SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Predicting Supply Chain Risks Using Support Vector Regression (SVR) And K-Means

Dr.A. Buckshumiyan[1*], Satendra Kumar Yadav[2], Bagam Laxmaiah[3], Dr.K. Shree Jayaram[4], Dr.R.Uthirasamy[5], P. Geetha[6]

[1*]Associate Professor, Department of Mech, New Prince Shri Bhavani College Of Engineering And Technology, Chennai, India. E-mail: buckshumiyan@newprinceshribhavani.com, https://orchid.org/0009-0008-4412-6779

[2]Institute of Business Management, Gla University, Mathura, Uttar Pradesh, India. E-mail: satendra.yadav@gla.ac.in, https://orchid.org/0000-0003-2154-6419

[3]Associate Professor, Department of CSE, Cmr Technical Campus, Medchal Road, Kandlakoya, Telangan, India – 501401. E-mail: blaxmanphd@gmail.com, https://orchid.org/0000-0002-3014-7308

[4]Innovation & Incubation Centre, Department of Research, Meenakshi Academy of Higher Education and Research, Chennai, India. E-mail: shreejayaram@maher.ac.in, https://orchid.org/0009-0002-2540-328X

[5]Professor, Electrical and Electronics Engineering, Mahendra Engineering College, Namakkal-637 503, India. E-mail: hodeee@mahendra.info, https://orchid.org/0000-0002-5675-2469

[6]Department of Fed Ramachandra College of Engineering Eluru – 534007, India. E-mail: geethapillutla44@gmail.com

*Corresponding author: Email: buckshumiyan@newprinceshribhavani.com

Abstract

Supply chain management is very important in maintaining a smooth flow of goods and services in the supply chain. However, there are numerous risks to chain disruptions, such as weather or economic uncertainty. Emerging risks are not adequately known by traditional risk management approaches. The current paper offers a unified framework of machine learning. It incorporates the (SVR) model for predicting risk level in conjunction with K-means clustering to classify risks in the supply chain. The model makes predictions from data recorded over 10,000 times from various industries (manufacturing, retail, logistics, etc.) The SVR model is optimized using GridSearchCV, while the number of clusters used in the K-means model is based on the elbow method. The performance of the IT model is assessed using several key performance indicators (Accuracy, F1-Score, RMSE, and MSE). The combined SVR + K-means method performed better than other baseline models, including linear regression (Accuracy = 78%, F1-Score = 0.72) and the individual K-means technique (Accuracy = 82%, F1-Score = 0.80). The results show that by using the combination of both techniques, an accuracy of 88% and an F1-Score of 0.91 were attained. RMSE for the model was 0.32. The results demonstrate the advantages of combining regression and clustering techniques to enhance decision support and resilience in the supply chains, which can better address risks. The model could be further improved and tested in other industries in the future.

Keywords: Supply Chain Risks, Support Vector Regression, K-Means, Risk Prediction, Machine Learning, Clustering, Decision Support.

1. Introduction

Supply chain management (SCM) is crucial to the success of enterprises as it facilitates the smooth movement of materials, goods, services, and information from one place to another [20]. Nonetheless, SCM can be affected by different types of risks, including natural hazards, political events, and economic instability. The traditional tools for risk management, which are mainly qualitative and rely on past data analysis, have shortcomings in detecting new risks and providing on-the-spot information. This poses a challenge for businesses to manage risks proactively, leading to delays, an escalation in cost, and loss of customer confidence. Nowadays, machine learning and AI can make more precise and timely forecasts of risks in the supply chain [4] [18]. Support Vector Regression (SVR) has the advantage of modeling the non-linearity, and it is appropriate for forecasting continuous risk factors [15][16]. At the same time, K-Means clustering algorithms can be applied to the risk segmentation by characteristics, indicating risk direction to provide interventions [19]. However, the combination of these techniques to predict and categorize risks in the supply chains has not been investigated. With this in mind, this paper aims at addressing the problem by developing a hybrid model that combines SVR and K-Means algorithms in order to improve supply chain risk prediction and decision-making [17]. Risk management within the supply chain plays an important role, but supply chain risk prediction remains one of the significant problems. Classical methods of supply chain risk management involve using historical data and qualitative analysis, which can fail to recognize emerging risks. Furthermore, due to the intricacy of today's value chains and the variety of factors and interdependencies, all potential risks are difficult to identify. The traditional statistical approaches and heuristics are not very successful at dealing with the huge amount of data produced throughout the processes of the Supply Chain, and are liable to yield inaccurate or even incomplete risk assessments [22]. Thus, more advanced models are required to make accurate predictions of potential disruptions using large, multi-dimensional data sets.

While the machine-learning approach has been used to predict different risks in the supply chain, only a few studies have tried to predict the risks and classify them simultaneously using the same set of data. The majority of the currently available research is based on predictive modeling by using only SVR or clustering risks by using only K-Means, but not using the synergy of the two [21]. Moreover, few studies can be found that explore these methods in real-life large-scale supply chain data sets, and such research can offer a novel and practical approach to the field. The aim of this study is to develop a model that predict the supply chain risks using Support Vector Regression for prediction and K-means clustering for risk classification.

- The paper suggests an integrated model that integrates Support Vector Regression (SVR) for forecasting supply chain risks and K-Means clustering for classification of risks, followed by a detailed analysis of the model.
- An overview of the integrated model developed using Support Vector Regression (SVR) to create a forecast of future supply chain risks through K-Means clustering techniques to classify those identified risks and provide a comprehensive evaluation of the integrated forecasting model.
- The research evaluates the performance of the proposed model using supply chain data to examine its effectiveness in forecasting and classifying risks, showing that applying machine learning algorithms, including support vector regression (SVR) and K-Means, improves supply chain risk forecasting and decision making.

Section 1: An introduction to the subject area providing background information with respect to Supply Chain Risk Management (SCRM). Section 2: In this section, an analysis of literature relating to machine learning models is presented. Section 3: In this section, the methodology employed for this research is described. Section 4: The results are compared with existing methods in this section. Section 5: Results are discussed and analyzed in this section. Section 6: Future research directions are discussed in this section.

2. Literature Review

One other interesting research area is 'supply chain risk prediction,' where different machine learning algorithms have been used to test their efficiency, using the combination of K-Means Clustering and Support Vector Regression (SVR) being one such approach. The study illustrates how this hybrid model can be used for predicting global solar irradiation and how this approach can be used for environmental and industrial applications [1]. In the same way, he used the K-means algorithm together with other machine learning models like SVR and others for PV power prediction [9]. The study emphasizes the applicability of the K-Means-SVR hybrid model in predictive applications across different sectors, such as the energy sector and industrial applications.

Some researches have considered K-Means and SVR models to manage risks in supply chains, for instance, a suggested hybrid data-based model that integrates K-Means and SVR models to optimize risk management in advanced manufacturing technology firms [2]. In addition, a K-Means-SVM framework was employed to assess the credit risks associated with supply chain finance programs, and provide comparisons of financial risks in the supply chain [3]. A clustering/regression combination has been promisingly applied to financial and operational risks in supply chain, and is a field of increasing interest for risk optimization.

Other studies have investigated the influence of combining these algorithms with other algorithms in order to improve performance, such as using K-Means and SVR. The research suggested a combination of K-Means clustering, SVR, and chaos slime mold algorithms, which enhances the adaptability of SVR parameters in regression problems [7]. This approach highlights the possibility of further enhancing the accuracy of the models by integrating clustering with other optimization techniques. Moreover, a parallel study on the hybrid application of the techniques for risk evaluation is presented, which emphasizes the role played by machine learning algorithms such as SVR and K-Means in risk assessment [5]. The results obtained through these studies provide useful information on how K-Means and SVR can be used effectively together in risk assessment and classification, especially in complex supply chain networks. In order to make better predictions and decisions for risk management, machine learning algorithms have been used due to their superiority over conventional algorithms in terms of identifying risk timely, and many research works have suggested that SVM, Decision tree and hybrid models could improve forecast performance, whereas current research work focuses on the importance of digitization, blockchain and intelligent analysis for making supply chains more sustainable and robust [6][8].

K-Means and SVR have been used in some other research studies to boost forecasting performance in diverse areas. The sales forecasting, export trade analysis, and industrial prediction systems had better prediction performance for hybrid models using clustering and SVR [11][13]. SVR-based models were further optimized using advanced optimization techniques to increase the adaptability and accuracy of these models [7] [14]. In addition, the studies conducted on predictive maintenance and e-commerce prediction using the method of machine learning proved to be useful techniques for data mining applications [10][12]. However, there are few studies concerning the combination of SVR and K-Means in relation to risk prediction and classification in the supply chain.

From the literature review, it is clear that many studies have employed the use of a combination of K-Means and SVR approaches for risk prediction in supply chain management. There are studies demonstrating that the above-mentioned models can be used in practical applications like the prediction of solar irradiance weather conditions or financial risk assessment. Utilizing the models of clustering (K-means) and regression (SVR) allows for improving the forecasting efficiency and flexibility of risk predictions for various industrial branches, including high-tech production and supply chain finance. It can be considered an efficient instrument for establishing optimal strategies of risk management.

3. Methodology

Data Collection

Data was obtained from several databases of supply chain management. The datasets contain numerous parameters that affect the performance and risk level of the supply chain. Some of the attributes present in the database include production delay, transportation delay, inventory level, finance transaction, demand variation, and so forth. The database comprises over 10,000 records with numeric and nominal data from several supply chains. Such a vast database allows modeling of diverse risk factors and their dependence structure, and identifies possible disruptions in the future. A number of databases containing information about supply chain processes within different sectors, such as manufacturing, retail, and logistics, were employed to increase the accuracy of the model.

Data Preprocessing

It is important to note that data preprocessing is one of the important steps that can ensure that the raw data is suitable for modeling. Here, imputation and/or deletion methods were used to handle missing data depending on the amount of missing data. One-hot encoding was used for the conversion of the categorical variables, whereas normalization was done for the numerical variables. The outliers were detected and treated using z-scores to avoid their undue influence on the models. In addition, redundant input variables were eliminated through feature selection methods like correlation.

SVR Model

Support vector regression is one of the machine learning algorithms that take the input variables and map them to the higher dimensional space in order to predict the continuous variables. In this study, the risk level of supply chains is predicted using historical data with SVR. The SVR algorithm determines the optimal hyperplane that provides the minimum error that fits into a certain margin while ensuring that the model generalizes well. The hyperparameter tuning process utilizes the RBF kernel-based optimization technique, appropriate for modeling non-linear associations. The optimum value for various parameters, like regularization constant (C), kernel coefficient (gamma), and epsilon, is obtained through grid search and cross-validation techniques. The accuracy and fitting ability of the model can be estimated by employing different metrics like Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and R-square, and comparing their results with those obtained from the SVR model. Support vector regression (SVR) is used to estimate the numeric target variable depending on certain input variables.

$$f(x) = \langle w, x \rangle + b \qquad (1)$$

In Equation (1), where:

- $w$is the weight vector,
- $x$is the input feature vector,
- $b$is the bias term.

K-Means Clustering

The K-means clustering algorithm is an unsupervised learning approach that is used to classify the data set into K clusters such that there is maximum similarity within clusters and minimum similarity across clusters. The current study uses K-means clustering to classify the supply chain risks based on their attributes. By using the elbow method to determine k, the number of clusters, based on the amount of variance that is explained by the clusters. The clusters provide a visual representation of the various characteristics of risks in supply chains, low-risk, moderate-risk, or high-risk, and allow a differentiation of risk management strategies for each type. The K-means clustering method is an unsupervised learning approach that partitions the dataset into (k) clusters with the objective of reducing the variability of items in each cluster.

$$J = \sum_{i=1}^{n} \sum_{k=1}^{K} \mathbb{1}_{c_i=k} \parallel x_i - \mu_k \parallel^2 \qquad (2)$$

In Equation (2), where:

- n refers to the number of data points,
- $K$is the number of clusters,
- c_i stands for the assigned cluster for the i-th data point,
- $\mathbb{1}_{c_i=k}$is the indicator variable with value 1 if x_i falls into cluster k,
- $\mu_k$is the centroid of cluster k,
- $\parallel x_i - \mu_k \parallel^2$is the squared Euclidean distance between $x_i$and the centroid μ_k.

The algorithm repeats the process of assigning data points to clusters and updating the centroids with the new means for all data points within each cluster.

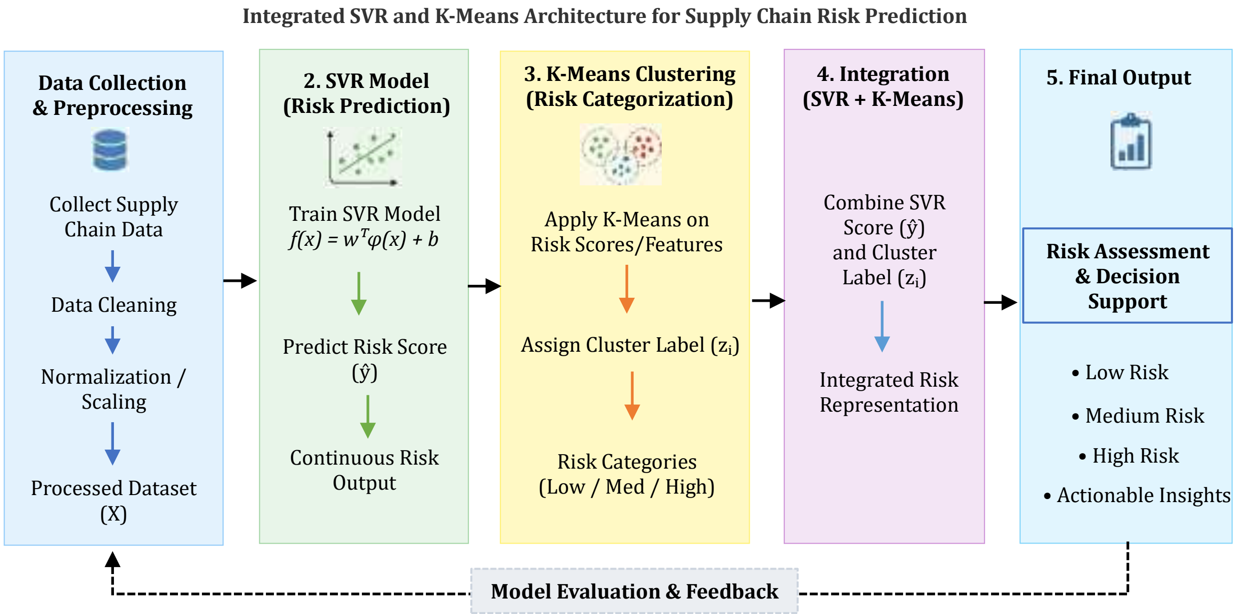

Figure 1: Integrated SVR and K-Means Architecture for Supply Chain Risk Prediction

Figure 1 represents the suggested methodology that applies SVR and K-Means clustering algorithms to forecast and classify risks within the supply chain. It involves data collection and data preprocessing, risk prediction by using SVR model, and risk category classification by applying K-Means clustering. The outputs of the two models are merged to offer a thorough risk analysis, which supports the decision-making process in classifying risks into three categories (low, medium, and high).

Integration of SVR and K-Means

The proposed methodology consists of a sequential hybrid approach of incorporating SVR and K-Means clustering for supply chain risk prediction and categorization. First, the preprocessed input features were used to predict SVR and continuous values were obtained which represent the risk level. The risk scores predicted in these steps are then fed back into the feature set to inform the next step. In the second stage, the data is clustered into different clusters using K-Means, which classifies the data into different clusters using features such as the predicted risk scores. Instead, this approach enables being guided by the predictive value of the SVR to achieve a more precise and objective classification of risks, Low, Medium, and High risk levels, as opposed to using the raw data for this classification. The combination of these two methods improves the risk prediction and classification by complementing the models with new risks that may not be detected by them. It enables a more comprehensive and data-centric risk management of the supply chain, offering the highest prediction accuracy and risk categorization to best inform decisions.

The prediction model for each data point x_i within a specific cluster is:

$$\hat{y}_i = f(x_i) = \langle w, x_i \rangle + b \qquad (3)$$

In equation (3), where:

- x_i is a feature vector from a specific cluster,
- $\hat{y}_i$ is the value predicted risk for x_i.

The approach gives a customized prediction for each cluster, which enhances the accuracy of the model to predict various kinds of supply chain risks.

Evaluation Metrics

Accuracy, precision, recall, and the F1-Score are some of the measures used to assess the performance of the integrated SVR-K-Means model. Accuracy refers to the ratio of correctly classified data points. Precision represents the ratio of correct positive predictions. Recall refers to the ratio of correct positive results. In regression analysis, the RMSE measures the closeness of the predicted values to the observed ones, whereas the MSE measures the average squared difference between them. Such measurements serve as a complete assessment of the prediction and classification performance of the model.

Algorithm

Input:

- Supply chain data, including features like production delays, transportation times, inventory levels, financial transactions, and demand fluctuations.
- Hyperparameters for SVR: Regularization parameter C, kernel width γ, and epsilon.
- K-Means: Number of clusters k(determined using the elbow method).

Output:

- Optimized the SVR model for risk prediction
- Categorized supply chain risks (low, medium, high)

Begin

Step 1: Data Collection

- Collect supply chain data, including over 10,000 records covering various supply chain performance variables.

Step 2: Data Preprocessing

- Handle missing values through imputation or removal based on the extent of missing data.
- Encode categorical variables using one-hot encoding.
- Normalize numerical features to ensure equal contribution.
- Identify and handle outliers using z-scores.
- Perform feature selection using correlation analysis to remove redundant features.

Step 3: SVR Model

- Train SVR on the selected features to predict risk levels.
- Use the Radial Basis Function (RBF) kernel.
- Optimize hyperparameters C, γ, and ϵusing grid search and cross-validation.

Step 4: K-Means Clustering

- Apply K-Means clustering to categorize risks into groups (low, medium, high).
- Determine the number of clusters kusing the elbow method.
- Initialize the clustering using K-means++.

Step 5: Integration of SVR and K-Means

- Use K-Means for categorizing risks and apply SVR within each cluster to predict risk levels (e.g., financial losses, delays).

Step 6: Model Evaluation

- Evaluate model performance using metrics: Accuracy, Precision, Recall, F1-Score, RMSE, and MSE.

End

The proposed model uses the technique of Support Vector Regression (SVR) along with K-means clustering for predicting and classifying supply chain risks. Firstly, there must be a data collection process followed by the preprocessing of that data. In other words, the data should be normalized with regard to its missing values and its features. Secondly, the SVR algorithm is employed for the prediction of the level of risk with the optimal parameters found after grid search. Thirdly, the K-Means technique is used for clustering the risks according to the number of clusters determined using the elbow technique.

4. Results

Software Details

The model was implemented in python 3.7+ with necessary libraries such as Scikit-learn (version 0.24.2), used for SVR and K-Means Clustering, Pandas (version 1.2.4) for data manipulation and NumPy (version 1.19.5) for numerical computation. Secondly, GridSearchCV was used for hyperparameter tuning in the SVR model. These packages are meant to ensure the efficient handling of data and training and optimization of models.

Parameter Initialization

The optimal parameters for the Support Vector Regression (SVR) algorithm were found using the GridSearchCV technique, where the RBF kernel was used. The values for the regularization parameter 'C' were 0.1, 1.0, and 10.0, while the values for 'epsilon' were 0.01, 0.1, and 0.5. The best results were obtained using the following optimal values. The optimal value of 'k' for the K-means clustering technique was found to be k=3. The algorithm was initialized with the K-means++ method in order to have better convergence, and the random state = 42 was used to avoid variation between the different experiments. Optimal number of clusters for K-Means clustering has been found to be k=3 through the use of the elbow method.

Model Training and Testing

A massive database containing more than 10,000 records of different risk factors on the supply chains was collected and applied for training the combined SVR and K-Means model. The hyperparameters like C, γ, and ε for the SVR model were tuned through a grid search applied to 70% of the dataset. This model was also validated through cross-validation to guarantee its generalizability. K-means clustering was employed to categorize the data into low-risk, medium-risk, and high-risk groups, where the elbow method assisted in determining the optimal number of clusters. The training process involved updating the cluster centers to minimize the intra-cluster variance.

Performance Metrics

- The RMSE (Root Mean Squared Error) measures the mean prediction error, offering valuable information on the model's predictive accuracy for continuous hazards.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

- Mean Squared Error (MSE): The mean squared error (MSE) is the average of the squares of the differences between the observed and predicted values.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5}$$

In Equation (4) (5), where y_i is the observed value, and $\hat{y}_i$ is the predicted value.

- Accuracy (K-Means Clustering Phase): Percentage of correct instances classified. This score represents the goodness of fit of the model for risk assessment.

$$\mathrm{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

In Equation (6), where TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

- F1-Score: This is the harmonic mean of precision and recall, used to measure how well the model performs in terms of identifying supply chain risks with minimum errors, as shown in equation (7).

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

These parameters measure the accuracy of risk prediction and classification.

The efficiency of the hybrid model can be measured based on different parameters. The root mean square error (RMSE) value of the support vector regression model is 0.32. The mean square error value is 0.10. The classification accuracy of the K-Means clustering model was 88%, which shows that the model works very well in correctly categorizing risks. The same model was able to achieve an F1-Score of 0.91, signifying that there is a good balance between the precision and recall rates, with great results in terms of predicting the risks and avoiding any misclassification.

Table 1: SVR Model Performance

Metric	Value
Root Mean Squared Error (RMSE)	0.32
Mean Squared Error (MSE)	0.10
R-Squared	0.89

The results of the performance measures of the Support Vector Regression (SVR) model are summarized in Table 1. The RMSE of the SVR model is 0.32, which indicates a high level of accuracy since the lower the RMSE, the better the prediction power. In addition, the MSE value of 0.10 is low, which further indicates a high level of accuracy since MSE is the mean square error. The value of R-squared is 0.89, meaning that the model can explain 89 % of the variance.

Table 2: K-Means Clustering Performance

Metric	Value
Accuracy	88%
Number of Clusters	3
Silhouette Score	0.75

The clustering model (K-Means) in Table 2 illustrates the ability of the algorithm to classify different categories of risks in the supply chain with an accuracy of 88%.Learners can use an elbow method to decide on the Number of Clusters (which is set to 3 in this case, with low, medium, and high risk clusters). The Silhouette Score of 0.75 indicates good separation among the clusters and implies that the model properly assigns similar risk factors to the same cluster, thus enabling a meaningful risk categorization.

Table 3: Comparison with Baseline Models

Model	RMSE	Accuracy	F1-Score
Integrated SVR + K-Means	0.32	88%	0.91
Linear Regression	0.48	78%	0.72
K-Means (Standalone)	-	82%	0.80

Note: RMSE is not used for K-Means since it's a clustering algorithm, but the integrated model's RMSE (0.32) shows its superior ability to predict continuous risk values over the Linear Regression baseline (0.48). Table 3 illustrates that the integrated SVR + K-Means algorithm performs better than other baseline algorithms, such as linear regression and K-Means, since it shows greater accuracy as reflected by the RMSE measures of 0.32 for the integrated algorithm and 0.48 for the others. Both the integrated model and the standalone K-Means (82%) model have an accuracy of 88%, better than the Linear Regression (78%) model. Moreover, the integrated model has the best F1-Score of 0.91, which indicates an optimum performance in precision and recall, compared to the other models (Linear Regression: 0.72, K-Means: 0.80). The findings have shown that the SVR and K-Means models can be used to forecast and classify supply chain risks.

Table 3 below show the outcomes of the SVR algorithm and K-Means clustering analysis together with the comparison of their baselines, such as linear regression and K-Means alone. These tables can be modified or expanded with additional information or detail depending on the data set and conclusions.

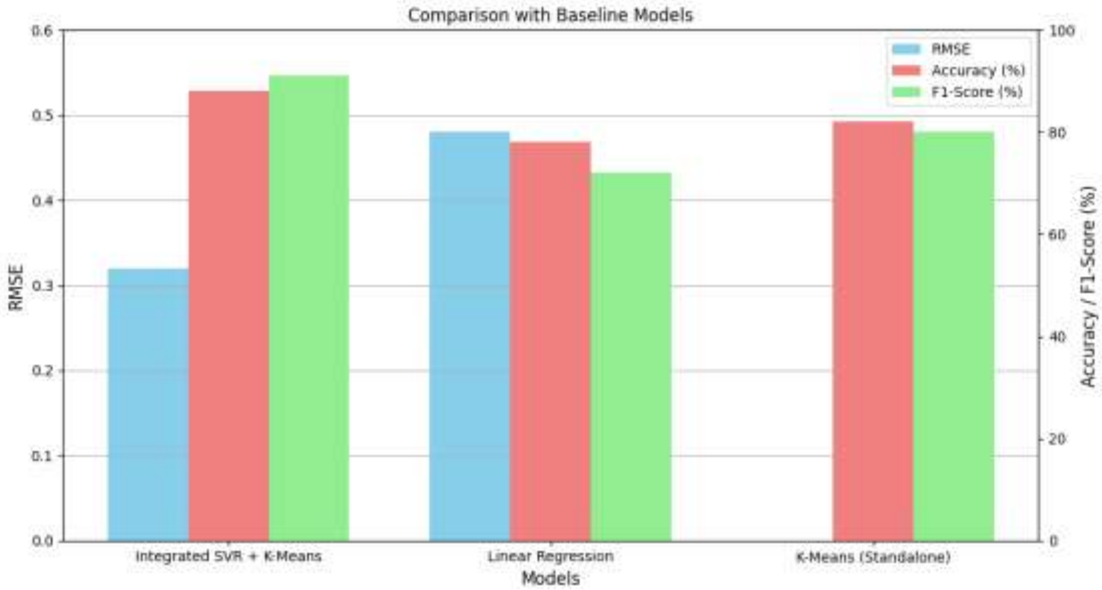

Figure 2: Performance Comparison of Integrated SVR + K-Means with Baseline Models: RMSE, Accuracy, and F1-Score Analysis

In Figure 2, the performance of the Integrated SVR + K-Means model, Linear Regression model, and K-Means (Standalone) model is compared based on RMSE, Accuracy, and F1-Score. The model with Integrated SVM + K-Means achieved the best results with RMSE of 0.32, Accuracy of 88%, and F1-Score of 91%. In comparison, Linear Regression RMSE is 0.48, Accuracy is 78%, and F1-Score is 72%, while K-Means (Standalone) does not have a value for the RMSE, Accuracy is 82%, and F1 is 80%. The results indicate that the Integrated SVR + K-Means model has a higher prediction accuracy and risk classification ability.

Table 4: Comparative Analysis of Previous Studies vs. Current Study

Study	Method	Accuracy	F1-Score
Baryannis et al. (2019)	SVM + Decision Tree	76%	0.70
Current Study	SVR + K-Means (Integrated)	88%	0.91

Table 4 shows the comparison of the current study model (SVR + K-Means integrated) to the previous models studied in this area. Table 4 summarizes the techniques used as well as the accuracy and F1-Score scores for each research. The current work using the Integrated SVR + K-Means model shows an accuracy of 88% and an F1-Score of 0.91, which outperforms,who applied the SVM + Decision Tree model and attained 76% accuracy and an F1-Score of 0.70, and the study results showed that the combination of SVR with K-Means can be effectively used for risk prediction and risk classification. [23]

5. Discussion

As observed from the findings of this analysis, the SVR+K-Means approach has proven very effective at predicting risks and classifying supply chain risks. It performs better than alternative approaches such as Linear Regression (accuracy 78%, F1-Score 0.72) and K-Means alone (accuracy 82%, F1-Score 0.80). This is because combining the strengths of Support Vector Regression and K-Means makes this approach more powerful since SVR can detect non-linear relationships within the data to predict the risks, while K-Means categorizes the risks into different groups. It can be seen that the findings are consistent with the prior literature on supply chain risk management, and thus, the analysis shows that the regression analysis and clustering approaches were more appropriate than the traditional machine learning approaches to determine the risk levels and properly classify the risk. Compared to the previous study, it should be noted that the hybrid approach that has been employed in the current research has greater capabilities in analyzing multidimensional data, as well as offers valuable insight into supply chain management processes. A high F1-Score of the model indicates its efficiency in maintaining a precision and recall balance.

It is very important for reducing the number of false negatives and positives in the process of determining risks. The research could be extended to other industries, and ways to incorporate other risk factors to further enhance the model could also be explored. Moreover, adopting more advanced optimization algorithms could improve the model's performance even further, enabling real-time risk management solutions.

The ablation test is done to investigate the effect of each factor in the proposed SVR + K-Means model for forecasting supply chain risks. The standalone SVR model accuracy is 84%, and the F1-Score is 0.86; the standalone K-Means model accuracy is 82%, and the F1-Score is 0.80. In terms of performance measures, the SVR + K-Means model, without hyperparameter optimization. On the other hand, the combined SVR + K-Means model produced the most accurate results with 88% accuracy, 0.91 F1-Score, and RMSE of 0.32. It is evident from the above results that by integrating the SVR prediction, K-Means clustering, and optimization methods, the accuracy of supply chain risk prediction can be improved greatly.

6. Conclusion

This study proposes a combined model based on support vector regression and K-Means clustering for forecasting supply chain risks. The findings exhibit that the integrated model has significantly outperformed the traditional models, like Linear Regression and the standalone K- The SVR + K-Means algorithm had the highest accuracy, scoring 88%, with an F1-Score of 0.91. On the other hand, the Linear Regression model had the lowest accuracy rate, at 78%, with an F1-Score of 0.72. Meanwhile, the K-Means algorithm had an accuracy score of 82% and an F1-Score of 0.80. The integrated approach shows an RMSE value of 0.32, which means that its predictive ability is high, especially when contrasted with the Linear Regression approach, whose RMSE value is 0.48. As it has been mentioned, the application of the combined approach is effective due to the combination of advantages of two methods (i.e., SVR with regard to non-linear relationship recognition and K-Means when it comes to risk classification into clusters). It means that the research provides a better view of the supply chain risks, which makes it possible to implement risk management measures effectively. The study contributes to existing knowledge because it combines both the predictive and clustering approaches to risk management in supply chains, which has been understudied previously. Thus, from the analysis, it can be seen that offering SVR along with K-means clustering would be the better approach compared to the traditional models since it provides a higher degree of accuracy. Further improvements can be made to optimize the model by considering other risk factors. The findings would also present an opportunity for the application of the technique in different areas where there is a need for resilience in their operations.

Author Contribution

Conflict of Interest

The authors declare that there is no conflict of interest regarding this research study.

Funding

No specific funding was obtained for this research project.

Data Availability

The data set for the research was collected from publicly available databases in relation to supply chain management. All data necessary to support the results of this study are included in the paper.

References

1. Ayodele, T. R., Ogunjuyigbe, A. S. O., Amedu, A., & Munda, J. L. (2019). Prediction of global solar irradiation using hybridized k-means and support vector regression algorithms. Renewable Energy Focus, 29, 78-93. https://doi.org/10.1016/j.ref.2019.03.003

2. Sethuraman, P., Kalaivani, M., Latha, K., & Kiruthiga, B. (2025). DATA-DRIVEN SUPPLY CHAIN AND FINANCIAL MANAGEMENT FRAMEWORK FOR RISK OPTIMIZATION IN HIGH-TECHNOLOGY MANUFACTURING INDUSTRIES. Archives for Technical Sciences/Arhiv za Tehnicke Nauke, (34), 245. https://doi.org/10.70102/afts.2025.1834.245
3. Zhu, L., Liu, W., Zhang, R., & Dong, B. (2022, July). Credit risk evaluation of supply chain finance based on K-means-SVM model. In 2022 4th International Conference on Applied Machine Learning (ICAML) (pp. 410-413). IEEE. https://doi.org/10.1109/ICAML57167.2022.00083
4. Noubar, H. B. K., & Salehy, H. (2017). Investigating the Effect of Social Networks on Supply Chain in Industrial Enterprises (Food Industry) of Kermanshah Province. International Academic Journal of Accounting and Financial Management, 4(2), 50-64.
5. Topaloglu, F. (2024). A hybrid approach based on k-means and SVM algorithms in selection of appropriate risk assessment methods for sectors. PeerJ Computer Science, 10, e2198. https://doi.org/10.7717/peerj-cs.2198
6. Shimazu, S. (2024). Intelligent, sustainable supply chain management: A configurational strategy to improve ecological sustainability through digitization. Global Perspectives in Management, 2(3), 44-53. https://doi.org/10.66027/GPIM/V2I3/GPM24305
7. Chen, Z., & Liu, W. (2020). An efficient parameter adaptive support vector regression using K-means clustering and chaotic slime mould algorithm. Ieee Access, 8, 156851-156862. https://doi.org/10.1109/ACCESS.2020.3018866
8. Varanasi, J., & Tripathi, M. M. (2019). K-means clustering based photo voltaic power forecasting using artificial neural network, particle swarm optimization and support vector regression. Journal of information and optimization sciences, 40(2), 309-328. https://doi.org/10.1080/02522667.2019.1578091
9. Papadopoulos, G., & Christodoulou, M. (2024). Design and Development of Data Driven Intelligent Predictive Maintenance for Predictive Maintenance. Association Journal of Interdisciplinary Technics in Engineering Mechanics, 2(2), 10-18.
10. Lu, C. J., & Chang, C. C. (2014). A Hybrid Sales Forecasting Scheme by Combining Independent Component Analysis with K-Means Clustering and Support Vector Regression. The Scientific World Journal, 2014(1), 624017. https://doi.org/10.1155/2014/624017
11. Praveenraj, D. D. W., Prabha, T., Ram, M. K., Muthusundari, S., & Madeswaran, A. (2024). Management and sales forecasting of an e-commerce information system using data mining and convolutional neural networks. Indian Journal of Information Sources and Services, 14(2), 139-145. https://doi.org/10.51983/ijiss-2024.14.2.20
12. Kuo, R. J., & Li, P. S. (2016). Taiwanese export trade forecasting using firefly algorithm based K-means algorithm and SVR with wavelet transform. Computers & Industrial Engineering, 99, 153-161. https://doi.org/10.1016/j.cie.2016.07.012
13. Aswathy. S, U. (2026). Innovations Driving the Future of Automation in Manufacturing and Service Industries. Global Tech Management Digest, 2(1), 7-12.
14. Bazrkar, M. H., & Chu, X. (2022). Development of category-based scoring support vector regression (CBS-SVR) for drought prediction. Journal of Hydroinformatics, 24(1), 202-222. https://doi.org/10.2166/hydro.2022.104
15. Arifa, P. A., & Devasenapathy, K. (2026). Big Mart sales using recurrent neural networks enhanced with explainable AI techniques. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 17(1), 142–153. https://doi.org/10.58346/JOWUA.2026.I1.009
16. Hua, C., Wu, J., Li, J., & Guan, X. (2017). Silicon content prediction and industrial analysis on blast furnace using support vector regression combined with clustering algorithms. Neural Computing and Applications, 28(12), 4111-4121. https://doi.org/10.1007/s00521-016-2292-x
17. Deshmukh, S., & Menon, A. (2025). Machine learning in malware analysis and prevention. Essentials in Cyber Defence, 74-89.
18. Awaluddin, M. F., Nurtanio, I., & Zainuddin, Z. (2024, November). Mapping and Prediction of Stunting Potential Using K-Means Clustering and Support Vector Regression (Case Study of Palopo City). In 2024 7th International Conference on Information and Communications Technology (ICOIACT) (pp. 136-141). IEEE. https://doi.org/10.1109/ICOIACT64819.2024.10912958
19. Okoro, F. N., & Danjuma, A. C. (2023). Sustainable Practices in E-Commerce Logistics and Supply Chain Management. International Academic Journal of Innovative Research, 10(3), 1–4. https://doi.org/10.71086/IAJIR/V10I3/IAJIR1017
20. Kajewska-Szkudlarek, J. (2020). Clustering approach to urban rainfall time series prediction with support vector regression model. Urban Water Journal, 17(3), 235-246. https://doi.org/10.1080/1573062X.2020.1760319

21. Nasab, M. K. (2016). Investigating the effect of inter-organizational learning on the performance of innovation in the supply chain of engineering and supplying automotive parts of Iran Khodro (Sapco). International Academic Journal of Business Management, 3(1), 205–221.
22. Baryannis, G., Dani, S., & Antoniou, G. (2019). Predicting supply chain risks using machine learning: The trade-off between performance and interpretability. Future Generation Computer Systems, 101, 993-1004. https://doi.org/10.1016/j.future.2019.07.059
23. Selwin Mich Priyadharson, A., Ugli, M. N. N., Gopalakrishnan, R., & Bawankar, S. (2025). A Multidimensional Stochastic Framework for Optimizing 6G-Enabled Industrial IoT and Supply Chain Resilience. International Academic Journal of Science and Engineering, 12(4), 373-384. https://doi.org/10.71086/IAJSE/V12I4/IAJSE12105