Read Smarter, Learn Faster

A NLP Project

Presented by Group C

Guillermo Chacon, Diego Cuartas, Esteban Delgado, Aayush Kerjiwal, Mohamed Khanafer, Mariana Naváez, and Valentina Premoli
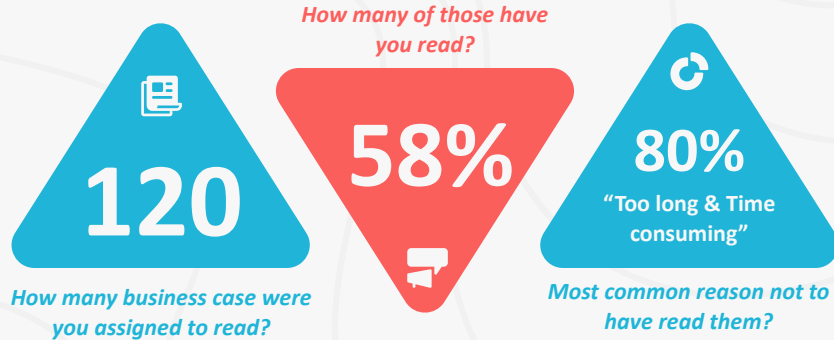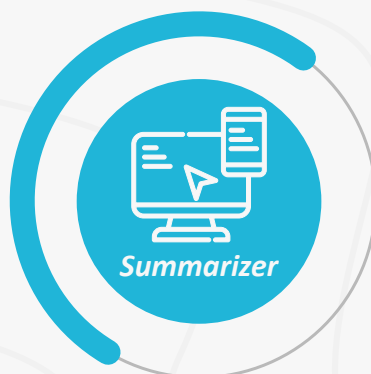
1

Problem = Opportunity

Introduction

2

Use Case - Validation

We surveyed 20 former MBA students

How many of those have you read?

120

58%

80%
"Too long & Time consuming"

How many business case were you assigned to read?

Most common reason not to have read them?

3

Application Idea

Summarizer

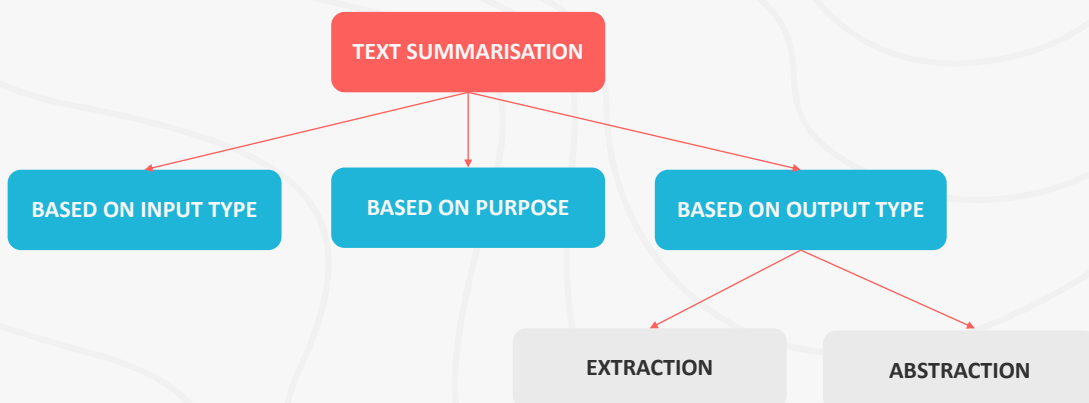**IE MBA Summarizer Assistant**

Helping you work smarter, not harder

4

What is text Summarization?

Automatically creating a compressed version of a given document preserving its information content.
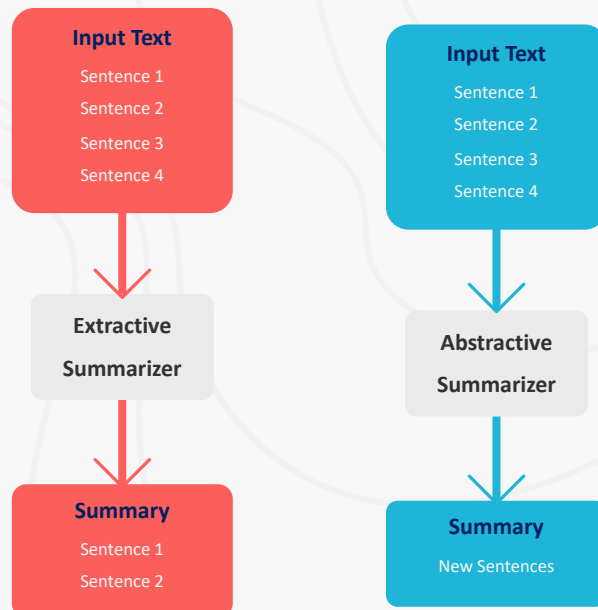
1 — **2** — **3**

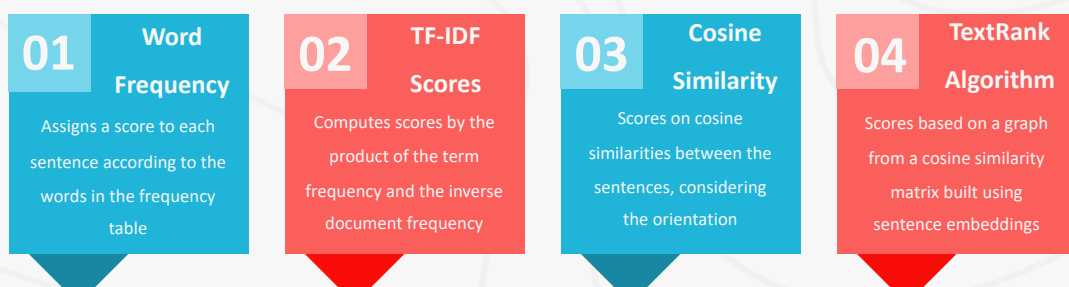**Analysis**   **Transformation**   **Synthesis**

5

Types of Summarization

TEXT SUMMARISATION

BASED ON INPUT TYPE   BASED ON PURPOSE   BASED ON OUTPUT TYPE

EXTRACTION   ABSTRACTION

6

Types of Summarization

Input Text

Sentence 1

Sentence 2

Sentence 3

Sentence 4

↓

Extractive Summarizer

↓

Summary

Sentence 1

Sentence 2

Input Text

Sentence 1

Sentence 2

Sentence 3

Sentence 4

↓

Abstractive Summarizer

↓

Summary

New Sentences

7

Way to our final model

01 Word Frequency

Assigns a score to each sentence according to the words in the frequency table

02 TF-IDF Scores

Computes scores by the product of the term frequency and the inverse document frequency

03 Cosine Similarity

Scores on cosine similarities between the sentences, considering the orientation

04 TextRank Algorithm

Scores based on a graph from a cosine similarity matrix built using sentence embeddings

8

The TextRank Model Pipeline

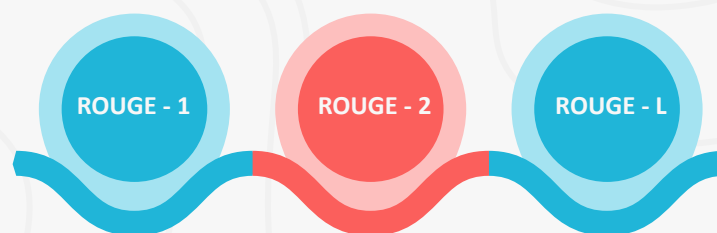|	Text Processing		Cosine Similarity Matrix		Summary Creation	
GloVe to generate word embeddings	Data Cleaning	By using GloVe word vector	Scoring each pair of sentences	Converting the matrix into a graph	Taking top n sentences	Model with the highest score
Input Text		Sentence Vector Representation		PageRank Algorithm		Summary
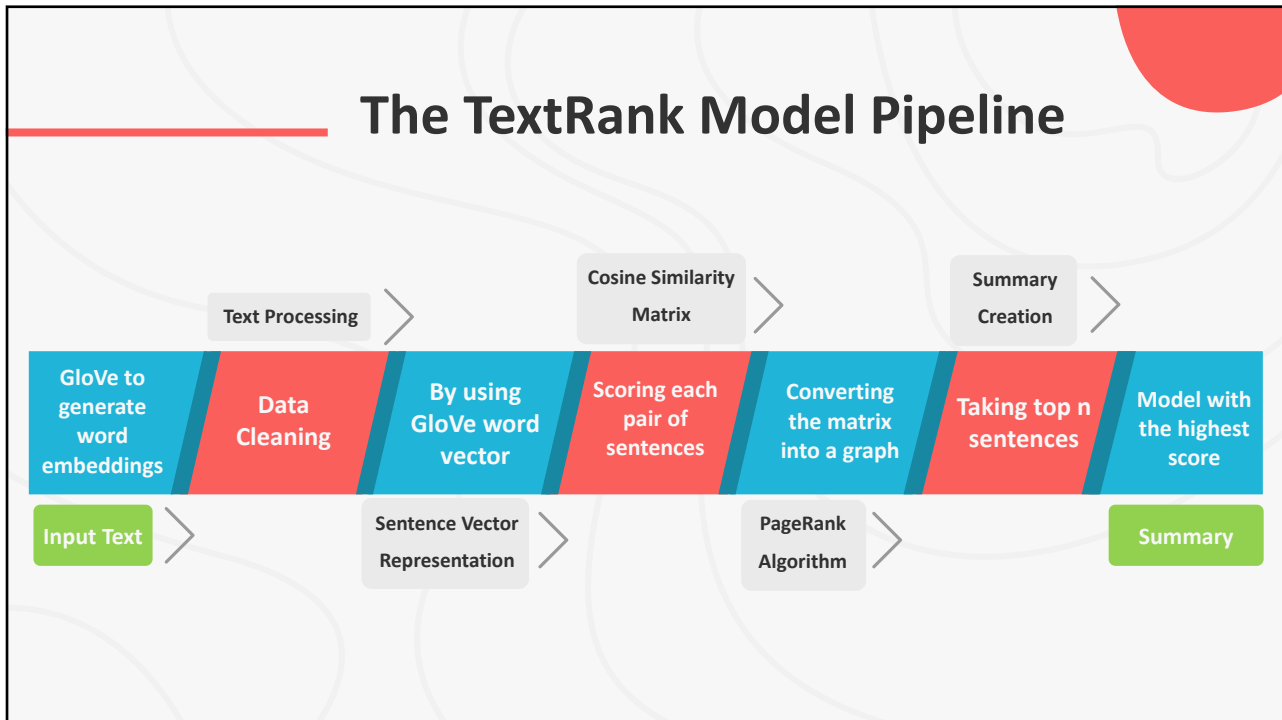
9

A note on the Evaluation Metrics

For evaluating the summaries, we used **ROUGE**
(Recall-Oriented Understudy for Gisting Evaluation)

Measure of Recall, Precision, and F1 for NLP tasks

ROUGE - 1   ROUGE - 2   ROUGE - L

10

Application Demo

11