ROHAN SAXENA 2015A7TS0017P
NAVEEN VENKAT 2015A7PS0078P

Problem 2: Pair popular products with less popular ones

Data Preprocessing and Selection:

The transactions were aggregated on the basis of the student ID, date and time. So, if the same student bought multiple items together at the same time (as part of the same order), then these were aggregated into one single record for the student. Further, we observed that a customer often buys products in multiple transactions across the course of his/her meal. To address this issue, we clubbed all transactions by a student - having less than a 2 hour interval between them - into a single transaction. We then selected only the set of items in the transactions for further processing, ignoring the student ID, date, etc.

Data Mining technique:

We applied Apriori algorithm for mining association rules, where we paired popular items with non-popular items, and found rules of the form: Popular ➔ non-popular items. This implies that when a customer buys the popular item, he/she also buys the non-popular item. We define a “popular” item as one having popularity greater than 3, while one having popularity less than 3 is non-popular. Since we expect the sales of non-popular items to drop, we will pair these items with popular items they were bought with earlier. Subsequently, out of the rules we obtained which satisfied the min_support and min_confidence criteria, we handpicked the rules which contained “sensible” combinations as required by the problem.

Interpretation / Evaluation:

We found 25 association rules in which the itemset was of cardinality 2, which passed our min_support and min_confidence criteria. We found that for itemsets having cardinality greater than 2, the support and confidence of such rules were very low. Thus, we neglected them and focused on providing combo offers of a popular item clubbed with one or more non-popular items based on the association rules obtained above.

We mined association rules from the transactions of months August - November, and tested our designed combo offers for December’s transactions. Also, we calculated loss using the maximum possible decrement possible. For instance, in a particular order, say 2 combos are applicable. That is, the customer wishes to order A, B, C, D and the 2

combos available are A-B and B-C. Then, we assume that the customer is rational and will choose that combo for which he/she gets maximum discount. So when multiple combos are applicable, we apply the combo offering the maximum discount to calculate the loss incurred by ANC.

We obtained this result:
M = 3
Loss = Rs 14598

We obtained the following association rules from minimum support of 1%:
[These are rules of the form ("Popular" -> "Non-popular") and with support > 0.01].

164 -> 151
Kadhai Paneer ----> Butter Naan

146 -> 151
Butter Chicken ----> Butter Naan

123 -> 2
Pastry ----> Veg Petty

148 -> 155
Tandury Chicken ----> Pepsi

12 -> 155
Pasta ----> Pepsi

148 -> 151
Tandury Chicken ----> Butter Naan

146 -> 155
Butter Chicken ----> Pepsi

164 -> 155
Kadhai Paneer ----> Pepsi

130 -> 134
Masala Dosa ----> Plain Maggi

123 -> 155
Pastry ----> PEPSI

169 -> 134
Pulpy Orange ----> Plain Maggi

169 -> 151
Pulpy Orange ----> Butter Naan

159 -> 134
MyCan ----> Plain Maggi

12 -> 134
Pasta ----> Plain Maggi

107 -> 155
Bingo chips ----> PEPSI

135 -> 155
Fried Maggie ----> PEPSI

138 -> 155
Paneer Franky ----> Pepsi

130 -> 155
Masala Dosa ----> Pepsi

107 -> 134
Bingo chips ----> Plain Maggi

124 -> 155
Cheese Burger ----> Pepsi

179 -> 2
Mix Juice ----> Veg Petty

142 -> 134
Chicken Sandwich ----> Plain Maggi

188 -> 2
Samosa Chat ----> Veg Petty

1 -> 134
Samosa ----> Plain Maggi

177 -> 151
Chilly Chicken ----> Butter Naan

An interesting observation that we can make is that there are various rules of the form A ➜ B and A ➜ C. Since A is a popular product which goes well with non-popular products B and C, we can introduce a new combo offer A, B, C apart from the offers A, B and A, C. That is how we introduce new combo offers like “Butter chicken, butter naan, pepsi” from association rules of just 2 itemsets.

The fact that there are only A ➜ B and A ➜ C rules, and not A ➜ B and A ➜ C and A ➜ D means that the data is only suitable for combo offers of 3 items.