SOLUTION TO MACHINE LEARNING CHALLENGE

Submitted By:
Ankit K. Sood
M.E (Software Systems)
BITS Pilani
Intern,
HERE Maps, Mumbai.

ASSUMPTIONS

1. Counting of days taken to prepare and dispatch fulfillments is based solely on the dates; the time (in hours and minutes) has been ignored.
2. Once the order has been dispatched, it is the courier partner's concern, not directly of paytm.

THE ALGORITHM

1. Read through the Transactions' data document and populate a HashMap, 'Transaction' of type

HashMap<Merchant id, HashMap <T1, HashMap <T2, HashMap <T4, HashMap <Product, Transaction >>>>>

Where Transaction is a data structure of type:

```
class Transaction
{
        int quantity;
        int fullfillmentCreated;
        int fullfillmentShipped;
        double rating;
}
```

2. Read through the profits' and cancellations' data documents and populate a HashMap,' completeRecord' of type

HashMap<Merchant id, HashMap <T1, HashMap <T2, HashMap <T4, OverallRating>>>>

Where overallRating is a data structure of type:

```
class overallRating {
    double commissionPercent;
    double cashbackPercent;
   double discountPercent;
   double profitRating;
   double cancelRating;
   int cancelNumber;
   int returnNumber;
   double transactionRating;
   double totalRating;
   int quantity;
}
```

3. Combine the Transactions data with the OverAll data.

4. Maintain a HashMap,' finalRating' of type
   HashMap<Merchant id, Rating>

   This data structure is useful for storing ratings for all merchants, which will be our final result.

5. Write a method (function) to return the rating of a merchant pertaining to given T1, T2 and T4.

6. Maintain a HashMap, 'merchantRating' of type

   HashMap<T1, <T2, <T4, <Merchant id, OverallRating>>>>

   This would return category wise rating of a merchant, which can even be aggregated along T2 and T1.
   (This wasn't a part of the requirements but serves as an important information for data analytics)

7. Record the quantity of products sold by every merchant for all categories and sub categories, such that the quantities sold can be aggregated along both T2 and T4.

ANALYZING THE ALGORITHM

1. HashMap proves to be useful in such a situation, as transaction rating, merchant rating and overall rating can be determined in constant time O(1), for given merchant and product categories.

2. The data structures used also enable one to aggregate data over any category or subcategory.

RATING A MERCHANT

1. For transactions dataset:
    a. A penalty of 0.5 for each day taken to begin the creation of fulfillment by the merchant, after 1 day of order being placed.
    b. A penalty of 1 day for each day delay in shipping the fulfillment.
    c. Add both the penalties to obtain the transaction rating for a merchant. Transactions rating is a negative metric.

2. For profits data set:
    a. Record commission percent factor as 10 times the commission percent offered by the merchant to paytm.
    b. Record discount percent as 10 times the discount offered by the merchant.
    c. Add both the commission percent and discount percent to obtain the profits rating of each merchant. Profits rating is a positive metric.

3. For cancellations dataset:
    a. A penalty of twice the number of orders cancelled.
    b. A penalty of thrice the number of order returned.
    c. Sum of both the above penalties give the cancellations dataset rating. This is a negative metric.

The weightage given to each attribute in every dataset is relative to the others, thereby reflecting a relative importance of each attribute.

There is no upper or lower limit to the score given to a merchant. A rating of 0 for a merchant means he's doing exactly what he is supposed to do, no added benefits to the company or consumers and no detriments either. A positive rating means the merchant is adding value to the company either in terms of customer considerations or in terms of profits to the company, while a negative rating indicates that his work is unfavorable.