Medium Article Recommendation System

Arth Akhouri J003 | Gaurav Ankalagi J004 | Parisha Bhatia J008

Problem Statement

- Companies use recommender systems to increase sales and provide customers personalized offers and an enhanced customer experience.
- Providing recommendations can help content writers widen and reach their audience effectively.
- Medium articles have started gaining popularity among not only the data science community but also in different fields like literature, marketing and coding.

Solution

Our WebApp is able to achieve 3 things together

Recommendation System

Keyword Based Article Finder

Insights About Articles Via Visualizations

Scraping

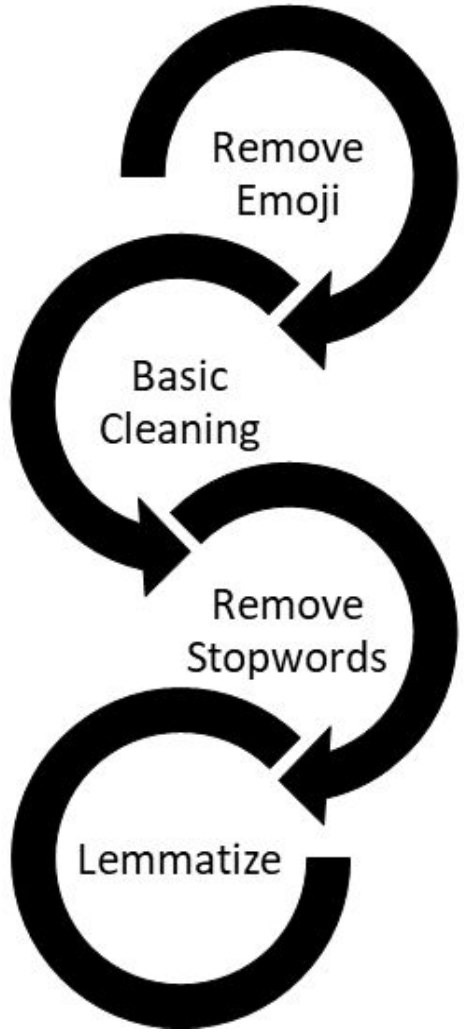
We used BeautifulSoup package to manually scrape approximately ~8300 articles spanning 6 months in around 5.5-6 hours.

The Following Information about each Article Was Collected

Date Published	Main Title	Subtitle
No. of Claps	No. of Responses	Article Author URL
Actual Article URL	Reading Time	Number of Sections
Section Title	Number of Paragraphs	Actual Text

Cleaning

- Remove emojis
- Removing links non alphanumeric characters, punctuations and convert to lowercase and remove newline characters
- Remove english and domain based stopwords
- Lemmatize the text to get root words



Topic Modelling

- TF-IDF
- Non-negative matrix factorization (NMF)
- Divides articles by words (original) into articles by topics and topics by words
- Output 10 different topics based on 15 different words

Topics

Business & Startups	Fundamental Coding
Data Science	Literature
Web Development	Social Media & Branding
Marketing & Sales	Team Dynamics
Cloud Development	Machine & Deep Learning

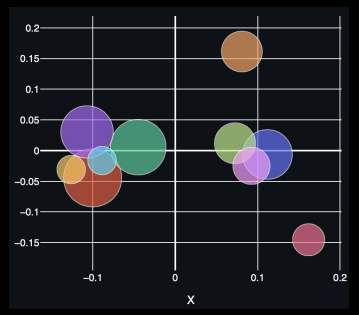
Recommendation System

- After topic modelling, we get scores for each article for each topic.
- We use sliders like these to accept scores for all the topics from the user.
- We then compare these scores with all the article scores.
- On the basis of the comparison, you are recommended an Article

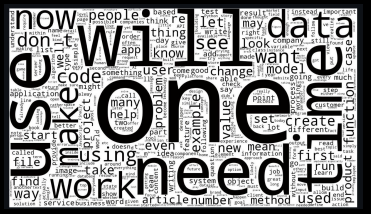
Visualizations

Our App Has A Range Of Visualizations Which Can Be Used To Derive Insights From The Various Articles

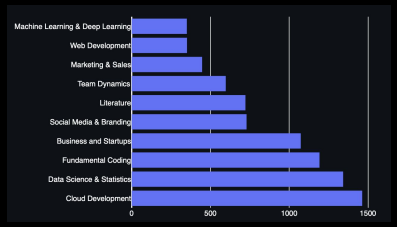
Radar chart



Bubble chart



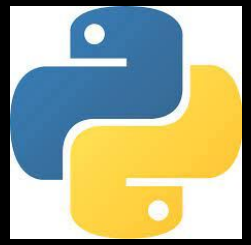
Word Clouds



Bar Graphs

Tech Stack

These are few of the important resources which we used to create the app.



Python

Scikit Learn

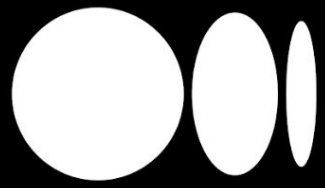
Plotly

Vaex

Heroku

Streamlit

Thank You



Questions?

Suggestions?