Chapter 5 - NLP operations

Text Mining in NLP

The process of converting unstructured text into a structured format with the purpose of identifying significant patterns and fresh insights is known as text mining, also known as text data mining. Companies are able to investigate and find hidden links within their unstructured data by utilizing advanced analytical approaches like Naive Bayes, Support Vector Machines (SVM), and other deep learning algorithms.

Although in everyday speech the phrases text mining and text analytics are mostly interchangeable, they can also indicate different things. Text mining and text analysis combine machine learning, statistics, and linguistics to find textual patterns and trends in unstructured data. More quantitative insights can be discovered using text analytics by putting the data into a more structured format through text mining and text analysis. Then, you can use data visualization techniques to share your findings with more people.

Methods and techniques

There are different methods and techniques for text mining.

1. **Basic Methods**

    ▪ *Word frequency*
    The most frequent terms or ideas in a batch of data can be found using word frequency. When examining customer reviews, social media interactions, or customer feedback, discovering the most frequently used words in unstructured text can be especially helpful.
    For instance, it may be a sign that you need to modify your rates (or your target market!) if the words pricey, overpriced, and overrated frequently appear in your customers' reviews.

    ▪ *Collocation*
    A group of words that frequently appear together are referred to as a collocation. Bigrams (a pair of words that are likely to go together, such as start, save time, or make a decision) and trigrams are the most prevalent kinds of collocations (a combination of three words, like within walking distance or keep in touch).
    Collocations should be recognized and counted as a single word to increase the text's granularity, better grasp its semantic structure, and ultimately produce more accurate text mining results.

- *Concordance*

When a word or group of words appears in a specific context or instance, concordance is utilized to identify it. It is common knowledge that a word can have multiple meanings depending on the situation. Understanding a word's precise meaning depending on context can be achieved by reviewing its concordance.

Here are a few examples from reviews that included the term "work," for instance:

Preceding context	Target	Following context
It saves time and helps teams	work	more efficiently.
Some advanced features only	work	in one language (English)
It enables us to	work	towards better conversion and retention.
We recommend this to several of the small businesses we	work	with, and they are all happy with the results.

2. Advance Methods

- *Text Classification*

Giving unstructured text data categories (tags) is the process of text classification. Complex text may be easily arranged and structured into useful data thanks to this crucial Natural Language Processing (NLP) activity.

Businesses may analyze a variety of data, including emails and support tickets, using text classification, and quickly and affordably gain insightful information.

We'll discuss some of the most well-liked text classification assignments below:

- *Topic Analysis***:** One of the most common methods of organizing text data is through topic analysis, which enables you to comprehend the primary ideas or subjects of a text. For instance, a support request informing me that an online order hasn't arrived falls within the category of shipping issues.

- *Sentiment Analysis*: It involves examining the feelings behind each particular text. Consider that you are studying a number of user reviews of your mobile application. You might discover that UI-UX or ease of use are commonly discussed issues in those assessments, but that information is insufficient to draw any conclusions. Understanding the opinions and feelings expressed in a text and categorizing them as positive, negative, or neutral is made possible through sentiment analysis. Sentiment analysis has several practical commercial uses, such examining social media posts, reviews, or support tickets. For example, in customer service, you might be able to spot irate consumers right away and give their issues first priority.

▪ *Language Detection***:** It allows you to group texts according to their language. Routing support tickets to the appropriate team's location automatically is one of its most practical applications. Teams can save precious time by automating this process, which is rather simple.

▪ *Intention Detection***:** To automatically determine the goals or purpose of a text, you may use a text classifier. When examining client chats, this can be quite helpful. By sorting through various outbound sales email answers, you might, for instance, separate the prospects who are interested in your product from those who are not or who want to unsubscribe.

Text Extraction

A text analysis approach called text extraction pulls out specific data from a text, such as keywords, entity names, addresses, emails, etc. Businesses can save themselves the effort of manually combing through their data to find the important information by employing text extraction.

The majority of the time, combining text extraction and text categorization in the same analysis can be effective.

- *Keyword Extraction***:**
  The terms that are most pertinent to a text are its keywords, which can be used to sum up its content. You can index data for searching, sum up a text's substance, or produce tag clouds, among other things, by using a keyword extractor.

- *Named Entity Recognition***:**
  It gives you the ability to locate and extract names of businesses, organizations, or people from a text.

- *Feature Extraction***:**
  It aids in locating specific traits of a good or service in a collection of data. For instance, you might quickly extract features like color, brand, model, etc. from product descriptions.

Need for Text Mining

Unstructured text data in a company setting can include emails, comments made on social media, conversations, support tickets, surveys, etc. Manually sorting through all of this information frequently fails. Not only is it costly and time-consuming, but it is also imprecise and impractical to scale.

However, text mining has shown to be a dependable and economical method for achieving accuracy, scalability, and quick reaction times. More specifically, the following are some of its key benefits:

- *Scalability***:**
  Text mining makes it feasible to quickly analyze huge amounts of data. Companies can save a lot of time that can be used to concentrate on other things by automating particular processes. Businesses as a result become more successful. Text mining makes it feasible to quickly analyze huge amounts of data. Companies can save a lot of time that can be used to concentrate on other things by automating particular processes. Businesses as a result become more successful.

- *Real-time analysis***:**
  Thanks to text mining, businesses may identify impending crises; identify product faults in real time, and priorities essential concerns accordingly. Why is this big deal? Because it enables businesses to act quickly.

- *Consistent criteria***:**
  People are more prone to error when performing repetitive, manual jobs. Additionally, they struggle with consistency and unbiased data analysis. Take tagging as an illustration. For the majority of teams, adding categories to emails or support tickets takes effort and frequently results in mistakes and inconsistencies. Automating this process not only provides for more accurate results and ensures that a consistent set of standards is applied to every ticket, but it also saves a lot of valuable time.