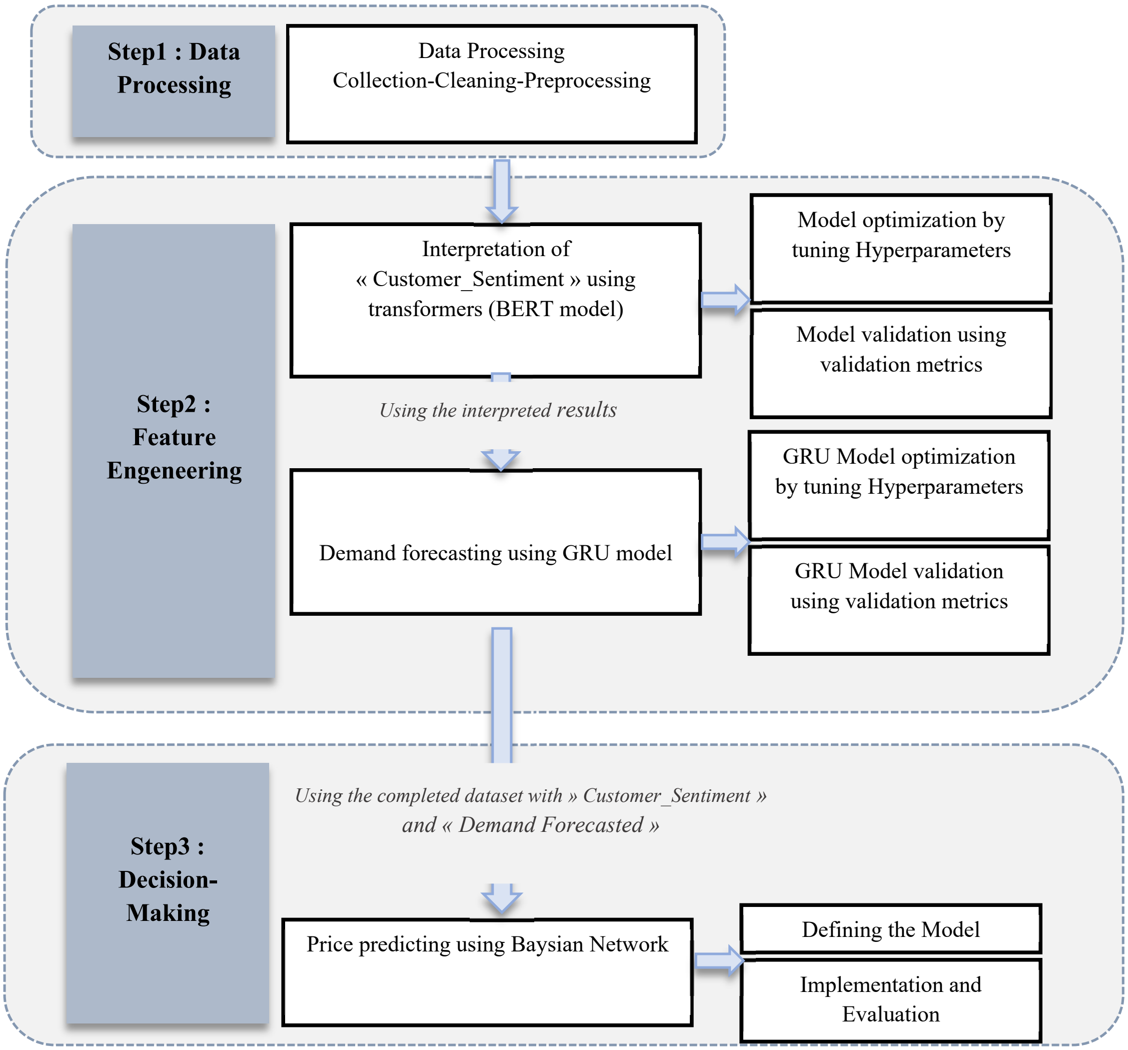

Fig. 1. Proposed Framework

Data Collection: The study used data from the ERP system of a Moroccan car distribution company, focusing on after-sales service visits from 2019 to 2022. The dataset contains 111,236 rows and 10 columns with information about customers, service dates, mileage, branch workshops, types of repairs, invoice prices, invoice numbers, margins, customer satisfaction survey comments, and vehicle age, brand, and model.
Data Cleansing: After removing 191 duplicate rows, negative margin values, and outliers, the final dataset had 91,769 rows and 12 columns. Empty cells were replaced with zeros, and data types and formats were standardized for consistency.
Data Preprocessing: Categorical variables were encoded as numerical values, and normalization was applied using Min-Max normalization to standardize the data scale. The formula for Min-Max normalization is:

$$x_{\min} = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} \tag{1}$$

The data was then split into a training set with 73,415 samples and a test set with 18,354 samples.

Interpretation of « Customer_Sentiment » using BERT model

The BERT model, based on the Transformer architecture (Vaswani *et al.*, 2017), employs multi-head self-attention, position-wise feed-forward networks (FFNs), and layer normalization. In multi-head self-attention, query (*q*), key (*k*), *d_k* the dimensionality of the key vectors, and value (*v*) vectors are computed using learned weight matrices (*Wq*, *Wk*, and *Wv*), and attention scores are calculated as follows:

$$Attention\ (q, k, v) = Softmax\ (q * k^T / \sqrt{d\_k}) * v \quad (2)$$

Multi-head attention combines multiple attention heads:

$$MultiHead\ (q, k, v) = Concat\ (head\_1, ..., head\_h) * Wo \quad (3)$$

Position-wise FFNs consist of two linear transformations with a ReLU activation function:

$$FFN(x) = max\ (0, x * W1 + b1) * W2 + b2 \quad (4)$$

Layer normalization stabilizes the learning process:

$$LayerNorm(x)=(x-mean(x)) / \sqrt{(var(x)+\xi)} * \gamma+\beta \quad (5)$$

In the above equations, *b1* and *b2* represent the bias terms for the position-wise feed-forward networks (*FFN*s), *W1* and *W2* are weight matrices, while ξ, γ, and β are parameters employed in the layer normalization procedure.
BERT is pre-trained using masked language modeling and next sentence prediction before fine-tuning on specific tasks (Devlin *et al*., 2018 ; Sirusstara *et al*., 2022).

Proposed model

The Algorithm 1 presents the process of sentiment analysis we applied using the BERT model.

Algorithm 1 *BERT model process*

Step 1 : Load BERT model and tokenize	**Step 2** : Create sentiment analysis pipeline
→ Input: Pre-trained BERT model and tokenizer	→ Input: BERT model, BERT tokenizer
→ Process: Load BERT model and tokenizer	→ Process: Build sentiment analysis pipeline
→ Output: BERT model, BERT tokenizer	→ Output: Sentiment analysis pipeline
Step 3 : Define analyze_sentiment function	**Step 4 :** Apply analyze_sentiment to DataFrame
→ Input: Text	→ Input: "Comment" column
→ Process:	→ Process: Apply analyze_sentiment function
a. Check if input text is valid	→ Output: "Customer_sentiment" column
b. Apply sentiment analysis pipeline	
c. Convert sentiment label to numerical value	
→ Output: Numerical sentiment value	

In the proposed model, the BERT model is utilized for sentiment analysis through a streamlined four-step process as described in Algorithm 1. Firstly, a pre-trained BERT model and tokenizer are loaded, with the tokenizer playing a key role in converting input text into processable tokens for the BERT model. The subsequent step involves the creation of a sentiment analysis pipeline, leveraging the loaded BERT model and tokenizer. This pipeline is designed to streamline the sentiment analysis process, allowing the input text to be efficiently tokenized, processed through the BERT model, and interpreted for sentiment. The third step outlines the functionality of an "analyze_sentiment" function, which processes valid input text through the aforementioned sentiment analysis pipeline, translating the output sentiment labels into numerical values to quantify the sentiment. Finally, in step 4, the "analyze_sentiment" function is applied to a DataFrame to derive the "Customer_sentiment" column from the "Comment" column, enabling each customer comment's sentiment to be assessed and quantified. This utilization of the BERT model thus facilitates an efficient, precise, and quantifiable sentiment analysis of the provided text data.

Before applying the BERT model to the comments in our dataset, we performed several preparation steps:

- Text Cleaning: The text data was cleaned to remove irrelevant information. Given that the comments are in different languages (including German and French), language-specific cleaning tasks were handled, such as removing language-specific stop words, punctuation, and unnecessary spaces.
- Tokenization: The BERT tokenizer, specifically designed for multilingual models (like 'bert-base-multilingual-cased'), was used to break down the text into smaller units or tokens, which is essential as BERT models interpret text data at the token level.
- Adding Special Tokens: Special tokens were added at the beginning ([CLS]) and end of a sentence ([SEP]) for each comment in the column, which is crucial for classification tasks as BERT uses the [CLS] token as the aggregate sequence representation for classification tasks.
- Attention Masks: Attention masks were created for each comment to help BERT differentiate the actual tokens from padding tokens. This is necessary as all comments need to be of the same length when inputted into BERT, and padding tokens are added to shorter comments to meet this requirement.

- Input Length: All comments were truncated or padded to ensure that they were of the same length before being inputted into BERT. This means that comments longer than a certain length (often 512 tokens for BERT) were truncated, and shorter ones were padded with special [PAD] tokens.

Gated Recurrent Units (GRU)Model for Demand forecasting

Gated Recurrent Units (GRU) were introduced by (Cho *et al*., 2014), as an optimization of the long short-term memory (LSTM) architecture, designed to combat the vanishing gradient problem in traditional Recurrent Neural Networks (RNN). The GRU model simplifies the LSTM by combining the forget and input gates into a single "update gate". It also merges the cell state and hidden state, reducing the overall complexity of the model.

Like LSTM, GRU units also possess a form of memory, capable of learning to retain information over time, which is crucial for processing sequential data. However, GRU's fewer tensor operations make it computationally more efficient, a critical advantage for training deep learning models.
Fig. 2 represents the GRU structure

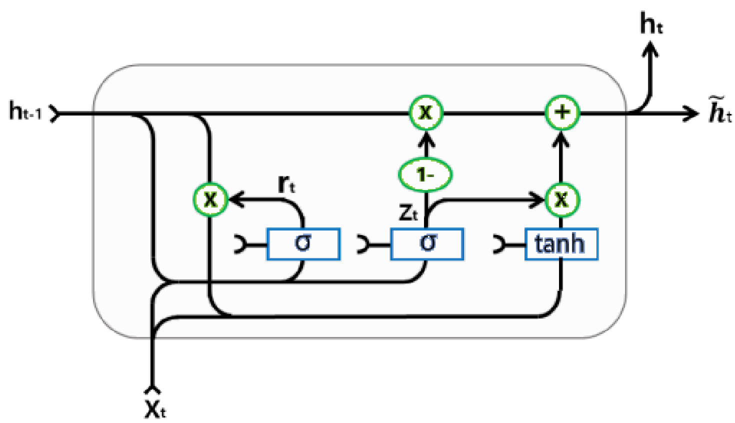

Fig. 2. *GRU Structure*

In terms of mathematical formulation, the GRU computes the hidden state h at each time step t using the following equations:

Update gate: $z(t) = \sigma(W(z)x(t) + U(z)h(t-1) + b(z))$ (6)
Reset gate: $r(t) = \sigma(W(r)x(t) + U(r)h(t-1) + b(r))$ (7)
Candidate hidden state: $h\sim(t) = \tanh(Wx(t) + U(r(t) * h(t-1)) + b)$ (8)
Hidden state: $h(t) = (1 - z(t)) * h\sim(t) + z(t) * h(t-1)$ (9)

where *x(t)* is the input at time step *t, σ* is the sigmoid activation function, * denotes element-wise multiplication, *W* and *U* are weight matrices, b are bias vectors, *z(t)* and *r(t)* are the update and reset gate activations respectively, and *h~(t)* is the candidate hidden state.

Proposed model

Table 1 provides a summary of the hyperparameters used in the GRU proposed model, including the number of units, activation functions, and learning rate, among others.

Table 1
GRU Model summary

Component	Hyperparameter	Value
GRU	Units	Tuned, Min: 32, Max: 512, Step: 32
Dropout	Rate	Tuned, Min: 0.0, Max: 0.5, Step: 0.1
GRU	Units	Tuned, Min: 16, Max: 256, Step: 16
Dropout	Rate	Tuned, Min: 0.0, Max: 0.5, Step: 0.1
Dense	Units	1
Dense	Activation	Linear
Model	Loss	Mean Squared Error
Adam Optimizer	Learning Rate	Tuned, Options: [0.01, 0.001, 0.0001]

The proposed Gated Recurrent Unit (GRU) model is specifically designed to perform sophisticated analyses of sequential data. Key to its operation is the tuning of various hyperparameters, which are methodically selected to optimize the model's performance. The number of units in each GRU layer, the rate of dropout, and the learning rate of the Adam optimizer are all dynamically adjusted within specific ranges, using a stepwise method to fine-tune the model. This allows us to ensure that the model best fits our specific dataset and task, striking a balance between computational efficiency and predictive accuracy. The model also employs a linear activation function in the dense layer and uses Mean Squared Error as the loss function. This setup contributes to the model's ability to make continuous predictions, which are crucial in various real-world applications such as time-series forecasting.

Bayesian Networks for Price Prediction

Mathematically, a Bayesian network is a directed acyclic graph (DAG) where each node X_i has an associated conditional probability distribution P(X_i | Parents(X_i)) that quantifies the effect of the parent variables on the child variable (Koller and Friedman, 2009). The joint probability distribution of a set of variables X_1, ..., X_n can be decomposed into a product of conditional probabilities:

This factorization enables efficient computation and inference, even for complex and large networks.

$$P(X_1, ..., X_n) = \prod P(X_i \mid \text{Parents}(X_i)) \quad \text{for} \;\; i = 1 \text{ to} \tag{10}$$

$$P(H \mid E) = P(E \mid H)\, P(H) / P(E) \tag{11}$$

Bayesian inference, which leverages Bayes' theorem, allows the updating of probabilities based on new evidence. Given a hypothesis H and evidence E, the posterior probability P(H | E) can be calculated as :

In the context of Bayesian networks, this updating process can be used to infer the probable values of unobserved variables based on observed variables.

Proposed model

The relationships between the variables was modeled using conditional probability distributions (CPDs), The modelization steps given Bayesian Network are as follow:

The marginal probabilities represent the individual probability distributions for each independent variable, and they are defined as follows:

- P(X1): The probability distribution for visits (X1).
- P(X2): The probability distribution for sentiment (X2).
- P(X3): The probability distribution for availability (X3).
- P(X4): The probability distribution for demand (X3).

The Conditional Probability Distributions (CPDs) define how the dependent variables are influenced by the independent variables:

- P (Y | X1, X2, X3): This CPD shows the probability distribution for 'pricing' (Y) given specific states of 'visits' (X1), 'sentiment' (X2), and 'availability' (X3).
- P (X4 | Y): This CPD gives the probability distribution for 'demand' (X4) conditional on a specific 'price' (Y).

The actual probability values for the distributions are denoted as follow:

- P (X1 = x1), P (X2 = x2), and P (X3 = x3) denote the probability of specific states for 'visits', 'sentiment', and 'availability', respectively, and the conditional probabilities:

Similarly, the conditional probabilities can be notated for specific combinations of variables:

- P (Y = y | X1 = x1, X2 = x2, X3 = x3) denotes the probability of a specific 'price' (Y) given particular states of 'visits', 'sentiment', and 'availability'.
- P (X4 = x4 | Y = y) represents the probability of a specific 'demand' (X4) given a particular 'price' (Y).

The proposed Bayesian Network model is a statistical framework that effectively models the dependencies between the variables: visits, sentiment, availability, price, and demand. The model's structure provides a representation of these relationships, employing Conditional Probability Distributions (CPDs) to illustrate how dependent variables are influenced by independent ones.

3. Results

The implementation of the models was executed in the Python programming language, utilizing the Jupyter platform, which is renowned for its interactive computing and comprehensive documentation capabilities. This enables easy tracking of the data processing steps and fosters reproducibility of the research.

The computation was performed on a Jupyter Notebook platform equipped with an Intel(R) CORE(TM) i5 processor and 8 GB of RAM, GPU: Intel UHD Graphics, Operating System: Windows 10 Pro 64-bit

BERT model evaluation

To evaluate the performance of our model, we used the following metrics:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{TP + FP + TN + FN} \tag{13}$$

$$Recall = \frac{\mathrm{TP}}{TP + FN} \tag{14}$$

$$F - Score = \frac{2 * \mathrm{Precision} * \mathrm{Recall}}{Precision + Recall} \tag{15}$$

$$Precision = \frac{\mathrm{TP} + \mathrm{TN}}{TP + FP} \tag{16}$$

TP, or True Positive, stands for the count of abnormal incidents that were correctly identified as such. TN, or True Negative, is the tally of ordinary occurrences accurately labelled as ordinary. FP, or False Positive, denotes the count of ordinary incidents mistakenly flagged as abnormal. FN, or False Negative, is the count of abnormal incidents misidentified as normal. The performance of a model in predicting results is evaluated by its accuracy. Precision measures how many positive outcomes (either true or false) actually align with the real-world scenario from all positive predictions made. Recall, on the other hand, shows the model's proficiency in forecasting positive outcomes relative to the overall quantity of positive instances. Lastly, F-Score is a metric employed to evaluate both precision and recall collectively. The performance of our BERT model, tested using a confusion matrix depicted in Fig. 3, was measured through an evaluation of consumer emotions in the "Customer Comment" field of our data set. The attained metrics present robust results:

• Accuracy: 0.809 - Our model exhibits an accuracy of 80.9%, implying its adeptness at correctly identifying more than 80% of the consumer sentiments. This implies a considerable proficiency in distinguishing positive and negative emotions drawn from consumer comments.
• Precision: 0.901 - The precision rate of 90.1% signifies that when the model predicted positive sentiment, it was correct in 90.1% of instances. This highlights the model's dependable capability of pinpointing positive consumer comments while keeping the number of false positives to a minimum.
• Recall: 0.695 - The model's recall rate of 69.5% shows that it managed to correctly identify 69.5% of all actual positive consumer comments within our data set. Although this performance is quite satisfactory, there is potential for enhancement, specifically in minimizing false negatives.
• F1-score: 0.785 - As a balanced measure of precision and recall, the F1-score provides a comprehensive assessment of the model's performance. An F1-score of 0.785 suggests a balanced performance in terms of precision and recall, effectively catering to the dual aspects of sentiment classification.

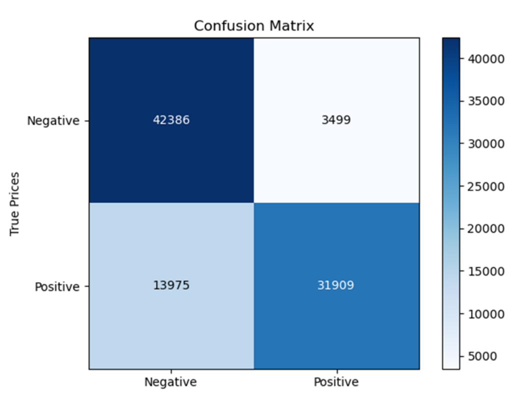

Fig. 3. BERT model's confusion matrix

These metrics illustrate the model's competency in discerning customer sentiments using the "Customer Comment" field. The impressive precision rate coupled with a respectable recall rate suggests the model's capability of accurately identifying positive customer sentiments while maintaining equilibrium between these two critical measures.

GRU model evaluation

Model performance is assessed using two metrics: Mean Squared Error (MSE) and R-squared (R2). MSE measures the average squared difference between the predicted and actual values, providing insight into the model's accuracy. R-squared, on the other hand, illustrates the proportion of the variance in the dependent variable that is predictable from the independent variables. A high R-squared indicates a high level of prediction accuracy.

$$MSE = \frac{\sum(y_i - y_p)^2}{n} \tag{16}$$

$$R^2 = 1 - \frac{\sum(y_i - y_p)^2)/n}{\sum(\hat{y}_i - y_n)^2)/n} \tag{17}$$

where y_i is the predictive value ; y_p is the truevalue ; $\hat{y}_i$ is the average value ; and n is the number of observations or rows. As illustrated in Fig. 4, the Gated Recurrent Unit (GRU) model displays robust performance, highlighted by a substantial decrease in the loss function values, which is measured by Mean Squared Error (MSE), over a few epochs for both training and validation datasets. This reduction in the loss function signifies the model's improving accuracy as it learns during training. The trend in this figure underscores the effectiveness of the GRU model for our specified objective, which is forecasting time series data. This strong performance, illustrated by the downward trajectory of the loss function, demonstrates the model's aptitude for generating reliable predictions based on historical trends.

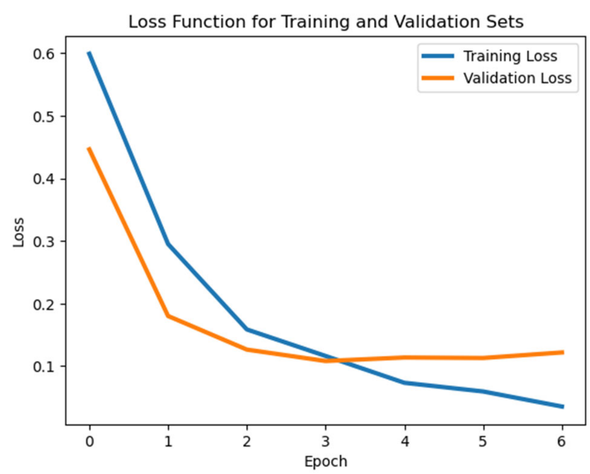

Fig. 4. Line plot of train and validation loss from the GRU

To further underscore the proficiency of the Gated Recurrent Unit (GRU) model in forecasting prices, its performance was juxtaposed with the Long Short-Term Memory (LSTM) model. This comparative analysis was conducted employing the Root Squared Error (R^2) and Mean Squared Error (MSE) as key performance indicators.
The outcomes of this comparative evaluation between the GRU and LSTM models are encapsulated in Table 2.

Table 2
Criteria metrics results depending on model

Model \ Metric	***MSE***	***R^2***
GRU	0.015	0.982
LSTM	0.019	0.975

Looking at the Mean Squared Error (MSE), which measures the average squared differences between the estimated and actual values, the GRU model outperforms the LSTM model with a lower value of 0.011 compared to the LSTM's 0.019. The lower MSE indicates that the GRU model makes fewer errors, thereby suggesting a higher accuracy in its predictions. In terms of the coefficient of determination (R^2), the GRU model again demonstrates superior performance with a score of 0.982 against the LSTM's 0.975. An R^2 value closer to 1 implies that the model accounts for a greater portion of the variance in the dependent variable. In this case, the GRU model can explain more of the variability in the data compared to the LSTM model.

Taken together, these results suggest that the GRU model performs better than the LSTM model in terms of both overall error rate (as indicated by MSE) and its ability to explain the variability in the data (as measured by R^2). These findings highlight the effectiveness of the GRU model for the task of price prediction in this context.

Baysian Network model evaluation

Before delving into the results, it is essential to outline the evaluation methodology employed in this study. We utilize error distributions of a Bayesian Network model, a robust tool for appraising predictive performance. This method involves comparing the predicted outcomes of the model with the actual values, the discrepancy of which constitutes the error. These errors, spanning a certain range, are then aggregated and analyzed statistically to understand their distribution. Central to this evaluation approach is the assumption of the errors following a Normal distribution, a key characteristic that upholds the reliability and stability of our model. This methodology offers a comprehensive view of the model's predictive accuracy, enabling us to identify systematic biases and random errors in predictions, thereby guiding improvements to refine the model. The following section presents the results of this evaluation process. Fig. 5 Exhibits the error distributions for our Bayesian Network model, providing key insights into its predictive accuracy. The error values range across the interval from -2 to 2, with a pronounced center at 0. This zero-centered distribution signals an exceptional balance in the model's predictions, implying an absence of consistent over- or under-estimations. The errors follow a Normal distribution, which reinforces the reliability and consistency of the model's performance. This conformity to the Normal law accentuates the model's robustness, suggesting a minimal amount of randomness in the errors. In summary, the characteristics of the error distributions depicted in the figure confirm the robustness and predic tive precision of our Bayesian Network model.

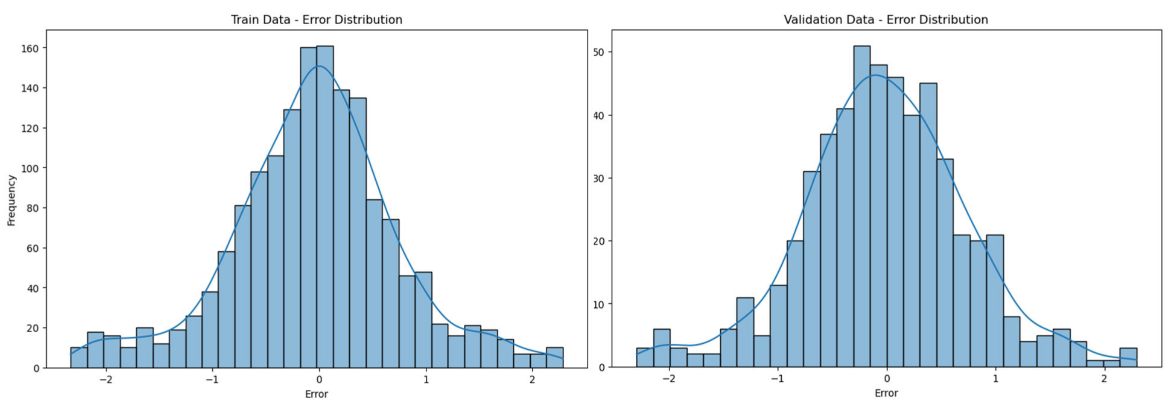

Fig. 5. Error Distribution for Training and Validation Data in Bayesian Network Model

To enhance our model assessment, Figure 6 presents the predicted prices against the actual prices extracted from the test set.

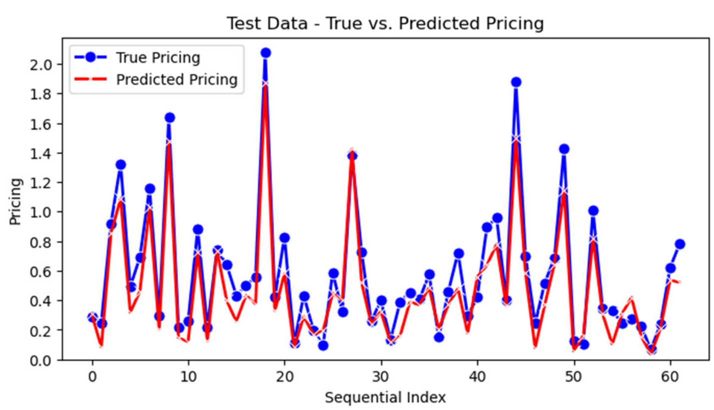

Fig. 6. True VS Predicted pricing

This comparison furnishes a tangible and visual metric to scrutinize the model's predictive prowess. The outcomes are commendable, with the predicted values closely aligning with the actual data points. This close correspondence attests to the efficacy of our proposed Bayesian Network model, showcasing its reliability in producing accurate predictions. Additionally, our proposed model demonstrates a distinct advantage over other probabilistic models like the Deep Gaussian Process (DGP) when it comes to handling large datasets. In our experiments, we encountered memory issues when attempting to apply these models, indicating their limitations in processing substantial volumes of data. Moreover, compared to traditional forecasting models like the AutoRegressive Integrated Moving Average (ARIMA), our approach proves more capable and versatile. ARIMA's approach is inherently linear and lacks the ability to effectively capture uncertainty parameters. In contrast, our model, with its non-linear capabilities, can better understand and encapsulate the inherent uncertainties and complexities found in real-world data. In sum, our model excels where others falter, showcasing its strength in managing large datasets and its aptitude in addressing the complexities of uncertainty, thereby providing a more robust and reliable tool for businesses.

4. Discussion

This article's primary contribution is the design and implementation of a novel, integrated framework that combines machine learning, deep learning, and probabilistic models. Specifically, the BERT, GRU, and Bayesian Network models have been woven together to create an effective solution for businesses, encompassing customer sentiment interpretation, demand forecasting, and price prediction. This represents a significant leap forward from the existing literature that usually discusses these models in isolation.

A noteworthy highlight of our approach is the BERT model's exceptional proficiency in understanding customer sentiment, achieving an outstanding precision rate of 90.1%. This figure confirms its exceptional ability to accurately identify positive sentiments, thus providing businesses with an invaluable tool for gauging customer satisfaction.

In the challenging field of demand forecasting, we employ the Gated Recurrent Unit (GRU) model, which outperforms the widely used Long Short-Term Memory (LSTM) model. It excels in predicting time series data, as demonstrated by its superior performance in both Mean Squared Error (MSE) and R^2 metrics, marking a significant advancement in the field of demand prediction.

In addition, our application of the Bayesian Network model for price prediction underscores the power of probabilistic models. This model exhibits a robust error distribution, with a range from -2 to 2 and a median of 0, underscoring its reliability and the precision of its predictions. This precision represents a marked advancement over previous methods, effectively eliminating the issue of systematic over- or under-estimations that are common pitfalls of many traditional models. Moreover, our model shows superior performance compared to other probabilistic models such as the Deep Gaussian Process (DGP), especially when handling large datasets. It also surpasses linear models like ARIMA, which often struggle with capturing uncertainty parameters. Our model's ability to handle these challenges represents a significant improvement, offering businesses a tool that provides more balanced, reliable, and nuanced forecasting in complex and uncertain market conditions. In essence, our adoption of the Bayesian Network model reaffirms the value of probabilistic models while also advancing the capabilities of price prediction methodologies. It distinguishes itself by providing a more comprehensive, versatile, and reliable approach for businesses navigating dynamic market conditions.

In sum, our integrated approach not only leverages the individual strengths of each model but also presents a comprehensive framework for businesses, giving them a more holistic view of their customers and market dynamics, thus distinguishing our contribution from existing literature.

5. Conclusions

In conclusion, our integrated approach that amalgamates the strengths of the BERT, GRU, and Bayesian Network models paves the way for businesses to delve deeply into their customer base and decipher market dynamics. Each model has its own unique strengths, and at the same time, highlights opportunities for enhancement. Collectively, they contribute towards the broader objective of facilitating well-informed business decisions, enhancing profitability, and enabling businesses to effectively navigate through supply chain risks and uncertainties. As we gaze into the future, an intriguing direction of research lies in revisiting and potentially redesigning the framework of our integrated model. While we will continue to optimize the existing models for improved predictive accuracy, like enhancing the BERT model's recall rate, mitigating potential overfitting in the GRU model with advanced regularization techniques, and investigating the effects of potential outliers in the Bayesian Network model's error range, it could be equally beneficial to explore how the synergy between these models within the framework could be enhanced. This could involve evaluating different combinations of models, the introduction of newer modeling techniques or even altering the sequence of application of these models to leverage their strengths more effectively. Ultimately, our aim remains to continually hone these models to augment their predictive accuracy, thereby increasing their value in diverse business applications.

References

Aamer, A., Eka Yani, L., & Alan Priyatna, I. (2020). Data analytics in the supply chain management: Review of machine learning applications in demand forecasting. *Operations and Supply Chain Management: An International Journal*, *14*(1), 1-13.

Amellal, A., Amellal, I., & Seghiouer, H., & Ech-Charrat, M. (2023). Improving Lead Time Forecasting and Anomaly Detection for Automotive Spare Parts with A Combined CNN-LSTM Approach. *Operations and Supply Chain Management: An International Journal*, 16(2), 265-278.

Avilés, A., Célleri, R., Solera, A., & Paredes, J. (2016). Probabilistic forecasting of drought events using Markov chain-and Bayesian network-based models: A case study of an Andean regulated river basin. *Water*, *8*(2), 37.

Aydinonat, N. E., & Köksal, E. (2019). Explanatory value in context: the curious case of Hotelling's location model. *The European Journal of the History of Economic Thought*, *26*(5), 879-910.

Balta, G. K., Dikmen, I., & Birgonul, M. T. (2021). Bayesian network based decision support for predicting and mitigating delay risk in TBM tunnel projects. *Automation in Construction*, *129*, 103819.

Boutselis, P., & McNaught, K. (2019). Using Bayesian Networks to forecast spares demand from equipment failures in a changing service logistics context. *International Journal of Production Economics*, *209*, 325-333.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Buntine, W. (1996). A guide to the literature on learning probabilistic networks from data. *IEEE Transactions on knowledge and data engineering*, *8*(2), 195-210.

Chickering, D. M. (2002). Learning equivalence classes of Bayesian-network structures. *The Journal of Machine Learning Research*, *2*, 445-498.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Chopra, S., Meindl, P., & Kalra, D. V. (2007). *Supply Chain Management by Pearson*. Pearson Education India.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dou, Z. Y., Yu, K., & Anastasopoulos, A. (2019). Investigating meta-learning algorithms for low-resource natural language understanding tasks. *arXiv preprint arXiv*:1908.10423.

Durairaj, A. K., & Chinnalagu, A. (2021). Transformer based Contextual Model for Sentiment Analysis of Customer Reviews: A Fine-tuned BERT. *International Journal of Advanced Computer Science and Applications*, *12*(11).

Esmaeili, P., Makui, A., Seyedhosseini, S., & Ghousi, R. (2022). The effect of probabilistic incentives to promote cooperation during the pandemics using simulation of multi-agent evolutionary game. *International Journal of Industrial Engineering Computations*, *13*(3), 319-328.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, *25*(2), 383-417.

Feizabadi, J. (2022). Machine learning demand forecasting and supply chain performance. *International Journal of Logistics Research and Applications*, *25*(2), 119-142.

Fera, M., Fruggiero, F., Lambiase, A., Macchiaroli, R., & Miranda, S. (2017). The role of uncertainty in supply chains under dynamic modeling. *International Journal of Industrial Engineering Computations*, *8*(1), 119-140.

Friedman, N., Goldszmidt, M., Heckerman, D., & Russell, S. (1997). Where is the impact of Bayesian networks in learning. In *International Joint Conference on Artificial Intelligence*.

Geetha, M. P., & Renuka, D. K. (2021). Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. *International Journal of Intelligent Networks*, *2*, 64-69.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Honjo, K., Zhou, X., & Shimizu, S. (2022, July). CNN-GRU Based Deep Learning Model for Demand Forecast in Retail Industry. In *2022 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

Huang, H., Zhang, B., Jing, L., Fu, X., Chen, X., & Shi, J. (2022). Logic tensor network with massive learned knowledge for aspect-based sentiment analysis. *Knowledge-Based Systems*, *257*, 109943.

Ivanov, D., & Dolgui, A. (2020). Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. A position paper motivated by COVID-19 outbreak. *International Journal of Production Research*, *58*(10), 2904-2915.

Kelleher, J. D., Mac Namee, B., & D'arcy, A. (2020). *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT press.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

Ma, Z., Wang, C., & Zhang, Z. (2021, September). Deep Learning Algorithms for Automotive Spare Parts Demand Forecasting. In *2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI)* (pp. 358-361). IEEE.

Noorbeh, P., Roozbahani, A., & Kardan Moghaddam, H. (2020). Annual and monthly dam inflow prediction using Bayesian networks. *Water Resources Management*, *34*, 2933-2951.

Noh, J., Park, H. J., Kim, J. S., & Hwang, S. J. (2020). Gated recurrent unit with genetic algorithm for product demand forecasting in supply chain management. *Mathematics*, 8(4), 565.

Pathak, U., Kant, R., & Shankar, R. (2020). Price and profit decisions in manufacturer-led dual-channel supply chain configurations. *International Journal of Industrial Engineering Computations*, *11*(3), 377-400.

Pauwels, S., & Calders, T. (2020). Bayesian network based predictions of business processes. In *Business Process Management Forum: BPM Forum 2020, Seville, Spain, September 13–18, 2020, Proceedings 18* (pp. 159-175). Springer International Publishing.

Pearl, J. (1985, August). Bayesian netwcrks: A model cf self-activated memory for evidential reasoning. In *Proceedings of the 7th conference of the Cognitive Science Society, University of California, Irvine, CA, USA* (pp. 15-17).

Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Rychłowska-Musiał, E. (2020). Real options games between two competitors: the case of price war. *International Journal of Computational Economics and Econometrics*, *10*(1), 92-110.

Saini, U., Kumar, R., Jain, V., & Krishnajith, M. U. (2020, July). Univariant Time Series forecasting of Agriculture load by using LSTM and GRU RNNs. In *2020 IEEE Students Conference on Engineering & Systems (SCES)* (pp. 1-6). IEEE.

Seuring, S. (2013). A review of modeling approaches for sustainable supply chain management. *Decision support systems*, *54*(4), 1513-1520.

Sirusstara, J., Alexander, N., Alfarisy, A., Achmad, S., & Sutoyo, R. (2022, September). Clickbait Headline Detection in Indonesian News Sites using Robustly Optimized BERT Pre-training Approach (RoBERTa). In *2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)* (pp. 1-6). IEEE.

Tao, X., & Yang, W. (2022, May). Prediction of machine production state based on Bayesian Network. In *ISCTT 2022; 7th International Conference on Information Science, Computer Technology and Transportation* (pp. 1-4). VDE.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wang, X., Chaolu, T., Gao, Y., Wen, Y., & Liu, P. (2023). Marketplace channel encroachment under private brand introduction of online platform. *International Journal of Industrial Engineering Computations*, *14*(2), 403-414.

Wu, S., Liu, Y., Zou, Z., & Weng, T. H. (2022). S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis. *Connection Science*, *34*(1), 44-62.

Yarullin, R., & Serdyukov, P. (2021). Bert for sequence-to-sequence multi-label text classification. In *Analysis of Images, Social Networks and Texts: 9th International Conference, AIST 2020, Skolkovo, Moscow, Russia, October 15–16, 2020, Revised Selected Papers 9* (pp. 187-198). Springer International Publishing.