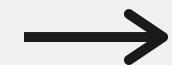


DILLON GONSALVES



KNOWLEDGE GRAPH USING SOCIAL MEDIA POSTS



Personal Information

Name - Dillon Gonsalves

College - Fr. Conceicao Rodrigues College of Engineering

Education stream - B.E Computer Science

Email ID - gonsalvesdillon17@gmail.com





Key takeaways:



Things to discuss:

- Problem statement
- Steps to solve the Problem Statement
- Skills required
- Outcome



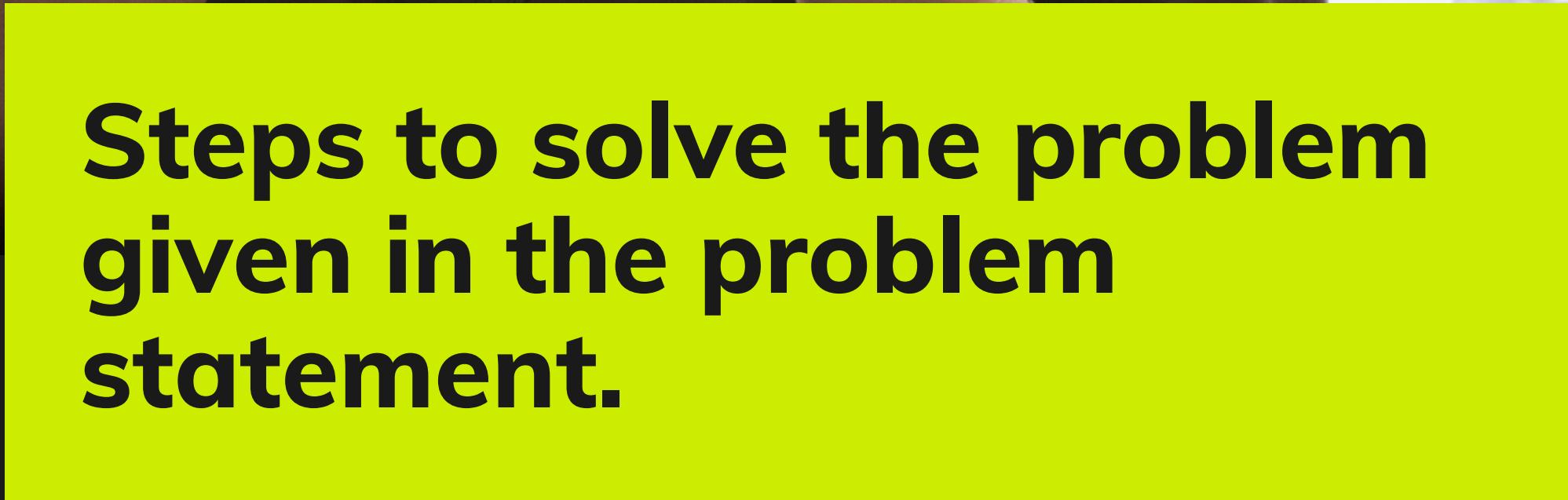


Problem Statement



Understanding customer sentiments about products is of utmost priority for any product company including HP. Consumers are more vocal on social media platforms and expressing their feedback and experience about any product. HP too need to know consumer sentiments firsthand so that it can make better products with great user experience and resolve customer issue faster. We need a one stop knowledge store, which can store reviews, suggestions, complaint and sentiments for all HP consumer printers/PC/laptop. This will help HP understand the consumers better and improve brand value and NPS score.





**Steps to solve the problem
given in the problem
statement.**





Step 1

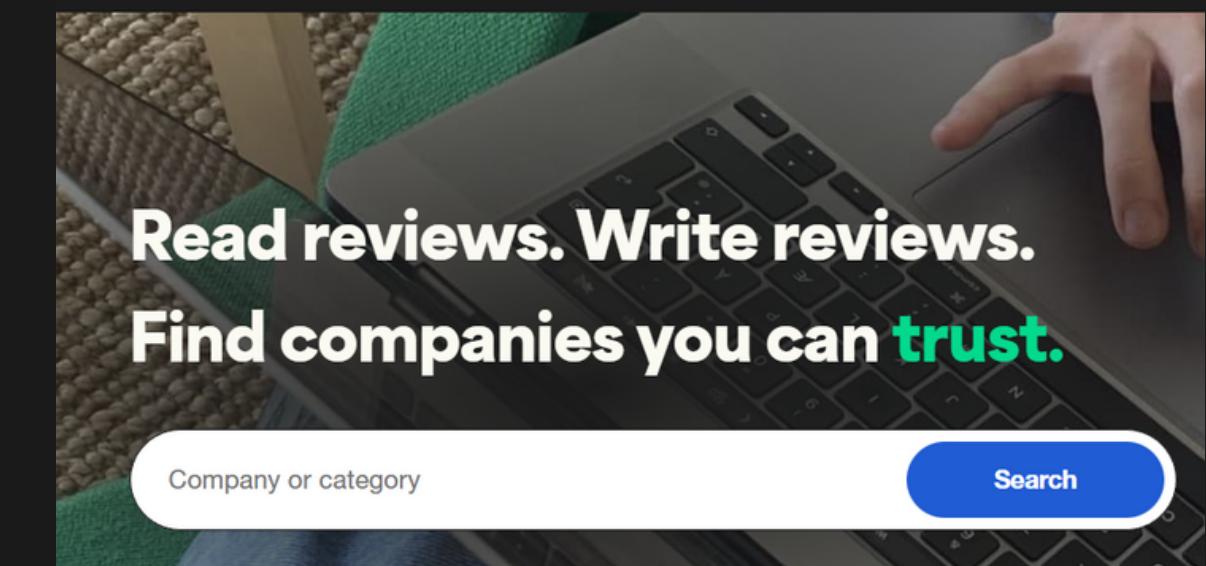
Popular web scraping tool BeautifulSoup was used to Scrape [Trustpilot](#) website which lets you write reviews about a products

```
import requests  
from bs4 import BeautifulSoup  
import re
```

Step 2

Set up accounts on the social media platforms you want to crawl. Trustpilot was used which lets user write reviews about any companies product.

★ Trustpilot





Step 3

Use the web scraping tool to scrape the social media platforms for posts containing specific keywords such as "HP PC" , "HP Printer" or "Model."

```
r = requests.get('https://www.trustpilot.com/review/hp.com')
soup = BeautifulSoup(r.text, 'html.parser')
regex = re.compile('.reviewContent.')
results = soup.find_all('p',{'class': 'typography_body-1__KUYFJ typography_appearance-'
reviews = [result.text for result in results]
reviews
```

- review
- 0 I purchased a £800 HP pavilion pc when I came...
 - 1 A TOTAL WASTE OF TIME HP.COM FAULTY...A TOTAL WA...
 - 2 Glad to see HP is doing anything they can to a...
 - 3 The customer service was just awful for our ag...
 - 4 I have been a customer of HP for years but tod...





Step 4

Used SentimentAnalyzer to find if review about a product is negative,positive or neutral

Id	neg	neu	pos	compound	review
0	0.078	0.862	0.060	-0.2737	Yet another complaint about HP Smart Printer s...
1	0.027	0.851	0.122	0.6955	HP MFP 4101fdwe DO NOT PURCHASEThis product do...
2	0.120	0.880	0.000	-0.7430	2 Broken printers from manifature. 5 hours of...
3	0.000	0.780	0.220	0.8520	Following up from my last review, case was acc...
4	0.119	0.753	0.129	0.0000	I have bought onsite support. My laptop will n...

Step 5

Data Preprocessing performed on data such as

- 1) Punctuation removing
- 2) Stopword
- 3) Tokenization
- 4) Lemmatization
- 5) LowerCase





Step 7

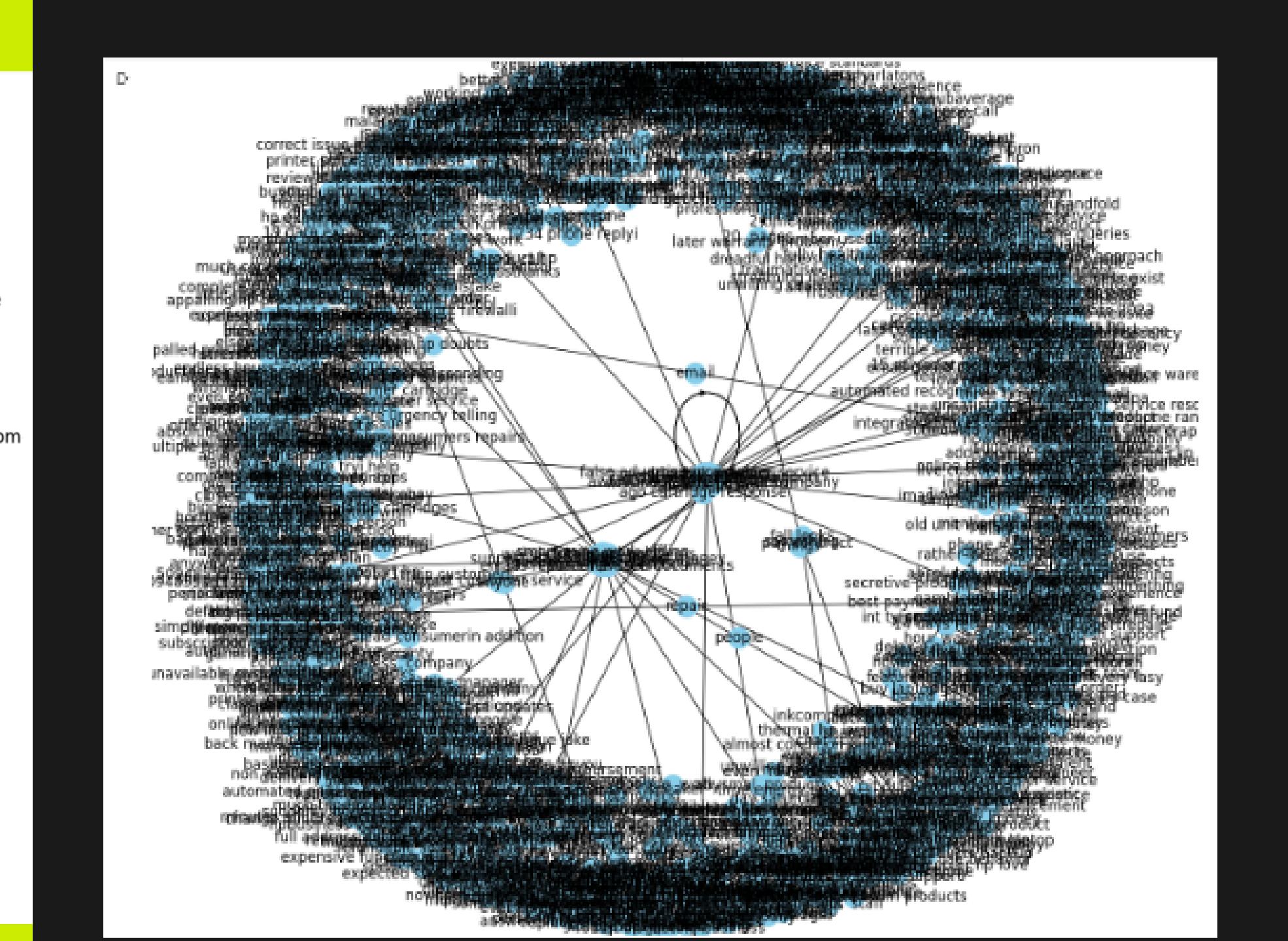
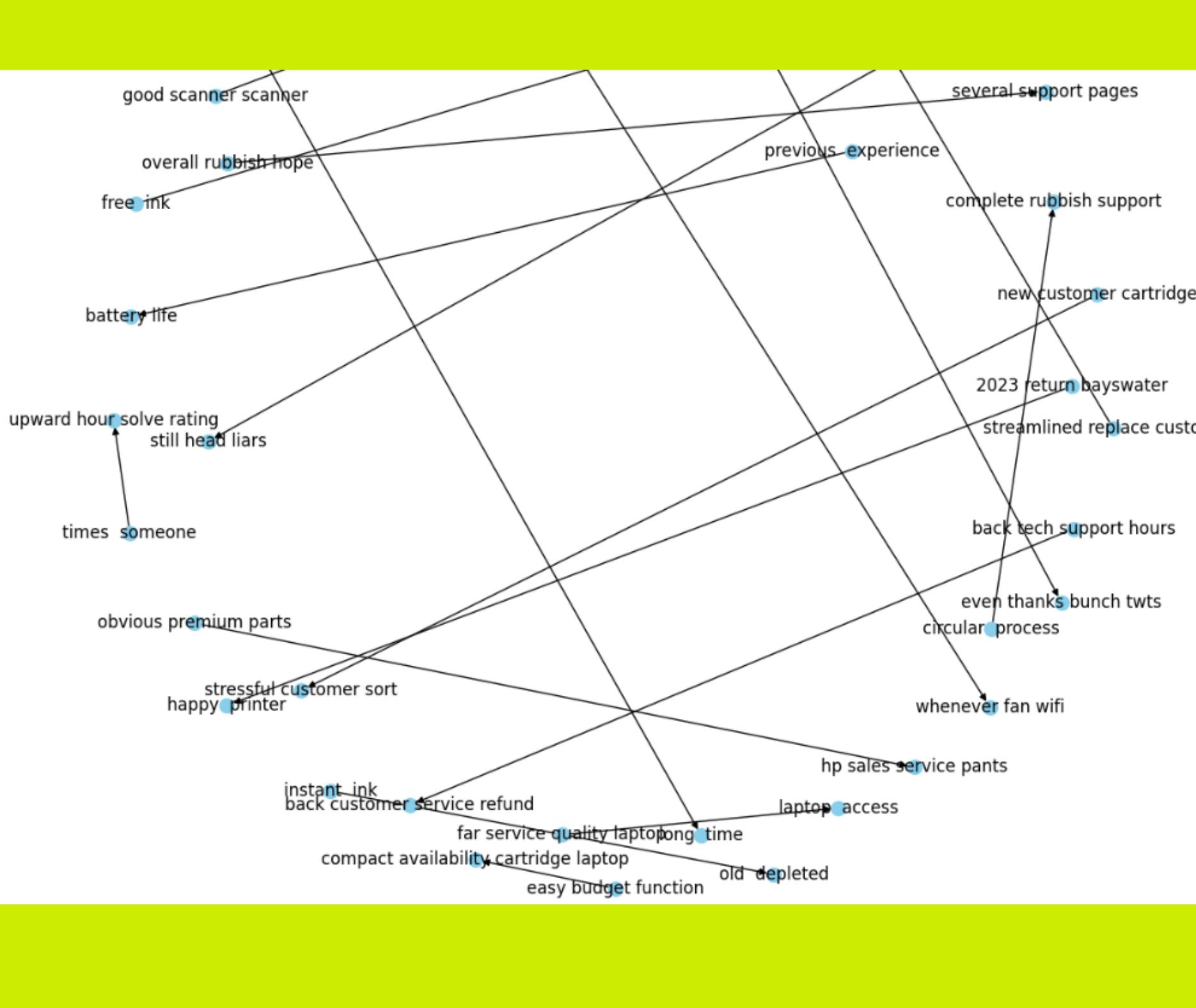
Final Preprocessed data is stored in
`FinalData.csv` file

<code>Id</code>	<code>review</code>	<code>token_review</code>	<code>token_review_lemm</code>
0	Yet another complaint HP Smart Printer softwa...	Yet another complaint HP Smart Printer softwa...	yet another complaint hp smart printer softwar...
1	HP MFP 4101fdwe DO NOT PURCHASEThis product ...	HP MFP 4101fdwe DO NOT PURCHASEThis product ...	hp mfp 4101fdwe purchasethis product connect h...
2	2 Broken printers manufacture 5 hours time w...	2 Broken printers manufacture 5 hours time w...	2 broken printers manufacture 5 hours time was...
3	Following last review case accelerated Con...	Following last review case accelerated Con...	following last review case accelerated contact...
4	I bought onsite support My laptop start I ...	I bought onsite support My laptop start I ...	bought onsite support laptop start contacted h...

Step 8

Create a knowledge graph that
maps the relationships between
brands, models, features,
problems, and sentiments.



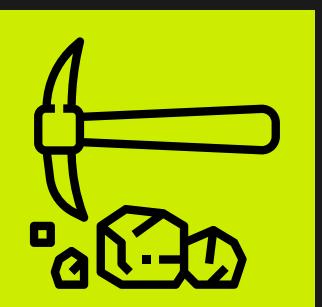




Skills Required



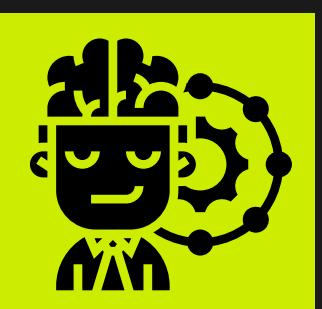
Web Scraping



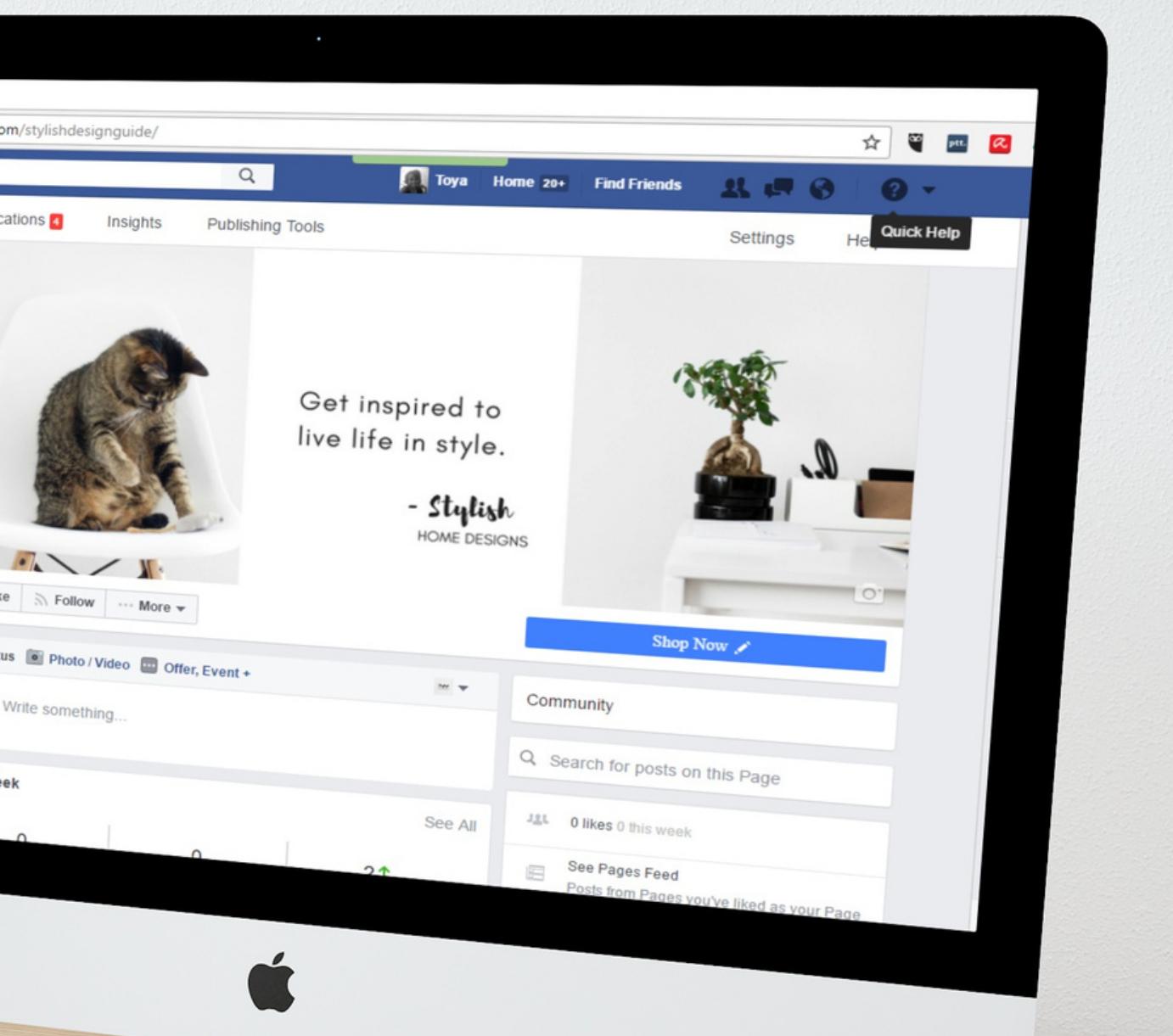
Web Data Mining



Natural Language Processing(NLP)



Machine Learning





Outcome:

The solution should be able to crawl social media platforms such as (facebook, twitter, LinkedIn etc.) for posts talking about HP PC and Printers.

- It should be able to classify/tag posts with a HP PC/printer brand or model, detect feature or problem the post is talking about and identify sentiments (complaint, suggestion or appreciation).
- The output should be in the form of a **knowledge graph** which can be queried by stakeholders for better customer service, product improvement and faster resolution.
The query can be like « List all posts talking about wifi issue in printer model X or brand Y ».





Brief Overview



To build the knowledge base first webscraping tool like BeautifulSoup was used to scrape Trustpilot website which gave all the reviews related to the HP products . After many Data preprocessing steps were taken to preprocessed the data like removinf stopwords, punctuations, tokenization , lemmatization etc. Sentimental Analysis was performed on the data to get Positive , Negative , Neutral reviews. To draw the knowledge graph Networkx library was used to show connections and NLP algorithm was used to display knowledge graph.



Thank

you
