

Prepared by Dr. Karim Ahmed

MKT631

Digital Analytics



The background features a series of vertical panels with rounded ends. The first panel on the left shows a person's hands typing on a laptop. The subsequent panels are semi-transparent and contain various white line-art icons: a graduation cap, a play button, a target with an arrow, an open book, a magnifying glass, and a group of people. The overall color scheme is warm, with shades of orange and yellow.

MKT631 **Digital Analytics**

Lecture 10

Prepared by
Dr. Karim Ahmed

Advanced AI and Algorithms

Objectives:

1. Understand the latest advancements in Natural Language Processing (NLP) and AI algorithms.
2. Explore how models like GPT-3, GPT-4o, and GPT-5 process and respond to human input.
3. Examine the role of text analytics in digital marketing and data analysis.
4. Learn key techniques such as sentiment analysis, topic and trend mining, intention mining, and concept mining.
5. Understand the text analytics process, including tokenization, stemming, n-grams, and similarity matrices.

Lecture Content:

- Introduction
- How GPT-3 (and Future Versions) Responds to the Questions Humans Ask
- GPT-4o
- GPT-5
- Algorithms and Text Analysis

One of the most exciting aspects of digital analytics today is the rapid advancement of **natural language processing (NLP)** techniques powered by deep learning and artificial intelligence. These advancements enable more sophisticated and accurate analysis of text data, leading to exciting possibilities.

- **Contextual Understanding:** Modern NLP models, such as GPT, and their variants, have significantly improved the ability to understand the context and meaning of the text. These models can capture nuanced relationships, contextual clues, and subtle nuances, enabling more accurate sentiment analysis, entity recognition, and language understanding tasks.

- **Multilingual and Cross-Lingual Analysis:** With the advancements in NLP, there has been progress in developing models that can handle multiple languages and perform cross-lingual analysis. The multilingual analysis opens up opportunities for analyzing and extracting insights from text data in various languages, facilitating global applications and communication.
- **Domain-Specific Text Analytics:** NLP models can be fine-tuned or specialized for specific domains or industries, such as healthcare, finance, legal, or customer service. This domain-specific focus enhances the accuracy and applicability of text analytics techniques in specific contexts, leading to more tailored and meaningful insights.

- **Ethical and Bias Mitigation:** Text analytics researchers and practitioners are increasingly addressing concerns related to bias, fairness, and ethical considerations in NLP models and applications. Efforts are being made to develop methods that reduce biases, promote inclusivity, and ensure the responsible use of text analytics in sensitive domains.
- **Interactive Visualization and Explainability:** As text analytics techniques evolve, so do the tools and platforms that enable interactive visualization and explainability of the results. This visualization allows users to understand better and interpret the outcomes of text analytics models, enhancing transparency, and fostering trust in the analysis.

How GPT-3 (and Future Versions) Responds to the Questions Humans Ask

Communications between humans and these new intelligent chatbots are created and delivered through a web-based interface or API (Application Programming Interface). When a person interacts with GPT-3 through an interface, the input provided by the user is sent to OpenAI's servers, where the language model resides. The input can be in the form of text, typically a prompt or a question. It could be a single sentence or a paragraph, depending on the complexity of the desired interaction.

Once the input is received, GPT-3 processes it and generates a response based on its understanding of the input and the patterns it has learned from its training data. The generated response is then sent back to the user through the interface, typically as text.

How GPT-3 (and Future Versions) Responds to the Questions Humans Ask

The process of generating a response involves several steps. GPT-3 uses its understanding of the context provided by the input and attempts to generate a coherent and contextually relevant response. It can consider a wide range of factors, including the wording, tone, and intent of the input, as well as any prompts or instructions that may have been given.

It's important to note that GPT-3's responses are generated based on statistical patterns learned from training data and do not have true understanding or consciousness. The model does not have access to real-time information or personal experiences beyond what it has learned during training. It operates solely based on the patterns and information present in the text it was trained on.

GPT-4o builds on GPT-3's method but adds new abilities. It still takes your text and returns a text reply, but it can also work with images, and in some apps it can talk back to you with real-time voice. GPT-4o understands context better, so it keeps track of longer conversations more reliably. It is faster and more accurate than GPT-3, and its replies feel more natural. Like GPT-3, GPT-4o does not have true awareness or access to live information unless it is connected to extra tools.

Feature	GPT-3	GPT-4o
Input/Output Types	Text only	Text, images, and (in some apps) real-time voice
Context Handling	Good for short to medium texts	Excellent for very long or complex conversations
Speed	Fast, but can slow on long prompts	Faster overall, low latency even with complex requests
Accuracy	High, but occasionally makes mistakes	Higher, with more coherent and on-topic responses

Feature	GPT-3	GPT-4o
Live Information Access	No (limited to training data)	No by default (can connect to plugins for live data)
Understanding “Awareness”	Pattern-based, no real understanding	Still pattern-based, no true consciousness
Special Capabilities	Strong text generation, translation	Multimodal (text, image, voice), better code and math
Ideal Use Cases	Chatbots, content drafting, summaries	Interactive assistants, multimedia content, advanced chat

August 7, 2025 Release Product

Introducing GPT-5

Our smartest, fastest, most useful model yet, with built-in thinking that puts expert-level intelligence in everyone's hands.

Try on ChatGPT ↗

GPT-5 is a next-generation large language model that combines advanced multi-step reasoning with expanded context windows and multimodal input capabilities. It can generate coherent, context-aware text, interpret images and structured data, and perform complex, chained tasks useful for business applications. In digital marketing and data analytics, GPT-5 accelerates content production, enables deep personalization at scale, and helps produce reproducible analysis pipelines. Its practical benefits require rigorous validation, strong governance, and continuous human oversight to manage hallucinations, bias, and privacy concerns.

Advantages of GPT-5 for digital marketing and digital data analytics:

- Advanced multi-step reasoning

GPT-5 can follow and execute complex, chained analytical instructions, enabling it to decompose marketing problems (e.g., attribution) into actionable steps. This reduces the manual orchestration required to turn a business question into a reproducible analysis plan.

- Expanded context window

The model can ingest much longer campaign histories, customer journeys, and aggregated reports in a single prompt, which supports more coherent cross-period insights and trend detection. Marketers can therefore ask higher-level questions that require seeing the full context rather than fragmented snippets.

- Multimodal understanding

GPT-5 can interpret and combine text, images, and structured data, allowing it to analyze creative assets alongside performance metrics. This capability is valuable for diagnosing which visual elements correlate with higher engagement or conversion.

- High-quality content generation at scale

It produces polished marketing copy, subject lines, landing page drafts, and social posts that maintain brand tone, enabling rapid iteration and multivariate testing of creatives. This shortens campaign production cycles and supports continuous optimization.

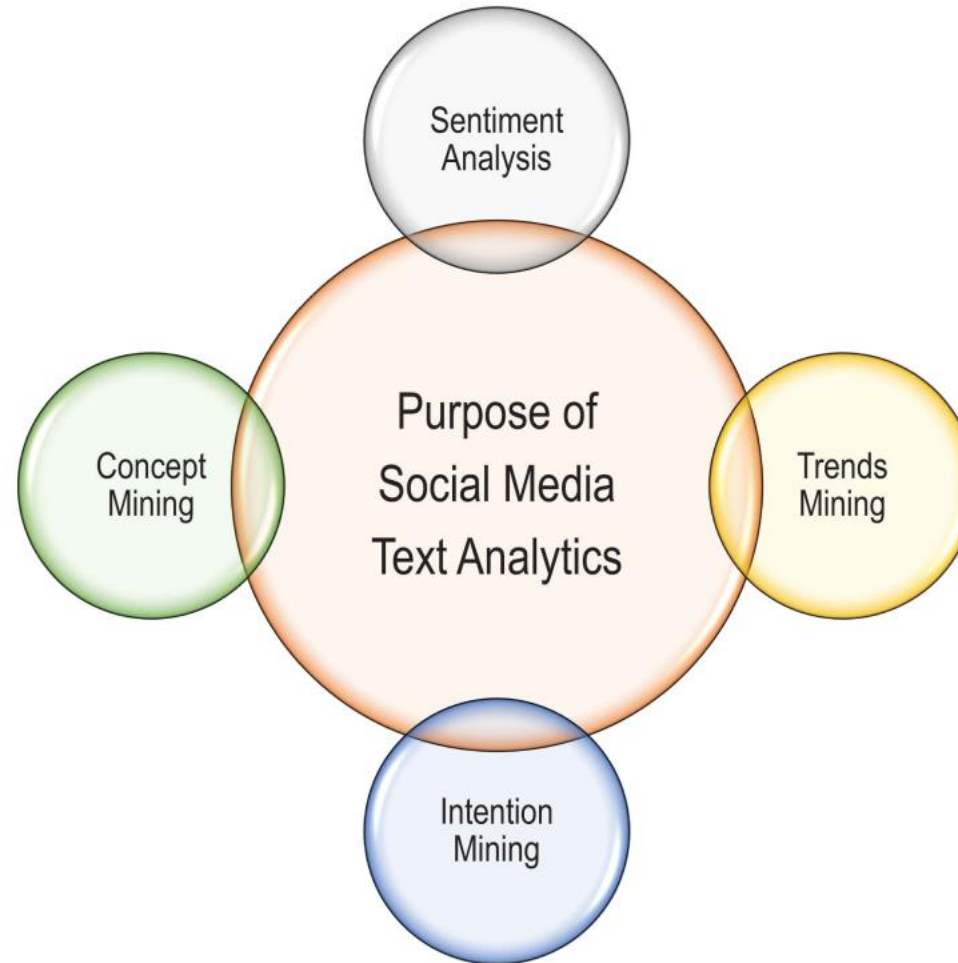
- Personalization and dynamic messaging

By conditioning on individual-level attributes and interaction history, GPT-5 can generate tailored messages and recommendations for diverse audience segments.

Reasons for Using Text Analytics:

- Hearing from people provides valuable insights.
- Text can be utilized alongside statistical measures to build predictive models.
- Textual data can give us very useful insights in a variety of use cases such as customer sentiment, healthcare (patient and hospital), financials (stock trading is especially hot), voluntary opinion mining (with and without survey data).
- Text analytics can be used to get customer sentiment in ways customers are reluctant to share voluntarily by mining their textual data.

Purposes of Text Analytics



Algorithms and Text Analysis

Purposes of Text Analytics

1- Sentiment Analysis

Sentiment analysis is an aspect of text analytics that focuses on opinion mining. Sentiment analysis uses computational methods to gather sentiment, opinions, appraisals, and emotions from the text.

Sentiment analysis can be used to understand how others see the world regarding the data being studied. Sentiment analysis works by comparing the words in documents to external tag lists of words (called lexicons) that are classified as positive or negative.

However, the meaning and sentiment of the terms can change, depending on the context they are used in.

Purposes of Text Analytics

1- Sentiment Analysis

The overall balance of positive to negative words determines the sentiment score of a document (this can be done on the sentence level as well). The type of list (lexicon/dictionary) is often subjective and applied arbitrarily by whoever composes the list; these definitions may or may not work well in every situation.

Purposes of Text Analytics

2- Topic and Trends Mining

Another reason to use text analytics is to understand what the text in the document is about. Use cases span from scholarly analysis of Shakespeare plays to risk detection, and everything in-between. We could read a document and form an opinion about topics, but someone else could read the same document and come away with an entirely different set of topics and meanings. Besides, when there are thousands of documents to understand, manually reading large swaths of data is not feasible. Consequently, topic and trend mining are usually performed on large sets of data.

Purposes of Text Analytics

2- Topic and Trends Mining

Researchers can use deep learning to automatically try to make sense of themes and the trends that lie within a document or set of documents, but that doesn't work too well unless the text analytics software is manually trained beforehand on documents that are very close in subject, language, syntax, and form to the documents being analyzed.

Purposes of Text Analytics

3- Intention Mining

Intention mining is what the Google search engine does when you enter a search query into Google, and it figures out what you mean based on your query. Google doesn't stop there, it also analyzes queries and webpages for what people did not explicitly say but what they intended to say, to understand the underlying opinions expressed in the document. Intention mining is a Text analytics operation that is a big part of search engine optimization.

Purposes of Text Analytics

3- Intention Mining

Intention mining can be done in several ways and uses some of the following text analytics operations:

- **Named entity recognition: NER** looks at recognizing nouns and could be used to extract persons, organizations, geographic locations, dates, monetary amounts, or the like from text, this works by looking at the words surrounding them.
- **Event extraction:** Determines the relationships and the events connected with named entities (NER). Establishes the kinds of inferences that can be made from the text.

Purposes of Text Analytics

4- Concept Mining

Text mining is the discipline of extracting information from a document. Concept mining attempts to do the same thing using the concepts in the document. Mapping words to concepts is ambiguous. Each word in a document may relate to several possible concepts.

Purposes of Text Analytics

4- Concept Mining

For example, the Google search engine indexes billions of new documents each day; using concept mining, it will automatically group similar documents together, as representative of a specific idea, along with other documents it has already categorized with the same or similar concept. Google can further refine its content classification by clustering the documents its index by topic; it uses this information to rank search results for a searcher's query better.

The Text Analytics Process

In the process of transforming text to numbers, there are many manipulations on the data, including tokenization, so we can find named entities in the text.

- Stop words (such as “the”, “of”, “and”, “a”, “to”, and so on) are removed.
- All remaining words are made regular by a process sometimes called stemming.

Lemmatization is a stemming operation that regularizes words while trying to figure out their part of speech. Caveat: How words get reduced to a stem may differ depending on how the words are used. For instance, the noun “moped” should not be stemmed into the form “mope”, while the verb “moped” should.

The Text Analytics Process

- Spelling errors are corrected using a dictionary and plurals must be singularized.

Idiomatic expressions are resolved.

- Tenses of words are made uniform so that the same word does not reappear many times with minor variations. Comments, such as “nothing” or “what?” or “no comment” or “mpmpmpmpmp”, often appearing in the commentary, especially social media, are removed (however, for sentiment analysis these are retained and analyzed).

The Text Analytics Process

Tokenization

Tokens are words taken from a block of text that has been cleaned. This process is called tokenization, and it is often being used to identify people or organizations (named entity extraction). Named entity extraction aims to overcome a problem created when separating words by using blank spaces; IBM SPSS and SAS Data Miner both do this operation.

- For instance, an expression such as “Mr. Sand” contains two tokens (Mr. and Sand), and the computer must recognize that they belong together.

The Text Analytics Process

Turning Text to Numbers, and Numbers into Data

It is very hard to take what people say and turn it into useful information. Humans understand many of the structures and meanings of language that computer software does not, making even seemingly simple sentences such as “A dog is chasing a boy on the playground” tough to decode. As text is processed, tagged, entities extracted, sentiment applied, much of the intrinsic meaning we humans take for granted, is lost in translation. Turning text into numbers also involves semantics to derive meaning, involving the ability to disambiguate the meaning of words in various contexts.

Turning Text to Numbers, and Numbers into Data

0. The Sky is blue and the grass is green.

1. Pattern Tokenizer

2. Trim Token Filter

3. Lowercase Token Filter

1. "Sky"¹ "is" "blue"² "and" "the"
"grass"³ "is" "green"⁴ " " "

2. "Sky"¹ "blue"² "grass"³ "
"green"⁴

3. "sky"¹ "blue"² "grass"³ "green"⁴

The Text Analytics Process

Sliding Windows and N-Grams

While searching for information within a document, an n-gram or sliding window (of varied character lengths) is used to pick up the occurrences of words and to find named entities. Sometimes the simplest models are also the most efficient – take the n-gram; it is a method of scanning information much the way our eyes scan. If we want to see what is around us, we should systematically move our eyes in a zig-zag fashion across the area of view, the point where vision is focused would correspond to the n-gram or the window from which we are looking at details. Text analytics works much the same way, zig-zagging across a text document with a window size (n-gram) of a certain number of characters.

The Text Analytics Process

Sliding Windows and N-Grams

A sliding window (an n-gram window) is one way to address the question of which words occur together.

	<u>N-Gram</u>
Paul Cezanne is a great painter.	'Paul Ce'
Paul Cezanne is a great painter.	'aul Cez'
Paul Cezanne is a great painter.	'ul Ceza'
Paul Cezanne is a great painter.	'l Cezan'

The Text Analytics Process

Sliding Windows and N-Grams

Think of the sliding window like a box that is a certain length (say seven characters) that moves through the text, one word at a time, and which keeps a count of how many times words fall together in that moving box. An approach like this would be useful for analyzing a collection of documents and looking for the strongest themes expressed across this set.

The Text Analytics Process

Similarities Matrix

Turning text into numbers is hard work, and one step in the process of running equations of the resulting data is to organize it into a similarity matrix. A similarities matrix shows how often words occur together. With a sliding window, the text takes a roundabout route to a table but finally appears in one. This chart, in turn, can lead to other, still more compact forms of data representation.

The Text Analytics Process

Deep Learning and Neural Networks

If there is a large enough sample of text, even if the opinion holders are not clearly segmented, it is possible to forecast the sentiment (and opinions) of opinion holders with increasing accuracy. Deep learning neural networks are now being used for sentiment analysis with some interesting results.

The Text Analytics Process

Counting Words, Centricity, and Betweenness

Another way of determining what and who is important in a corpus of text is by looking at the concepts of “centricity” and “betweenness”. Counting the number of words in a document is a way to understand the content as well as the words that are occurring more than the average.

- The more the betweenness of a word, the more it appears connected with other words that represent different concepts.
- Centricity has more to do with the way that betweenness is represented in a network diagram. The idea of centricity has often been used with influencers’ network mapping.

The Text Analytics Process

Predictive Text Analytics

Describing relationships (a word map or word cloud) is useful, yet predicting which routes in a map will bring us to our destination more quickly is far more actionable. Text analytics uses the methods (such as linear regressions) to determine probabilities. By using probabilities, decisions can be made on the probabilities calculated from the text being analyzed.

In this lecture, we discussed:

- Introduction
- How GPT-3 (and Future Versions) Responds to the Questions Humans Ask
- GPT-4o
- GPT-5
- Algorithms and Text Analysis



Thank You