International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 08 | August -2020 ISSN: 2582-3930

Decision Support System for Supplier Selection using Machine Learning Algorithms and Ensemble Techniques

Ruchita Bhojane [1], Prof. Kavita Kelkar [2]

[1] *Student, Department of Computer Engineering, K.J. Somaiya College of Engineering, Vidyavihar, Mumbai, India*
[2] *Department of Computer Engineering, K.J. Somaiya College of Engineering, Vidyavihar, Mumbai, India*

---***---

Abstract - It is widely believed that supplier selection decision is the key challenge in the supply chain management, while looking for corporate competitive advantage. Decision making is often based on the experience and insight of managers instead of the analysis and interpretation of data and information, which with today's IT systems can be made readily available. Also considering multi-criteria for decision-making for supplier selection is the need rather than just depending on price factor. In this paper, we have proposed a decision support system which facilitates the manager to select the best supplier which will have a great influence on company's performance by using machine learning and ensemble techniques. Different ML classification algorithms such as Naive Bayes, Decision Tree, XGB and ensemble techniques are discussed and implemented. A comparative analysis has been conducted of the selected algorithms and the results suggest that these models can be used for the decision making. Aim of this paper is to select the best supplier based on multiple criteria's by using ML models on historical supplier evaluation data with multi-classification objective. This study contributes to the development of automated decision support system which is fast, reliable, flexible, efficient and effective for the best supplier selection.

Key Words*: *Decision support system, Supplier Selection, Multi-criteria decision making, Machine learning, Ensemble Methods, Supplier Evaluation.

1. INTRODUCTION

Supply chain management is the process of management of the flow of services and goods and includes all processes that transform raw materials into final products. Goal of supply chain management is to maximize company's profit and gain a competitive advantage in the marketplace. To develop and run supply chains in the most effective and efficient ways possible, a lot of efforts by the supply chain firms are required. Supply chain activities involves everything from product development, sourcing, production, logistics and also the information systems that are required to coordinate all these activities.

To be competitive in the market along with maximum performance of organization, it is very crucial for the supply chain management to be effective and efficient. It is very important to stay competitive because the competition among the organizations is increasing day by day and is affected by supply chain management. One of the important process in SCM is Sourcing. Sourcing involves all the activities within the procurement process which deals with identification and evaluation of potential suppliers, engaging with selected suppliers and selecting the best supplier which will bring value to the organization.

In Sourcing and especially in supply chain management Supplier selection decision is the key challenge. Supplier selection process is identification of the suitable or best value supplier out of a number of potential suppliers. Supplier selection in a broader way can be viewed as a process in which potential suppliers are evaluated and one of the suppliers is selected. The selection process requires to evaluate suppliers based on various supplier-specific data, e.g. product prices, quality, delivery, reliability, etc.

Because the trend of long-lasting and win-win relationships within supply chains is becoming more and more important, supplier selection should not limit on cost only and should consider many other evaluation factors. The success of supplier selection impact effective SCM practices. To maintain a competitive position in the global market, organizations have to follow good and effective strategies and practices and hence selecting the right suppliers is a vital component of these new strategies. Supplier selection is one of the most crucial components of purchasing and supply chain management for many companies.

Currently, decision making is based on experience and insight instead of the analysis and interpretation of data and information, which with today's IT systems can be made readily available. Availability to explore various options in selecting suitable supplier that best fulfill the requirements in a competitive environment will be beneficial and easy for Supply chain managers. Also considering multiple criteria for supplier selection is the need rather than just depending on price factor. Additionally, selecting the best supplier has a great impact on a company's performance.

This problem motivated us to develop a decision support system for ease of use to logistics managers for supplier selection process. For taking decision, a comprehensive and system driven process view will enable the business users or decision makers to handle his day-to-day operations in smooth and cost-effective manner.

In this paper, we have proposed a decision support system (DSS) which facilitates the business user or manager to select the best supplier which will have a great influence on company's performance by using machine learning and ensemble techniques. The DSS is developed using different machine learning classification algorithms such as Naive Bayes, Decision Tree, XGB, and ensemble classification models.

Aim of this paper is to select the best supplier based on multiple criteria's by using Machine Learning models on historical supplier evaluation data with multi-classification objective, so that algorithms can learn the supplier ranking classification.

A comparative analysis has been conducted of the selected algorithms and the results or outcome of the research suggests that these models can be employed for the decision support system by business users.

2. RELATED WORK

Selection of supplier is the crucial and most challenging part of supply chain management as its performance is largely dependent on effective supplier selection process and incorrect choice can lead the supply chain as a whole to suffer losses and hence this would directly affect the performance of the organizations involved.

Several previous studies using decision-making models such as data envelopment analysis (DEA), analytic hierarchy process (AHP), analytic network process (ANP), simple multi-attribute rating technique (SMART) and fuzzy technique has been carried out to perform supplier evaluation and analysis. However, decision making related to supplier selection lacks the application of machine learning/ data mining methods. Machine learning techniques warrants an automated decision support system wherein machine learning models can be trained on historical data of supplier evaluation to predict the best supplier.

This paper solves the research question of how we can predict suppliers based on historical data considering multi-criteria by using ML methods. Our aim is to predict supplier by applying classification models on the historical data. A development of automated decision support system which is reliable and efficient for the supplier prediction is much needed.

To utilize the available information for the decision making is often complex and laden with many difficulties. Supplier selection in fashion industry is highly critical and complex process as it entails multiple qualitative and quantitative criteria, and the participation of many managers making decisions in the supply chain management of companies. To predict supplier for customers specific order in the future, authors of [1] aims to train data mining models on historical customer order data that includes information of products customized features.

[1] used historical customer order data with labels, meaning that customer order attributes were tagged with the label of suppliers who fulfilled these orders; therefore, supervised machine learning approach has been adopted in this study. Classification models was selected to predict the suitable supplier by identifying and including appropriate predictors in the dataset. From Implementation and results it shows that the models are overfitting to the trained data and can be ridden with biasedness.

To address this problem, k-fold cross validation on trained data with k=10 was applied. It was observed that model accuracy is reduced after applying cross validation on trained data and it implies that models are not overfitting to the data. After applying KNN, RF, NN, NB classification models on the customer order data to predict suppliers, it was found that the performance of RF and NN models is better than KNN and NB models.

In paper [2] least Square method is used to predict the stock needs of goods and C4.5 algorithms to provide supplier selection solution. Accuracy of 60% was achieved when C4.5 algorithm was tested using cross validation also it showed good accuracy for decision support settlement. Unstructured historical data of 23-month period was processed into structured data in accordance with the needs of further modeling process. The transforming stage was needed to classify the values of the variables required for the modeling process using the C4.5 algorithm. The 4 criteria considered were: Sales Trend, Shipping Estimation, Type of Delivery and Product Quality. The shape of the decision tree had 3 nodes that were formed namely Sales Trend, Product Quality and Delivery Estimation. The root of this tree was Sales Trend. Then if the value of Sales Trend attribute was medium, then the attribute that was seen next was Product Quality. Whereas if the value of Sales Trend attribute was less, then the attribute to be seen next was Estimated Delivery. From the classification result of the decision tree, it was used to assist in the determination of supplier selection with accuracy of 60%.

In [3] supplier decision making criteria used are price, delivery order, delivery speed, stock availability, quality of service and purchase method. Research data consisted of 6 goods with 13 suppliers. The best supplier from many suppliers of one good was selected using combined Intuitionistic Fuzzy Topsis and K-Means for help clustering data with number value. The result of this research has accuracy level of 64% from clustering.

The literature and studies on supplier selection criteria and methods are full of various analytical approaches. Hybrid models has been developed by some researchers by combining more than one type of selection methods. An overall summary of research on supply chain management, supplier selection criteria and supplier selection evaluation methods for multi-criteria decision making is given by authors of [4]. In order to improve success and competitiveness, summary of the supplier selection process will be helpful for companies to have a clear understanding of the concept resulting to improve their success and competitiveness. The results of the research show that the application of a structured decision-making technique is vital, especially when considering the complex conditions that include both qualitative and quantitative criteria.

In [5] authors have developed a model for steel pipe and steel sheet supplier's evaluation by applying Analytic Hierarchy Process (AHP). The aim of the models is to maximize overall part makers' satisfaction.

The first part of research is the evaluation of raw material suppliers using AHP and the importance of decision criteria complying with part makers' satisfaction are weighted. A decision model for supplier selection using integer programming is proposed in the second part. Suitable suppliers are selected which will maximize overall company's satisfaction considering the weight of each criteria and raw material consumption. The results from the first part of this study shows that the most important criteria are cost, quality, delivery, service. The proposed model helps to increase satisfactions between car seat makers and their suppliers.

Hence, in this project the aim is to select the best supplier based on multiple criteria's by using Machine Learning models on historical supplier evaluation data with multi-classification objective, so that algorithms can learn the supplier ranking classification.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 08 | August -2020 ISSN: 2582-3930

3. METHODOLOGY

In this paper, a two- stage approach is proposed to build a decision support system for supplier selection Architecture of proposed methodology is shown in Fig. 1.

In stage 1: Data Preprocessing, Feature Engineering and target class (final rating of supplier) creation or data classification in 1 to 5 class is done.

In stage 2: Model Building using Implementation of Machine Learning and Ensemble Methods Techniques for Multi-Classification.

A User View is present where User will upload list of suppliers as input and output in terms of best selected supplier is displayed on GUI.

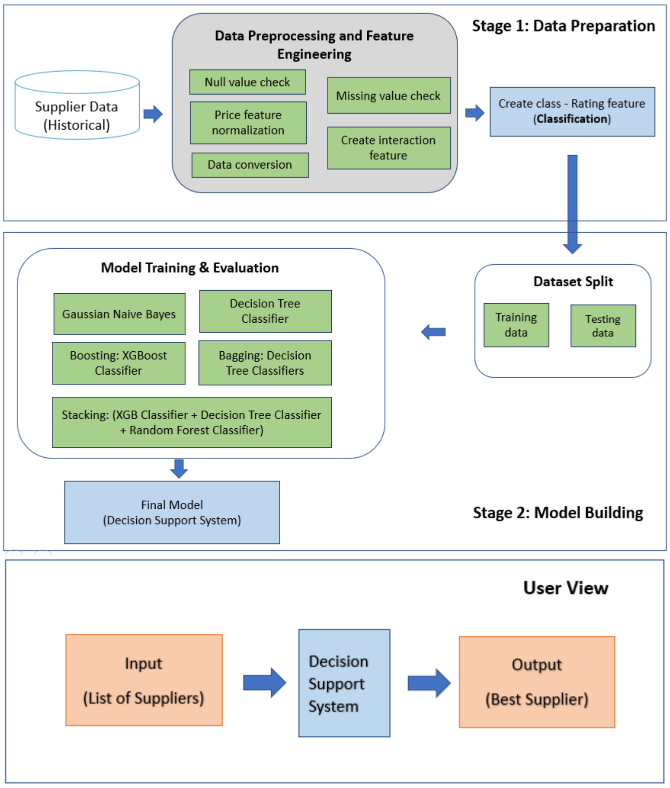


Fig. 1: Architecture of proposed methodology

Stage 1: Dataset Preparation

A. Dataset

Considering the inputs from industry experts and literature reviews 13 features are used to perform supplier evaluation. The input dataset consists of historical supplier's evaluation data. Initially the data size is 8982 records X 13 attributes.

In dataset, 25 supplier's evaluation is done for 6 Products (Product A, Product B, ...Product F) based on most significant supplier performance measurements /features such as price, quality, flexibility, reliability, support etc.

Attributes used in this study along with its possible values are described in Table 1.

Sample Dataset is shown in Fig. 2

For Rating Factor Scale (1-5), 5 is considered as best rating and 1 as poor.

B. Data Preprocessing and Feature Engineering:

The supplier's data has been taken for preprocessing step.

Step 1: **Not null value check:** Not null values are checked and if present are removed from dataset.

Step 2: **Missing value check:** Missing values are checked and if present are removed from dataset.

Table -1: Description of dataset (Attributes List)

Feature	Description	Datatype	Possible Values
supplier_id	Unique supplier id of supplier.	String	Supplier 01, ..., Supplier 25
location	Location of supplier.	String	Mumbai, Pune, etc.
product	Product supplied by supplier.	String	Product A, Product B, . Product F
price	Total price of product. Price criteria includes unit price, pricing terms, exchange rates, taxes and discounts.	Float	e.g. 5500.00
estimated_date	Estimated date of product delivery.	Date	10/4/2018
delivered_date	Actual date of product delivery.	Date	10/11/2018
deliverly_days	Total delivery days to deliver product.	Integer	e.g. 5
quality	Rating of Quality of product.	Integer	1,2,3,4,5
flexibility	Rating of supplier's flexibility. Acceptability of supplier regarding changes of orders.	Integer	1,2,3,4,5
reliability	Rating of supplier's reliability. Supplier's quality of being trustworthy and dependable.	Integer	1,2,3,4,5
support	Rating of support post-delivery. Support provided by supplier for the product delivered.	String Categorial Data	Poor Not sufficient Satisfactory Good Excellent
warranty	Rating of warranty policies post-delivery. Ability to return or replace product.	Integer	1,2,3,4,5
sustainability	Rating of supplier's sustainability.	Integer	1,2,3,4,5

supplier_id	location	produ	price	estimated_date	delivered_date	delivery_days	quality	flexibility	reliability	support	warranty	sustainability
S25	Navi Mumbai	Product C	5,374.10	10/4/2018	10/11/2018	10	3	3	3	Not sufficient	1	1
S01	Mumbai	Product F	3,093.00	4/16/2017	4/18/2017	5	4	4	5	Good	5	5
S04	Navi Mumbai	Product E	5,273.00	6/29/2018	7/5/2018	10	1	2	1	Satisfactory	2	1
S13	Navi Mumbai	Product D	4,178.00	7/18/2018	7/21/2018	7	4	2	3	Satisfactory	5	5
S01	Mumbai	Product C	5,081.00	9/23/2018	9/24/2018	5	4	4	4	Excellent	4	4
S20	Nagpur	Product B	3,825.00	3/15/2018	3/18/2018	9	4	4	5	Satisfactory	4	3
S16	Satara	Product B	3,390.00	11/20/2019	11/21/2019	6	3	2	4	Good	3	4
S04	Navi Mumbai	Product F	3,086.00	7/1/2018	7/4/2018	6	4	4	4	Not sufficient	2	4
S02	Navi Mumbai	Product E	5,448.00	2/9/2019	2/13/2019	8	4	4	4	Poor	3	1
S07	Pune	Product E	5,499.00	9/5/2019	9/8/2019	7	4	4	4	Good	5	4
S21	Pune	Product C	5,336.00	9/6/2018	9/7/2018	6	4	4	5	Good	5	4
S24	Navi Mumbai	Product B	3,725.00	9/27/2017	9/29/2017	10	4	4	5	Good	4	5
S20	Nagpur	Product F	3,274.00	9/27/2019	9/30/2019	8	3	1	1	Satisfactory	3	3
S19	Mumbai	Product C	5,521.00	9/25/2017	9/27/2017	8	3	2	2	Not sufficient	2	3
S05	Mumbai	Product C	5,460.00	7/1/2018	7/3/2018	6	2	2	4	Satisfactory	2	4
S13	Navi Mumbai	Product B	3,793.00	9/16/2018	9/17/2018	7	4	4	4	Satisfactory	4	4

Fig. 2: Sample Dataset

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 08 | August -2020 ISSN: 2582-3930

Step 3: **Create interaction feature:**

- delay_days: Total delayed days in delivery after estimated delivery date.
  Delay days is derived from delivered_date and estimated_date difference in dataset.

- ontime_delivery
  ontime_delivery rating (1-5) is derived from delay_days.
  e.g. If delay_days = 0, ontime_delivery=5

Step 4: **Normalization using Decimal scaling.**

- price
  price feature is normalized using decimal scaling

Step 5: **Categorical data conversion to numerical type Ordinal Encoding technique**

- support

support feature has categorical data (Poor, Not sufficient, Satisfactory, Good, Excellent) which is converted to numerical data (1,2,3,4,5)

C. Create class (Classification)

Rating feature (Final rating of supplier) is calculated and used as class for training algorithms

- rating: Final rating of supplier
  Each instance in the data model is classified, on a rank of 1 to 5, based on the supplier performance measurements features.

Score is calculated by multiplying the weights and the feature values and based on score classification into class 1-5 is done.

Weights of features considered are:

wt_price = 0.1, wt_ontime_delivery = 0.13, wt_quality = 0.24, wt_flexibility = 0.20, wt_reliability = 0.11, wt_support = 0.06, wt_warranty = 0.07, wt_sustainability = 0.09

After Feature Engineering:
Summary of data is shown in Fig. 3.

```
----------data info----------
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8982 entries, 0 to 8981
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   supplier_id      8982 non-null   object
 1   location         8982 non-null   object
 2   product          8982 non-null   object
 3   price            8982 non-null   float64
 4   estimated_date   8982 non-null   datetime64[ns]
 5   delivered_date   8982 non-null   datetime64[ns]
 6   delivery_days    8982 non-null   int64
 7   quality          8982 non-null   int64
 8   flexibility      8982 non-null   int64
 9   reliability      8982 non-null   int64
 10  support          8982 non-null   int64
 11  warranty         8982 non-null   int64
 12  sustainability   8982 non-null   int64
 13  delay_days       8982 non-null   int64
 14  ontime_delivery  8982 non-null   int32
 15  price_norm       8982 non-null   float64
 16  weight_score     8982 non-null   float64
 17  rating           8982 non-null   int32
dtypes: datetime64[ns](2), float64(3), int32(2), int64(8), object(3)
memory usage: 1.2+ MB
```

Fig. 3: Summary of data

Unique values of features in dataset to train the models, shown in Fig. 4.

```
Unique suppliers in the dataset :  ['S01' 'S02' 'S03' 'S04' 'S05' 'S06' 'S07' 'S08' 'S09' 'S10' 'S11' 'S12'
 'S13' 'S14' 'S15' 'S16' 'S17' 'S18' 'S19' 'S20' 'S21' 'S22' 'S23' 'S24'
 'S25']

Unique locations in the dataset :  ['Navi Mumbai' 'Mumbai' 'Nagpur' 'Satara' 'Pune']

Unique products in the dataset : ['Product A' 'Product B' 'Product C' 'Product D' 'Product E' 'Product F']

Unique quality values in the dataset : [1 2 3 4 5]

Unique flexibility values in the dataset : [1 2 3 4 5]

Unique reliability values in the dataset : [1 2 3 4 5]

Unique support values in the dataset : ['Excellent' 'Good' 'Not sufficient' 'Poor' 'Satisfactory']

Unique warranty values in the dataset : [1 2 3 4 5]

Unique sustainability values in the dataset : [1 2 3 4 5]

Unique rating values : [1 2 3 4 5]
```

Fig. 4: Unique values of features in dataset

Histogram of features in dataset is shown in Fig. 5.

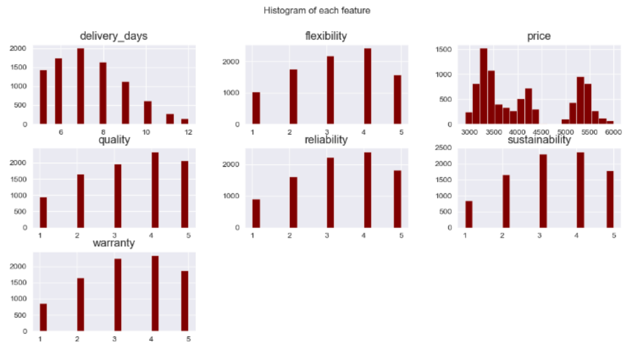


Fig. 5: Histogram of features

Visualization plot of all the suppliers count in the dataset is shown in Fig. 6.

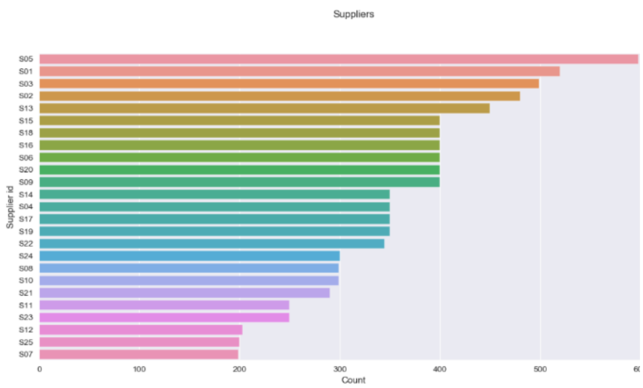


Fig. 6: Suppliers Count plot

Stage 2: Model Building using Implementation of Machine Learning and Ensemble Methods Techniques for Multi-Classification.

The data is split into 70 percent into the training set (n=6287) and 30 percent into test set (2695). Next, k-fold cross validation was applied on trained data and the default value for k is set to 10. Each classification model is then applied on training dataset and prediction is done on test dataset. Prediction accuracy and the model performances were evaluated according to the common metrics.

Machine Learning (ML) is a highly relevant technique to find innovative ways to use and understand their data in order to get a better overview of their business. With a constant flow of new data, the ML models can make sure that the models are up to date.ML utilize many different algorithms that repetitively learn from data to improve and describe data while ultimately predicting possible outcomes. The outcomes are also called outputs and comes from the ML model, which is generated when the algorithm is trained by data. When the model is trained, it can be given certain input and then provide the mentioned output in return.

Dataset consists of historical supplier's evaluation data with class, indicating suppliers final rating (1 to 5); therefore, supervised machine learning approach has been adopted. Since the supplier's ratings are unique labels in the dataset, classification models were used to predict the suitable supplier by appropriate predictors in the dataset. Classification models then perform the task of drawing conclusions by observing data points, and subsequently predict the label value for each label in the dataset.

Model training is done using following methods:

1. Gaussian Naive Bayes

Naive Bayes is fast classification algorithm for binary and multi-class classification, which is suitable for a large volume of data. Bayes theorem of probability is used to predict the unknown class with the assumption of independence among predictors. Gaussian Naive Bayes is the extension on Naive Bayes algorithm and follows Gaussian normal distribution.

2. Decision Tree Classifier

Decision Tree Classifier is a classification algorithm for binary and multi-class classification A decision tree can be described as a flowchart-like tree structure in which each leaf node represents the outcome, the internal node represents feature and the branch represents a decision rule. The topmost node of the decision tree is known as the root node. Tree learns to do recursive partition on the basis of the attribute value. This tree structure helps in decision making as its it's visualization tree diagram easily mimics the human level thinking

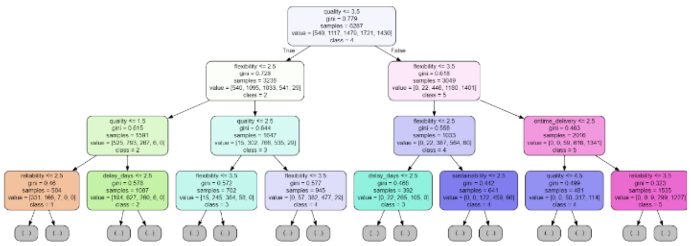


Fig. 7: Decision tree plot for depth=3

Fig. 7, illustrates the shape of the decision tree for the supplier selection case.
There are 3 nodes that are formed namely Sales Trend, Product Quality and Delivery Estimation.
The root of this tree is the quality feature. Then if the value of quality attribute is <=3.5, then the attribute that will be seen next is flexibility.
Similarly, the tree is read and the accuracy of tree obtained is 84.6 %.

Ensemble Methods:

Ensemble methods are learning technique wherein the predictions of multiple models are combined to acquire better prediction accuracy than any individual model.

3. Boosting: XGBoost Classifier

Boosting is a sequential ensemble technique. It improves prediction accuracy by combining a set of weak learners.

XGBoost means Extreme Gradient Boosting and this algorithm is used for supervised learning problems by implementing gradient boosting decision tree algorithm. In Gradient boosting new models are created that predict the errors of previous models and then added together to make the final prediction. It uses a gradient descent algorithm in order to minimize the loss when adding new models. The two reasons to use XGBoost are also the two goals of this project: 1. Execution Speed. 2. Model Performance

4. Bagging: Decision Tree Classifiers

Bagging is a parallel ensemble technique where the base learners are generated in parallel and intends to improve the strength and accuracy of ML algorithms used in classification and regression purpose. It additionally minimizes fluctuation of data (variance) and prevent from over-fitting. Several subsets of data are created from training sample which are chosen randomly with replacement.

The decision trees are then trained on each subset of data collected and as a result ensemble of different models are formed. The individual decision trees are grown deep without pruning and These trees will have both low bias and high variance. Predictions from different trees are averaged are used which is more robust than depending on single decision tree.

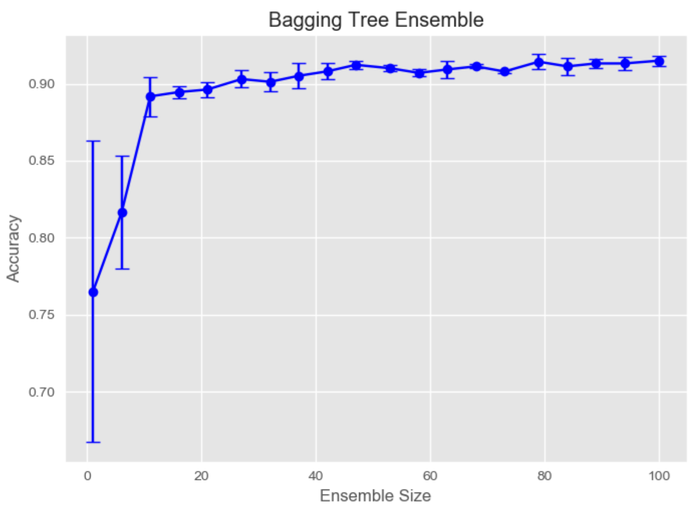


Fig. 8: Bagging Tree Ensemble plot of Accuracy vs Ensemble Size

The Bagging Tree Ensemble accuracy keeps on increasing as the ensemble size is increasing as seen in Fig. 8. We can see the improvement of performance when Bagging Ensemble

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 08 | August -2020 ISSN: 2582-3930

Techniques was performed on Decision Tree compared to single Decision Tree Classifier.

5. Stacking: (XGB Classifier + Decision Tree Classifier + Random Forest Classifier)

Stacking is an ensemble learning technique in which multiple models are combined via a meta-classifier which can be any classifier the predictions of multiple classifiers get stacked and are further used as new features to train a meta-classifier.

Unlike boosting, stacking is not normally used to combine models of the same type for example a set of decision trees, rather it is applied to models built by different learning algorithms. Stacking tries to learn which classifier are the reliable ones, by using another ML algorithm i.e. The meta-learner to find out how best to combine the output of the base learners.

We have implemented a stacking model which allows to train XGB classifier, Random Forest with all of the training data, then those outputs are merged with a Decision Tree. The idea is to improve accuracy and reduce overfitting

Decision Support System for Supplier Selection Prototype

The GUI shown in Fig. 9 and Fig. 11 is used for supplier selection process. With the help of file picker attach the input dataset containing the suppliers to be ranked or predicting the best supplier and click submit.

Input File attachment GUI shown in Fig. 9 and sample input format shown in Fig. 10.

For deciding the best suppliers from the input suppliers for particular Product, input provided is trained for the supplier's particular product records. The suppliers with the highest rank are displayed and sorted w.r.t input price.

Output GUI with the best supplier is shown in Fig. 11

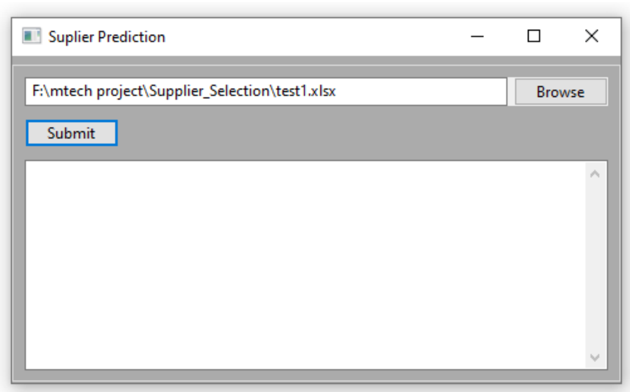
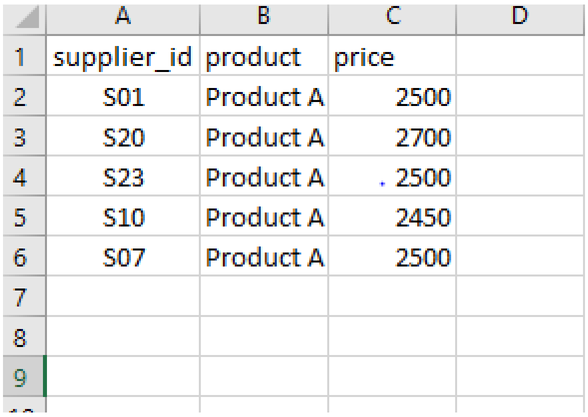


Fig. 9: Input for Supplier Selection



	A	B	C	D
1	supplier_id	product	price	
2	S01	Product A	2500	
3	S20	Product A	2700	
4	S23	Product A	2500	
5	S10	Product A	2450	
6	S07	Product A	2500	
7				
8				
9				

Fig. 10: Sample test data for input

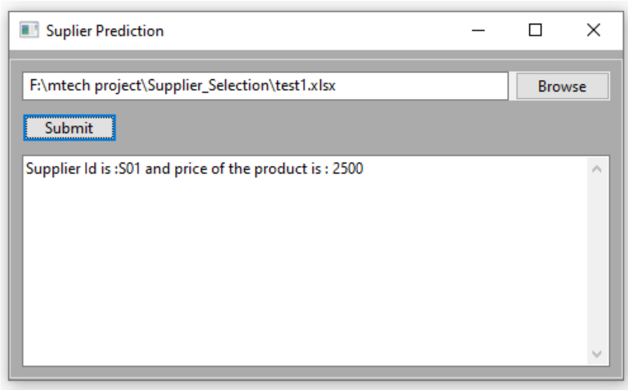


Fig. 11: Output of Supplier Selection

4. RESULTS

Classification models to predict supplier are applied on trained dataset (70% of the original dataset); k fold crossed validation data (k = 10) and finally on test dataset (30% of the original dataset).

It is observed that, on the trained dataset model performance of XGBoost, Stacking, Decision tree and Bagging gives 100% accuracy results, while Gaussian Naïve Bayes model gives 86% accuracy.

The accuracy values of classification models applied on test dataset can be seen in Table 2 and their performances are compared in Fig. 12.

Comparison of classifiers

Comparison of implemented classifiers with the evaluation metrics is shown in Table 2.

Table -2: Comparison of Classifiers

Metrics	Classifiers				
	Gaussian	XGB	Stacking	Decision tree	Bagging
Accuracy	85.45454545	93.20964749	92.83858998	84.60111317	89.313543
roc_auc_score	0.979941518	0.996282712	0.955241187	0.903756957	0.988880
Precision Score	0.906975298	0.985742965	0.881930696	0.748810738	0.953370
Cross-validation Score	0.864798	0.932560	0.928742	0.856209	0.889454
Training Time	0.015625	12.265625	14.453125	0.0625	0.265625

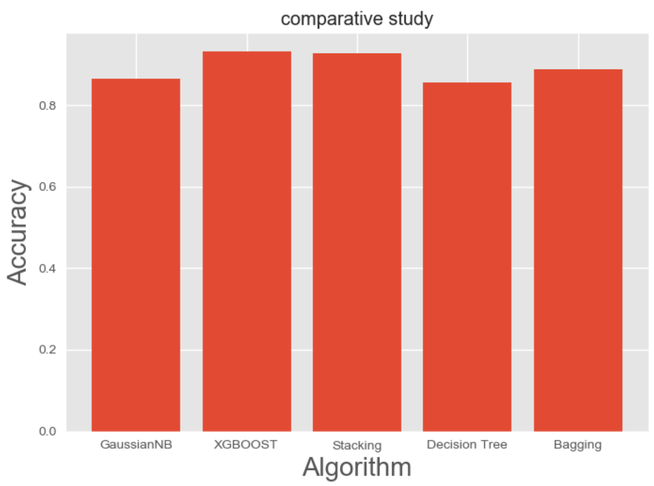


Fig. 12: Accuracy Comparison

According to evaluation metrics, the ranking of models based on performance (higher to lower)

1. XGBoost
2. Stacking
3. Bagging
4. Gaussian Naive Bayes
5. Decision Tree

The test results of implemented techniques suggest that performance of XGBoost and Stacking is consistent and are appropriate methods to perform complex classification tasks leading to successful classification of suppliers.

As XGBoost performed best, XGBoost model was selected for decision support system for supplier selection process.

The ROC curve is the plot between True Positive rate (y-axis) and False Positive rate (x-axis). The ideal score is considered as TPR = 1 and FPR = 0, which corresponds to the point on the top left. Typically, we calculate the area under the ROC curve, and greater the AUC-ROC the better.

From the Fig. 13 we can see that XGBoost has the best AUC-ROC curve compared to other classifiers.

5. CONCLUSION AND FUTURE WORK

The present paper proposes a decision support system for supplier selection using supplier's historical performance evaluation data. Two stage approach was followed for the DSS. Data Preprocessing is carried out along with cross validation to improve the performance of algorithm for Decision Support System. The main motive behind this is to feed the overall pattern of the data to the model.

In this paper five Algorithms (Gaussian Naïve Bayes, XGBoost, Stacking, Decision Tree, Bagging) were implemented and evaluated with various Model Evaluation Metrics such as

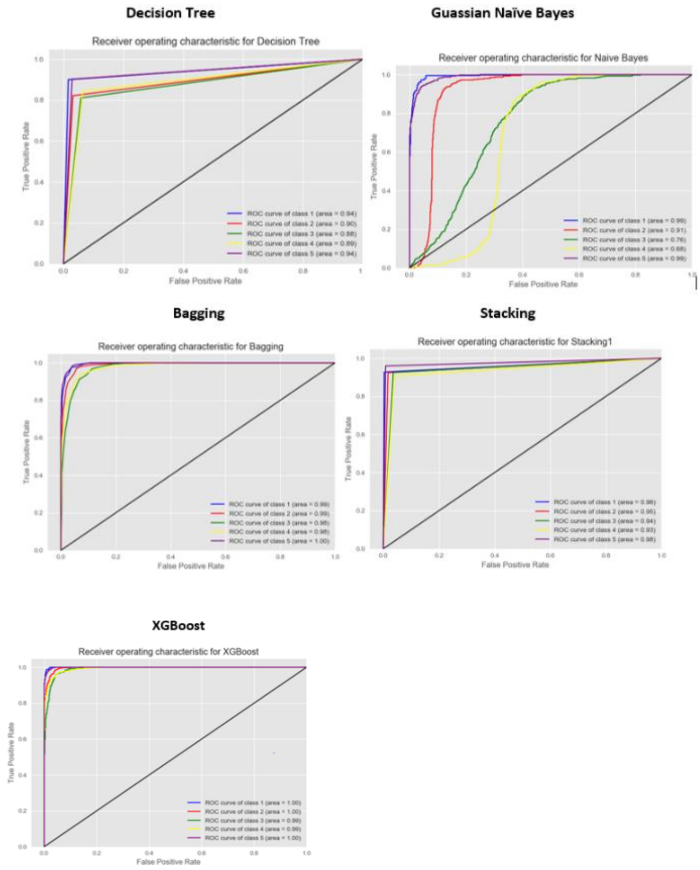


Fig. 13: ROC Curve plots of all classifiers

Learning Curve, Confusion Matrix, Model Accuracy, roc_auc_score, Precision Score, Classification Report, F1 score, Support, ROC Curve, Cross-validation score. It is observed that the performance of XGBoost and Stacking is consistent. After Model Evaluation XGBoost Classifier performed the best and gave the best accuracy compared to other algorithms. The ROC and AUC of XGBoost was best then other classifiers.

Also, we can see the improvement of performance when Bagging Ensemble Techniques was performed on Decision Tree compared to single Decision Tree Classifier. Decision Tree when used as meta classifier in Stacking Ensemble Technique, the model gave the second-best accuracy and performance after XGBoost.

As XGBoost performed best, XGBoost model was selected for decision support system for supplier selection process. This DSS can now be used for decision making in supplier selection process and will be beneficial for SCM companies as this DSS is fast, reliable, flexible, efficient and effective for the best supplier selection.

Further research can be done by considering the review of supplier or supplier's behavior in supply chain posted on social networking sites.

REFERENCES

1. Nitin Harale, Sebastien Thomassey, Xianyi Zeng, "Supplier Prediction in Fashion Industry Using Data Mining Technology", IEEE International Conference on Industrial Engineering and Systems Management (IESM), 2019.
2. Beda Puspita Candra, Eko Rachmat Slamet Hidayat Saputra2, Ruhamah, Kurniawan Wicaksono Kusrini "Decision Support System for Stock Prediction and Supplier Selection Using Least Square and C4.5 Algorithm", IEEE 2018 3rd International Conference on Information Technology, Information Systems and Electrical Engineering, 2018.
3. Melany Mustika Dewi, Kusrini, Hanif Al Fatta , " Supplier Selection Using Combined Method of K-Means and Intuitionistic Fuzzy Topsis", 2018 International Seminar on Application for Technology of Information and Communication, 2018.
4. Hamed Taherdoosta, Aurélie Brardc, "Analyzing the Process of Supplier Selection Criteria and Methods ", 12th International Conference Interdisciplinarity in Engineering, (INTER-ENG),2019.
5. N. Phumchusri, S. Tangsiriwattana, P. Luangjarmekorn, "Supplier Selection Method: A Case-study on a Car Seat Manufacturer in Thailand", 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM 2018).
6. Sudhir Anakal, Sandhya P, "Clinical Decision Support System for Chronic Obstructive Pulmonary Disease using Machine Learning Techniques", 2017 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques, 2017.
7. James Scott, William Ho, Prasanta K. Dey, Srinivas Talluri, "A decision support system for supplier selection and order allocation in stochastic, multi-stakeholder and multi-criteria environments", International Journal of Production Economics,2015.