

MMA's Cleanipedia Ideathon

Proposed Solution:

We propose to build a platform for **Hindustan Unilever** which could serve as:

- Analytical Dashboard for the Organization to monitor how it's promotional Campaigns are doing on different social media platforms like Facebook, YouTube and mainly on Twitter.
- User forum - For the users to read more about Company's products. The idea is to come up with a platform (like Quora) wherein we can keep track of the kinds of articles a user is reading, and the Recommendation engine could churn out articles based on that to increase customer engagement.
- Central Hub for Unilever Products – To provide an interface wherein a user can search a product (say Surf Excel) and the platform will come up with links of Ecommerce Websites wherein the product is available for sale.

Technology Details:

Module 1: User Forum

- Assuming Hindustan Unilever would have articles about its products describing its benefits and usage. Even If there are not many, Content writers can help to come up with articles that are user-engaging.
- Also, let us assume that when users register on the platform, they help us with the geographical information like Country and State to which they belong to.
- At first time, when user logs into the platform we present him few random articles from the corpus of articles to read.
- Recording of Click-Stream Data: We keep track of which articles he has clicked on to read them to build his profile in terms of preferences.
- Now comes the Recommendation Engine which would decide which articles are to be presented to a user when he logs in. Now the Question is how a machine will decide the same?
- Machine Learning and Natural Language Processing comes to rescue:
 - **LDA (Latent Dirichlet Allocation) Model:** This is a probabilistic model which is fed with the corpus of articles and it comes up with “topics” that are discussed in the articles and the topic distribution in each of the articles. This model belongs to the category of Topic Modelling.
 - Let us take an example to understand it better. In case of articles from Hindustan Unilever we would have articles talking about Detergents, Hand Wash products, Hair related products. So LDA model will basically come up with categories like these from the corpus that it is fed with.
 - **TF-IDF Model and NER Model:** NER model which stands for Named Entity Extraction basically extracts out the Named Entities that an article is talking about. TF-IDF Model stands for Term Frequency – Inverse Document Frequency model wherein each article/document is represented by vector and the size of vector is the number of unique words in the corpus and weights are calculated based on frequency counts of a word in an article and across all articles. Then use Cosine Similarity to find how similar two articles are.
 - **Spectral Clustering:** Here we will cluster users based on the articles that they have read. Once that is done, for a given user say U1 we identify it’s cluster and then find the article that is trending in that cluster and that article gets recommended to user.

- **K-prototypes Clustering:** This model focusses on clustering users based on geographical region why is it important is because products' success also depends on geographical regions. So lets identify the articles that are getting read a lot in the region to which the user belongs to and recommend it to him.

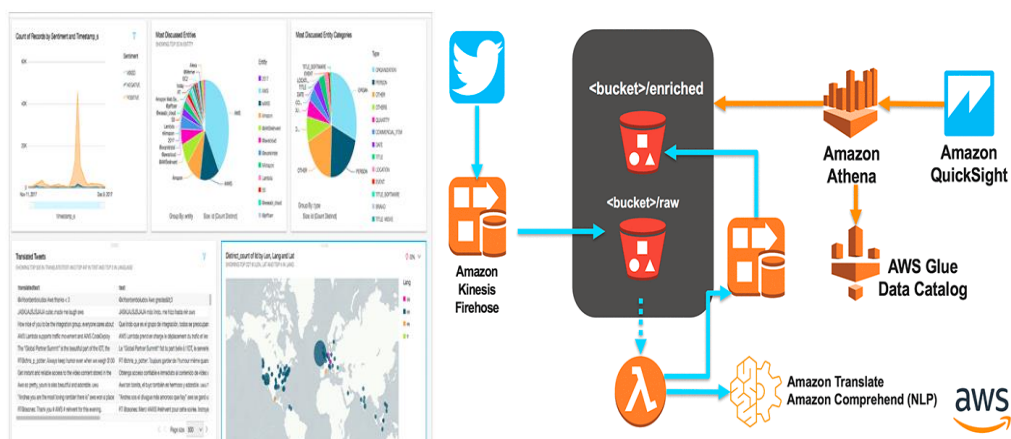
- The Recommendation Engine basically churns out articles in following ways:
 - **Content Based Filtering:** Based on what you read (LDA, TF-IDF, NER)
 - **Collaborative Filtering:** Based on what people like you read (Spectral)
 - **Collaborative Filtering:** Based on the region you belong to (K-prototypes)

Note:

- Data for Machine Learning models, we plan to have synthetic data for presenting our prototype in Hackathon.
- Programming Language: Python
- NLP and ML Models from built-in Libraries

Module 2: Analytical Dashboard for Organization

- To monitor and analyse how well your social campaigns are doing is very important for any organization.
- We plan to exploit the platforms like twitter and youtube mainly.
- How shall we do it?
 - **Twitter:**
 - We use AWS services like Kinesis, Athena, Amazon Translate and Lambda Functions to build a Real Time analytical dashboard which would analyse tweets about the campaigns at run time. [\[Link\]](#)



- **Youtube:**
 - Using Youtube Analytics APIs to extract features like likes, comments, views, portion of video clip people are watching etc [\[Link\]](#)

Module 3: Central Hub for Unilever Products

- The architecture for this module is as follows:
 - Come up with an auto-complete search field for facilitating easy search of products.
 - Suppose the user types in “Surf Excel” then we make use of Google Search API in background to get search results based on keywords like “Buy Surf Excel online” and filter out the results wherein the link of Json Object has domains like Flipkart/Amazon/Snapdeal and provide those results with links to users. So that they can go on their trusted e-commerce website and buy it.
- Technologies used:
 - Environment: Node JS
 - Backend: Express Server
 - Google Search APIs