RECOMMENDER SYSTEM FOR E-COMMERCE WEBSITE

- **PROBLEM STATEMENT:** An e-commerce website thrives on the number of customers who make purchases from its website and an integral part of increasing the number of customers on the website is customer satisfaction. And customer satisfaction is rooted in the type of purchases that the customers make i.e.(A customer would be more satisfied if he buys products that he likes.). It's not too difficult to think that a customer who is not satisfied by the service that online portal offers is likely to cease any further business with that e-commerce company. There are a plethora of products in an e-commerce company's inventory and when a customer visits the website to buy some product, it's highly unlikely that he or she would find the respective product simply by browsing. If the company maintains such an approach for its website then it is likely to lose its customer base. It's website should possess such a predictive power that it recommends products to the customer depending on the customer's history, liking and disliking. This approach would not only reduce the hassle for the customer and keep him loyal to the company, but also encourage the customer to buy more things as the things predicted are as per his or hers' liking, thereby improving the profits of the e-commerce company and also of the sellers selling products on the company's website. *As a result of this I will try to build a predictive model which not only recommends an existing customer the products of his or her liking, but will also recommend the best-selling products and products in trend to new customers. The model will also try to recommend new things to a customer based on the reviews he or she gives to a particular product.*

- **DATA:**
  - **Review data:**

```
{ "reviewerID": "A2SUAM1J3GNN3B",
"asin": "0000013714",
"reviewerName": "J. McDonald",
"helpful": [2, 3],
"reviewText": "I bought this for my husband who plays the piano.
He is having a wonderful time playing these old hymns. The music
is at times hard to read because we think the book was published
for singing from more than playing from. Great purchase
though!",
"overall": 5.0,
"summary":
"Heavenly Highway Hymns",
"unixReviewTime": 1252800000,
"reviewTime": "09 13, 2009" }
```

where

- `reviewerID` - ID of the reviewer, e.g. A2SUAM1J3GNN3B
- `asin` - ID of the product, e.g. [0000013714](0000013714)
- `reviewerName` - name of the reviewer
- `helpful` - helpfulness rating of the review, e.g. 2/3
- `reviewText` - text of the review
- `overall` - rating of the product
- `summary` - summary of the review
- `unixReviewTime` - time of the review (unix time)
- `reviewTime` - time of the review (raw)

- **Metadata od product:**

```
{ "asin": "0000031852",
"title": "Girls Ballet Tutu Zebra Hot Pink", "price": 3.17,
"imUrl": "http://ecx.images-
amazon.com/images/I/51fAmVkTbyL._SY300_.jpg", "related": {
"also_bought": ["B00JHONN1S", "B002BZX8Z6", "B00D2K1M3O",
"0000031909", "B00613WDTQ", "B00D0WDS9A", "B00D0GCI8S",
"0000031895", "B003AVKOP2", "B003AVEU6G", "B003IEDM9Q",
"B002R0FA24", "B00D23MC6W", "B00D2K0PA0", "B00538F5OK",
"B00CEV86I6", "B002R0FABA", "B00D10CLVW", "B003AVNY6I",
"B002GZGI4E", "B001T9NUFS", "B002R0F7FE", "B00E1YRI4C",
"B008UBQZKU", "B00D103F8U", "B007R2RM8W"], "also_viewed":
["B002BZX8Z6", "B00JHONN1S", "B008F0SU0Y", "B00D23MC6W",
"B00AFDOPDA", "B00E1YRI4C", "B002GZGI4E", "B003AVKOP2",
"B00D9C1WBM", "B00CEV8366", "B00CEUX0D8", "B0079ME3KU",
"B00CEUWY8K", "B004FOEEHC", "0000031895", "B00BC4GY9Y",
"B003XRKA7A", "B00K18LKX2", "B00EM7KAG6", "B00AMQ17JA",
"B00D9C32NI", "B002C3Y6WG", "B00JLL4L5Y", "B003AVNY6I",
"B008UBQZKU", "B00D0WDS9A", "B00613WDTQ", "B00538F5OK",
"B005C4Y4F6", "B004LHZ1NY", "B00CPHX76U", "B00CEUWUZC",
"B00IJVASUE", "B00GOR07RE", "B00J2GTM0W", "B00JHNSNSM",
"B003IEDM9Q", "B00CYBU84G", "B008VV8NSQ", "B00CYBULSO",
"B00I2UHSZA", "B005F50FXC", "B007LCQI3S", "B00DP68AVW",
"B009RXWNSI", "B003AVEU6G", "B00HSOJB9M", "B00EHAGZNA",
"B0046W9T8C", "B00E79VW6Q", "B00D10CLVW", "B00B0AVO54",
"B00E95LC8Q", "B00GOR92SO", "B007ZN5Y56", "B00AL2569W",
"B00B608000", "B008F0SMUC", "B00BFXLZ8M"], "bought_together":
["B002BZX8Z6"] },
"salesRank": {"Toys & Games": 211836},
"brand": "Coxlures",
"categories": [["Sports & Outdoors", "Other Sports", "Dance"]] }
```

where

- `asin` - ID of the product, e.g. 0000031852

- `title` - name of the product
- `price` - price in US dollars (at time of crawl)
- `imUrl` - url of the product image
- `related` - related products (also bought, also viewed, bought together, buy after viewing)
- `salesRank` - sales rank information
- `brand` - brand name
- `categories` - list of categories the product belongs to

• BASIC METHODOLOGY

A.) When a person logs in/signs up.
1. A person logs-in to the website and based on the person's history of liking and disliking, products are recommended.
2. A new person signs-up. The highest rated products/best-selling products are recommended.

B.) A person has already bought a product, and he writes a review about that product. Instantly recommend him products based on the review.
3. A review written by the customer is classified to a rating and turns out that it is a bad review. Do the following for bad review:
   ♦ Take into consideration the category of the product. Since the review was bad we do not want similar products for the customer and hence we look for the best products from the category and recommend those products to the customer.

4. A review written by the customer was classified as good.
   ♦ Since the customer liked the product we would recommend products similar to that product. We would also recommend the best products from products bought together, viewed together and other things that are included in the metadata.

C.) Finally since we are also creating a model for product popularity/trend for recommending products to new customer. This model can be used for recommending which products should a seller target for maximum profits.

• Possible Algorithm methodology for part A, B and C:

A.)   As mentioned in part A above we will be recommending things to customers during log-in and sign-up.
1.) When an old user logs-in:
    Create the following table(For more info regarding features see the data section above)

history [table name]

reviewerID	asin(prod. ID)	categories(from metadata)	review_time	overall
Unique customer (categorical)	Product can repeatfor the same or different customer (categorical)	Categories. Also there are multiple categories for every product. (categorical)	Time may repeat. I am not sure. (Date Time object)	Rating(5) (categorical)

- ◆ When the customer log-in we can create an entry for the customer by creating a table such that it contains every product purchased by the customer in the following way.
  [SELECT * FROM history WHERE reviewerID = <our customer> and overall >= 4] = current_customer

  - ➤ Then for every entry in the current_customer table we can run an algorithm like K-means to get K similar results and then select the result which is most similar to our entry.

2.) When new user signs up:
  - ◆ Using the same table above we can present products with highest frequency of purchase also taking into consideration the sales rank and the average rating of that particular product. The model will return few products of few categories depending on the frequency of purchase and the average rating to that product.

B.) When a person writes a review about a bought product.
  - ◆ For this we first need to classify our reviews into particular reviews. Since reviews for a particular product will be very less we will train our model for review based on every category. So first thing in this prediction model would be to train it on the reviews with label as rating using some classification algorithm like Naïve Bayes, Neural network , perceptron or SVM. We will use this trained model then to predict the new reviews.

3.) **If review is classified as bad:** A customer writes a review and it is classified as bad. Therefore we use the table from part A and extract all the products belonging to the category of the product and give recommendation of similar products but with better overall rating. Or the other option is to go through the metadata of this product and recommend products with the best average rating.

4.) **If review is classified as good:** We go through the metadata of the product and recommend the best products.

C.) Finally from the sellers point of view we can ask the seller to choose from the category and depending on the categories we can recommend products with highest purchasing frequency and rating so that the seller can target on similar products.