**Merchandise Recommendations**

Dataset Considered: We took the amazon review dataset for Clothes, shoes and jewelleries and beauty products. We considered the reviews and ratings given by the user to different products as well as his/her reviews about his/her experience with the product(s).

[ <https://www.kaggle.com/druss4/amazon-data> ]

Approach: We performed sentiment analysis for predicting the helpfulness of reviews and then designed item based collaborative filtering model on KNN to find 5 most similar items.

Models used / compared:

Sentiment analysis:

1. Logistic Regression
2. Naïve Bayes – Multinomial
3. Naïve Bayes – Bernoulli

Recommendation system:

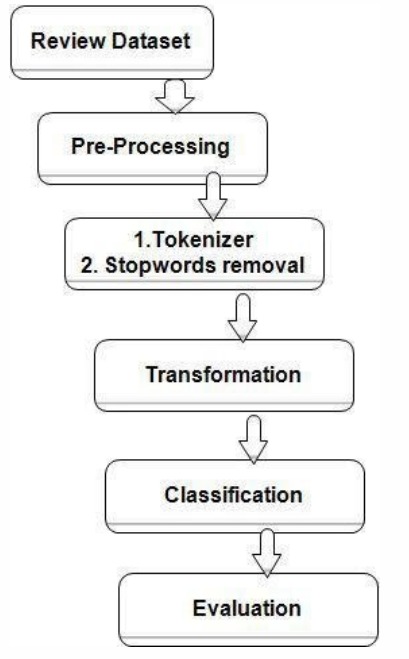
1. k-Nearest Neighbours

Sentiment Analysis: Sentiment Analysis or Emotion AI refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.[3] Sentiment analysis aims to determine the attitude of a speaker, writer, or other subject with respect to some topic or the overall contextual polarity or emotional reaction to a document, interaction, or event. The attitude may be a judgment or evaluation, affective state (that is to say, the emotional state of the author or speaker), or the intended emotional communication (that is to say, the emotional effect intended by the author or interlocutor).

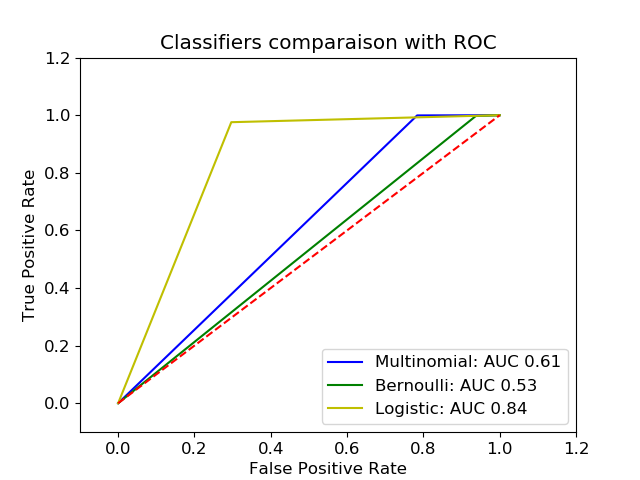
In our case for this project we treat sentiment analysis as a model to predict the attitude of the user for reviews for a particular product in terms of usefulness / likeliness of the product to the user. The sentiment analysis can be broadly classified into following approaches:

1. Rule-based systems: perform sentiment analysis based on a set of manually crafted rules.
2. Automatic systems: rely on machine learning techniques to learn from data.
3. Hybrid systems: combine both rule based and automatic approaches.[4]

For the Sentiment Analysis we implemented the following Automatic approach based on ML.

1. Review Dataset: We used the dataset from the amazon which has around 300k reviews for Clothes, Jewellery, Beauty, Shoes products
2. Pre-Processing: This step is to clean the data. Since the review data is in the format containing special characters, etc. we need to remove the and convert the dataset to a clean summary of the review.
3. Tokenizer / Stopwords removal: Here we use the CountVectorizer model to first convert the review dataset into vectors of useful words / phrases in the review. We also use STOPWORDS from python’s nltk library to help remove the unnecessary words / characters from the data.
4. Transformation: We used the TfidfTransformer to convert the tokenized words from the above step to form the weighted tf-idf vectors. Term frequency–inverse document frequency model was applied for the purpose of finding strongly related words for relevant documents.
5. Classification: This is the most important step of the sentiment analysis where we train an ML classifier which can further be used to predict sentiment of any given review. In this step we used 3 different classifiers, Naïve Bayes – Bernoulli, Naïve Bayes – Multinomial and Logistic regression. The input to these models are the tf-idf vectors formed in the above step. And the output is a positive or a negative sentiment.
6. Evaluation: For the evaluation of the above 3 models we divided the dataset into 80% training and 20% testing. The accuracy scores were 89% for NB – Bernoulli, 91% for NB – Multinomial, 94% for Logistic Regression. The figure below shows the comparison of the three models in the terms of the true positive rate vs the false positive rates.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1- score | support |
| Positive | 0.78 | 0.70 | 0.74 | 5291 |
| Negative | 0.97 | 0.98 | 0.97 | 44360 |
| Total/avg | 0.94 | 0.95 | 0.95 | 49651 |

We can clearly infer from the ROC chart that the True positive rate is much higher (with AUC = 0.84) for Logistic Regression model, when compared to other models.

Thus from all the results shown above we can infer that Logistic regression is the best model for sentiment analysis using the automatic approach and on our review dataset. These are only the results for review dataset, however the sentiment analysis using Logistic Regression may perform differently using the other datasets (including conversational sentences).

In the next part of the system we will see how these sentiment analysis results help in building a recommender system.

Recommendation System: Recommender systems are an important part of the information and e-commerce ecosystem. They represent a powerful method for enabling users to filter through large information and product spaces. A recommender system / engine is a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item[6]. Recommender systems are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general. In our case we are going to use the recommender system for the products[7].

The user-generated texts in form of reviews are implicit data for the recommender system because they are potentially rich resource of both feature/aspects of the item, and the user’s evaluation/sentiment to the item.[3] Features extracted from the user-generated reviews are improved meta-data of items. In our work we see sentiments extracted from the reviews as user’s rating scores on the corresponding features.

The recommender system that we have built uses KNN algorithm as its base model to give recommendations to the user. We will use the helpfulness factor of the review available in the dataset along with the feature tf-idf vectors extracted in the above sentiment analysis model as an input feature set to the KNN model. To improve the accuracy of the model we have considered only those reviews of the products for which at least 80% of the users have voted as “helpful”. This was done in order to eliminate the fake / useless reviews from the dataset.

**Results:**

We ran our model on different settings. Firstly we tried to vary the training and the testing set size (85% - 90% training data) and also the k value for KNN from (3 - 5). And the best results were obtained when the setting was training set of 85% and the k value = 5.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1- score | support |
| 3 | 0.50 | 0.33 | 0.40 | 9 |
| 4 | 0.93 | 0.97 | 0.95 | 31 |
| Total/avg | 0.89 | 0.91 | 0.90 | 40 |

The accuracy observed was 0.906 and the MSE was 0.094.

**References:**

1. Wikipedia
2. <https://monkeylearn.com/sentiment-analysis/>
3. Machine Learning-Based Sentiment Analysis for Twitter Accounts Ali Hasan 1, Sana Moin 1, Ahmad Karim 2 and Shahaboddin Shamshirband 3,4,\*
4. Michael D. Ekstrand, John T. Riedl and Joseph A. Konstan (2011), "Collaborative Filtering Recommender Systems", Foundations and Trends® in Human–Computer Interaction: Vol. 4: No. 2, pp 81-173. http://dx.doi.org/10.1561/1100000009
5. Pazzani M.J., Billsus D. (2007) Content-Based Recommendation Systems. In: Brusilovsky P., Kobsa A., Nejdl W. (eds) The Adaptive Web. Lecture Notes in Computer Science, vol 4321. Springer, Berlin, Heidelberg