e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:09/September-2024 Impact Factor- 7.868 www.irjmets.com

ARTIFICIAL INTELLIGENCE AND BIG DATA ANALYTICS FOR SUPPLY CHAIN MANAGEMENT

Jubin Thomas*1, Kirti Vinod Vedi*2, Sandeep Gupta*3

*1Media, Pennsylvania, USA.

*2Highland Park, New Jersey, USA.

*3SATI, Vidisha, M.P., India.

DOI : https://www.doi.org/10.56726/IRJMETS61488

ABSTRACT

Companies are reconsidering their positions on the data analysis and supply chain changes that have occurred with the onset of AI and its associated features. Supply chain management scholars have attempted a prodigious endeavour in identifying and addressing danger to the supply chain because of ML and DL and their exclusive properties for future data prediction and data classification. To fully tap DL, this study provides several DL approaches towards DL. One of them is to ascertain if the late delivery of a certain product is due to a failure to identify an unforeseen factor in the complex SC system. Thus, the following research questions guide this paper in assessing an effectiveness of Bi-LSTM DL models to enhance a demand forecasting aspect of SC management. The Kaggle dataset titled "Datacom Smart Supply Chain for Big Data Analytics" is used in the SC research. It has been cleaned and scaled before being split into training and testing sets. A BiLSTM model is used to forecast a most important features of late delivery for treatment. Data displaying an F1-Score of 90. 11%, AUC-ROC of 14. 01, and an accuracy of 97. 59% prove the BiLSTM models are effective. Comparison with other models, such as K-Nearest Neighbours (KNN) and Decision trees, also put into more light the efficiency of BiLSTM in handling SC risks. Overall, the findings suggest that BiLSTM models hold significant promise for optimising logistics operations and mitigating SC risks through more accurate demand forecasting.

Keywords: Supply Chain Management, Big Data Analytics (BDA), Demand Forecasting, BiLSTM, Deep Learning, Late Deliveries.

I. INTRODUCTION

Contemporary business leadership and management have reacted with vigour to the changing face of the SC through the use of predictive BDA in demand prediction. Using data from various fields, from the company's performance records, patterns in the market[1][2], customer engagement in social media platforms, and even the climatic conditions of specific regional areas, it becomes easier for a business venture to forecast demand levels[3][4]. Such a feature can be quite beneficial in the decision-making process of businesses since it will provide accurate figures as to which inventory can be expected to be in high demand in the near future [5][6].

Forecasts on demand are very important because they help reduce the risk of having too much stock or having no stock at all, which has a way of affecting the earnings of a business organisation. Overstocking makes the holding costs high[7][8], and the inventory might become obsolete on the other hand, stockout means that customers are lost or have to wait for the product, and this would not be well acceptable[9][10]. Through rationing methodologies like the predictive model, firms are able to some extent predict demand and thus optimise inventory levels, minimise operating expenses, and improve organisational performance [11][12]. Businesses and other organisations work together in a structured framework called a supply chain to meet consumer expectations while making the most efficient use of available resources[13][14]. Global supply chains have been established due to the fact that global competitiveness has affected every business. This has led to a dramatic increase in the complexity of SC management on a worldwide basis[15][16]. Management of supply chains has also been much more efficient in recent years, thanks to developments in transportation and communication technology [17][18]. Moreover, predictive big data analytics facilitates more agile and responsive SC strategies [19]. Companies can adjust their procurement, production, and distribution processes in near real-time, ensuring that they remain aligned with current demand patterns [20][21].

The rationale for using predictive BDA and machine learning in supply chain demand forecasting is the ever-complicated and dynamic nature of market scenarios that need to be managed effectively. As consumers' behaviour and other market factors change, existing approaches to forecasting are frequently insufficient, and managers face challenges in terms of inventory management costs, including overstocking or stockout and imbalanced inventory levels. Thrive in the utilisation of these innovative tools to sort out accomplishments that can deliver definitive and tangible values in making prediction and demand forecasting much more effective. In summary, the use of advanced techniques of forecasting employing predictive analytics and machine learning tools is critical in the attainment of competitive advantage, innovation, and sustainable growth in today's complex and evolving global markets.

Contribution of the study

The study aims to enhance the predictive capabilities and operational efficiencies in supply chain management by leveraging advanced deep learning (DL) methodologies. These are the contributions made by this study:

- The study introduces a comprehensive methodology for big data analytics in SCM, leveraging a dataset from Kaggle and employing advanced techniques like Bi-LSTM for classification.
- A study meticulously addresses data preprocessing challenges, including converting diverse data types into numeric forms, handling duplicates and missing values, and eliminating irrelevant features, ensuring the integrity and quality of the dataset.
- The research employs EDA techniques, including graphical representations like pie charts and line graphs, to elucidate key patterns and trends in delivery status categories and shipping performance over time, providing valuable insights for supply chain optimisation.
- The study rigorously evaluates model performance using metrics like accuracy, F1 score, and AUC ROC, showcasing the superiority of the proposed BiLSTM approach over conventional techniques like KNN and DT, thus contributing to advancements in predictive analytics for supply chain management.

This study is organised as follows: An extensive assessment of prior research on a topic is given in Section 2. In Section 3, the methodology for the research used in this study is described. The study's analysis,results and discussion are covered in Section 4. The study's conclusions and future goals are presented in Section 5.

II. LITERATURE REVIEW

This literature review compiles every piece of published research on SCM to date, including a broad variety of methods from DL and ML to neural network models.

In Huang, 2023, suggests that the best ways to guarantee the dynamics, effectiveness, and agility of SCM are to approach the problem from three different angles: the first is to create a collaborative mechanism for data-driven supply chain risk management (SCRM); the second is to create a two-factor mechanism for risk assessment in data-driven supplier risk management; and the third is to establish a big data-driven process restructuring for SCRM. As a conclusion, the article confirms that the big data–driven collaborative operation method is beneficial[22].

In Hu et al. (2023), This article examines China's SCM system's digital application state first. After that, it highlights the issues with the SCM system in China's areas of product development, strategic procurement, manufacturing, storage, logistics, and transportation, as well as sales and after-sales support. It then suggests solutions for these issues through big data analysis technology[23].

In Bag et al., (2021), investigates, using the COVID-19 pandemic as an example, how innovation leadership moderates the relationship between big data analytics (BDA) and innovation, responsiveness, and resilience in the healthcare supply chain (HSC). The impact of BDA skills on HSC innovation is amplified by highly innovative leadership. Leadership that prioritises innovation also amplifies the impact of BDA skills on responsiveness. The second step was to use the data from 30 semi structured qualitative questionnaires to confirm and enhance the study's empirical results[24].

In Ghouati et al., (2024), this research explores current artificial intelligence applications within supply chain processes and pinpoints areas for additional exploration investigation. We conducted a review of 80 research articles published during the period from 2020 and 2023 retrieved from the Scopus database Our analysis outlines critical features regarding artificial intelligence's role in improving supply chain resilience. Ultimately, we assert that judicious utilization of AI across diverse supply chain processes facilitates a development of a resilient supply chain, underpinning a robust decision support system[25].

In Effah et al., (2024), emphasises the use of AI to reduce vulnerabilities in the supply chain caused by EW. Initially, a cognitive mapping method is used to catalogue EW-related dangers, their interconnections, EW instances, and AI capacities that might lessen their destructive effects. Secondly, these dangers caused by EW are ranked using the best-worst method (BWM). According to BWM, the midstream supply chain is the most vulnerable to EW-induced threats, whereas the upstream and downstream chains are less affected. Out of the eleven risk variables, the most common ones are those related to EW-induced transportation, farms, and demand; the least common ones are those related to psychological stress, market share risk, and customer discontent[26].

In, Koç, Erdinç et al., (2022) DL methods suggested in the paper include a form of the TCN and three RNN structures: LSTM, Bi-LSTM, and GRU. The numerical results show that TCN, one of the proposed models, can estimate the danger of shipping to a given site under COVID-19 limitations with an accuracy of about 100% [27].

In Shah et al. (2021), Using striped datasets, the present study has developed a variety of ML multiclass classification techniques, such as linear SGD, RF, XRT, DT, MLP, XGB, Cat Boost (CB), and linear NB. Findings The results show that medicinal supply logistics suppliers only give an accurate turn-around time of 62.91%. The study's proposed answer, on the other hand, is up to 48.62% more accurate than reality and up to 93.5% more correct overall [28].

In this study, Velasco-Gallego and Lazakis (2022) Construct a regression model using ML to forecast the severity of suppliers' delivery delays. Our model remains unreliable despite achieving R^2 values of 92% for the "order request" prediction and 98% for the "order placement" prediction[29].

In, Rattanakul and Lenbury, (2020) The study's goal was to forecast how the GDP will react to interruptions in the supply chain using methods from DL and ANN. The results highlight how important Google Trends is for consistently predicting GDP figures, which might have serious consequences. In conclusion, the comparative discussion regarding the superior predictive performance of DL in comparison to ANN experiments concludes with the implications for global policy, decision-makers, and business managers [30].

In, Xu and He, (2020) The supply chain's financial credit risk may be evaluated online using a DBN, according to the study. This approach outperforms the SVM and Logistic methods in terms of assessment accuracy and rationality, with a score of 96.04%[31].

In, Islam and Amin, (2020) the Study relies on ML, which is selected for its capacity to improve the model's explain ability. DRF and GBM are used to forecast product backorders in this research. They demonstrate how this algorithm may anticipate likely backorder goods prior to sales. A 20% improvement was made to the accuracy of the back-order forecast[32].

In this study, Abbasi et al., (2020) The suggested approach is implemented to determine the movement of blood units throughout an institutional network. Applying a trained NN model lowers the average daily cost by about 29% compared to the previous policy; the precise optimal policy further cuts it by 37%. A rise in the cost of an average daily education has been realised to be 29%. Mean reduction in cost is the lowest for ANN among all types of residential buildings provided it is used optimally[33].

Table 1: Review of relevant literature on utilisation of big data analytics for managing supply chain

Reference	Methodology	Dataset	Performance	Limitations & Future Work
[22]	Proposes a three-level data-driven collaborative mechanism for supply chain risk management.	-	Ensures the efficacy of the collaborative operation mechanism that is proposed and driven by big data.	Future work can involve testing in various supply chain environments and sectors for a more comprehensive validation.
[23]	Analyzes the	China's SCM data	Provides insights	Future research could

	digital application status of China's SCM system, identifies issues, and proposes optimisation measures.		on issues in product design, procurement, logistics, and offers big-data-driven solutions for SCM optimization.	involve empirical validation of these optimizations in other regions and industries.
[24]	Empirical analysis of BDA on healthcare supply chain (HSC) innovation, responsiveness, and resilience.	30 semi-structured qualitative questionnaires	Shows BDA positively affects HSC innovation and responsiveness, moderated by innovation leadership, with validation through both qualitative and quantitative methods.	Focuses only on healthcare supply chains, future work should explore other sectors and additional moderating factors.
[25]	Review of 80 articles on AI applications in supply chain resilience.	80 research articles from Scopus (2020-2023)	AI facilitates resilient supply chain development and supports decision-making processes.	Need for further exploration in specific supply chain areas, such as AI implementation for real-time responsiveness in high-risk environments.
[26]	Focuses on EW-induced risks in supply chains and AI capabilities to mitigate these risks, using cognitive mapping and BWM.	-	BWM ranks EW-induced risks, identifies most prevalent risks in midstream, and highlights AI's role in mitigating these risks.	Future work could focus on exploring more EW-induced risks in other supply chain phases and expanding the application of AI beyond midstream.
[27]	DL models: LSTM, BiLSTM, GRU, TCN	Shipment data during COVID-19 restrictions	TCN model: ~100% accuracy in predicting shipment risk	Further testing is required to validate in different contexts; extend to other types of disruptions
[28]	ML models: RF, DT, MLP, XGB, CB, SGD, NB; XRT, stacked meta-models; model zoo	During COVID-19, almost three million shipments of medicinal supplies were tracked in the real world.	Proposed solution: 93.5% accuracy (48.62% improvement)	Performance improvement with more historical data; exploration of additional features
[29]	ML-based regression model	High-dimensional input features data related to late	R^2 of 92% at 'order request'; 98% at 'order placement'	Inclusion of more specific component identifiers may improve accuracy;

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:09/September-2024 Impact Factor- 7.868 www.irjmets.com

		deliveries		address remaining inaccuracy
[30]	DL and ANN algorithms	Monthly data from 2008 to 2022, including recession episodes	Demonstrated sensitivity of GDP variations to supply chain	Comparative analysis of DL vs. ANN; policy implications for decision-makers
[31]	Evaluating the financial risk of the supply chain online using DBN	Three kinds of datasets with 21 indicators	Evaluation accuracy: 96.04%	Comparison with SVM and Logistic methods; potential application to other financial risk assessment contexts
[32]	Tree-based ML: DRF and GBM	Details including inventory levels, revenue, sales forecasts, and lead time measurements	20% performance improvement for backorder prediction	Analysis of decision trees for business decisions; more scenarios for backorder prediction
[33]	Neural network models for decision-making on blood unit transshipment	Network of hospitals' blood unit transshipment data	Average daily cost reduction: 29% compared to current policy	Comparison with commercial optimisation solvers; potential for wider application despite not achieving exact optimal policy; exploration of use in not-for-profits and SMEs

Research gaps

It has also revealed that there is great potential for optimising the SCRM approaches by combining AI and big data. To fill gaps within a current accumulated note, efforts should go into proving models such as Temporal Convolutional Networks (TCN) across various settings, and more features and history to improve the models. The incorporation of detailed operational data into regression models is critical for enhancing the models' accuracy. If DL and ANN models are compared suitably, it can provide optimum insights for policy-making purposes as well as at the firm level. Using approaches like Deep Belief Networks (DBN) with a larger set of financial risks and working on the prediction scenarios of backorders helps in fine-tuning the business decisions and, at the same time, making optimisation techniques feasible for small-scale organisations to get better business operations. In turn, organisations can use AI and big data to enhance the reliability as well as utility of SCRM models in practice, performing better decision-making and managing logistics chains.

III. METHODS AND MATERIALS

The research steps that are included in the current research methodology can be outlined as follows. To begin with, the collection of data from the Kaggle database involves choosing the database called 'Datacom Smart Supply Chain for BDA with 52 column characters and 180,519 rows with data of different types. Cleaning is next which involves converting various data types into numeric forms, handling duplicate values, removing features with missing data and errors, and eliminating duplicates and irrelevant information in the context and where the 'late delivery task' column is established as the output variable. Data partition to training and testing sets is subsequently performed after normalisation stage in preparation for algorithm training and testing. Next, feature selection is performed from the 13 feature columns, including the target one, whereas the classification model of choice is a Bi-LSTM. Then lastly evaluate a model. Figure 1 shows the whole process of implementation.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:09/September-2024 Impact Factor- 7.868 www.irjmets.com

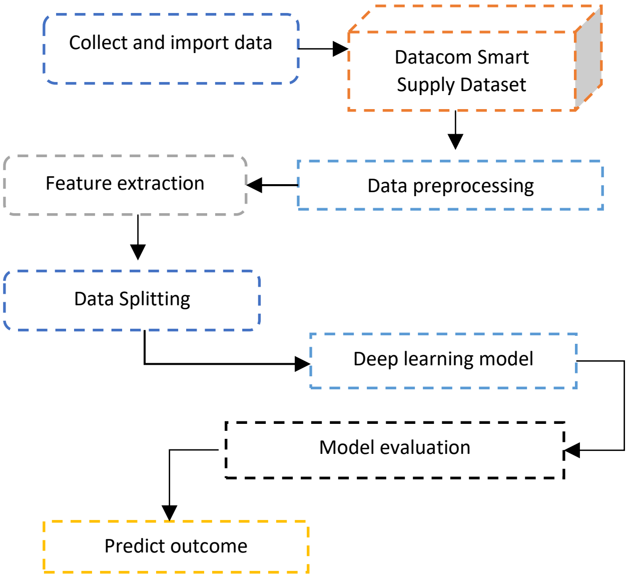

Figure 1: Data Flow Diagram for big data analytics in SCM

Dataset collection

The SC data comes from Kaggle, a well-known public database in the fields of data science and ML. This investigation makes use of the "Datacom Smart Supply Chain for BDA" dataset, which integrates SC data used by Datacom Global. There are 180,519 rows and 52 columns in the dataset. These records are assembled in several formats, including numerical, text, and date/time type, just like any other SC data collection[34]vn. Preprocessing to categorise the data or extract data characteristics to determine whether a certain activity is postponed is therefore essential to handle the complexity of varied data formats. In the event that an order is delivered late, it is recorded as 1, whereas if it is not, it is recorded as 0.

Dataset Preprocessing

This phase employs the columns that were acquired in the preceding phase. I have previously stated that text, date, and time data must be converted to numerical categories due to the diversity of data types[35]. Furthermore, it is necessary to rectify any errors, misspellings, duplicate data, or lacking features in the data. Data sets that have duplicate entries are also removed. Such that both the "Customer Id" and the "Order Customer Id" columns have the same values. They keep the first column and remove the second one as redundant if we absolutely must keep both columns' contents. Another useful feature is that the "days for shipment and dispatch" column may be used in place of the "shipping date" column since both columns include the same information. Furthermore, the risk concern is unrelated to a number of unique identifiers, thus we have eliminated them. These include "Customer Email" and "Product or Order's Image," for example. In addition to removing any pre-existing predicting data, such order data, any possible bias will be eliminated. Therefore, the order status information columns are configured to be 13. Predictions regarding the "late delivery assignment" column are forthcoming. The following subsections offer an explanation of the techniques that derive robust features to distinguish between on-time and delayed orders.

Data Splitting

Data normalisation comes after pre-processing in both the training and testing stages of model construction. Once data is spit, we use it to train an algorithm, setting away the test data set for later use. The training data will be used to build the training model, which will include the feature values, related logic, and algorithms. Basically, normalisation is done to ensure that all the attributes are of the same size. There are two sets of data used in this experiment: one for training and one for testing.

Feature Extraction

Following a pre-processing phase, a most notable characteristics from those 13 columns are recorded in this step. We will go over the specifics of the suggested DL techniques in this paragraph. Additionally, we provide the layers, their total number, and each layer's properties [36].

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:09/September-2024 Impact Factor- 7.868 www.irjmets.com

Classification with BiLSTM MODEL

Four layers make up the fundamental unit of BiLSTM, which is made up of several smaller structures: an input layer; an output layer; a forward propagation layer; and a backward propagation layer. The backward transmission layer extracts features from the input sequence in reverse order while the forward propagation layer extracts vector forward characteristics in reverse order[37]. The output layer is responsible for combining the data that has been generated by the forward and backward propagation levels. Our goal is to determine the vectors' forward-backwards correlation [38]. The following equations (Equations 1–5) describe the functioning of the input gate I, the output gate O, and the forget gate F at time t in the section of the BiLSTM neural network:

$$i_t = \sigma(W^i x_t + U^i h_{t-1})i_t = \sigma(W^i x_t + U^i h_{t-1}) \quad (1)$$

$$f_t = \sigma(W^f x_t + U^f h_{t-1} \quad (2)$$

$$o_t = \sigma(W^o x_t + U^o h_{t-1})o_t = \sigma(W^o x_t + U^o h_{t-1}) \quad (3)$$

$$c_t = \tan h\,(W^c x_t + U^c h_{t-1})c_t = \tan h\,(W^c x_t + U^c h_{t-1}) \quad (4)$$

$$c_t = i_t \times c_t + f_t \times c_{t-1})c_t = i_t \times c_t + f_t \times c_{t-1}) \quad (5)$$

Among them, Wi, Wf, Wo, Wc, Ui, Uf, Uo, and Uc are all weight matrices. A fundamental concept behind the generation of BiLSTM models is the use of feature data gathered at time t to concurrently integrate information by the past and the future [39].

IV. RESULT ANALYSIS AND DISCUSSION

In this section, presents a result analysis information about the model experiments and their outcomes are presented according to accuracy and f1-sore.

EDA (Exploratory data analysis)

A solution to this problem is the development of exploratory data analysis methods. To some extent, these methods achieve their goals by masking certain facts while highlighting others. Two common forms of cross-classification are used in exploratory data analysis. To start, there are graphical and non-graphical approaches to each process. Second, although most methods are bivariate, others are univariate. Summary statistics are often computed using non-graphical techniques, but graphical approaches clearly use diagrams or pictures to summarise the data.

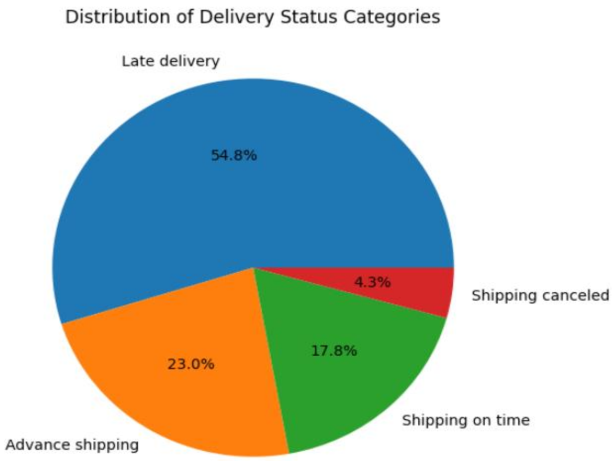

Figure 2: Distribution of delivery status categories

Figure 2 depicts a pie chart titled “Distribution of Delivery Status Categories.” The Late delivery which is largest section in the pie chart, colored blue, represents late deliveries. It accounts for approximately 54.8% of all

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:09/September-2024 Impact Factor- 7.868 www.irjmets.com

deliveries. The Advance shipping represent by orange section, comprising about 23.0%, corresponds to advance shipping. The Shipping on time represented by green segment, accounting for 17.8%, indicates deliveries that were on time. The Shipping canceled showed in smallest section, colored red, represents canceled shipments, making up 4.3% of the total.

Figure 3 shows a line graph of Average Days for Shipping Over Time." The graph represents the average shipping time per month over the course of a year. January through December are shown on the horizontal axis, while the average number of shipping days are displayed on the vertical axis. The fluctuating line indicates variability in shipping performance throughout the year, allowing organisations to identify trends and make informed decisions to improve their delivery services.

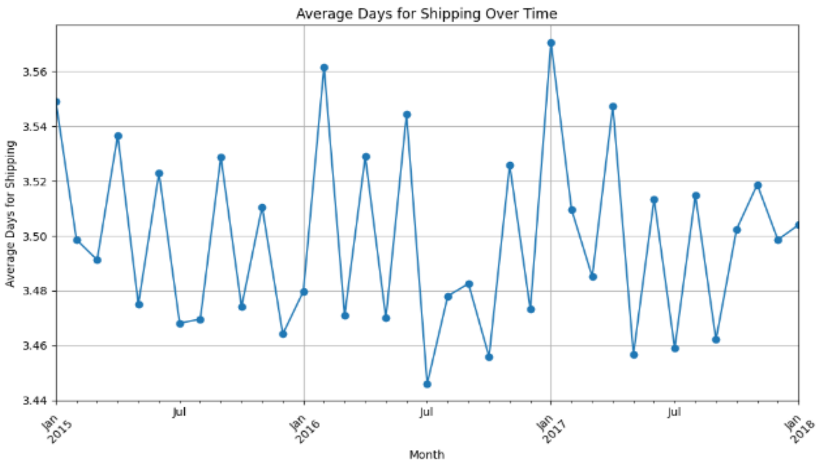

Figure 3: Average days for shipping over time

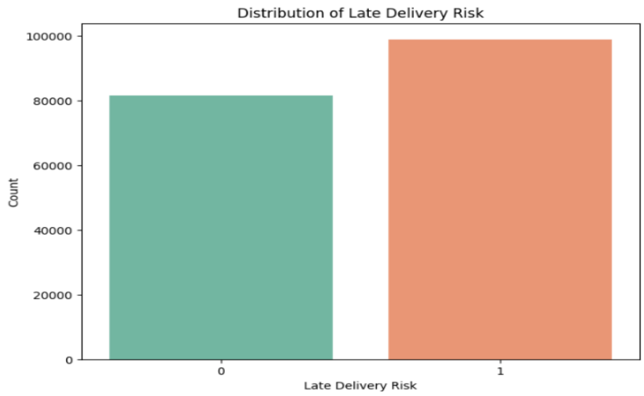

Figure 4: Late delivery risk

Figure 4 shows a bar graph of Distribution of Late Delivery Risk. The graph compares two categories labelled '0' and '1' on the horizontal axis, representing different levels of late delivery risk. The vertical axis is labelled 'Count,' ranging from 0 to 100,000 in increments of 20,000. There are two bars: the green bar corresponding to '0' reaches approximately 80,000 on the count axis, while the orange bar for '1' extends just beyond 60,000. This graph provides insights into the frequency of late deliveries in a logistics context, comparing low-risk versus high-risk scenarios.

Performance matrix

This study measured the model's predictive capability using accuracy, F1-score and AUCROC. Performance metrics are as follows:

- **Confusion Matrix (CM)**

The categories seen are contrasted with the categories anticipated in a table known as the Confusion Matrix. An aggregate count of samples taken in each of the four orientations is shown. We talk about the model's accuracy while defining its components, which include False Positive, True Negative, False Positive, and True Negative.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:09/September-2024 Impact Factor- 7.868 www.irjmets.com

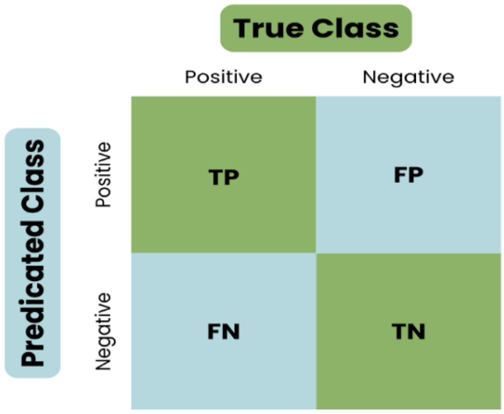

Figure 5: Confusion matrix

Figure 5 displays a confusion matrix for classifying four classes. Four-class classification is splitted into four classes. "Class A, Class B, Class C, and Class D" Negative (0) and positive (1) reflect expected values, while true (1) and false (0) represent actual values. From the confusion matrix expressions TP, TN, FP, and FN, classification model estimates are calculated.

- **Accuracy**

An accuracy measure is one that correctly predicts test results. This metric measures the accuracy of data testing predictions. To determine correctness in ML, one uses the following formula (Equ.6):

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+Fp+TN+FN}} \tag{6}$$

- **F1 Score**

A F1Score is calculated by summing a Precision and Recall scores harmonically. F1 Score will also decrease if any of these gets critically low. Finding a happy medium between the two metrics, Precision and Recall, is possible with the aid of F1 Score. The formula (Equ.7) are given as:

$$\text{F1} - \text{Score} = \frac{2(\text{Precision}*\text{Recall})}{\text{Precision+Recall}} \tag{7}$$

- **AUC ROC**

An overall assessment of performance overall potential classification levels is provided by the area under the ROC curve, which is known as the AUC. The Equation 8 is as follows:

$$AUC = \sum_{i=1}^{n=1}(FPR_{i+1} - FPR_i \times \frac{TPR_{i+1}+TPR_i}{2} \tag{8}$$

The value of AUC lies between 0 and 1. Assuming no discriminative ability, an AUC of 0.5 is the same as taking a random guess. Indicative of superior performance is an AUC that is closer to 1.

- **Experiment analysis**

This section provides the result of the technique on the dataset. In this step, determine how effectively the model can classify. It presents the results in the form of bar graphs, tables, and figures.

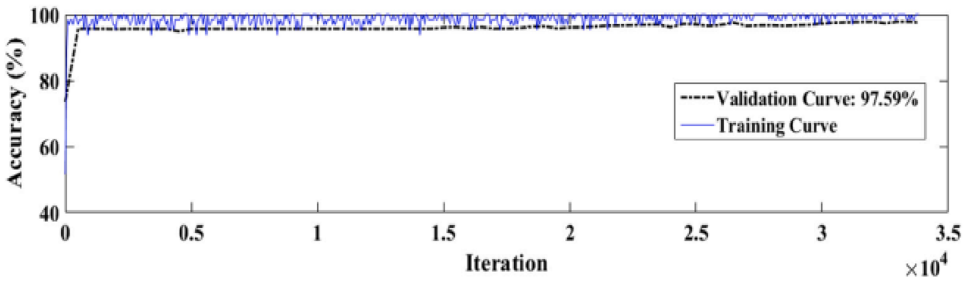

Figure 6: Training accuracy and validation accuracy of Bilstm

Figure 6 displays the accuracy for both training and validation. The accuracy is displayed on a y-axis, and the iteration is ilustrate on a x-axis. A Bilstm model achieve a Figure shows the training and validation accuracy. A Bilstm model achieves a validation accuracy of 97.59%.

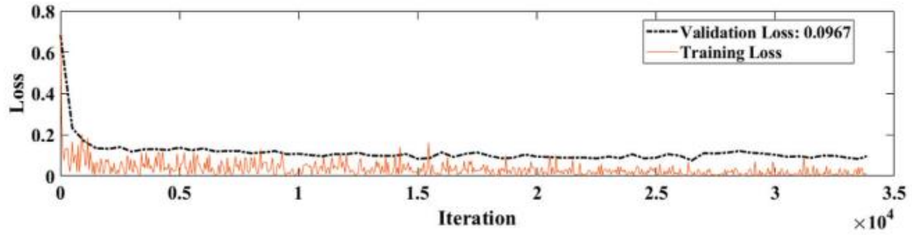

Figure 7: Training and loss curves for Deep BiLSTM.

Figure 7 displays the loss for both training and validation. Iteration counts and loss values are shown on the x-axis and y-axis, respectively. There is a loss during training and validation that the Bilstm model was able to accomplish. At 0.966, the Bilstm model passes the validation test.

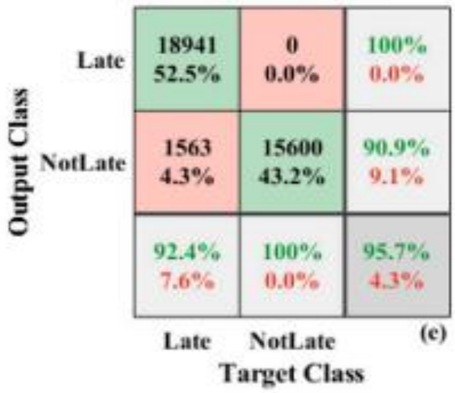

Figure 8: Confusion matrix of BiLSTM

The Bilstm model's confusion matrix is shown in Figure 8. It showed that diagonally showed valued are correctly predicted

V. COMPARATIVE ANALYSIS

The comparison of models according to performance parameters like F1-score, accuracy, and AUCROC for big data analytics using SCM is presented in this section.

Table 2: Comparison between various techniques for big data analytics using supply chain management

Model	Accuracy	F1 score	AUC ROC
BILSTM	97.59%	90.11%	14.01
KNN [40]	0.8542	0.8317	18.9
Decision tree [41]	0.6295	68.43	67.83

Table 2 presents a comparative analysis of various approaches for big data analysis using supply chain management. The BILSTM model achieves a highest accuracy 97.59%, f1-score of 90.11%, and AUC ROC is 14.01.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:09/September-2024 Impact Factor- 7.868 www.irjmets.com

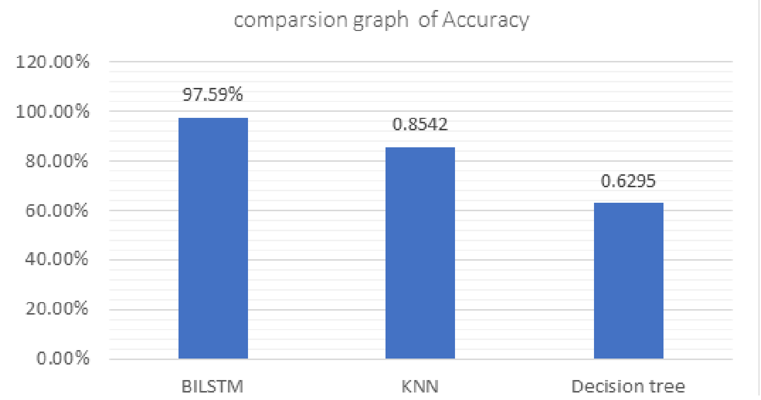

Figure 9: Accuracy comparison of models

Figure 9 shows the “Comparison Graph of Accuracy" illustrates the accuracy rates of three different machine learning models: BiLSTM, KNN, and Decision Tree are some of the basic algorithms for making predictions related to public policy. The highest accuracy is for the BiLSTM model which is 97.59%, the kNN with an accuracy of 85.43. The Decision Tree model appears to be less accurate compared with the other two models and will estimate a 62% accuracy when applied. Delayed delivery or early delivery occurrences within the dataset of Supply Chain Management are predicted accurately by the BiLSTM model with an accuracy of 95% higher than other models.

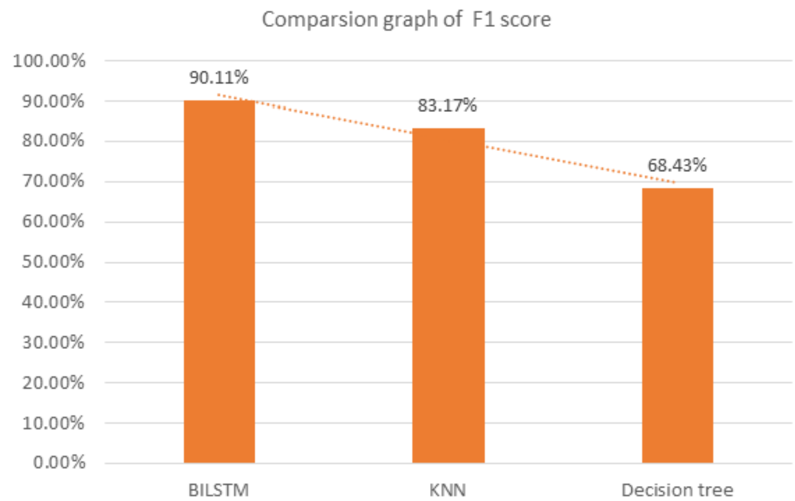

Figure 10: F1-score comparison of models on dataset

Figure 10 shows the "Comparison Graph of F1 Score" displays the F1 scores of three different machine learning models: BiLSTM, KNN, and Decision Tree. The BiLSTM model achieves the highest F1 score at 90.11%, indicating superior performance in balancing precision and recall. The K-Nearest Neighbors (KNN) model follows with an F1 score of 83.17%. The DT model has the lowest F1 score among the three, at 68.43%. This comparison demonstrates that the BiLSTM model provides the most effective trade-off between precision and recall for predicting late deliveries in the supply chain data.

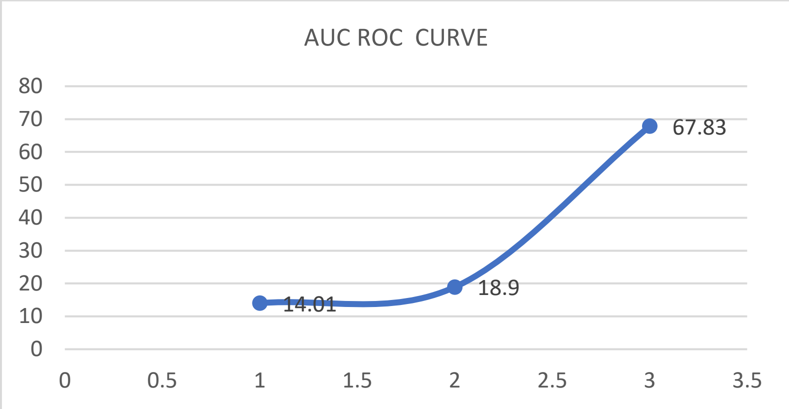

Figure 11: AUC ROC CURVE OF BILSTM

Figure 11 shows a line graph of AUC ROC CURVE" with the x-axis ranging from 0 to 4 and the y-axis from 0 to 80. It plots three key data points: (1, 14.01), (2, 18.9), and (3, 67.83). The blue line connecting these points indicates a gradual increase at first, followed by a steep rise between the second and third points.

VI. CONCLUSION AND FUTURE SCOPE

AI and big data analytics as applied to SCM have been closely linked to completely re-envisioning the ways that demand forecasting and risk management are conducted by going beyond traditional approaches. This study uses DL to improve the forecast and actual functionality of an SC, paying specific attention to the Bi-LSTM approach. Using the 'Datacom Smart Supply Chain for BDA dataset from Kaggle, a state-of-art BiLSTM model was applied to carry out significant intensive training on the provided data and the model delivered an accuracy level of 97, in precision of 59% and F1- score of 90. All the results were slightly low, with sensitivity at 11%, specificity at 87%, and an AUC-ROC of 14. 01.Comparison with other related models BiLSTM performs better in decision making for supply chain risk management. The BiLSTM model provides these benefits over the traditional methods of analysis; therefore, it is more precise and has a good rate of recall and precision of the predicted data.

Future researches could focus on the following areas to build upon the current findings and progress an employ of AI and BDA in SCM. First, it is necessary to reproduce and enhance various deep learning models, in different scenarios and datasets to verify their effectiveness and identify their limitations. However, the use of other characteristics and past data might enhance reliability of models for prediction. In addition, there is potential to look into the existing and new methods of data preprocessing and feature selection which could help to improve the performance and shorten the model making process.

VII. REFERENCES

[1] M. Hussain, M. Ajmal, G. Subramanian, M. Khan, and S. Anas, "Challenges of big data analytics for sustainable supply chains in healthcare – a resource-based view," *Benchmarking*, 2023, doi: 10.1108/BIJ-06-2022-0390.

[2] J. Thomas, "Enhancing Supply Chain Resilience Through Cloud-Based SCM and Advanced Machine Learning: A Case Study of Logistics," *J. Emerg. Technol. Innov. Res.*, vol. 8, no. 9, 2021.

[3] T. Nguyen, L. ZHOU, V. Spiegler, P. Ieromonachou, and Y. Lin, "Big data analytics in supply chain management: A state-of-the-art literature review," *Comput. Oper. Res.*, 2018, doi: 10.1016/j.cor.2017.07.004.

[4] J. Thomas, "Optimizing Bio-energy Supply Chain to Achieve Alternative Energy Targets," pp. 2260–2273, 2024.

[5] I. Lee and G. Mangalaraj, "Big Data Analytics in Supply Chain Management: A Systematic Literature Review and Research Directions," *Big Data and Cognitive Computing*. 2022. doi: 10.3390/bdcc6010017.

[6] J. Thomas, "Optimizing Nurse Scheduling : A Supply Chain Approach for Healthcare Institutions," pp. 2251–2259, 2024.

[7] S. K. R. Anumandla, V. K. Yarlagadda, S. C. R. Vennapusa, and K. R. V Kothapalli, "Unveiling the Influence of Artificial Intelligence on Resource Management and Sustainable Development: A Comprehensive

Investigation," *Technol. \& Manag. Rev.*, vol. 5, no. 1, pp. 45–65, 2020.

[8] K. R. V. K. Raghunath Kashyap Karanam, Dipakkumar Kanubhai Sachani, Vineel Mouli Natakam, Vamsi Krishna Yarlagadda, "Resilient Supply Chains: Strategies for Managing Disruptions in a Globalized Economy," *Am. J. Trade Policy*, vol. 11, no. 1, pp. 7–16, 2024.

[9] M. D. R. Pérez-Salazar, A. A. A. Lasserre, M. G. Cedillo-Campos, and J. C. H. González, "The role of knowledge management in supply chain management: A literature review," *Journal of Industrial Engineering and Management*. 2017. doi: 10.3926/jiem.2144.

[10] K. V. V. and S. G. Jubin Thomas , Piyush Patidar, "An analysis of predictive maintenance strategies in supply chain management," *Int. J. Sci. Res. Arch.*, vol. 06, no. 01, pp. 308–317, 2022, doi: DOI: https://doi.org/10.30574/ijsra.2022.6.1.0144.

[11] S. R. Thota and S. Arora, "Neurosymbolic AI for Explainable Recommendations in Frontend UI Design - Bridging the Gap between Data-Driven and Rule-Based Approaches," no. May, pp. 766–775, 2024.

[12] M. Seyedan and F. Mafakheri, "Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities," *J. Big Data*, 2020, doi: 10.1186/s40537-020-00329-2.

[13] B. P. Kishore Mullangi, Niravkumar Dhameliya, Sunil Kumar Reddy Anumandla, Vamsi Krishna Yarlagadda, Dipakkumar Kanubhai Sachani, Sai Charan Reddy Vennapusa, Sai Sirisha Maddula, "AI-Augmented Decision-Making in Management Using Quantum Networks," *Asian Bus. Rev.*, vol. 13, no. 2, pp. 73–86, 2023.

[14] A. J. K. and R. N. N.-S. Sini Sunny, jennifer Houg, Shibu Navaneeth, Sayed Aniqa, "Abstract P2073: Hyperbaric Oxygen Therapy Protects The Myocardium From Reductive Stress-induced Proteotoxic Remodeling," 2023, doi: https://doi.org/10.1161/res.133.suppl_1.P2073.

[15] V. Kumar, V. V. Kumar, N. Mishra, F. T. S. Chan, and B. Gnanasekar, "Warranty failure analysis in service supply Chain a multi-agent framework," in *SCMIS 2010 - Proceedings of 2010 8th International Conference on Supply Chain Management and Information Systems: Logistics Systems and Engineering*, 2010.

[16] V. V. Kumar, A. Sahoo, S. K. Balasubramanian, and S. Gholston, "Mitigating healthcare supply chain challenges under disaster conditions: a holistic AI-based analysis of social media data," *Int. J. Prod. Res.*, 2024, doi: 10.1080/00207543.2024.2316884.

[17] G. Hanifan and K. Timmermans, "New Supply Chain Jobs Are Emerging as AI Takes Hold," *Harv. Bus. Rev.*, 2018.

[18] R. Tandon, "Face mask detection model based on deep CNN techniques using AWS," *Int. J. Eng. Res. Appl.*, vol. 13, no. 5, pp. 12–19, 2023.

[19] V. Rohilla, S. Chakraborty, and R. Kumar, "Deep learning based feature extraction and a bidirectional hybrid optimized model for location based advertising," *Multimed. Tools Appl.*, 2022, doi: 10.1007/s11042-022-12457-3.

[20] F. Naz, A. Kumar, A. Majumdar, and R. Agrawal, "Is artificial intelligence an enabler of supply chain resiliency post COVID-19? An exploratory state-of-the-art review for future research," *Oper. Manag. Res.*, 2022, doi: 10.1007/s12063-021-00208-w.

[21] S. Tiwari, H. M. Wee, and Y. Daryanto, "Big data analytics in supply chain management between 2010 and 2016: Insights to industries," *Comput. Ind. Eng.*, 2018, doi: 10.1016/j.cie.2017.11.017.

[22] L. Huang, "Operation Mechanism Management of Parts Supply Chain Driven by Big Data," in *2023 6th International Conference on Artificial Intelligence and Big Data, ICAIBD 2023*, 2023. doi: 10.1109/ICAIBD57115.2023.10206325.

[23] J. Hu, "Optimization of Supply Chain Management System Based on Big Data Analysis Technology," in *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques, EASCT 2023*, 2023. doi: 10.1109/EASCT59475.2023.10393709.

[24] S. Bag, S. Gupta, T. M. Choi, and A. Kumar, "Roles of Innovation Leadership on Using Big Data Analytics to Establish Resilient Healthcare Supply Chains to Combat the COVID-19 Pandemic: A Multimethodological Study," *IEEE Trans. Eng. Manag.*, 2021, doi: 10.1109/TEM.2021.3101590.

[25] S. Ghouati, L. Rizi, S. Oulfarsi, and A. El Amri, "The Impact of Artificial Intelligence on Supply Chain Resilience: bibliometric analysis and systematic literature review," in *2024 IEEE 15th International Colloquium on Logistics and Supply Chain Management (LOGISTIQUA)*, 2024, pp. 1–8. doi: 10.1109/LOGISTIQUA61063.2024.10571409.

[26] D. Effah, C. Bai, W. A. Asante, and M. Quayson, "The Role of Artificial Intelligence in Coping With Extreme Weather-Induced Cocoa Supply Chain Risks," *IEEE Trans. Eng. Manag.*, 2024, doi: 10.1109/TEM.2023.3289258.

[27] E. Koç and M. Türkoğlu, "Forecasting of medical equipment demand and outbreak spreading based on deep long short-term memory network: the COVID-19 pandemic in Turkey," *Signal, Image Video Process.*, 2022, doi: 10.1007/s11760-020-01847-5.

[28] P. Shah, "Post COVID-19 Supply Chain Optimization for the Indian Pharmaceutical Industry using AI Techniques," *Intersect Stanford J. Sci. Technol. Soc.*, 2021.

[29] C. Velasco-Gallego and I. Lazakis, "A novel framework for imputing large gaps of missing values from time series sensor data of marine machinery systems," *Ships Offshore Struct.*, 2022, doi: 10.1080/17445302.2021.1943850.

[30] C. Rattanakul and Y. Lenbury, "Model analysis and simulation on impacts of COVID-19 pandemic on the economy: A case study of Thailand's GDP and its lock down measures," *Int. J. Biol. Biomed. Eng.*, 2020, doi: 10.46300/91011.2020.14.24.

[31] R. Z. Xu and M. K. He, "Application of deep learning neural network in online supply chain financial credit risk assessment," in *Proceedings - 2020 International Conference on Computer Information and Big Data Applications, CIBDA 2020*, 2020. doi: 10.1109/CIBDA50819.2020.00058.

[32] S. Islam and S. H. Amin, "Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and Gradient Boosting Machine learning techniques," *J. Big Data*, 2020, doi: 10.1186/s40537-020-00345-2.

[33] B. Abbasi, T. Babaei, Z. Hosseinifard, K. Smith-Miles, and M. Dehghani, "Predicting solutions of large-scale optimization problems via machine learning: A case study in blood supply chain management," *Comput. Oper. Res.*, 2020, doi: 10.1016/j.cor.2020.104941.

[34] A. P. A. Singh, "STRATEGIC APPROACHES TO MATERIALS DATA COLLECTION AND INVENTORY MANAGEMENT," *Int. J. Bus. Quant. Econ. Appl. Manag. Res.*, vol. 7, no. 5, 2022.

[35] K. Patel, "An Analysis of Quality Assurance Practices Based on Software Development Life Cycle (SDLC) Methodologies," *J. Emerg. Technol. Innov. Res.*, vol. 9, no. 12, pp. g587–g592, 2022.

[36] A. Kumar, V. T. Narapareddy, V. A. Srikanth, A. Malapati, and L. B. M. Neti, "Sarcasm Detection Using Multi-Head Attention Based Bidirectional LSTM," *IEEE Access*, 2020, doi: 10.1109/ACCESS.2019.2963630.

[37] H. Sinha, "ANALYZING MOVIE REVIEW SENTIMENTS ADVANCED MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING METHODS," *Int. Res. J. Mod. Eng. Technol. Sci. (*, vol. 06, no. 08, pp. 1326–1337, 2024.

[38] M. Rhanoui, M. Mikram, S. Yousfi, and S. Barzali, "A CNN-BiLSTM Model for Document-Level Sentiment Analysis," *Mach. Learn. Knowl. Extr.*, 2019, doi: 10.3390/make1030048.

[39] L. Xiaoyan, R. C. Raga, and S. Xuemei, "GloVe-CNN-BiLSTM Model for Sentiment Analysis on Text Reviews," *J. Sensors*, 2022, doi: 10.1155/2022/7212366.

[40] A. Boiko, V. Shendryk, and O. Boiko, "Information systems for supply chain management: Uncertainties, risks and cyber security," in *Procedia Computer Science*, 2019. doi: 10.1016/j.procs.2019.01.108.

[41] K. S. Liu and M. H. Lin, "Performance assessment on the application of artificial intelligence to sustainable supply chain management in the construction material industry," *Sustain.*, 2021, doi: 10.3390/su132212767.