SUMMARY

PROBLEM 3

Q1. Classification in preference of Vegetarian and Non- Vegetarian food item depending on the class of students and the hour they come to ANC. This will help to build preference and discounts and diff hours.

Bayesian Network Classification has been used to solve this problem. We classify the food items as veg and nonveg. For the food items for which data for non-veg is not available is taken as NULL, eg Veg burger. Also items available at pit-shop like biscuits, chips, coke, etc. are removed. We remove the items with NULL value and apply the Bayes Net Classifier to the training data.

Results for output field veg/nonveg

Individual Models

Comparing $B-veg/nonveg with veg/nonveg

Correct	1,675,536	73.65%
Wrong	599,586	26.35%
Total	2,275,122	

Performance Evaluation

0	0.004
1	0.813

Evaluation Metrics

Model	AUC	Gini
$B-veg/nonveg	0.668	0.335

Q2. Market Segmentation. We try to group customers into related sets, in our case based upon revenue generated in the given period. It is an important tool for applied Marketing.

Two Step Clustering has been used to solve this problem. The revenue generated by each transaction is found using Price * Quantity. Then we aggregate the data based upon StudentID, such that, we get the revenue generated by each student. We need to remove the revenue generated by the cash transactions. Then after normalizing the attributes StudentID and Revenue we apply Two Step Clustering to it. The result is as shown below: