

COMPARISON OF THE TEXT SUMMARIZATION ALGORITHMS

Arda Cem Özmen - 115200046

Tuğra Burak Çakıcı – 117200073

09.01.2021

Advisor: Uzay Çetin

CONTENTS

- Basic Summary
- Problem Definition
- Design of Solution
- Some Components
- Methodology
- Conclusion



BEFORE WE
START...

Homeschooling: pros and cons.

With more than 2.3 million students that got home in 2016 alone, more and more parents wonder whether they should opt for home education for their kids too. They wonder whether it is a better choice concerning values and beliefs spread at public schools and whether this education style won't affect their kids' ability to socialize. There are two sides to this medal. And I will look at both of them in this article.

Homeschooling advantages

First of all, homeschooling is a great way to protect your kids from increasing bullying which, unfortunately, is getting out of control these past years. Parents who do not want their kids to become victims of peer pressure and active bullying, often decide to raise and educate their kids themselves. And it is a great solution!

Second of all, home-based education gives you freedom. Students have schedules, but they are much more flexible than those at public schools. They can easily reschedule their class if unexpected guests come to visit or their family decides to go on a spring vacation earlier this year.

Finally, homeschooling allows parents to share their core values and beliefs with their kids as they stay side by side at home daily. With the world getting more and more liberal in many senses, it is great that at least some families choose to lay a firm foundation in their kids' beliefs system.

However, there are also several downsides of homeschooling.

Homeschooling disadvantages

We would like to state that homeschooling can be quite expensive. With all the books to buy and software to install, parents often get to spend a lot of money on supplies. And this can be viewed a disadvantage.

Besides, home-based education does not give as many choices to socializing. Parents often eliminate this by offering their kids to join a local sports team or join some class. That is where homeschooled kids can get all the communication and socialization they want. But if such things are nowhere close to their neighborhood, it can be a problem.

Last but not least home-based education means that children do not have access to skilled and well-trained teachers. They have their parents to help them, but as long as their parents are not professionals in this area, there might be some troubles along the way.

As you see homeschooling has both pros and cons one has to take into account when making a final decision. Personally, I believe that homeschooling is great, and parents who have a chance to educate their kids this way should definitely go for it. But I understand the concerns of people not supporting this concept at all. What side are you on?

The Long March episode - Democracy and Dictatorship- Two sides of the same coin?

by MAJED IQBAL

The weekend in Pakistan saw the much awaited long march take its turn to add itself onto the list of the on-going Political crisis plaguing the country.

President Zardari had pledged to reinstate the deposed Chief Justice Mr Chaudhry, who had been removed by Musharraf in the November state of emergency, within 30 days of taking office, but reneged on the promise, fearing the justice would re-open the case of alleged corruption by his late wife Benazir Bhutto and him over which immunity was granted by Musharraf in a potential power sharing deal.

Power

The election of the PPP in power in January and the 'in-house' move of Zardari to Presidency saw him back down on his promise during the elections of re-instating the deposed judges if the PPP won.

Whilst Musharraf was in power we saw parties rally around the message that dictatorship was bad for the country and that is should be removed to guarantee 'real' progress. The message was not just for Musharraf but anyone with similar ambitions for dictatorial rule which had caused havoc in the country through the pursuance of the war on terror with no consultation from the national assembly, raising Juniors in army ranks above seniors who were more

loyal to Musharraf, sending the heads of the two opposition parties into exile and sacking judges who could contest President Musharraf's authority as un-constitutional.

Democracy

Today, Democracy, through free and fair elections has bought Zardari, a renowned crook into the highest office of the country, allowed the bombing and killings in swat which has seen 500,000 people misplaced and thousands of innocent men, women and children killed by American drones, lying to the public that American strikes are unstoppable as they take place from Afghanistan.

Whereas new information proves that the Pakistani gave the green light to America to bomb the citizens of the country, banning media reporting against the government as was done with GEO and colluding with the army to crackdown on lawyers, activists and opposition parties calling for the re-instatement of the deposed chief justice.

Clampdown

What can the people of Pakistan expect next even after the Chief Justice has been sworn back into power?

How can people rest their confidence in Democracy for Pakistan after the information minister, Sherry Rehman resigned in the wake of the growing clampdown by government on protests by lawyers and opposition groups which speaks volumes on how many in the government see no difference in



• PROTESTS: Thousands of people took to the streets in Pakistan over the weekend

the previous regime and the new one?

it under Musharraf's rule, are they now willing to challenge democracy and ask for its removal after witnessing its fruits?



**Supporting your needs
is an Amanah**

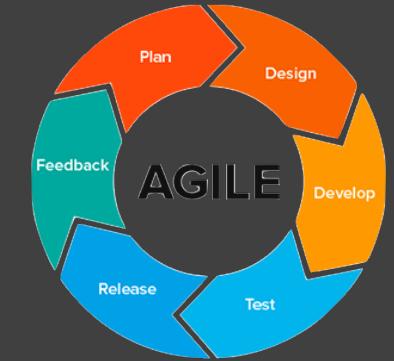
ABSTRACT

- We will try to find the best possible solution to summarize texts, by comparing text summarization algorithms
- We will also be able to observe how valid deep learning can get using different variables for their solutions.
- By the end, we will create our own summarization model that will serve for the general problem with most efficiency.

The Main Idea Behind Our Project

- People's time is important, more than ever.
- Main idea is almost always the most important part of the text.
- When the summarized version is enough, a tool is needed to do it in place of us.
- Must be efficient, precise and as fast as possible.

How We Will Operate?



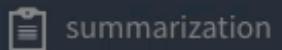
- Compare the different types of algorithms.
- Classify said algorithms and categorizing where to use, when to use.
- Create our own model powered by Fast API.
- Implement the model for Turkish language.

Abstractive Summarization

- The model will create its own summarization by using the info it gets from the text.

THE WORDS MUST BE ITS OWN!

- There are multiple ways to construct an abstractive model, so we will try to tackle the important ones: using Seq2Seq, LSTM and BERT(transformer-based) models.
- There are also some semi open-source or closed source models like VizSeq and Google abstractive summarization models.

Hosted inference API ⓘ

summarization

The tower is 324 metres (1,063 ft) tall, about the same height as an 81-storey building, and the tallest structure in Paris. Its base is square, measuring 125 metres (410 ft) on each side. During its construction, the Eiffel Tower surpassed the Washington Monument

Compute

This model can be loaded on the Inference API on-demand.

Contributed by

Google AI

company

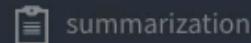
4 team members · 87 models

Extractive Summarization

- The model will salvage the summarization of the text from its main components.

THE WORDS WILL COME FROM THE TEXT

- The main idea will act as the summarization of the text.
- TextRank is its one of the most popular algorithms.

Hosted inference API ⓘ

summarization

The tower is 324 metres (1,063 ft) tall, about the same height as an 81-storey building, and the tallest structure in Paris. Its base is square, measuring 125 metres (410 ft) on each side. During its construction, the Eiffel Tower surpassed the Washington Monument

[Compute](#)

This model can be loaded on the Inference API on-demand.

Contributed byGoogle AI company

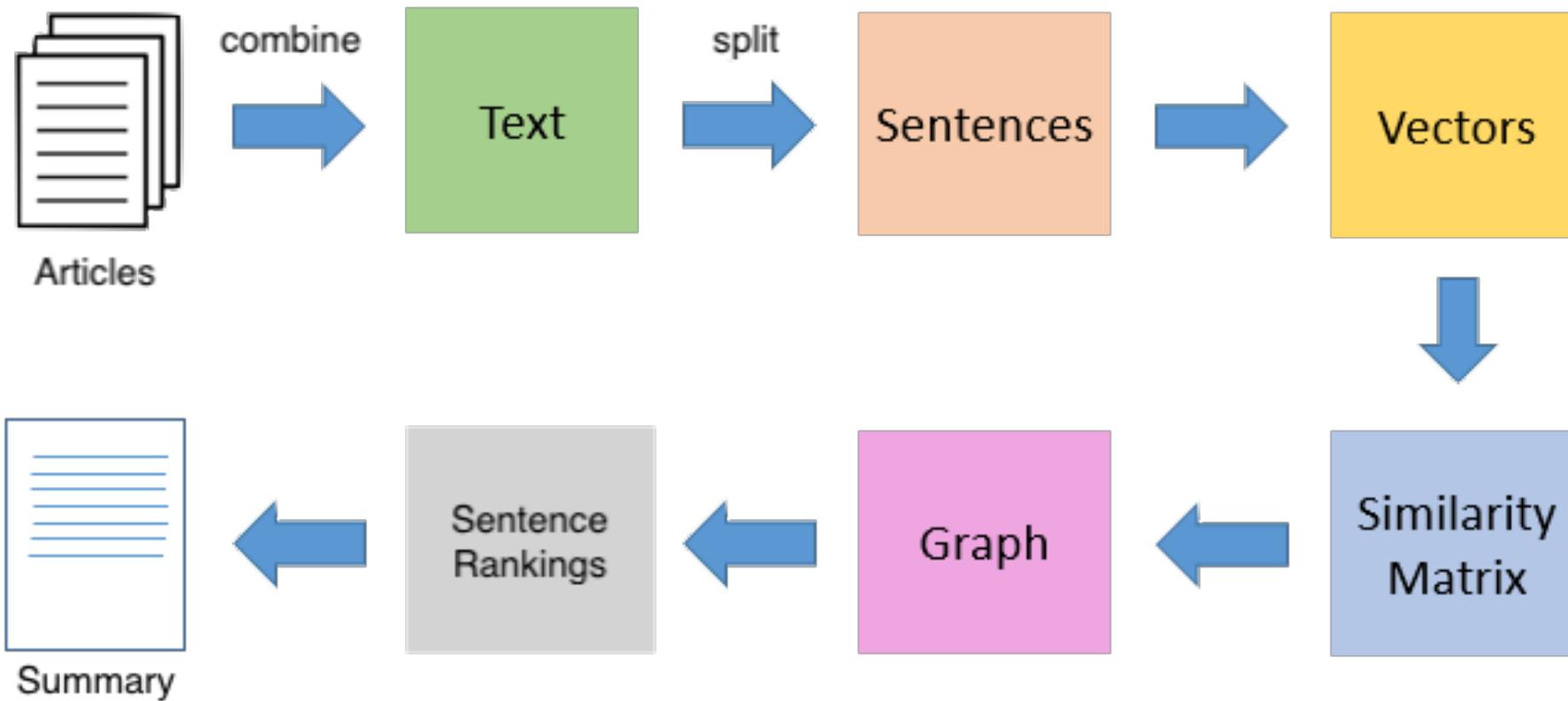
4 team members · 87 models

DATA PREPARATION

- The data that is used has gotten by using webscrapping(using BeautifulSoup, Requests, html.parser, lxml).
- One other way that has been used to pull data is taking it from kaggle(using pandas).
- In order to process the data, data manipulation(cleaning) has been done to all texts. This has been done to make text summarization go smoothly.

TEXTRANK

- TextRank is a graph-based ranking model for text processing
- It can be used to find the most relevant sentences by ranking its keywords.
- Uses the PageRank, which ranks the websites according to their reference-based citations, to rank the words within sentences.
- Some of the TextRank implementations can be "directed" by adding semantic relations



TEXTRANK

PROCESS

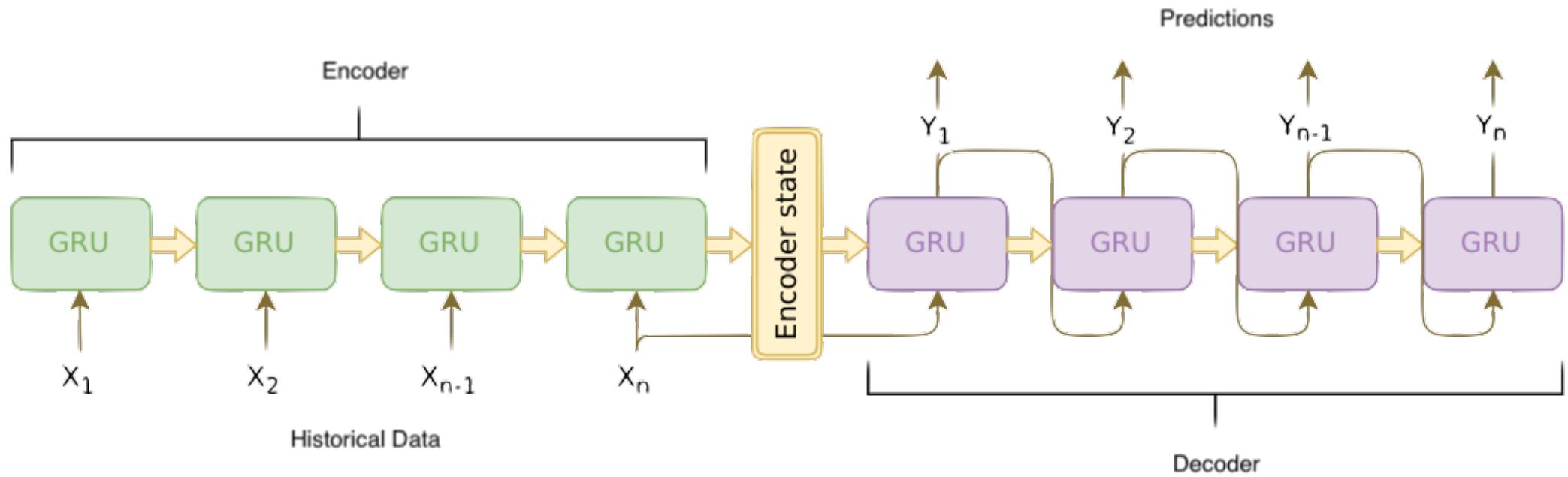
- Tokenize the text into sentences
- Tokenize each sentence into a collection of words
- Convert the sentences into graphs
- Score the sentences via pagerank

PROS

- Easy to build in terms of coding structure.
- TextRank implementations tend to be lightweight and can run fast even with limited memory resources compared to transformers models built like BERT and such.
- Data size can be comparatively smaller when compared to models like Seq2Seq.

CONS

- TextRank usually doesn't cover for semantic relations.
- Behaviour of the repeated words can disrupt the main idea.
- The varying size of texts can confuse the ranking.



SEQ2SEQ

SEQ2SEQ

- A Network system that is convenient for generative results.
- Recursive nature for long term inputs to be transformed into output.
- Mainly used in Speech Recognition, Translation and Text Summarization.

BASIC SUMMARY

- Seq2Seq turns the data into an input sequence to be processed as an output sequence
- It is built within LSTM based encoder and decoder models.
- The output sequence will be normalized to get the summary.

PROCESSION

- Text extraction and cleaning.
- Setting the rules for data processing.
- Vectorizing and tokenizing the data.
- Model Building and Optimization.

OUTPUT

Review: actress margot best known for playing the role of lois lane in the superman film franchise of the 1970s and 1980s passed away on sunday at the age of 69 an autopsy will be performed to determine the cause of her death margot who suffered from car accident in 1990 had experienced long term mental health issues

Original summary: start actress who portrayed lane in superman film passes away end

Predicted summary: start actress passes away at cannes end

PROS

- Generative Summary
- Flexible coding options.
- Dictation over Summary options.

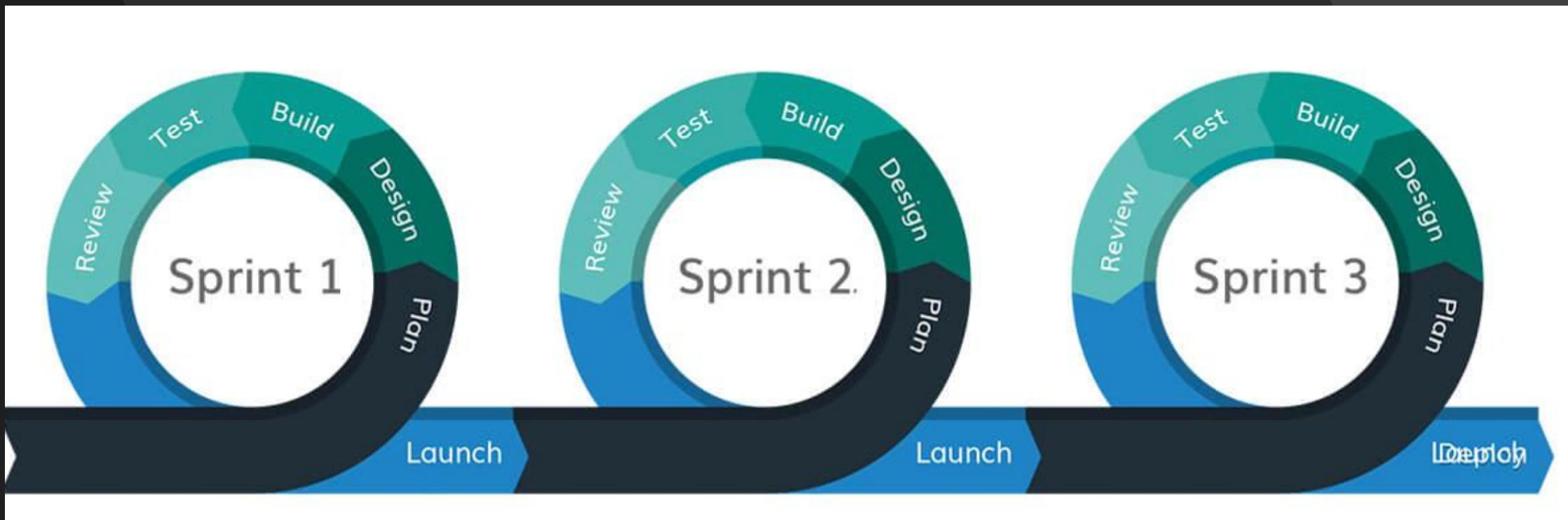
CONS

- Optimization cost.
- Data availability.

TRANSFORMERS

- A collection of pretrained models which can be used in text summarization, translation and generation and so on.
- Using pipelines, any pretrained model will be able to be used by system with quite a less cost.
- More work will be done about Transformers.

METHODOLOGY



AIMED ACHIEVEMENT

- Understanding the summarization process made by deep learning technology
- Creating a product that will be able to serve to people.
- Overall, gaining experience on a field that we are not familiar of.

ANY QUESTIONS?

THANKS FOR LISTENING..

Arda Cem Özmen / 115200046
Tuğra Burak Çakıcı / 117200073

References

- <https://huggingface.co/>
- <https://www.analyticsvidhya.com/blog/2018/11/introduction-text-summarization-textrank-python/>
- https://en.wikipedia.org/wiki/Agile_software_development
- <https://google.github.io/seq2seq/#introduction>
- <https://cran.r-project.org/web/packages/textrank/vignettes/textrank.html>