LIST OF FIGURES

Evaluation of text summarization

Twitter

LIST OF TABLES

LIST OF EQUATIONS

ABBREVIATIONS

ABBREVIATION	DESCRIPTION
TF-IDF	Term Frequency – Inverse Document Frequency
BLEU	Bilingual Evaluation Understudy
BP	Brevity Penalty
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
LCS	Longest Common Subsequence
WOEID	Where On Earth IDentifier
NLP	Natural Language Processing
NLTK	Natural Language Toolkit (Python)
API	Application Programming Interface
DOM	Document Object Model
GPU	Graphical Processing Unit
CPU	Central Processing Unit
RAM	Random Access Memory
XML	Extensible Mark-up Language
UML	Unified Modelling Language
JSON	JavaScript Object Notation
BBC	British Broadcasting Channel
URL	Uniform Resource Locator
TOI	Times of India

ABSTRACT

With the boom of social media in the 21st century, it has grown to impact nearly every aspect of our life. This widespread nature of social media makes it a critical platform for the spread of information. This information, more often than not, is filled with flaws such as incomplete information, uncredited sources and may even go as far as being downright false. Platforms such as Facebook, Twitter, Instagram, and many more receive more than a billion active users every day with double that active every month, making them excellent platforms for the spread of information. There have always been suspicions and even some verified examples of these platforms being used to influence the lives of people by feeding them news from biased sources in order to influence the decisions they make in real life.[14]

Twitter is a critical player in the propagation of information with 500 million tweets containing snippets of information being sent every day, proved to be the ideal platform on which we could implement a system to verify and then spread information. The decision was made to consider the trending topics on twitter to be the keywords on which more information must be found and propagated.

Summarization of the selected articles, however, proved to be the greatest challenge. We decided to focus on extractive methods of summarization where sentences are lifted from the source verbatim instead of abstractive methods, which generated sentences due to its high complexity and resource requirements. We compared a variety of methods such as word frequency, term frequency-inverse document frequency, TextRank, latent semantic analysis based on the intrinsic evaluation. We also modified word frequency to obtain better results.

Bots proved to be the ideal form with which could implement said system as it reduces the probability of human error, removes work redundancy, and is inherently unbiased. The bot would scrape data off articles from reliable sources, summarize them using the best method from those mentioned above, and then tweet them along with a link to the actual source.