Intelligent Supplier Selection - Capstone Project

Intelligent Supply Chain Strategies

Note: Provide justifications/calculations/steps along with each answer to illustrate how you arrived at the answer. State your arguments clearly, in logical sequence. You will not receive credit for giving an answer without sufficient explanation.

Problem Statement

Consider a multinational corporation that operates in multiple countries and has several departments. The company is looking to standardise its supply chain operations and improve its overall efficiency. To achieve this, it has decided to select suppliers for the goods it requires. The goal of the multinational corporation is to choose the best supplier on various criteria such as cost, quality, delivery time, and reliability. At the same time, the company would like to minimise the risk. This project will ask you to predict supplier's risk using supervised learning (classification) paradigm and incorporate this in an optimization problem to finally make efficient supplier selection.

Dataset

Please refer to the files `supplier_data.csv` and `New_supplier_data.csv` provided to you with this capstone project. The file `supplier_data.csv` provides the data for the risk of 1000 suppliers based on 20 different criterion listed below:

1. Technical expertise
2. Quality
3. Delivery time
4. Reliability
5. Flexibility
6. Capacity
7. Innovation
8. Reputation
9. Financial stability
10. Environmental impact
11. Social responsibility
12. Safety record
13. Technical expertise
14. Customer service
15. Lead time
16. Responsiveness

17. Contract terms

18. Diversity

19. Sustainability

20. Scalability

Dataset Description: Each row is about a supplier. We have details about their technical expertise, quality, delivery time etc. In default column (target column), we have value 1, if the supplier has not defaulted and the value 2, if the supplier has defaulted.

Tasks

1. In the target column, replace 1 with 0 and 2 with 1. So that the defaulters will become positive class.

2. As a Data engineer, your role is to conduct a detailed data preprocessing of `supplier_data.csv`. Identify if there are any missing values. If yes, do the appropriate imputation.

3. *Exploratory data analysis:* Perform exploratory data analysis to identify which factors differentiate/influences customers who are defaulting with others.

4. *Data Imbalance:* Identify whether the target column is balanced or imbalanced. Plot the same using a simple pie chart.

5. Convert all categorical columns to numerical columns using one hot encoding.

6. Divide the dataset in to training (80%) and testing (20%).

7. Use as many classification algorithms as possible to build your classifier. Use five-fold cross validation while building your machine learning model.

8. Using all the models Predict the target class for test dataset and populate the following metrics and report the same in a consolidated table

   (a) Accuracy
   (b) Sensitivity
   (c) Specificity
   (d) F1-score
   (e) AUC

9. Plot the ROC curves for all the models in a single plot and identify the best model using the same chart.

10. Justify your choice of classifier.

11. Use your classifier to predict supplier's risk of `New_supplier_data.csv`.

12. Incorporate this information in the objective function of Vivo Supplier Selection problem discussed in week 2 by assuming that the risk scales at a cost rate of Rs 80,000. What would be your new supplier selection plan?

13. Do sensitivity analysis by considering a few other cost rates such as Rs 60,000, Rs. 90,000 etc.

14. Prepare an executive report for this project.