

# DOCUMENT SUMMARISATION

Exploring Extractive Techniques to Save Time and Extract Key Points

# AGENDA

**01**

INTRODUCTION & WHAT IS THE  
PROBLEM

**02**

WHY IS IT INTERESTING AND  
IMPORTANT

**03**

WHY IS IT HARD

**04**

EXISTING APPROACHES & BENIFITS

**05**

OUR APPROACH

**06**

CONCLUSION



# INTRODUCTION

- We live in an era of information overload on the internet.
- Automatic text summarization is the solution to the challenge of processing vast amounts of information efficiently.
- The goal is to make our lives easier by summarizing information and distilling key content.
- Robots to the Rescue: Imagine helpers who read all the stuff and tell you the most important parts in a short, easy-to-understand way. That's what text summarizers can do.
- We'll learn how these robots work, what they're good at (and not so good at), and how they can help us escape the info-flood!
- Discover how these robots can save us time, make reading easier, and unlock all the treasures hidden in the internet ocean.



# PROBLEM

- In today's digital world, individuals and businesses are dealing with a huge amount of written stuff – articles, reports, research papers, and news. The challenge isn't just that there's a lot to read, but also that there's not enough time to read and understand it all.
- So, there's a need for something that can help make sense of all this information overload. We need a solution that can pick out the most important parts from long documents and give us quick summaries. This is important because, in our busy lives, we don't always have the time to go through a mountain of text to get to the essential points.

# WHY IS IT INTERESTING AND IMPORTANT

- In a world overflowing with data, recognizing the importance of this challenge is crucial.
- Two key issues: limited time and a large volume of content.
- Quick summaries become essential for effective information management





# WHY IS IT HARD

## 1. Information Overload:

- Coping with the sheer volume of information available on the internet is a daunting task.
- The deluge of data poses challenges in identifying, processing, and extracting relevant insights.

## 2. Diverse Information Formats:

- The internet presents information in various formats, including text, images, videos, and more.
- Summarizing content in diverse formats demands adaptability to effectively capture the essence.

### **3. Trust Issues:**

- Verifying the credibility of information sources is challenging in an environment where misinformation is prevalent.
- Ensuring the reliability of data requires additional efforts in fact-checking and source validation.

### **4. Time Constraints:**

- Limited time compounds the difficulty, as individuals cannot thoroughly engage with every piece of information.
- Quick decision-making becomes crucial, necessitating efficient methods to grasp key insights rapidly.

### **5. Cognitive Bias:**

- Information overload can lead to cognitive biases, influencing decision-making based on readily available or frequently encountered information.
- Overcoming biases requires conscious efforts to seek diverse perspectives, adding a layer of complexity

# EXISTING APPROACHES

1

When it comes to summarizing documents, we have two main ways of doing it: Extractive and Abstractive approaches.

2

## EXTRACTIVE APPROACH

**What it Does:** Picks out the important stuff from a text, like using a highlighter on key points.

**Why It's Used:** Simplifies summarization by directly pulling out the main sentences.

## ABSTRACTIVE APPROACH

**What It Does:** Creates a fresh summary using advanced language understanding, similar to how humans summarize.

**Why It's Used:** Adds a creative touch by generating new phrases, forming a unique representation of the original text.

# BENIFITS OF NLP SUMMARIZATION

## 1. Scalable and Quick

Manually summarizing a short document is fairly easy, but what if you have an article or paper that is hundreds or thousands of pages long?

Rather than having dozens of employees manually going through thousands of documents summarizers are used.

## 2. Leverages Existing Tools

To take advantage of text summarization, you don't have to build machine learning infrastructure in-house or hire a data science team. The technology exists and it's accessible to everyone.

### **3. Understand Your Customers Better**

In customer relationship management, text summarization refers to compressing all the above-mentioned customer data, and turning it into abridged summaries that you can present in a meeting or use in other business processes.

### **4. Summarize the Text in Different Formats**

Natural language processing helps you obtain summarized text extracted from your competitor's web pages, market research documents, industry-related articles, etc. Having a clear idea of the market and your competitors helps you determine actionable steps for presenting your product or refining your business strategy. This helps you stand out amongst the competitors and maintain a competitive advantage in the market

### **5. Ensure all Critical Information is Covered**

The human eye can overlook important details, while standalone software is more accurate. What every reader needs is the ability to extract what is useful to them from any piece of material.

The automated text summarizing approach makes it easy for the user to read all the most important sentences in a document.

# KEY COMPONENTS OF OUR APPROACH (EXTRACTIVE)

WORD  
TOKENIZATION

STOP WORD  
REMOVAL

TAKING  
FREQUENCY

NORMALIZING

SENTENCE  
TOKENIZATION

SCORING THE  
SENTENCES

COMBINING SENTENCES TO FORM A SUMMARY

# CONCLUSION

- We're drowning in information online. Summaries help us navigate through it. This project has explored how machines can automatically summarize text.
- Having the exact sentences that can influence your decision-making is possible through NLP text summarization.
- Innovations like NLP Summarizers make it easier now to stay focused on key insights no matter how large your information sources may be.
- Staying focused and up to date regardless of the velocity and volume of information is now easier than ever thanks to NLP technology.



# LIMITATIONS



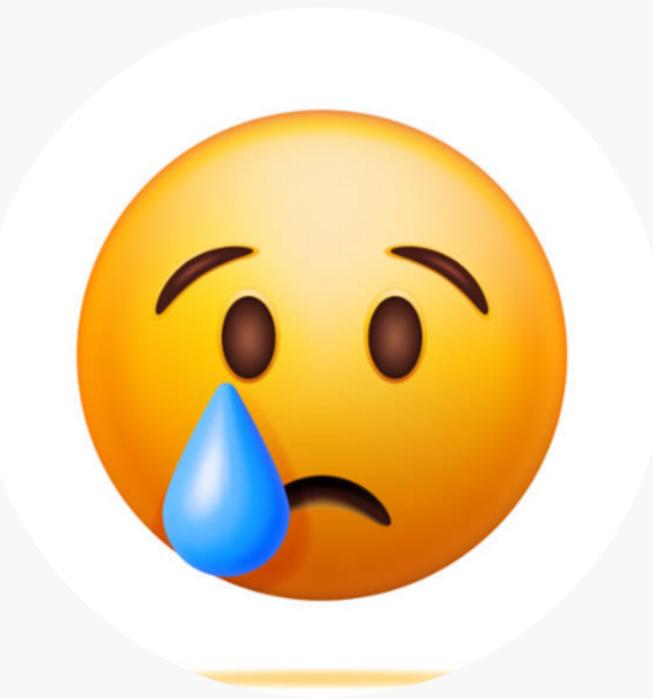
## LIMITED ABSTRACTION

Since the approach focuses on extracting sentences directly from the original text, it may not capture higher-level abstractions or generate entirely new expressions. The summary might lack a creative or interpretative element.



## SENSITIVITY TO WORD FREQUENCY

Relying on word frequency might result in a bias towards common words. Important but less frequently used terms might be overlooked, affecting the overall quality of the summary.



## DIFFICULTY WITH AMBIGUOUS TEXTS

The approach may struggle with ambiguous or context-dependent language. It might misinterpret the meaning of sentences, especially in cases where the context is not explicit.

# **THANK YOU!**

**LET'S  
WORK  
TOGETHER**

