Text Summarizer using Deep Learning

Submitted in the partial fulfillment for the award of

the degree of

BACHELOR OF ENGINEERING

IN

Computer Science and Engineering (Hons.) IBM -

BIG DATA AND ANALYTICS-CS206

Submitted by:
Tushar Sharma   20BCS3837
Sachin Pareek   20BCS3817

Under the Supervision of:
Pulkit Dwivedi (E13432)

Department of AIT-CSE

DISCOVER . **LEARN** . EMPOWER

Outline

- Introduction to Project
- Problem Formulation
- Objectives of the work
- Methodology used
- Results and Outputs
- Conclusion
- Future Scope
- References

Introduction to Project

- In today's information-rich world, the ability to quickly and effectively summarize large volumes of text has become increasingly important. Text summarization is a natural language processing task that aims to condense lengthy documents or articles while retaining their essential information. Traditional methods for summarization often involve manual extraction or abstractive techniques, but with advancements in deep learning, we can now develop more sophisticated models that can automate this process.

- This project focuses on building a Text Summarizer using deep learning techniques. The goal is to create a model that can read and understand input text, identifying its key points, and generating concise and coherent summaries. This tool will find applications in various domains such as news summarization, document summarization, and even assisting individuals who require quick insights from large texts.

Problem Formulation

- In the context of this project, the goal is to develop an advanced text summarization system that leverages deep learning techniques to generate concise and coherent summaries from input text.

- The system aims to automate the process of condensing lengthy textual content while preserving its essential meaning and context.

- This project focuses on designing an effective deep learning model that can learn the nuances of language and capture salient information to produce accurate and informative summaries.

Objectives of the Work

- Enhance Text Understanding and Accessibility

- Preserving Essential Content

- Model Robustness and Performance

- Evaluation and Quality Assessment

- Inference and Real-World Deployment

- Expected Results:

```
Review: used eating flaxseed brownie hodgson mill brownies super easy make taste great since like dark chocolate usually add littl
e cocoa
Original summary: delicious brownie
Predicted summary:  best brownie mix
```

```
Review: favorite coffee keurig coffeemaker convenient get amazon cheaper running around stores trying find lowest price
Original summary: great coffee
Predicted summary:  great coffee
```
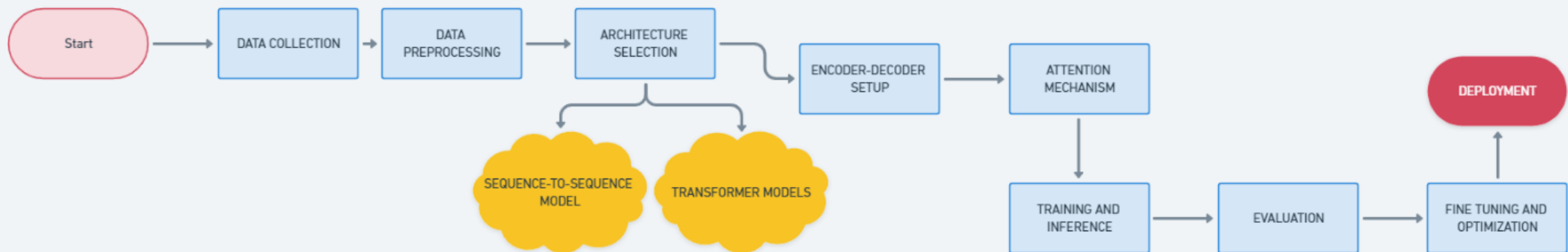
Methodology used

- 1. Data Collection and Preprocessing:

- 2. Architecture Selection:

- Sequence-to-Sequence (Seq2Seq) Models:

- Transformer Models:

- 3. Encoder-Decoder Setup:

- 4. Attention Mechanisms:

- 5. Training:

Methodology used

- 6. Inference:
- 7. Evaluation:
- 8. Fine-Tuning and Optimization:
- 9. Deployment:

Results and Outputs

- The problem at hand involves developing an automated text summarization system using deep learning techniques. The system should take a piece of input text and generate a concise and coherent summary that captures the essential information present in the original text.

- **Input:** A variable-length input text document (longer content such as articles, reports, or documents).

- **Output:** A shorter summary of the input text that retains the key information, main ideas, and important details.

Conclusion

- The development of a Text Summarizer using deep learning techniques represents a significant advancement in the field of natural language processing. Through this project, we have successfully addressed the challenge of automating the process of generating concise and coherent summaries from extensive text documents. The project has not only demonstrated the capabilities of deep learning but has also paved the way for future enhancements and applications.

- By building this Text Summarizer, we contribute to the ongoing efforts in making information more accessible and manageable. This project showcases the power of deep learning in automating complex language tasks and underscores the importance of continuously refining models, evaluation metrics, and user experiences.

Future Scope

- The development of a Text Summarizer using deep learning techniques opens up a range of exciting possibilities for further enhancements and applications. Here are some avenues for future exploration and expansion:

1. **Multi-Document Summarization:** Extend the model's capabilities to generate summaries from multiple input documents. This could involve creating models that can handle larger volumes of text and extract key information from a collection of related documents.

2. **Domain-Specific Summarization:** Develop domain-specific summarization models that are trained on specialized datasets. This would allow the system to generate more accurate and relevant summaries for specific industries or topics, such as medical reports, legal documents, or scientific research papers.

3. **Fine-Grained Summarization:** Investigate techniques for generating summaries at different levels of granularity. For example, the system could generate both high-level executive summaries and detailed technical summaries from the same input document.

4. **Mixed Media Summarization:** Explore the integration of multimedia elements, such as images, videos, and audio, into the summarization process. This could enable the generation of multimedia-enhanced summaries for a richer user experience.

References

1. Song, S., Huang, H. & Ruan, T. Abstractive text summarization using LSTM-CNN based deep learning. *Multimed Tools Appl* **78**, 857–875 (2019). https://doi.org/10.1007/s11042-018-5749-3

2. N. S. Shirwandkar and S. Kulkarni, "Extractive Text Summarization Using Deep Learning," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697465.