RANKING OF SUPPLIER PERFORMANCE USING MACHINE LEARNING ALGORITHM OF RANDOM FOREST

Vincent H Wilson
Professor, Department of Manufacturing Engineering, School of Mechanical Engineering (SMEC), Vellore Institute of Technology, Vellore, India.

Arun Prasad N S
UG Student, Department of Manufacturing Engineering, School of Mechanical Engineering (SMEC), Vellore Institute of Technology, Vellore, India

Aswin Shankaran
UG Student, Department of Manufacturing Engineering, School of Mechanical Engineering (SMEC), Vellore Institute of Technology, Vellore, India

Sushant Kapoor
UG Student, Department of Technology Management, School of Mechanical Engineering (SMEC), Vellore Institute of Technology, Vellore, India

John Rajan A
Professor, Department of Manufacturing Engineering, School of Mechanical Engineering (SMEC), Vellore Institute of Technology, Vellore, India

ABSTRACT

In this current data-driven world, corporate and industries need decisions driven by data for staying ahead of the competition and to satisfy the customers. To calculate the results using algorithms that could learn from the fluctuations and abnormalities has been sought after by tech companies as we are now started moving towards industry 4.0. At this time of advent, Machine Learning and Data Science are going to play a major role in shaping the industry and help in its operations. Furthermore, if we look into the industries, many tech and manufacturing conglomerates need suppliers to furnish them with services, materials, and semi-finished products for their operations. Finding the best supplier has been an age-old problem and several solutions have been introduced over the period. In this paper, we will look at how this age-old problem could be solved using modern techniques.

Keywords: MCDM (Multi-Criteria Decision Making), Fuzzy AHP, Machine Learning, Decision Tree, Random Forest, Supplier Performance
1. INTRODUCTION

Selecting the best supplier for a product is a very important task in the process of outsourcing and various MCDM (Multi-Criteria Decision Making) approaches have been used for this. The approaches that have been used are the Genetic algorithm, DEMATEL, ELECTRE, TOPSIS, VIKOR, Simple multi-attribute rating technique, Analytic Hierarchy Process, and so on. These methods require extensive manual calculations and if there is a large entry of data, then each method gives out different answers. Therefore, to make the calculations accurate and faster, thereby saving a lot of time, we use Machine Learning Algorithms to find out the best supplier by providing data.

In this paper, we are going to use a Machine Learning method called Random Forest to find the best supplier performance from a set of 20 suppliers. Each supplier is evaluated under different criteria and sub-criteria. These criteria which are used in this paper are already used by various experts in the manufacturing sectors and used up in research papers [1]. Several factors decide the performance of the supplier and each factor/criteria has a different weightage. To calculate this weightage the experts have used various methods and, in this paper, we will rank the suppliers based on the rank obtained by the Fuzzy AHP method [2].

1.1. Machine Learning

Machine Learning is used for two types of problems, namely Classification and Regression. In this context, we have to classify the data and no need for calculation of any unknown data. On the topic of classification if we further lookup, we would see that there are 179 classifiers subdivided from 17 families/groups (Bayesian, stacking, support vector machines, decision trees, rule-based classifiers, boosting, bagging, random forests and other ensembles, neural networks, generalized linear models, nearest K neighbors, partial least squares and principal component regression, logistic and multinomial regression, discriminant analysis, multiple adaptive regression splines, and other methods), implemented in Python, C, Weka, Apache Spark and so on, and also other classifiers being added daily through research[3].

Now when we look into the data we have, the grading of each supplier on each criterion needs a lot of decisions to be made in each level of the problem. So they should have decision trees in each level and a forest of random decision trees to connect all i.e. Random Forest. The methodology on how each decision tree works and how they contribute to the random forest will be explained afterwards in this paper. The reason why we chose using a machine learning algorithm rather than just programming with the formulas of any MCDM method is that we want the machine to understand the data and predict the best cause abnormalities arise and even humans make errors.

2. LITERATURE REVIEW

The selection of supplier is process involving lots of factors and there had been literature works in the past categorizing the factors. The basic criterion/factors are Quality, Delivery, Price and Service [4][5][6]. The set of criteria and sub criteria are usually determined by a group of specialists within the organization and can also be chosen from many studies and literature works conducted during this subject.
This paper discusses about the evaluation and ranking of 20 suppliers of a manufacturing industry based on their performance. The performances of these suppliers are evaluated based on several criteria and sub criteria. They are classified under 3 main Criteria –

- Quality, Delivery and Performance History
- Capacity and Capability
- Responsiveness, Service, Safety and Trust

Under each criterion there are few sub criteria. These criteria have been widely used for the performance evaluation of suppliers and have been discussed in several literatures as indicated in Table 1[2][1].

<table>
<thead>
<tr>
<th>Table 1 Criteria and Sub Criteria for Supplier (Performance) Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quality, Delivery &amp; Performance History (QDP)</strong></td>
</tr>
<tr>
<td>Quality – Product &amp; Manufacturing</td>
</tr>
<tr>
<td>Quality – Service &amp; Working Environment</td>
</tr>
<tr>
<td>QC &amp; QA Documentation – Certification &amp; Validity</td>
</tr>
<tr>
<td>Complaince to Delivery Commitments</td>
</tr>
<tr>
<td>Completions – PGMA/ DU/ Tonnage</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

2.1. Weightage of Criteria

Analytical Hierarchical Process (AHP) developed by Saaty in the 1970s is one of the most used methods for quantifying the weights of each criterion and thereby helping us understand which criteria are important over others. An AHP hierarchy could be a structured means that of modeling the choice at hand. It consists of associate degree overall goal, a bunch of choices or alternatives for reaching the goal, and a bunch of things or criteria that relate the alternatives to the goal, the factors may be additionally de-escalated into subcriteria, sub-sub criteria, and so on, on several levels because of the drawback needs. A criterion might not apply uniformly, however, it could have hierarchical variations sort of a very little sweetness is gratifying but an excessive amount of sweetness may be harmful. therein case the criterion is split into subcriteria indicating different intensities of the criterion, like very little, medium, high, and these intensities are prioritized through comparisons beneath the parent criterion, sweetness.

As these criteria’s are being used by researchers in the supply chain field for over a period of time [7], we could get the weightage of the sub-criteria which we have found in the literature review and therefore we could get the rank of sub-criteria under each criteria on which they are classified[2].
In table 2 we can see how the sub-criteria are ranked based on their weightage. Based on these as target values we would compare it with the values given to suppliers for each of the criteria to rank the supplier performance.

3. RANDOM FOREST METHOD

3.1. Decision Tree

It is a supervised machine learning approach which could be used for both classification and regression problems. The idea for this algorithm originated from an ordinary tree which has a root, that subdivide to branches and then leaves. It is far more easier to understand decision tree as compared to other classification algorithms. The tree representation of the algorithm makes it easier to understand with the leaf nodes corresponding to class label and the internal nodes corresponding to an attribute. The below figure will help us understand the decision tree better.

![Figure 1 Representation of Decision Tree for deciding whether a person is fit](https://dimensionless.in/building-blocks-of-decision-tree/)

3.2. Random Forest Algorithm

Random Forest Algorithm works best for classifiers involving a lot of criteria to be considered upon. As we know a forest consists of trees, even in this Random Forest Algorithm we have different decision trees on various sub-criterias of the given dataset and this algorithm takes the average out of them to increase the predictive accuracy [8]. Now that what it contributes to our paper is that we have different criteria’s and each criteria would

http://www.iaeme.com/IJARET/index.asp 301  
editor@iaeme.com
give out different vendors as the best based on its decision tree, now with the help of Random Forest Algorithm we can get to the best supplier out of the total decisions made. Moreover, to increase the accuracy of the results we can divide the dataset into training sets and testing sets for calculation purposes. The Random Forest method could be best explained in the below Figure 2.

![Figure 2 Working of Random Forest algorithm](https://www.javatpoint.com/machine-learningrandom-forest-algorithm)

### 4. CODE AND OUTPUT

We are using Python for this Machine Learning process as it is easy, cheap, robust and adaptable environment. As Random Forest is a classifier and is stored in packages. To access various such packages which would help us with our program we need a GUI (Graphic User Interface). Here we use Anaconda Navigator as it has lots of packages which could be used for Machine Learning and moreover it provides with various libraries and channels without using command-line commands. The python Random Forest Classifier uses an algorithm where it classifies the data with value 1 as more important than 5. But we have used a scale where 1 is least important and 5 is excellent.

**Table 3** Code and Output

<table>
<thead>
<tr>
<th>Comparison Scale</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Poor</td>
<td>1</td>
</tr>
<tr>
<td>Poor</td>
<td>2</td>
</tr>
<tr>
<td>Good</td>
<td>3</td>
</tr>
<tr>
<td>Very Good</td>
<td>4</td>
</tr>
<tr>
<td>Excellent</td>
<td>5</td>
</tr>
</tbody>
</table>

#### 4.1. Data

We have collected Data on 20 suppliers from BHEL (Bharat Heavy Electricals Limited), Trichy for each criteria. Each supplier is labelled from V1 to V20 and are evaluated under 20 criteria as mentioned above in the literature review section. The data collected is shown below in Table 4.
5. PHYTHON CODE

```python
import numpy as np
from matplotlib import pyplot
import pandas as pd

df=pd.read_csv("Supplier Data.csv")

inputs=df.drop('Alt/Criteria',axis='columns')
target=df['RANK']

from sklearn import datasets
import sklearn
import pandas as pd

inputs=inputs.drop('RANK',axis='columns')
target=df['RANK']

from sklearn import tree
classifier()
initial_model.fit(inputs,target)
initial_model.score(inputs,target)

from sklearn.model_selection
import train_test_split

X_train, X_test, y_train, y_test = train_test_split(inputs,target, test_size=0.2, random_state=20) print(X_test)
```

http://www.iaeme.com/IJARET/index.asp
from sklearn.ensemble import RandomForestClassifier

model=RandomForestClassifier(n_estimators=10, random_state=30)
model.fit(X_train,y_train)
prediction_test=model.predict(X_test)
prediction_test

feature_list=list(inputs.columns)

feature_imp=pd.Series(model.feature_importances_,index=feature_list).sort_values(ascending=True)

import seaborn as sns
import matplotlib

sns.barplot(x=feature_imp, y=feature_imp.index)
plt.xlabel('Supplier Rejection Score') plt.ylabel('Suppliers')
plt.title("Visualizing Performance Features") plt.legend()
plt.show()

5.1. Final Output

Figure 3 Supplier Ranking from Top to bottom
6. RESULT AND CONCLUSION
From Figure 3 it is could be found out that V15, V20, V18, V7 and V14 have less rejection score thereby ranked highest. We could therefore consider them as best options and reject V9 as it comes at the bottom with the highest rejection rate.

As machine learning has made a process which takes over a lot of time to finish and analyse, we have saved much of our time and cost involved in the calculation. We thereby have a competitive advantage over our competitors.

There are further more classifiers and data that could be used for calculating the best suppliers for an industry and would be an area of future scope for research.

REFERENCES


Ranking of Supplier Performance Using Machine Learning Algorithm of Random Forest


