**Name: Nikhil Jadhav**

**Student ID: 801075504**

**Cloud Computing for Data Analysis**

**VIDEO CASE 10 : Text Classification**

Watch following videos:

**Video 1:** <https://youtu.be/Czueje0eVzY>

**Video 2:** <https://youtu.be/tzfOsLeVoPI>

**Video Case Questions:**

1. What is text mining?
2. How text mining tools works on the given text for extracting information?
3. What are some applications of text mining?

Answers:

1. Text Mining (also known as Text Analytics) are Artificial Intelligence (AI) technologies that empower users to rapidly transform the key content in text documents into quantitative, actionable insights.

Text mining is the process of examining large collections of written resources to generate new information, and to transform the unstructured text into structured data for use in further analysis. Text mining identifies facts, relationships and assertions that would otherwise remain buried in the mass of textual big data. These facts are extracted and turned into structured data, for analysis, visualization (e.g. via html tables, mind maps, charts), integration with structured data in databases or warehouses, and further refinement using machine learning (ML) systems.

1. Following are the steps which are taken to extract information on the given text:

* **Step 1 : Information Retrieval**

This is the first step in the process of data mining. This step involves the help of a search engine to find out the collection of text also known as corpus of texts which might need some conversion. These texts should also be brought together in a particular format which will be helpful for the users to understand. Usually XML is the standard for text mining

* **Step 2 : Natural Language Processing**

This step allows the system to perform grammatical analysis of a sentence to read the text. It also analyzes the text in structures.

* **Step 3 : Information extraction**

This is the second stage where in order to identify the meaning of a particular text mark-up is done. In this stage a metadata is added to the database about the text. It also involves adding names or locations to the text. This step lets the search engine to get the information and find out the relationships between the texts using their metadata.

* **Step 4 : Data Mining**

The final stage is data mining using different tools. This step finds the similarities between the information that has the same meaning which will be otherwise difficult to find. Text Mining is a tool which boosts the research process and helps to test the queries.

Text Mining includes the following list of elements

1. Text Categorization
2. Text Clustering
3. Concept/entity extraction
4. Granular taxonomies
5. Sentiment Analysis
6. Document Summarization
7. Entity Relation Modelling
8. Applications of text mining are as follows:

* Analyzing open ended survey responses

Open ended survey questions will help the respondents to give their view or opinion without any constraints. This will help to know more about the customers’ opinions than relying on structured questionnaires. Text mining can be used to analyze such information in the form of text.

* Automatic processing of messages, emails

Text Mining is also mainly used to classify the text. Text Mining can be used to filter the unnecessary mail using certain words or phrases. Such mails will automatically discard such mails to spam. Such automatic system of classifying and filtering selected mails and sending it the corresponding department is done using Text Mining system. Text Mining will also send an alert to the email user to remove the mails with such offending words or content.

* Analyzing warranty or insurance claims

In most of the business organizations information is collected mainly in the form of text. For example in a hospital the patient interviews can be narrated briefly in text form and the reports are also in the form of text. These notes are now a day’s collected electronically so that it can be easily transferred into text mining algorithms. These records can then be used to diagnose the actual situation.

* Investigating competitors by crawling their web sites

Another important application area of Text Mining is processing the contents of web pages in a particular domain. Through this way the text mining system will automatically find a list of terms which is used in the site. Through this way one can find out the most important terms used in the website. By this way one can know the capabilities about the competitors which can help you to deliver business efficiently.

The other applications of Text Mining include the following

1. Business Intelligence
2. E Discovery
3. Bioinformatics
4. Records Management
5. National Security or Intelligence works
6. Social Media Monitoring