## **Info 7374 Final Project**

**Project Member:**

Yufan Yang

Vividh Talesara

**Project Description:**

**● Overview**

**A close up of text on a whiteboard

Description automatically generated**

This project is aimed to build a social media content-based recommendation system to enhance the user experience, which includes recommending or predicting some new items (articles) they will like based on their preference. This project includes the following functions: article classification, article topic keyword extraction, article similarity analysis, article recommendation, user recommendation, user preference analysis, article influence ranking, article influence prediction. This project explores the details and meaning of the marketing recommendation system through different angles and methods.

**● Project requirements and goals**

Use concepts covered in the class (See textbook and various class links) involving algorithmic marketing to perform data analysis based on large datasets.

Goals:

Build a social media content-based recommendation system to recommend, in other words, predict some new items (articles) users will like based on their preference, recommend new users to authors and so on.

Perform text analysis and keywords extraction based on related datasets contain the key interaction information with users and contents such as: user\_id, user\_action (view/like/comment/share), content\_id, content\_text.

**● Problems to be addressed**

Existing recommendation systems have imperfections, such as not considering different user behaviors representing different levels of interest. The new recommendation system will try to Improve the accuracy of recommendations and remove annoying recommendations. It’s important to use preference levels which means the record we have for users who interact with the social media posts (in our case including view, like, share, etc) can have a decent effect on the recommendation.

Explanation:

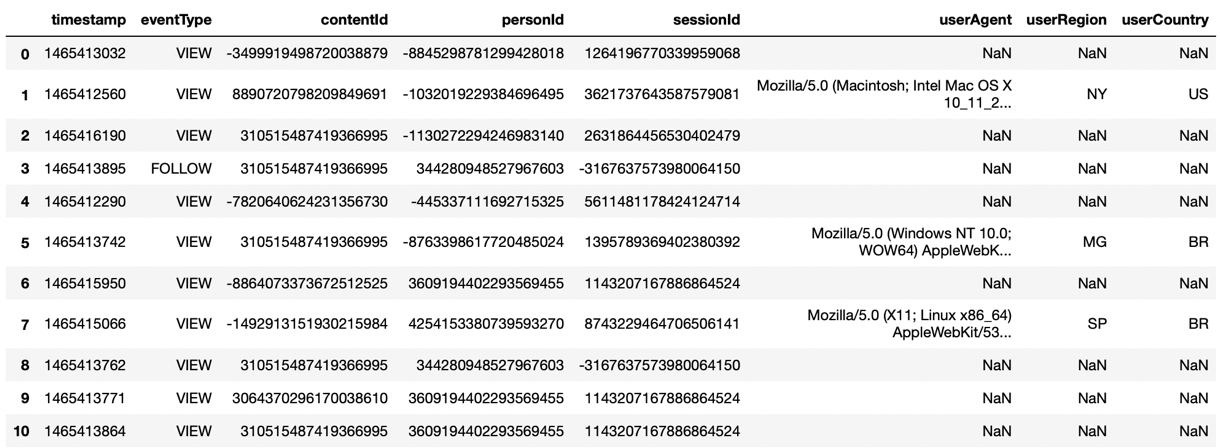
High Accuracy of recommendations means users received new items(contents) with similar interest belongs to contents that the user once liked, commented on, or shared.

Annoying recommendations means users received new items(contents) are only something they viewed before but not really had an interest in.

**Data:**

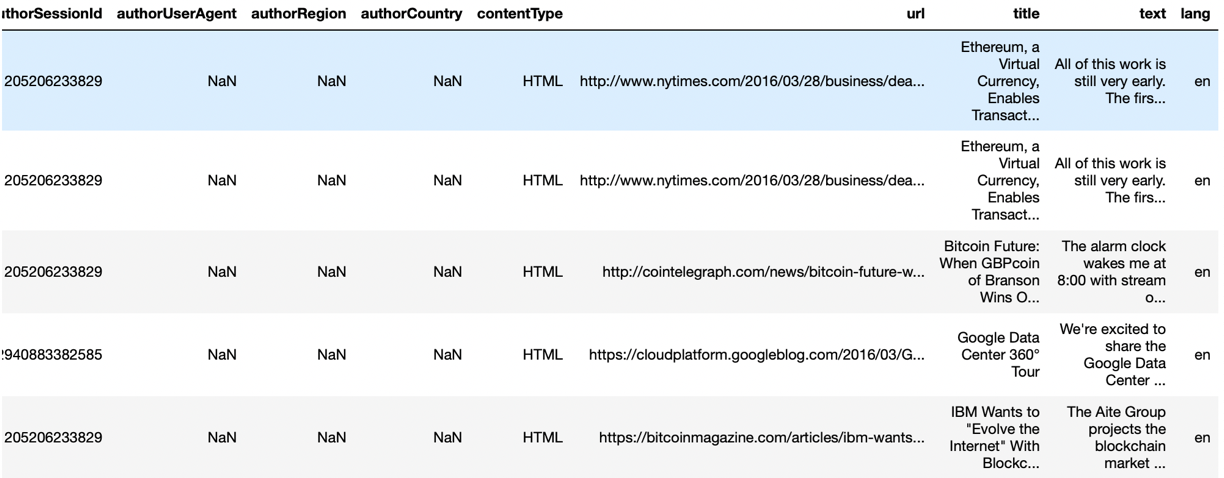
We mainly used the dataset from kaggle: https://www.kaggle.com/gspmoreira/articles-sharing-reading- from-cit-deskdrop. The two datasets are about the information of articles sharing and reading from CI&T DeskDrop. Deskdrop is an internal communications platform developed by CI&T, focused in companies using Google G Suite.Among other features, this platform allows companies employees to share relevant articles with their peers, and collaborate around them.

User Interaction dataset:



Articles dataset:





**Process Outline:**

1. Data Preprocessing

2. Exploratory Data Analysis

3. Text feature/keywords extraction and perform article auto-classification

4. Build recommendation system

5. Design of a pipeline and system to implement this approach and discussion on the system’s capabilities

6. Deploy the Model on Azure/AWS or Google Cloud Computing Platform

7. Build a web application to demonstrate the prediction and recommendation results.

**Deliverables:**

1.Generate a recommendation item list for a specific user; Generate a recommendation users list for a specific item;

2. Achieve article auto-classification and keywords extraction.

3. Further Steps: new article potential influence prediction and new contents suggestions for authors

**Use Cases:**

1. Select one topic of articles and get a list of top n popular articles on the platform

2. Get recommendation of articles based on the reading record and preference level.

3. Get recommendation of potential readers based on the reading record and preference level.

4. Check how popular an author’s articles are on the platform.

5. An author will get keywords extraction and article auto-classification to tel

**Milestones:**

|  |  |  |
| --- | --- | --- |
| No | Tasks | page2image49578176page2image51396128Timeframe |
| 1 | Environment set up & Data Preparation | Day 1-2 |
| 2 | Data Exploration & Set preference Level | Day 3-4 |
| 3 | Text classification & Keywords Extraction | Day 5-6 |
| 4 | Determine and implement algorithm to build the recommendation model and complete prediction | Day 7-8 |
| 5 | Evaluation & Presentation | Day 9-10 |