DEVELOPING A MACHINE LEARNING POWERED TEXT SUMMARIZER APP

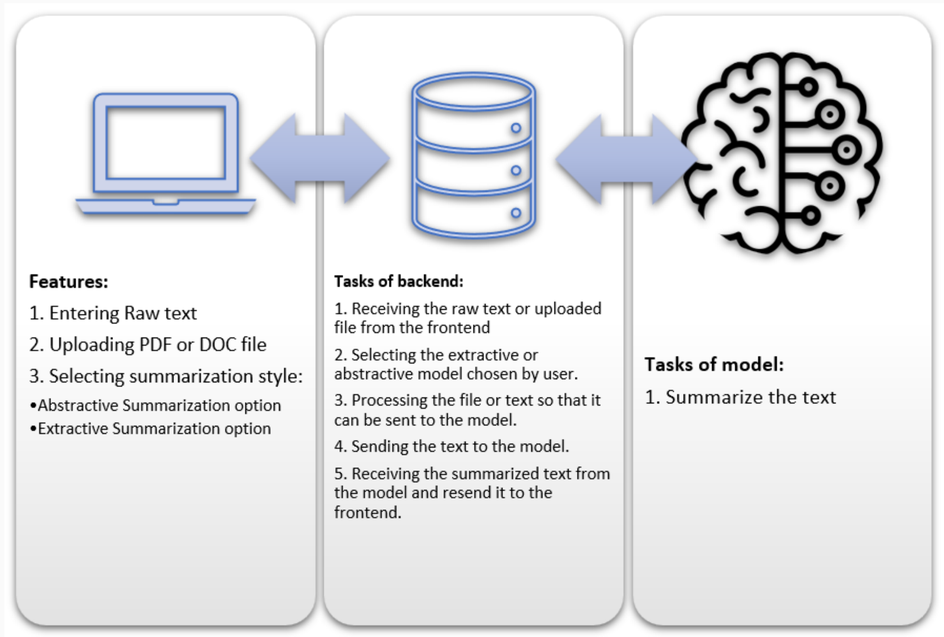
Student Name: Mohammed Rashed Alghaithi **Supervisor name: Quratul-ain**

Project Introduction

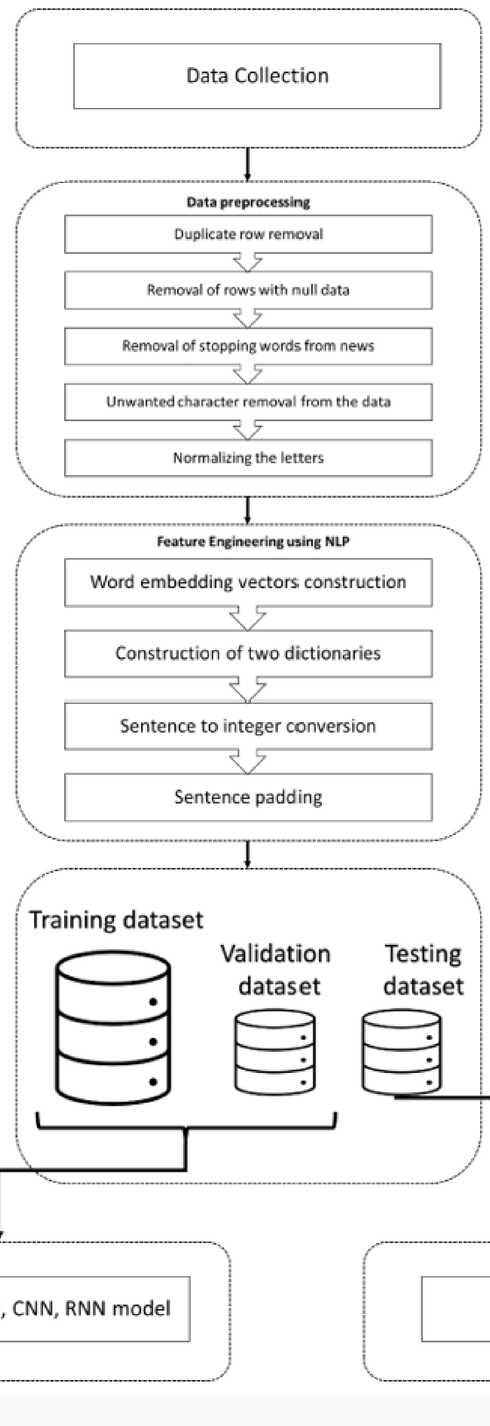
The aim of this project is to develop an AI powered text summarization application including the abstraction and extractive summarization model. This app will allow users to enter raw text, upload document and select what type of summarization he/she wants.

Literature Review Topic	Summary
Feature Engineering Methods	Techniques such as TF-IDF, TF-ISF, title/headline relevance, cues, position, similarity, sentence length, proximity, and proper nouns are utilized for summarization.
Techniques to Measure Model Effectiveness	Evaluation methods include ROUGE (ROUGE-1, ROUGE-2, ROUGE-L), BLEU, BERTScore, MoverScore, Pyramid Method, user-based, document-based, semantic similarity, diversity and coverage metrics, and human evaluation for readability and coherence.
Extractive Summarization Model	Extracts or copies parts from the original text based on scores computed using statistical or linguistic features.
Abstractive Summarization Model	Rephrases the original text to generate new phrases, requiring advanced machine learning and NLP techniques.
Evaluation Metrics	ROUGE, BLEU, BERTScore, MoverScore, Pyramid Method, user-based, document-based, semantic similarity, diversity and coverage metrics, human evaluation for coherence, readability, and inter-annotator agreement.

Product Architecture



Planned Methodology



Product Components

Frontend | Backend | Summarization model

Summarization Model

Abstractive summarization model	Extractive Summarization model

Future Work

1. Multi-document summarization capability can be added.
2. A mobile application can be developed based on the developed models in the project.

References

1. K. A. Hawelikar, “EXPLORING THE IMPACT OF PROMPT ENGINEERING ON CHATGPT 3.5 TEXT SUMMARIZATION: A BERT SCORE EVALUATION”, doi: 10.56726/IRJMETS45268.
2. W. Zhao, M. Peyrard, F. Liu, Y. Gao, C. M. Meyer, and S. Eger, “MoverScore: Text Generation Evaluating with Contextualized Embeddings and Earth Mover Distance,” EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference, pp. 563–578, Sep. 2019, doi: 10.18653/v1/d19-1053.
3. T. T. Hailu, J. Yu, T. G. Fantaye, T. T. Hailu, J. Yu, and T. G. Fantaye, “Intrinsic and Extrinsic Automatic Evaluation Strategies for Paraphrase Generation Systems,” Journal of Computer and Communications, vol. 8, no. 2, pp. 1–16, Feb. 2020, doi: 10.4236/JCC.2020.82001.
4. A. Nenkova, R. Passonneau, and K. Mckeown, “The Pyramid Method,” ACM Transactions on Speech and Language Processing (TSLP), vol. 4, no. 2, May 2007, doi: 10.1145/1233912.1233913.
5. P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching Word Vectors with Subword Information”, Accessed: Dec. 06, 2023. [Online]. Available: http://www.isthe.com/chongo/tech/comp/fnv