

Class BS Data Science

Subject: Text Mining

Day 1: Date: 11/10/2024

Topic : Introduction to Text Mining

Slide 2: What is Text Mining?

- **Definition:** Text mining is the process of extracting meaningful insights and patterns from unstructured text data.
 - **Goal:** Transform text into useful information for analysis, decision-making, or automated processes.
 - **Commonly used in** data science, linguistics, business, and research.
-

Slide 3: Why is Text Mining Important?

- **Data Abundance:** Over 80% of today's data is unstructured (e.g., social media posts, emails, articles).
 - **Business Value:** Provides insights into customer sentiment, brand perception, and more.
 - **Applications:** Used in marketing, social media monitoring, customer support, and beyond.
-

Slide 4: Key Steps in Text Mining

1. **Data Collection** - Gathering raw text data from sources (web scraping, APIs, databases).
 2. **Text Preprocessing** - Cleaning and preparing data (removing stop words, stemming, tokenization).
 3. **Feature Extraction** - Converting text into numeric data (e.g., TF-IDF, word embeddings).
 4. **Modeling & Analysis** - Applying algorithms to find patterns (e.g., clustering, sentiment analysis).
 5. **Interpretation & Visualization** - Making sense of the results through charts and reports.
-

Slide 5: Text Preprocessing Techniques

- **Tokenization:** Splitting text into smaller units (words, sentences).
 - **Removing Stop Words:** Removing common words (e.g., "the," "is") that add little meaning.
 - **Stemming & Lemmatization:** Reducing words to their root form (e.g., "running" to "run").
 - **Normalization:** Converting text to lowercase, removing special characters.
-

Slide 6: Popular Methods in Text Mining

- **Keyword Extraction:** Identifying significant words or phrases in a text.
 - **Sentiment Analysis:** Determining the emotional tone (positive, negative, neutral).
 - **Topic Modeling:** Discovering hidden themes in large volumes of text (e.g., LDA).
 - **Text Classification:** Assigning predefined labels to texts (e.g., spam detection).
 - **Named Entity Recognition (NER):** Identifying names, places, dates, etc.
-

Slide 7: Text Mining Tools & Libraries

- **Python Libraries:** NLTK, SpaCy, TextBlob, Gensim.
 - **Machine Learning Libraries:** Scikit-learn, TensorFlow, Keras.
 - **Software Platforms:** RapidMiner, IBM Watson, SAS Text Miner.
-

Slide 8: Real-World Applications of Text Mining

- **Customer Feedback Analysis:** Analyzing reviews and surveys to improve products.
 - **Healthcare:** Extracting patient insights from medical records for better treatment.
 - **Finance:** Sentiment analysis on news to predict stock trends.
 - **Social Media Monitoring:** Tracking brand sentiment, trending topics, and more.
-

Slide 9: Challenges in Text Mining

- **Data Quality:** Text data can be noisy and messy.
 - **Language Complexity:** Ambiguities, slang, and context make analysis complex.
 - **Scalability:** Processing large volumes of text efficiently.
 - **Privacy and Ethics:** Ensuring compliance with privacy laws and ethical standards.
-

Slide 10: Future of Text Mining

- **Integration with AI:** Enhanced insights with NLP and deep learning.
 - **Real-time Analysis:** Rapid processing for live data streams.
 - **Cross-lingual Capabilities:** Mining text in multiple languages.
 - **Ethical Text Mining:** Emphasis on privacy and responsible AI use.
-

Slide 11: Conclusion

- Text mining is a powerful tool for extracting insights from unstructured data.
 - From sentiment analysis to topic modeling, it has vast applications across industries.
 - With advancements in AI and NLP, the potential for text mining continues to grow.
-

Thanks

Assignment No. 1.

Q1. Is chatGPT relates to text mining ? if Yes! How?

Q2. What is NLP?

Q3. What is text classification?

Q4. What is features extraction and Features Representation?

Q5. What is sentiment analysis?