

According to industry estimates, only 21% of the available data is present in structured form. Data is being generated as we speak, as we tweet, as we send messages on Whatsapp and in various other activities. Majority of this data exists in the textual form, which is highly unstructured in nature.

Few notorious examples include - tweets / posts on social media, user to user chat conversations, news, blogs and articles, product or services reviews and patient records in the healthcare sector. A few more recent ones includes chatbots and other voice driven bots.

Despite having high dimension data, the information present in it is not directly accessible unless it is processed (read and understood) manually or analyzed by an automated system.