LAB
MANUAL
PART A
(PART A: TO BE REFFERED BY STUDENTS)

Experiment No-
05

A.1 Aim:
Develop Content( text, emoticons, image, audio, video) based social media analytics model for business. (e.g.Content Based Analysis: Topic, Issue, Trend, sentiment/opinion analysis, audio, video, image analytics)

| | |
|---|---|
| Lab Objective | To design and develop social media analytics models |
| Lab Outcome | Design and develop content and structure based social media analytics models |

A-2 Prerequisite
Python

A.3 OutCome
Students will able to Collect, monitor, store and track social media data.

A.4 Theory:

Content Based Analysis:
Topic: Through topic-based analysis, businesses can identify the topics that are being discussed about them on social media. This helps businesses to better understand their customers' interests,needs, and opinions.
Issue: Issue-based analysis helps identify areas of concern and potential improvement for businesses. It can also help businesses identify areas of strength and weaknesses.
Trend: Trend-based analysis helps businesses keep track of their product and services on social media. It can also be used to identify opportunities for new products or services.
Sentiment/Opinion Analysis: Sentiment/opinion analysis helps businesses understand their customers' feelings and attitudes towards their products or services. It can be used to improve customer service or develop new products or services.
Audio Analysis: Audio analysis helps businesses identify the types of conversations that customers are having about their products or services. It can be used to identify customer feedback and to improve customer service.
Video Analysis: Video analysis helps businesses understand the types of videos that customers are watching about their products or services. This can be used to identify customer preferences and tocreate better content for customers.

Image Analysis: Image analysis helps businesses identify the types of images that customers Amazon Product Reviews Analytics: Topic Modeling of Amazon Product Reviews using LDA

With the boom in the number of online buyers and the simultaneous influx of reviews, understanding user experience is becoming an increasingly challenging task. Reviews talk volumes about a product, the seller and local partners. However, scraping such a myriad of customer feedback can be a tricky task. This tutorial helps you understand better ways of retrieving and structuring reviews of products to draw powerful insights. For our use case here, we will be using reviews of Amazon Echo. This study aims to use Latent Dirichlet Allocation (LDA) to perform topic modeling on amazon echo reviews. Collect the reviews for Amazon echo. The reviews were pre-processed, and the top 50 most frequent words were extracted. The words were then used to create a corpus, which was used to train the LDA model. After the model was trained, it was used to generate the topics in the reviews. The results shows some insights from topic modeling. They are as follows:

1. We observe that Topics 4 and 5 have some reviews in common. Reviews that talk about how Amazon Echo involves in everyday tasks seem to frequently compare Echo with Google Home.
2. A small interaction between Topics 2 and 5 indicate Echo was compared with Google Home on issues with Wi-Fi connectivity too.
3. The interaction between Topics 2 and 3 supports that few of the top problems that customers complain and compare on are Wi-Fi connectivity, answering simple questions and helping users in everyday tasks.
4. 'Good' is the third biggest contributor to Topic 8. This shows that speaker sound quality is a strong point and could be used as a positive point in advertising.
5. Topic 1 suggests that the Echo makes great gifts, especially during the Christmas season. Increased attention to advertising with this perspective is suggested during the Christmas season.

.Latent Dirichlet Allocation (LDA) is a machine learning technique used to uncover the hidden topics in a collection of documents. It is a generative probabilistic model that attempts to explain the co-occurrences of words in a document in terms of latent topics. Each document is assumed to be a mixture of a number of topics, and each topic is assumed to be a mixture of words. LDA can be used for document clustering, feature extraction, document summarization, and topic modeling. It is typically used to discover the topics that are present in a collection of documents, and to determine the probability of each topic in each document. It can also be used to identify the topics that are most closely related to a particular document or group of documents. To use LDA, the data must first be preprocessed, such as stemming and lemmatization. Then, the data must be vectorized, such as using bag-of-words or TF-IDF. Finally, the LDA model can be trained and applied to the data. The output of the model is a set of topics and their associated words.

Steps of LDA (Latent Dirichlet Allocation):

1. Choose the number of topics (k) you want to extract from the corpus.
2. Preprocess the reviews corpus by removing stop words, punctuations, and converting words to their root forms using stemming or lemmatization.
3. Create a vocabulary list of all unique words in the corpus.

4. Convert each review in the corpus into a bag-of-words representation, where each word isrepresented by its index in the vocabulary list and the count of that word in the review.
5. Initialize the model by randomly assigning each word in each review to one of the k topics.

6. For each review 'r' in the corpus, iterate through each word w in the review and calculate the probability distribution over the k topics, given the current assignments of all other words in the document to their topics and the current topic-word distribution.
7. Sample a new topic assignment for word w based on the probability distribution calculated in step6.
8. Repeat steps 6 and 7 for all reviewss in the corpus until convergence is achieved.
9. Output the topic-word distribution and document-topic distribution as the final result.

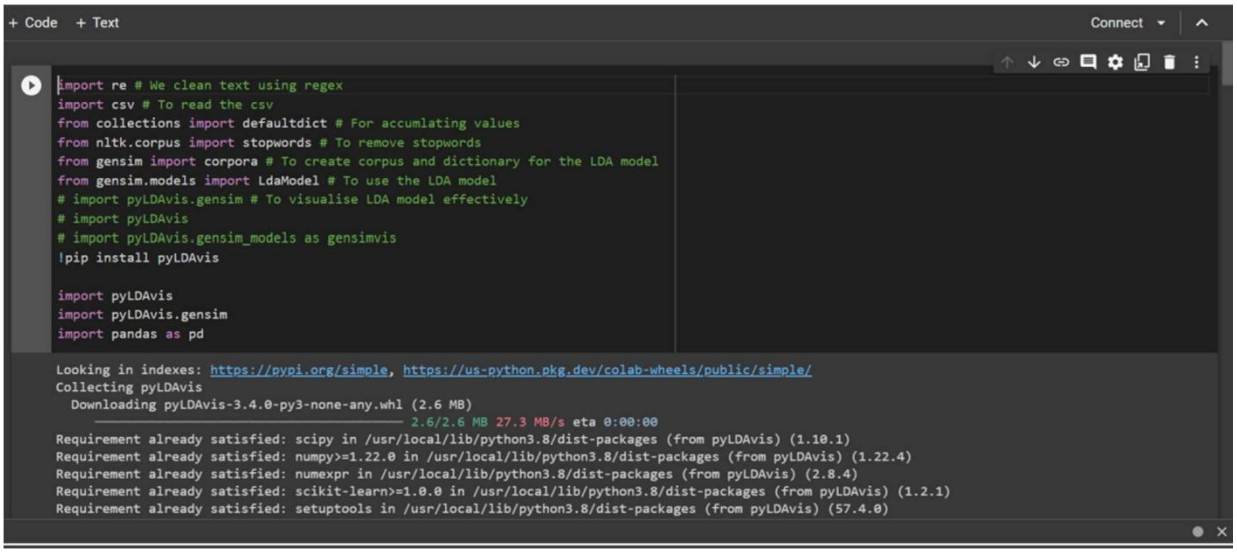
PART B

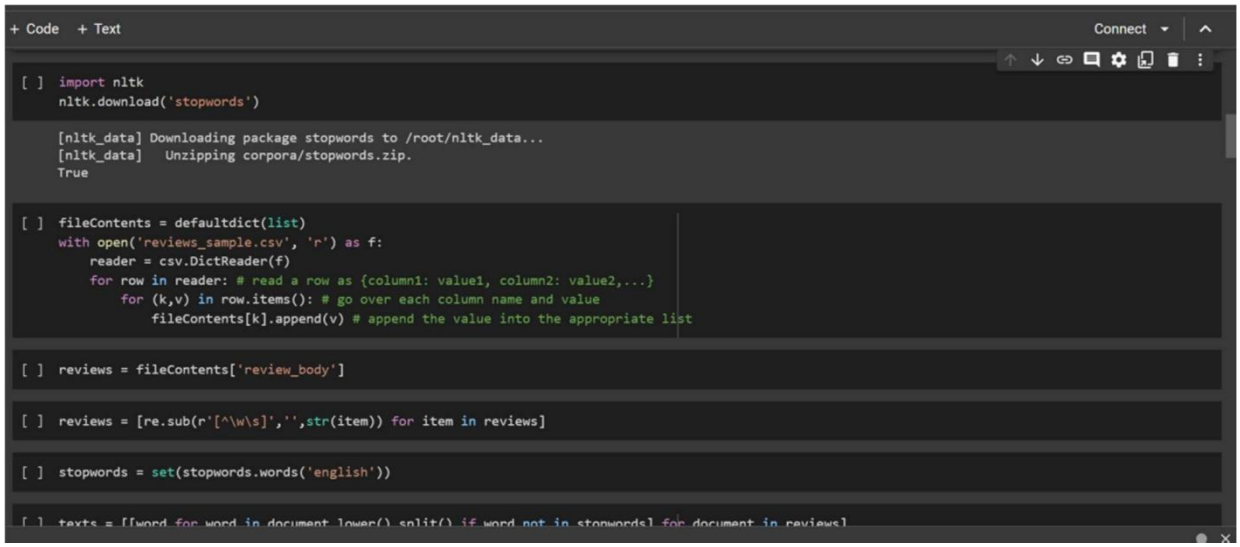
(PART B: TO BE COMPLETED BY STUDENTS)

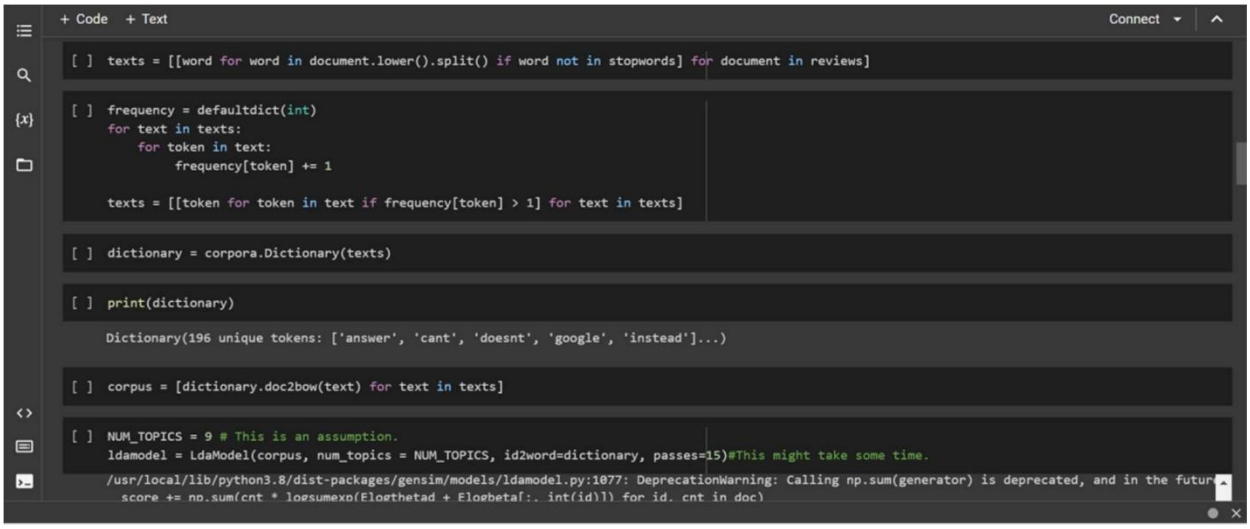
| | |
|---|---|
| **Roll. No.: C36** | **Name: Sanskruti Kadam** |
| **Class: BE** | **Batch: C2** |
| **Date of Experiment:** | **Date of Submission:** |
| **Grade:** | |

B.1 Study the fundamentals of social media platform and implement data cleaning,preprocessing, filtering and storing social media data for business:

B.2 Input and Output:



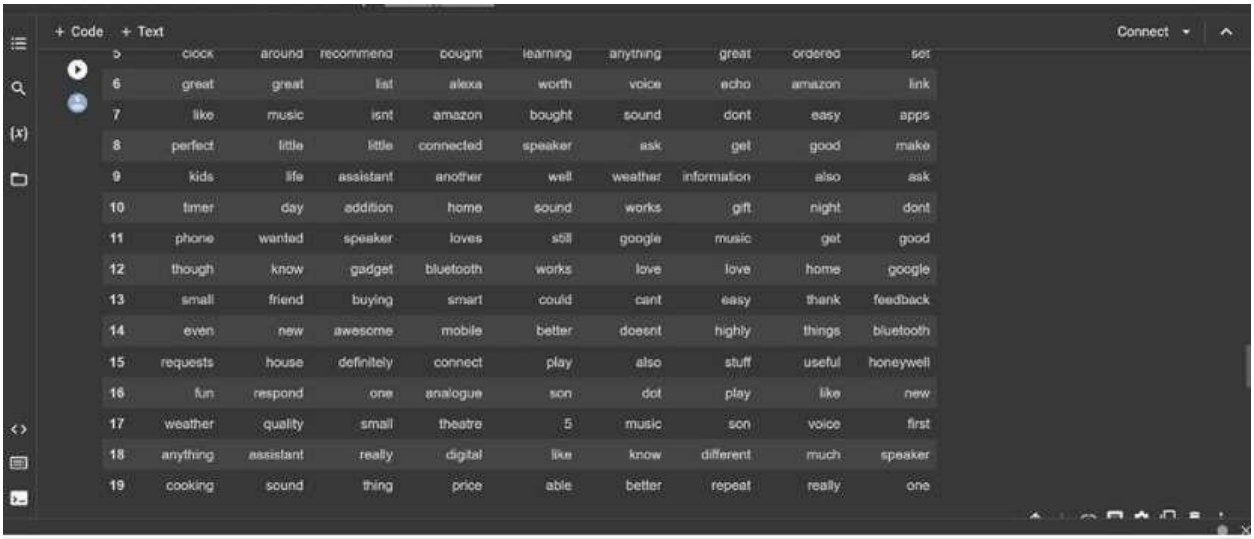





```
+ Code + Text                                                                    Connect

[ ] texts = [[word for word in document.lower().split() if word not in stopwords] for document in reviews]

[ ] frequency = defaultdict(int)
    for text in texts:
        for token in text:
            frequency[token] += 1

    texts = [[token for token in text if frequency[token] > 1] for text in texts]

[ ] dictionary = corpora.Dictionary(texts)

[ ] print(dictionary)

    Dictionary(196 unique tokens: ['answer', 'cant', 'doesnt', 'google', 'instead']...)

[ ] corpus = [dictionary.doc2bow(text) for text in texts]

[ ] NUM_TOPICS = 9 # This is an assumption.
    ldamodel = LdaModel(corpus, num_topics = NUM_TOPICS, id2word=dictionary, passes=15)#This might take some time.

    /usr/local/lib/python3.8/dist-packages/gensim/models/ldamodel.py:1077: DeprecationWarning: Calling np.sum(generator) is deprecated, and in the future
      score += np.sum(cnt * logsumexp(Elogthetad + Elogbeta[:, int(id)]) for id, cnt in doc)
```

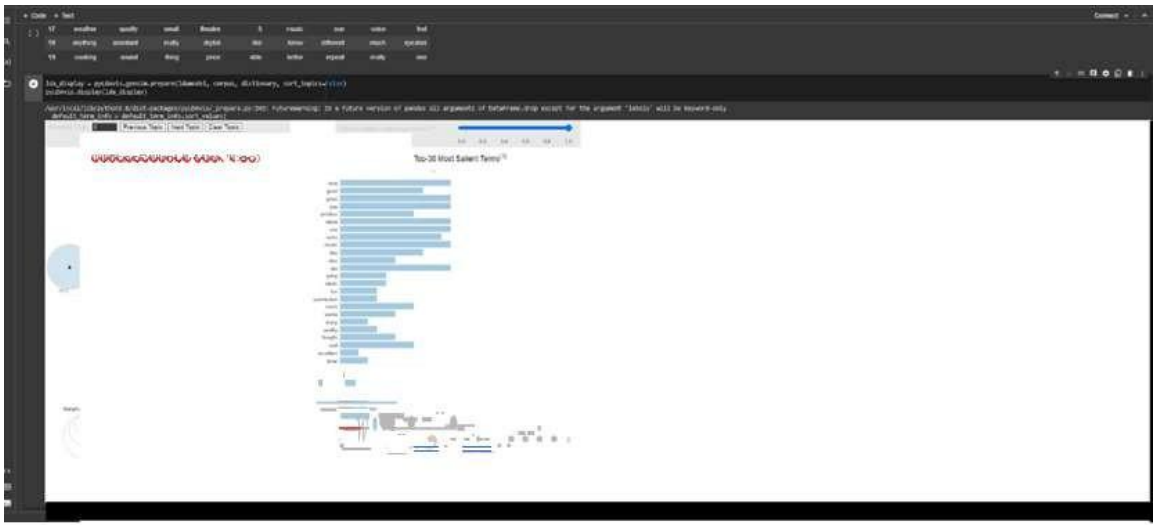
```
+ Code + Text                                                                    Connect

[ ] topics = ldamodel.show_topics()
    for topic in topics:
        print(topic)

    (0, '0.083*"love" + 0.051*"excellent" + 0.038*"music" + 0.035*"works" + 0.035*"alarm" + 0.035*"clock" + 0.035*"great" + 0.035*"like" + 0.035*"perfect"
    (1, '0.085*"alexa" + 0.057*"love" + 0.056*"echo" + 0.046*"dot" + 0.029*"well" + 0.029*"around" + 0.022*"great" + 0.020*"music" + 0.020*"little" + 0.02
    (2, '0.081*"good" + 0.066*"great" + 0.044*"price" + 0.044*"also" + 0.044*"quality" + 0.023*"recommend" + 0.023*"list" + 0.023*"isnt" + 0.023*"little"
    (3, '0.048*"dot" + 0.048*"connection" + 0.039*"echo" + 0.029*"laptop" + 0.029*"crackle" + 0.029*"bought" + 0.020*"alexa" + 0.020*"amazon" + 0.020*"cor
    (4, '0.054*"product" + 0.054*"love" + 0.054*"fun" + 0.054*"enjoy" + 0.050*"great" + 0.037*"learning" + 0.037*"worth" + 0.034*"bought" + 0.019*"speaker
    (5, '0.052*"much" + 0.042*"great" + 0.034*"alexa" + 0.033*"really" + 0.032*"use" + 0.030*"anything" + 0.028*"voice" + 0.027*"sound" + 0.025*"ask" + 0.
    (6, '0.071*"one" + 0.041*"like" + 0.038*"set" + 0.033*"alarm" + 0.031*"time" + 0.027*"great" + 0.026*"echo" + 0.026*"dont" + 0.026*"get" + 0.025*"info
    (7, '0.075*"use" + 0.054*"product" + 0.051*"music" + 0.031*"im" + 0.030*"great" + 0.027*"ordered" + 0.027*"amazon" + 0.027*"easy" + 0.021*"good" + 0.0
    (8, '0.056*"alexa" + 0.030*"amazon" + 0.025*"would" + 0.021*"well" + 0.021*"get" + 0.018*"set" + 0.017*"link" + 0.017*"apps" + 0.017*"make" + 0.016*"a

[ ] word_dict = {};
    for i in range(NUM_TOPICS):
        words = ldamodel.show_topic(i, topn = 20)
        word_dict['Topic # ' + '{:02d}'.format(i+1)] = [i[0] for i in words]
    pd.DataFrame(word_dict)
```

| | Topic # 01 | Topic # 02 | Topic # 03 | Topic # 04 | Topic # 05 | Topic # 06 | Topic # 07 | Topic # 08 | Topic # 09 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | love | alexa | good | dot | product | much | one | use | alexa |



```
+ Code + Text                                                                    Connect
```

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 5 | clock | around | recommend | bought | learning | anything | great | ordered | set |
| 6 | great | great | list | alexa | worth | voice | echo | amazon | link |
| 7 | like | music | isnt | amazon | bought | sound | dont | easy | apps |
| 8 | perfect | little | little | connected | speaker | ask | get | good | make |
| 9 | kids | life | assistant | another | well | weather | information | also | ask |
| 10 | timer | day | addition | home | sound | works | gift | night | dont |
| 11 | phone | wanted | speaker | loves | still | google | music | get | good |
| 12 | though | know | gadget | bluetooth | works | love | love | home | google |
| 13 | small | friend | buying | smart | could | cant | easy | thank | feedback |
| 14 | even | new | awesome | mobile | better | doesnt | highly | things | bluetooth |
| 15 | requests | house | definitely | connect | play | also | stuff | useful | honeywell |
| 16 | fun | respond | one | analogue | son | dot | play | like | new |
| 17 | weather | quality | small | theatre | 5 | music | son | voice | first |
| 18 | anything | assistant | really | digital | like | know | different | much | speaker |
| 19 | cooking | sound | thing | price | able | better | repeat | really | one |



### B.2 Observations and learning:

With the boom in the number of online buyers and the simultaneous influx of reviews, understanding user experience is becoming an increasingly challenging task. Reviews talk volumes about a product, the seller and local partners. However, scraping such a myriad of customer feedback can be a tricky task.

### B.3 Conclusion:

In this experiment we have develop Content based social media analytics model for business by analyzing Amazon Product Reviews Using LDA Topic Modelling

### B.4 Question of Curiosity

Q1. Explain in detail; why social media data collection is important?

Ans: By collecting data, it becomes easier to know which of the social media platforms are most popular among your audience. Social media data collection also ensures that you know exactly what type of content appeals to your audience, what time of the day you should post, and howto achieve maximum reach.

Q2. Explain: What kind of social media data should you track?

Ans: Your social media goals are what determine your metrics. For every goal, you need a related metric, which will help determine if your social strategy is hitting the mark or not. For example, your business goal may be to increase conversions. Therefore, your social media goal becomes increasing conversions from those that visit your site via posts that are part of your strategy. Now that you have a goal in mind, you can clearly identify which social media metrics to measure and a time frame in which to measure them.

Q3. What is social listening?

Ans: Social listening, also referred to as social media listening, is the process of identifying and assessing what is being said about a company, individual, product or brand on the internet. Conversations on the internet produce massive amounts of unstructured data.

Q4. What is Facebook pixel? Explain the working of Facebook pixel with suitable case study.

Ans: Facebook Pixel is a tracking code that is placed on a website to help businesses measure the effectiveness of their Facebook ads. It is a piece of code that is added to a website and is used to track conversions, optimize ads, and build target audiences. Essentially, it tracks user behavior on awebsite, and this information is then used to create targeted Facebook ads.

The Facebook pixel works by placing a cookie on the user's browser when they visit a website. This cookie tracks the user's behavior on the website and reports it back to Facebook. This information is then used to create targeted ads that are shown to the user on their Facebook feed.

A case study that illustrates the working of Facebook Pixel is the following:

Let's say that a fashion e-commerce website wants to increase its sales by reaching out to people who have shown an interest in their products. They can use Facebook Pixel to track the behavior ofpeople who have visited their website and create a targeted ad campaign to reach out to these people.

The first step would be to add the Facebook Pixel code to their website. This can be done by placingthe code in the header of the website. Once the code is added, the Pixel starts tracking the behavior of users on the website.

The next step would be to create a custom audience in Facebook Ads Manager. This audience would consist of people who have visited the website in the last 30 days. The custom audience canbe created using the data collected by the Pixel.

The final step would be to create a targeted ad campaign to reach out to this custom audience. Thead campaign can be optimized using the data collected by the Pixel, such as which products were viewed the most, which pages were visited the most, and so on.

By using Facebook Pixel, the fashion e-commerce website can create a targeted ad campaign that reaches out to people who have shown an interest in their products. This can lead to increased salesand better return on investment for the ad campaign.