Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Backorder prediction in the supply chain using machine learning

Lokesh Malviya [a,*], Pankaj Chittora [b], Prasun Chakrabarti [c], Raj Shekhar Vyas [c], Sandeep Poddar [d]

[a] *Department of Mechanical Engineering, Techno India NJR Institute of Technology, Udaipur, Rajasthan, India*
[b] *Department of Computer Science Engineering, Techno India NJR Institute of Technology, Udaipur, Rajasthan, India*
[c] *Techno India NJR Institute of Technology, Udaipur, Rajasthan, India*
[d] *Lincoln University College, Malaysia*

ARTICLE INFO

Article history:
Received 11 November 2020
Accepted 17 November 2020
Available online xxxx

Keywords:
Supply chain management
Backorder prediction
Inventory management
Machine learning algorithms

ABSTRACT

Unpredictable behaviour of customers' demands makes the conventional supply chain structure to be less valuable in predicting demand forecasting which results in back ordered problems to be encountered. To overcome shortages or backorder problem in supply chain, first, the industries must have to identify the parts with the premier probability of deficiency earlier to its incidence to meet a high prospectus to pick up an overall company's performance. In present paper, the various ML algorithms are analyzed and compared in context to identifying the parts with shortages based on previous sales and forecasting. This paper considers backorder predictions using various machine learning algorithms and provides simply comprehensible and feasible backorder decision scenarios.

Selection and peer-review under responsibility of the scientific committee of the Emerging Trends in Materials Science, Technology and Engineering.

1. Introduction

Artificial Intelligence becomes most revolutionary technology since its inception, due to its global acceptance in all major sectors including primary sectors (mining and agriculture), secondary or manufacturing sector, tertiary or service sectors and quaternary sector (education, public, research & development sector). The artificial intelligence has wide range of applications right from aircraft & aviation industries to FMCG industries, even more beyond our imaginations. There are key drivers or enablers such as Big Data, Machine Learning etc., which makes artificial intelligence more lucrative between all contributing sectors of the economy. These enablers are the foundation of artificial intelligence system and play a crucial role in its overall success. Earlier, we have been witnessed such acceptance with the optimization techniques. Hence the combination of artificial intelligence with optimization allows execution of supply chain model with more promising results and efficiency. The supply Chain Management is the linkage between uninterrupted supply of goods and services in terms of resources, organizations, individuals, activities and technologies involved. By effective execution of supply chain philosophy, industries are able to slice down surplus costs and quick product deliveries to the end users, which results to keep out businesses from headlines and away from pricey recalls and lawsuits. However, nowadays, the supply chain managers have seen ever-increasing challenges to make, and remain, competent and successful supply chain methods. There are certain challenges are associated with supply chain models, which are customer service, cost control, planning & risk assessment, supplier/partners relationships and product back orders. In this paper, our aim is to address the 'Product back orders" problem. When, an end user demands a product and due to unavailability instantly or provisionally out of stock scenario, the end user will have to wait until made available for him. These circumstances fall under the "back order" category for specific item [2,3]. Apparently, it can create troublesome for any kind of business in terms of revenue, customers, trust and share market price, if not controlled immediately and smartly. Conversely, immediate action place massive pressure on different echelons of existing SC, which may result in non-value addition activities and logistics expenses [4,5]. Unpredictable behavior of customers' demands makes the conventional supply chain structure to be less valuable in predicting demand forecasting which results in back ordered problems to be encountered [6,7]. Currently, with the advent of Industry 4.0, industries opted prediction of customers' demand by the application of Machine Learning forecasting classifiers algorithm to overcome misleading demand forecasting [8]. Backorder prediction may be viable for non-volatile market demand where

[*] Corresponding author.
E-mail address: lokesh.malviya@technonjr.org (L. Malviya).

https://doi.org/10.1016/j.matpr.2020.11.558
2214-7853/© 2020 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the Emerging Trends in Materials Science, Technology and Engineering.

EOQ, unit cost, lead time and inventory level are the key enablers [9]. On the other hand, abrupt fluctuating demands generates other threat ribbons with existing supply chain models and may leads to failure or loss [10,11]. The multi objective inventory models have suggested solving highly fluctuating demands problems. In present work, we reflect on a factual world dataset available on Kaggle's competition Can You Predict Product Backorders? [1] A We have separated model and test dataset to build model on IBM SPSS Modeler and predicts backorder for available datasets. We have separated data for anticipating appearances in diverse range, and we have conceded it to various algorithms like artificial neural network (ANN), random tree, logistic regression, C5.0, Bayesian network, SVM and discriminant analysis for forecasting. Moreover, the concrete data is fed to those models. However, the results obtained by these models are different from each other and we choose highest degree of accuracy between all algorithms. This paper considers backorder predictions using aforementioned algorithms and provides simply comprehensible and feasible backorder decision scenarios (See Fig. 1 Fig. 2 Table 1 Table 2).

2. Literature review

There are many tools available for demand forecasting in supply chain system, but having less accuracy in predicting demands of the customer. This forces researchers to identify more accurate technique to answer fluctuating demand in existing supply chain management system. Eventually, with the advent of AI System, Machine Learning (ML) methods permits us to estimate comparatively more accurate results in various aspects of the supply chain management system like production, profits, trade, demand and backorder. A several researchers applied various algorithms of Machine Learning techniques and finds different results with different algorithms.

Machine Learning approaches have been extensively used to forecast manufacturers' confused demands.

Babak Abbasi et al. [14] analyzed in his research and shared the promising results of ML models as executive tools in a blood supply chain network. They identify and executed highly cited ML algorithms to respond in day-to-day operational nature challenge to take optimal decisions. They used artificial neural networks (ANN), Classification and Regression Tree (CART), Random Forest algorithm and K-nearest neighbour (k-NN) algorithms. The test data's were successfully demonstrated to get optimized solutions, which can provide can provide well-organized decisions in a network of hospitals. Results in getting optimized solution on few clicks on software to answer fluctuating demands and orders.

George Baryannis et al. [15] proposed a risk assessment framework model for predictions of SCRM (Supply Chain Risks using Machine Learning), which connects the machine learning algorithms with SCRM objectives to formulate intelligibility in excess of forecasting or vice-versa. This work was demonstrated through a physical worlds' dossier of aerospace manufacturing supply chain exaggerated by the peril of postponed deliveries. This framework model has concluded to get better results by analyzing various metrics using machine learning algorithms and black box techniques. This results in favour of intelligibility over forecasting methods.

Real Carbonneau et al. [16] proposed the usefulness of forecasting the unclear claim or demand for the supply chain management through sophisticated non-linear ML techniques to address the situations where supply chain cannot work in partnership. For similar cases, the capability to boost forecasting accurateness will result in lesser costs and superior customer happiness because of punctual deliveries.

Rodrigo Barbosa de Santis et al. [13] analyzed the inventory model of the supply chain and identified few products, which have higher chances of shortages against demand and availability, which results in backorder problem. Santis et al. investigated various machine learning classifiers to suggest a prognostic model for imbalanced rank problem.

The prognostic model based on machine learning algorithms suggests for better inventory control to answer backorder issues. In view of the fact that the matter which cascade on backorder (positive deposit) are unusual compare to matter which doesn't

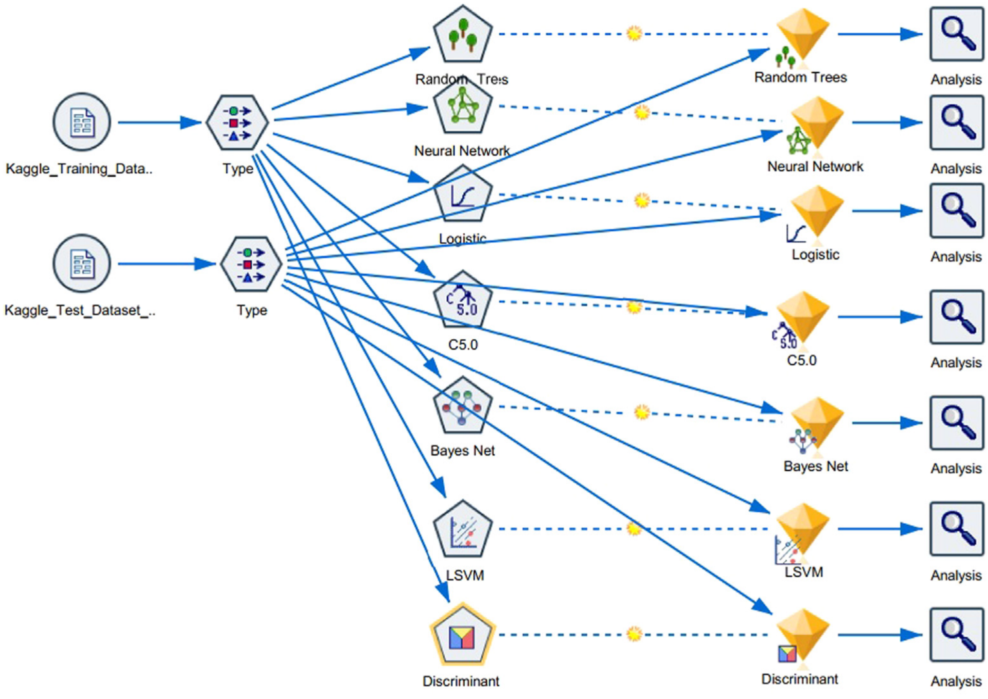

Fig. 1. Models preparation and analysis on IBM SPSS Modeler.

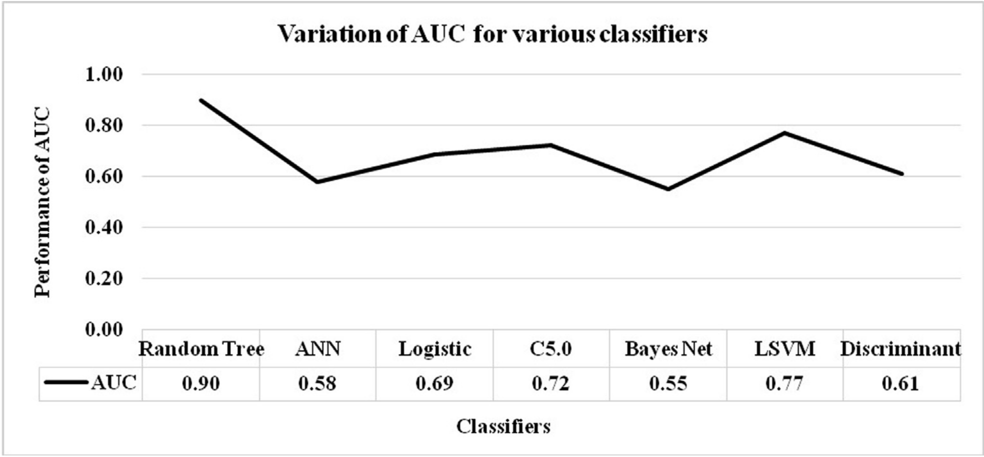

Fig. 2. Graphical representation of AUC score for different classifiers.

Table 1
Comparative study of related papers.

Authors	Forecast realm	ML Classifiers	Result metrics	Decision making conditions	Inventory control
Shin K, Shin Y, Kwon JH, Kang SH [17]	Backorder replacement			Yes	Yes
Islam and Amin [12]	Product backorder	DRF, GBM	Yes	Yes	Yes
Guanghui WANG [19]	Supply Chains' demand	SVR, RBF	Yes		
R B de Santis, E Pestana de Aguiar, E Pestana de Aguiar [13]	Material backorder	CART, LOGIST	Yes		
Prak D, Teunter R [18]	Forecast uncertainty			Yes	yes
RCarbonneau, K Laframboise, R Vahidov [16]	Confused/jumbled order	SVM, NN	Yes		

Table 2
Details of attributes.

SN	Attribute	Explanations
1	sku	Arbitrary product ID
2	national_ inv	Existing stocks in hand of components
3	lead _time	Transit time from source to destination
4	in_ transit_ qty	Pipeline quantity of the stock
5	forecast_3 month	Prediction of sales for the next 3 months
6	forecast_6 month	Prediction of sales for the next 6 months
7	forecast_9 month	Prediction of sales for the next 9 months
8	sales_1 month	Sales quantity for the past 1 month duration
9	sales_3 month	Sales quantity for the past 3 month duration
10	sales_6 month	Sales quantity for the past 6 month duration
11	sales_9 month	Sales quantity for the past 9 month duration
12	min_bank	Buffer stock
13	potential_issue	Uncertain issues attributes (Yes/No)
14	pieces_past_due	Parts overdue from source
15	perf_6 moth_avg	Source average performance in past 6 months
16	perf_12 moth_avg	Source average performance in past 12 months
17	local_co_qty	Amount of stock orders overdue
18	deck_risk	Common risk flags (Yes/No)
19	oe_constraint	Common risk flags (Yes/No)
20	ppap_risk	Common risk flags (Yes/No)
21	stop_auto_buy	Common risk flags (Yes/No)
22	rev_stop	Common risk flags (Yes/No)
23	went_on_backorder	Target Value- Backorder

(negative deposit), a few fastidious ways and metrics are engaged moreover in propose, expansion, and assessment of the tools in the considered imbalanced set problem. The maximum AUC score occurs in GBOOST, whereas BLAG carry out preferably when pleasing into analytically, augmented costs with precision-recall curves.

To assess which ML algorithm is predictable to implement best, we investigate relative studies in relation to classification algorithms and summarize the answer to sustain our conclusion on which models to include in our study.

Based on literature reviews, we choose to start creating the models and test the data using various algorithms including artificial neural network (ANN), random tree, logistic regression, C5.0, Bayesian network, SVM and discriminant analysis classifier models. We consider precision, recalls, F-measure, AUC, gini coefficient and accuracy of the dataset.

3. Methodology

3.1. Dataset

In present work, we reflect on a factual world dataset available on Kaggle's competition. The available dataset is further subdivided into training dataset of 1,687,862 observations and testing dataset of 242,077 observations in the ratio of 87.5:12.5. And the observations in these datasets are having mixed features in nature like integer, decimal values, flags etc. The available dataset is of provided dataset have the precedent data for the eight weeks previous to the week we are annoying to forecast. And these datasets were recorded on weekly basis at the beginning of each and every week. Based on these data, and constraints values of attributes, we will forecast the backorder and accuracy of forecasting. The attributes are given below:

For this work, we plan to use frequent attributes like inventory, sales, buffer stock, lead time, estimated sale, past sale, for predicting backorder scenario for any business.

3.2. Machine learning algorithms

The Machine Learning algorithms are exceptional techniques for creating the models on each dataset competent of performing multifarious tasks such as predictions, improvements in existing

Table 3
Comparative analysis of various classifiers.

	Precision	Recall	F-Measure	AUC	Gini Coefficient	Accuracy
Random Tree	0.997	0.877	0.933	0.90	0.792	87.60%
ANN	0.989	0.939	0.963	0.58	0.159	92.84%
Logistic	0.989	0.990	0.994	0.69	0.377	92.14%
C5.0	0.989	0.990	0.994	0.72	0.442	98.89%
Bayes Net	0.989	0.939	0.963	0.55	0.102	92.70%
LSVM	0.989	0.999	0.994	0.77	0.540	98.89%
Discriminant	0.990	0.712	0.828	0.61	0.223	70.80%

Table 4
Area under the curve for different classifiers.

Classifiers	AUC
Classifiers	AUC
Random Tree	0.90
ANN	0.58
Logistic	0.69
C5.0	0.72
Bayes Net	0.55
LSVM	0.77
Discriminant	0.61

Table 5
Precision, recall and accuracies for various classifiers.

Classifiers	Precision	Recall	Accuracy
Random Tree	0.997	0.877	87.60%
ANN	0.989	0.939	92.84%
ANN	0.989	0.939	92.84%
Logistic	0.989	0.990	92.14%
C5.0	0.989	0.990	98.89%
Bayes Net	0.989	0.939	92.70%
LSVM	0.989	0.999	98.89%
Discriminant	0.990	0.712	70.80%

situations through repeated trials using different algorithms. The ML algorithms create the models and trained throughout the supervised learning process, where variables and attributes are harmonized through repeated trials with keeping originality of dataset and generates the models based on each classifiers or algorithms used in that process.

There are still chances of having few errors, but keeping these limitations of ML, we can generate various models based on algorithms and compare the accuracies on various parameters like AUC, precision, recall, F score, GINI coefficients etc. Then these models were tested on test dataset for identifying the results. In this paper, multiple classifiers were used for predicting the backorder scenario based on available datasets. These included (i) Artificial Neural Network (ANN) (ii) Random Tree (iii) Logistics Regression (iv) C5.0 (v) Bayesian Network (vi) SVM and (vii) Discriminant Analysis. These algorithms are categorized below here.

3.2.1. Artificial neural network (ANN)

An ANN is computational digital edition of human neural network, which follows the functionality of human brain networks of neurons. Hence, an ANN is computing system is deliberated to do simulations of human brain and development into execution of activities. This is an outstanding technique to solve such problems, which would have been challenges for any statistician or machines. An ANN is a multilayer technique of statistical processing. The input layer collects numerous informations from the dataset. And there are many hidden layers, which converts these input data into the information, which an output layer can understand. These layers are completely connected to each other and arranged on weighted manner. Initially ANN is trained through supervised learning process in order to distinguish pattern of different subsets and then it compares repeatedly between original output and modeled output. The difference between outputs is tuned using return path, means output layer to input layer to accommodate weighted manner until it reaches to minimum error situation.

3.2.2. Random tree

Random forest or random tree classifier is an ensemble & supervised learning algorithms. It generates a forest to estimate results in terms of numerous decision trees. And each decision tree come up with results and thus results in many solutions. This algorithm picks nth data parameters from provided dataset and merges them to get précised, accurate and stable solution. This is mostly used classifiers for large dataset as well.

3.2.3. Logistic regression

Logistic regression establishes the relationship between the categorical dependent discrete variables and one or more independent discrete variables by calculating probabilities of binary responses. Based on the probabilities predictions and categorized identification can be done smoothly.

3.2.4. c5.0

This is also a decision tree classifier, which applicable for both discrete and continuous data. The dataset divided into discrete and continuous data and generates decision trees accordingly. This splits data multiple times and suggests prediction by minimizing the error.

3.2.5. Bayesian network

This technique categorized under Probabilistic Graphical Modeling (PGM), which is used to calculate the uncertainties by Directed Acyclic Graphs (DAG). Basically, DAG graph have nodes and links. DAG graphs model the uncertainty of an event depends on the Conditional Probability Distribution (CPD) of each random variable. A Conditional Probability Table (CPT) is used to correspond to the CPD of each variable in the network.

3.2.6. Support vector machine

This algorithm is used to find the decision boundaries between two parameters that are quite far away from any points in the training dataset. It forms a kernel, which transforms data into subsets of desired group depending upon their relationships.

3.2.7. Discriminant analysis

It identifies separate classes or groups among the dataset and figures out best combinations. The main objective is to eliminate the dimension of data and new features will form minimum dispersion of samples between the same groups and maximizes the dispersion between different groups.

Table 5
Precision, recall and accuracies for various classifiers.

Classifiers	Precision	Recall	Accuracy
Random Tree	0.997	0.877	87.60%
ANN	0.989	0.939	92.84%
ANN	0.989	0.939	92.84%
Logistic	0.989	0.990	92.14%
C5.0	0.989	0.990	98.89%
Bayes Net	0.989	0.939	92.70%
LSVM	0.989	0.999	98.89%
Discriminant	0.990	0.712	70.80%

4. Testing & result analysis using IBM SPSS Modeler

There are almost two million of data available for predicting the backorder scenarios for given dataset. We have created the partition for preparing the model first and then testing would be done for the remaining data. So, out of these two million of data, training dataset of 1,687,862 observations and testing dataset of 242,077 observations in the ratio of 87.5:12.5 is separated out. Using these training datasets, we have created various models based on machine learning classifiers and then analysis is being made after testing with testing dataset of 242,077 observations. We have considered various performance evaluation parameters to verify the results.

4.1. Comparative analysis

With the aforementioned dataset, we have prepared the models based on 7 different algorithms and then further analysis is being done using testing dataset. A comparative analysis is being made to evaluate the various classifiers through a systematic performance evaluation process. Table 3 shows the performance of classifiers based on precision, recall, F measure, AUC, Gini coefficient and accuracy.

4.2. Variation of AUC for various classifiers

More the AUC Score value, improved results of the model at characteristic between the positive and negative deposits. Table 4 shows the results of area under the curve, which clearly reflecting the results in favour of random tree classifiers having the maximum area of about 90%.

Random tree is providing the prediction based on AUC score of 0.90, which is quite acceptable and shows the power of the discrimination and can be titled under excellent category. However, the AUC score less than 0.6 have poor discrimination for the Baye-

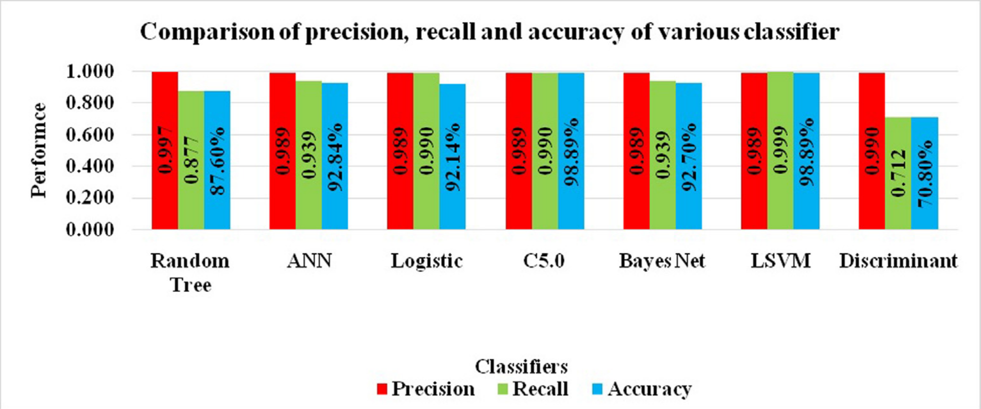

Fig. 3. Comparison of precision, recall and accuracy of various classifiers.

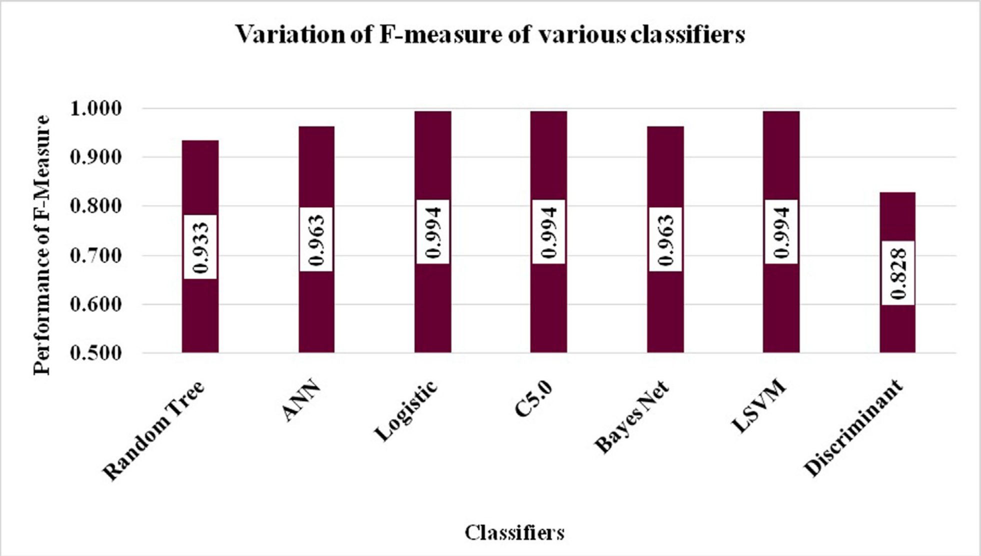

Fig. 4. Variation of F-measure of various classifiers.

sian network and artificial neural network. However, the accuracy of random tree classifier is found to be of 87.6%, but more accuracies (about 92.7%) have been achieved using Bayesian network and artificial neural network classifiers.

4.3. Comparison of precision, recall and accuracy of various classifiers

Degree of correct predictions can be made using Precision, Recall and Accuracy performance parameters. Table 5 represents the probabilities of predictions accurately for various classifiers. The precision results show amount of correct credentials returned by machine learning model after successfully execution of testing dataset. Eventually getting the results in the order of 0.989 to 0.997, this is highly précised. Similarly, the accuracy is the amount of correct predictions made by machine learning model. The best predictions made by C5.0 and SVM classifiers. Hence these classifiers are given priority based on maximum accuracy in predicting the product backorder Table 6.

Fig. 3 explains the comparative analysis for various classifiers numerically on precision, recall and accuracy. The below graphical representation is self sufficient to discriminate true positive values or correct predictions. The rank is provided for maximum accuracy and précised values for various classifiers and identifies that C5.0 and LSVM gives the best predictions among all (See Fig. 4).

4.4. F score values for different classifiers

It determines the model's accuracy based on dataset. The calculation of F Score is basically a pathway to merge the precision and recall of the model, which may be considered as harmonic mean of the model in terms of precision and recall. The following table shows the variations in F- score for different classifiers.

The graphical representation gives the discrimination values among all applied classifiers. Conversely, the classifiers, those have maximum precision and recall, also have the maximum F-score as well. The below graph confirms the maximum F score values for C5.0 and LSVM classifiers along with logistic regression classifier.

5. Conclusions

The presented paper shows backorder predictions using various machine learning classifiers in order to effective control of inventory management. Results show that backorder prediction is a rare incidence compared to the parts which does not belongs to backorder case. In current paper, seven different machine learning classifiers have been applied and concluded that there are always few possibilities exist for backorders or shortages. And varieties of factors involved with shortages. Hence the predictions using machine learning classifiers provides an edge to the industries, to make suitable changes for particular part, in order to being competent.

CRediT authorship contribution statement

Lokesh Malviya: Conceptualization, Methodology, Software, Formal analysis, Validation. **Pankaj Chittora:** Data curation. **Prasun Chakrabarti:** Resources, Formal analysis. **Raj Shekhar Vyas:** Supervision. **Sandeep Poddar:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] https://www.kaggle.com/tiredgeek/predict-bo-trial.

[2] K.B. Clark, T. Fujimoto. Product development performance: strategy, organization, and management in the world auto industry. 1991.

[3] L. Guo, Y. Wang, D. Kong, Z. Zhang, Y. Yang, Decisions on spare parts allocation for repairable isolated system with dependent backorders, Comput. Ind. Eng. 127 (2019) 8–20.

[4] C.R. Carter, D.S. Rogers, A framework of sustainable supply chain management: moving toward new theory, Int. J. Phys. Distrib. Logistics Manag. 38 (5) (2008) 360–387.

[5] F. Mohebalizadehgashti, H. Zolfagharinia, S.H. Amin, Designing a green meat supply chain network: a multi-objective approach, Int. J. Prod. Econ. 219 (2020) 312–327.

[6] D. Simchi-Levi, P. Kaminsky, E. Simchi-Levi, R. Shankar, Designing and Managing the Supply Chain: Concepts, Strategies and Case Studies, Tata McGraw-Hill Education, New York, 2008.

[7] L. Yu, Y. Duan, T. Fan, Innovation performance of new products in China's high-technology industry, Int. J. Prod. Econ. 219 (2020) 204–215.

[8] A. Mitra, Fundamentals of Quality Control and Improvement, Wiley, New York, 2016.

[9] J.A. Rodger, Application of a fuzzy feasibility Bayesian probabilistic estimation of supply chain backorder aging, unfilled backorders, and customer wait time using stochastic simulation with Markov blankets, Expert Syst. Appl. 41 (16) (2014) 7005–7222.

[10] M.P. De Brito, V. Carbone, C.M. Blanquart, Towards a sustainable fashion retail supply chain in Europe: organisation and performance, Int. J. Prod. Econ. 114 (2) (2008) 534–553.

[11] B.M. Tosarkani, S.H. Amin, An environmental optimization model to configure a hybrid forward and reverse supply chain network under uncertainty, Comput. Chem. Eng. 121 (2019) 540–555.

[12] Islam Amin, Jornals of Big Data: Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and Gradient Boosting Machine learning techniques, (2020) 7:65 https://doi.org/10.1186/s40537-020-00345-2.

[13] Rodrigo Barbosa de Santis, Eduardo Pestana de Aguiar, Eduardo Pestana de Aguiar: Predicting Material Backorders in Inventory Management using Machine Learning, 978-1-5386-3734-0/17/$31.00 c 2017 IEEE.

[14] Babak Abbasi, Toktam Babaei, Zahra Hosseinifard, Kate Smith-Miles, Maryam Dehghani, Predicting solutions of large-scale optimization problems via machine learning: A case study in blood supply chain management- https://doi.org/10.1016/j.cor.2020.104941 0305-0548 /© 2020 Elsevier Ltd.

[15] George Baryannis, Samir Dani, Grigoris Antoniou, Predicting supply chain risks using machine learning: The trade-off between performance and interpretability- https://doi.org/10.1016/j.future.2019.07.0590167-739 X /© 2019 Elsevier B.V.

[16] Real Carbonneau, Kevin Laframboise, Rustam Vahidov. Application of machine learning techniques for supply chain demand forecasting -0377-2217/$ - see front matter _ 2007 Elsevier B.V. doi: 10.1016/j.ejor.2006.12.004.

[17] K. Shin, Y. Shin, J.H. Kwon, S.H. Kang, Development of risk based dynamic backorder replenishment planning framework using Bayesian Belief Network, Comput. Ind. Eng. 62 (3) (2012) 716–725.

[18] D. Prak, R. Teunter, A general method for addressing forecasting uncertainty in inventory models, Int. J. Forecast. 35 (1) (2019) 224–238.

[19] W.A.N.G. Guanghui, Demand forecasting of supply chain based on support vector regression method, Procedia Eng. 29 (2012) 280–284.

Further Reading

[1] H. He, E. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (9) (2009) 1263–1284, https://doi.org/10.1109/TKDE.2008.239.