

Summarizing Text From Learning Material Using ML

HIGHLIGHTS



25% Reduction in Course
Creation Costs



30% Time Saved With Text
Summarization



Evaluation of Summary Without
Manual Effort

THE CUSTOMER

The customer creates learning material for students. This is done by manually referencing Wikipedia and then cleaning & trimming content.

THE CONTEXT

The customer was finding it challenging to select suitable abstractive and extractive summarization algorithms. They were using strong computing power to run memory and CPU-intensive algorithms. In addition to this, the customer was writing code from scratch in order to fine-tune it.

THE OBJECTIVE

The objective was twofold. First, to extract data from Wikipedia and clean it. Second, to print a summary based on word count and fine-tune algorithms.

THE SOLUTION

Here's the solution Team Grazitti implemented for the customer:

- Used Python to extract data from Wikipedia into a CSV file and clean it using regular expressions such as HTML tags & extra punctuation marks
- Used this data in a BERT-based (Bidirectional Encoder Representations from Transformers) summarization algorithm and printed a summary based on word count
- Enabled the fine-tuning of algorithms by using a file with pre-existing documents and their summary, and inserting it into the algorithm

THE OUTCOME

The customer now has the ability to summarize text from learning material by fine-tuning algorithms according to requirements. They are able to extract data from Wikipedia articles. In addition to this, they have the ability to summarize a big piece of text into an x% summary (x is a variable added by the customer into the system as per requirements). Furthermore, they are able to evaluate the summary without human effort.

