Text Summarization

Team Members:

Dharani Doppalapudi
Saiabhinav Chekka

Outline for Presentation:

- ➢ Project Description
- ➢ Project Abstract
- ➢ Workflow
- ➢ Summary of Tasks
- ➢ Evaluation Metrics
- ➢ Results & Observations
- ➢ Conclusion

Project Description

➢ We worked on DeepMind Q&A Dataset. The news stories and story highlights from the CNN website are included in this collection. There are approximately 90k documents in this dataset.

➢ Our project aims at generating summaries of news stories using existing text summarization methods.
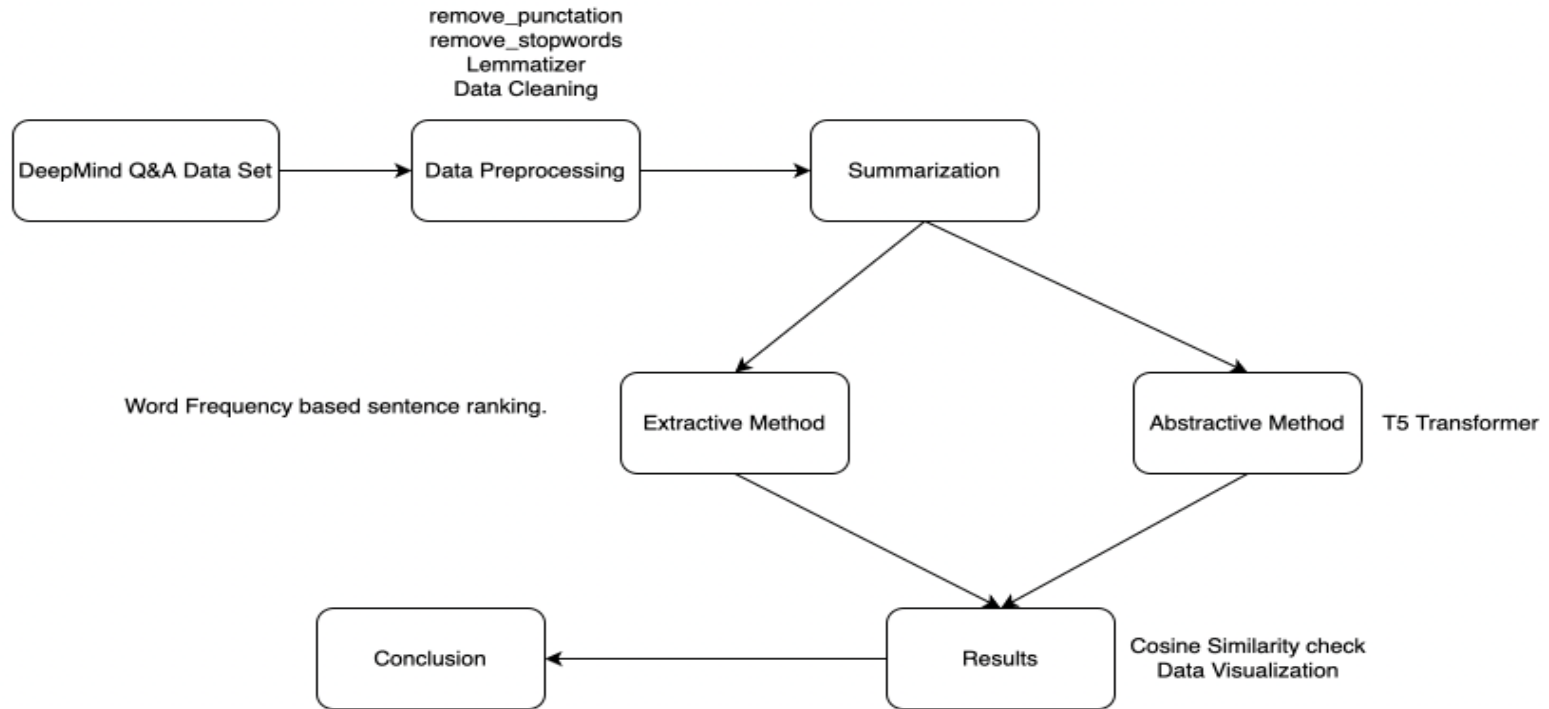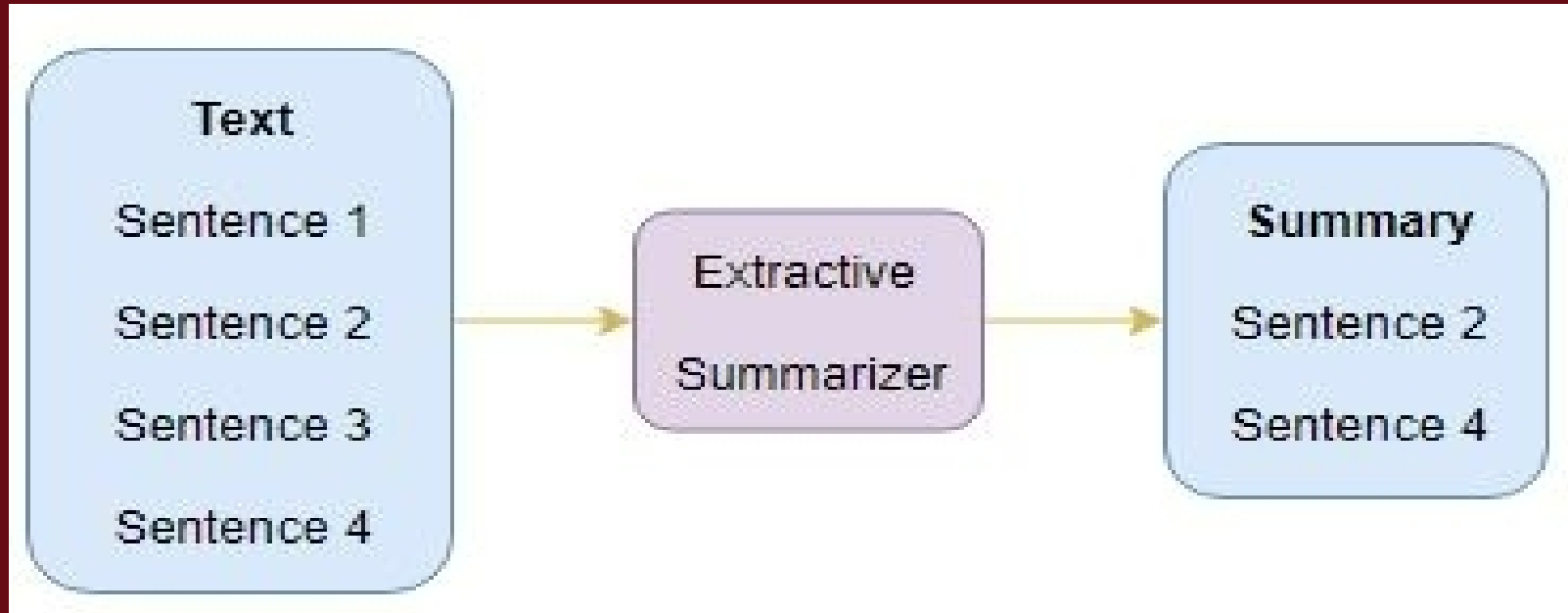
Project Abstract

➤ On the CNN news premise, we performed data preprocessing and summarized the data using Extractive and Abstractive methods.

➤ Evaluated the obtained summarizations using cosine similarity check and acquired similarity score of 75% with Extractive method and around 83% with Abstractive method.
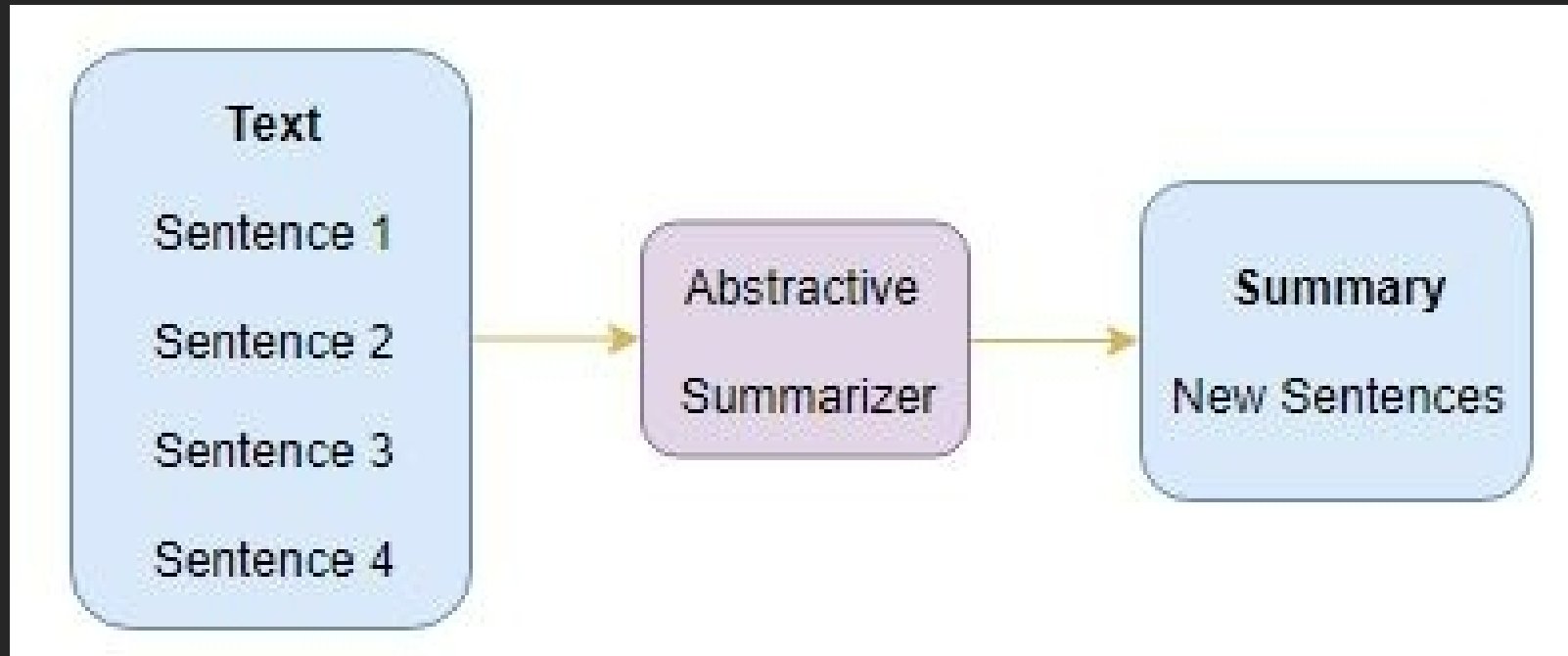
Workflow

remove_punctation
remove_stopwords
Lemmatizer
Data Cleaning

DeepMind Q&A Data Set → Data Preprocessing → Summarization

Word Frequency based sentence ranking.

Extractive Method

Abstractive Method — T5 Transformer

Conclusion ← Results — Cosine Similarity check / Data Visualization

Extractive Summarization:

Abstractive Summarization:

Evaluation Metrics

evaluation measures

intrinsic

extrinsic (task-based)

text quality evaluation
grammaticality
non-redundancy
referential clarity
structure and coherence

content evaluation

co-selection
precision, recall, F-score
relative utility

content-based
cosine similarity
unit overlap
longest common subsequence
n-gram matching (ROUGE)
Pyramids
LSA-based measures

document categorization
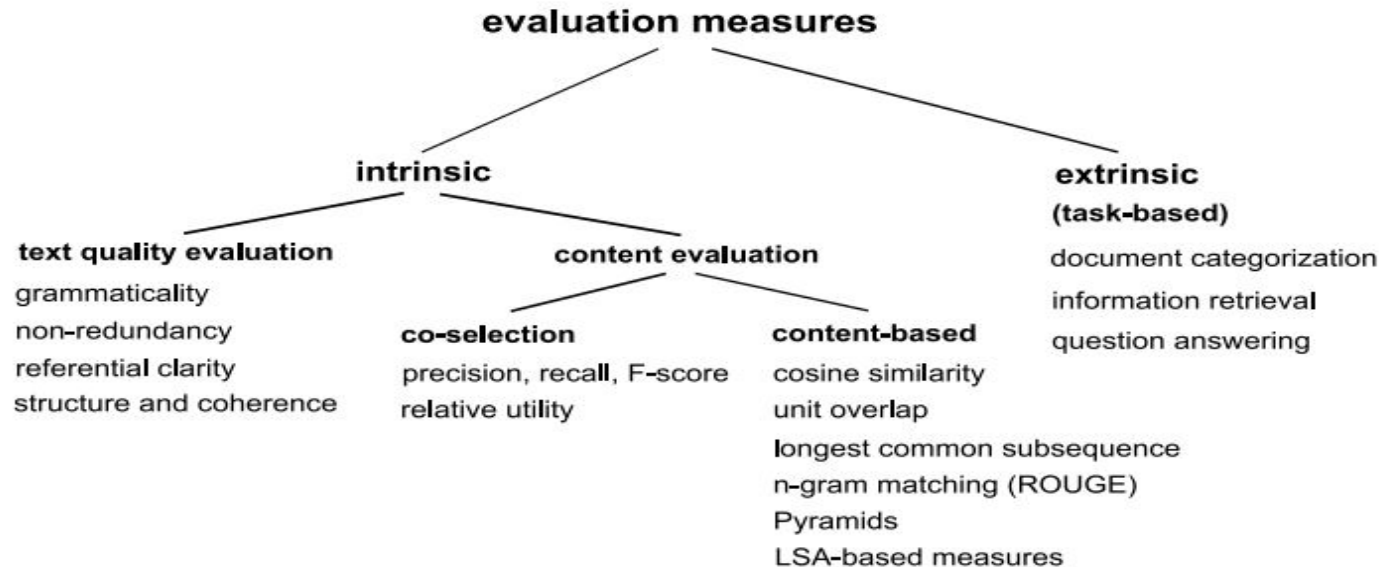information retrieval
question answering

Fig. 1. The taxonomy of summary evaluation measures

Generated Summaries

```
## Main story
subset_stories.story[13]
```

"(CNN) -- A single-engine airplane made an emergency landing on a California highway Sunday morning, though no major injuries were reported, authorities said.\n\nThe Piper Comanche 260 carrying a married couple landed on the southbound lanes of U.S. Highway 101.\n\nThe Piper Comanche 260 carrying a married couple landed on the southbound lanes of U.S. Highway 101 just outside Santa Barbara and a few miles from the airport, said California Highway Patrol spokesman Officer James Richards.\n\nThe plane's engine quit, and as the pilot descended, he lost control of the plane and landed in the southbound lane facing oncoming traffic. The plane struck two vehicles while landing, then spun and hit another one with its tail, Richards said.\n\nOne vehicle passenger was treated for minor injuries, he said. No other injuries were reported.\n\nThe landing happened at 10:36 a.m. (1:36 p.m. ET) and held up traffic for less than two hours, Richards said.\n\nHe added that the plane had departed Temecula in southwestern Riverside County, California, and was destined for the airport in Santa Barbara, a flight of about 180 miles.\n\nThe pilot told authorities that he attempted to switch fuel lines during the flight, but was unable to restore power to the plane. He said he alerted a tower at the airport that a landing on the highway was imminent, Richardson said.\n\n"

```
## Extractive Summary
subset_stories.summary[13]
```

' (CNN) -- A single-engine airplane made an emergency landing on a California highway Sunday morning, though no major injuries were reported, authorities said. The Piper Comanche 260 carrying a married couple landed on the southbound lanes of U.S. Highway 101 just outside Santa Barbara and a few miles from the airport, said California Highway Patrol spokesman Officer James Richards.'
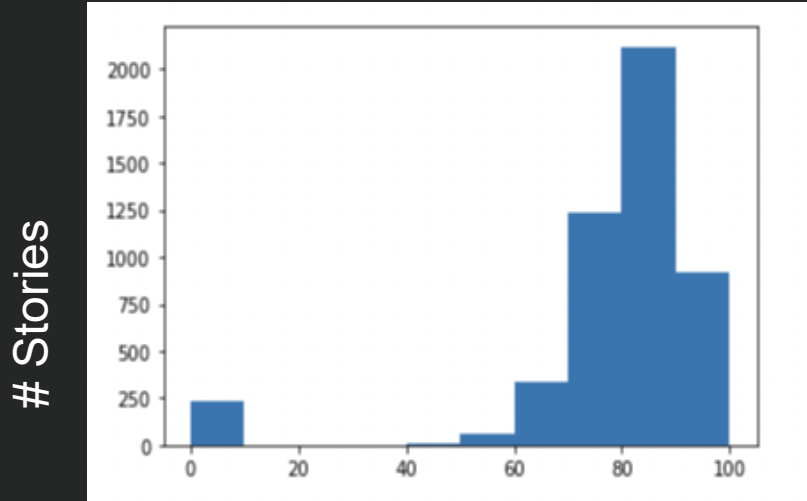
```
## Abstractive Summary
subset_stories.summary2[13]
```

'<pad> the single-engine plane made an emergency landing on a California highway. no major injuries were reported. the pilot lost control of the plane and landed in the southbound lane facing oncoming traffic. the plane struck two vehicles while landing, then spun and hit another one with its tail. the landing happened at 10:36 a.m. (1:36 p.m. ET) and held up traffic for less than two hours, authorities say.</s>'
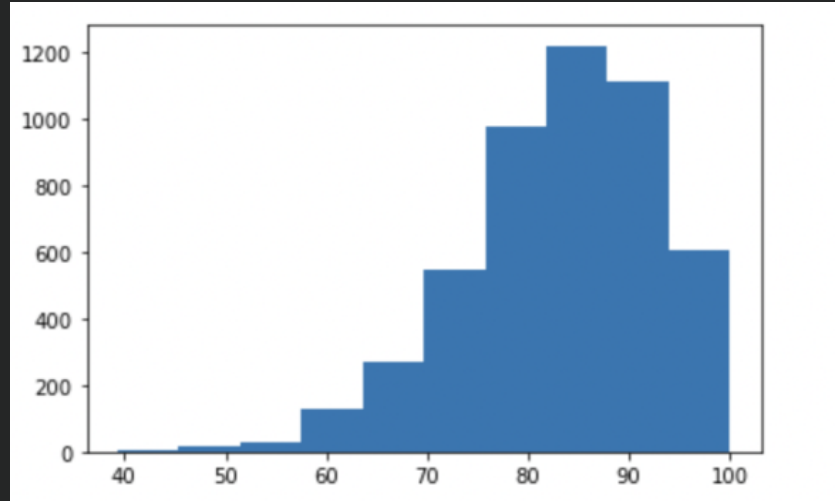
Observations

Extractive

Abstractive

Stories

Cosine similarity score distribution

CONCLUSION

➢ Our primary goal of this project is to summarize the long stories of CNN news stories and we accomplished the goal using two summarization techniques: Extraction and Abstraction Summarizations.

➢ We achieved a similarity score of 75% with Extractive and around 83% with Abstractive summarization.

THANK YOU

INDIANA UNIVERSITY BLOOMINGTON