

# Proposal - Bangkit 2023 Company-based Capstone

**Please make a copy of this document (do not ask for edit permissions)**

Replace the highlighted part of this proposal with your team's answers.

Please see the limits of each answer below. You may exceed one page if necessary.

**Company : Dicoding Indonesia**

**Problem Code : C23-DF**

## **Team Member:**

- (ML) M181DSX0351 - Carles Octavianus
- (ML) M017DSX0074 - Antonius
- (ML) M181DSX2305 - Agustinus Bravy Tetuko Ompusunggu
- (ML) M181DSX0290 - Jeremy Christianto
- (CC) C181DSX3235 - Aryoshi Wicaksono
- (CC) C181DSX3342 - Sabet Maulana Mubarak

## **Why is this problem/project interesting for your team?**

Enhancing discussion forum user-system interaction to be more human-like presents an NLP challenge that intrigues us to solve it.

## **What's your group's initial idea to work on this project?**

We propose two features for the forum: tagging suggestions and semantic search.

Tagging suggestions will automatically assign predefined tags, as a suggestion, to new discussions based on their title and description, improving visibility and aiding proper tagging for new users.

Semantic search will retrieve discussions and videos based on the similarity between the query and the contents (transcription in the case of videos), allowing users to find relevant discussions and related videos without the need for exact queries. Both features aim to enhance user experience on the forum and improve the accessibility and organization for all users.

## **Does your team have unique solutions to be proposed?**

The features we propose are essential for a discussion forum to have, yet still achievable within a one-month time span. Moreover, we will make use of the latest state-of-the-art (SOTA)

## Proposal - Bangkit 2023 Company-based Capstone

technology, such as transformer-based models for our tagging suggestions and semantic model, While not as complex or costly as the GPT text generation model, which is expensive to train and fine-tune. Additionally, we will ensure the scalability of the system, making it easy to scale up in the future. For detailed infrastructure and diagrams, please refer to the answer provided in the last question on the list.

### **Based on your knowledge, what tools/IDE/Library will your team use to solve the problem?**

Below is a list of the tools, IDEs, and libraries that we will be using (although this list is not exhaustive):

Wandb, Hugging Face, Tensorflow/pytorch, Docker, Label Studio, Gantry, Gradio, kaggle/colabs (cloud gpu), Google cloud platform.

### **Based on your knowledge and explorations, what will your team need support for?**

- Discussion forum data (title and description), and search query data.
- Mentors with expertise in natural language processing (NLP), particularly for Indonesian language
- Access to cloud GPU resources

Note: This list is not exhaustive, and additional items may be needed.

### **Any other notes/remarks we should consider on your team's application?**

While these are not the final system design diagrams, below are the proposed system design diagrams for each feature, as well as the overall project infrastructure diagram:

# Proposal - Bangkit 2023 Company-based Capstone



Rangga Raditya Delpiero • 22 Feb 2023

SELESAI

Kapan harus menggunakan lt, gt, le, dll?

Title (features)

Kapan kita harus menggunakan le, ge, gt, lt, dll? Kenapa tidak menggunakan >, < saja?

Description (features)

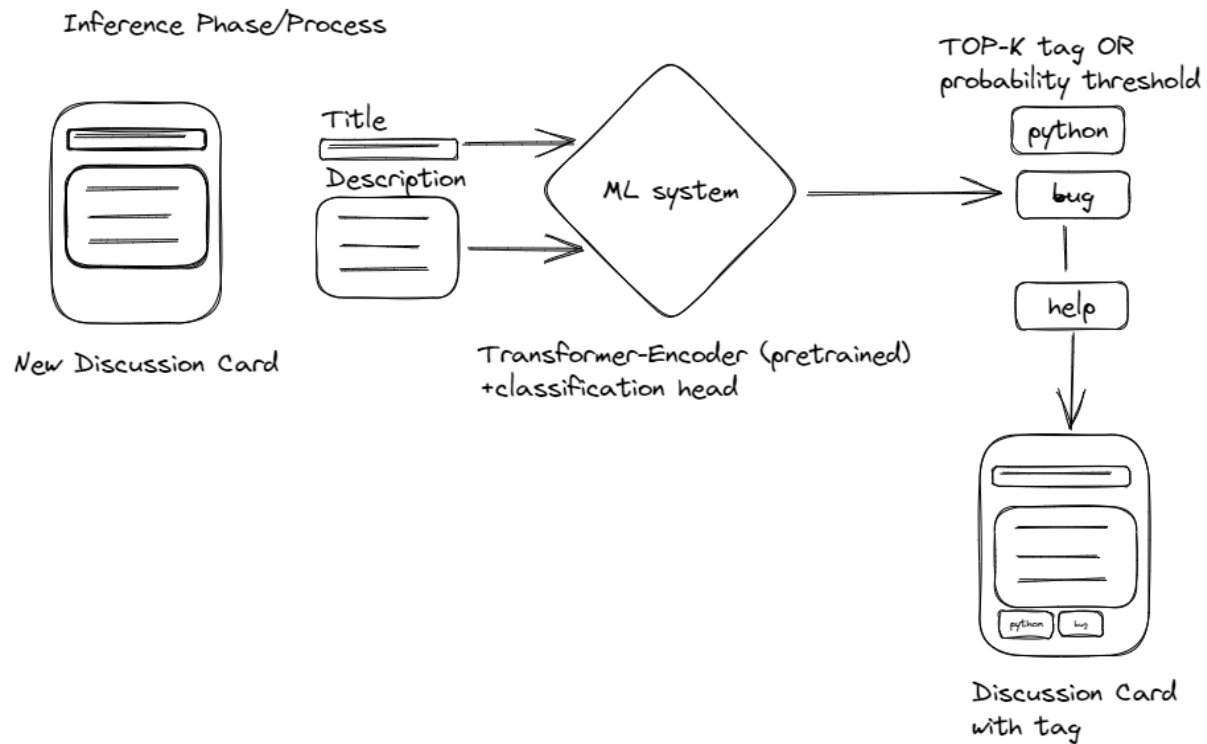
#python #operand

Tagging (target)

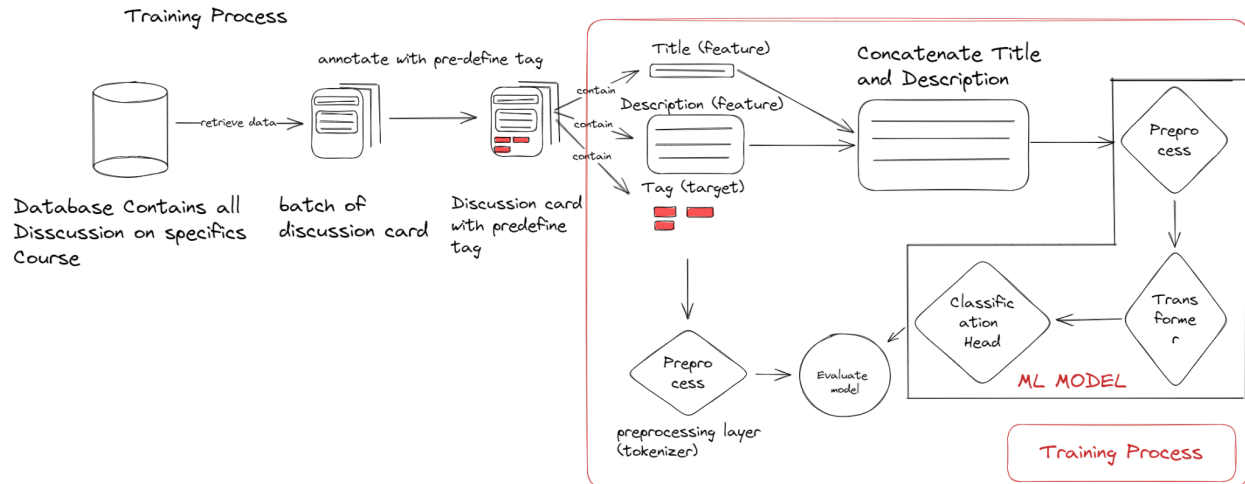
6 Pembahasan

Operator, Operands, dan Expressions

Discussion Auto-tagger

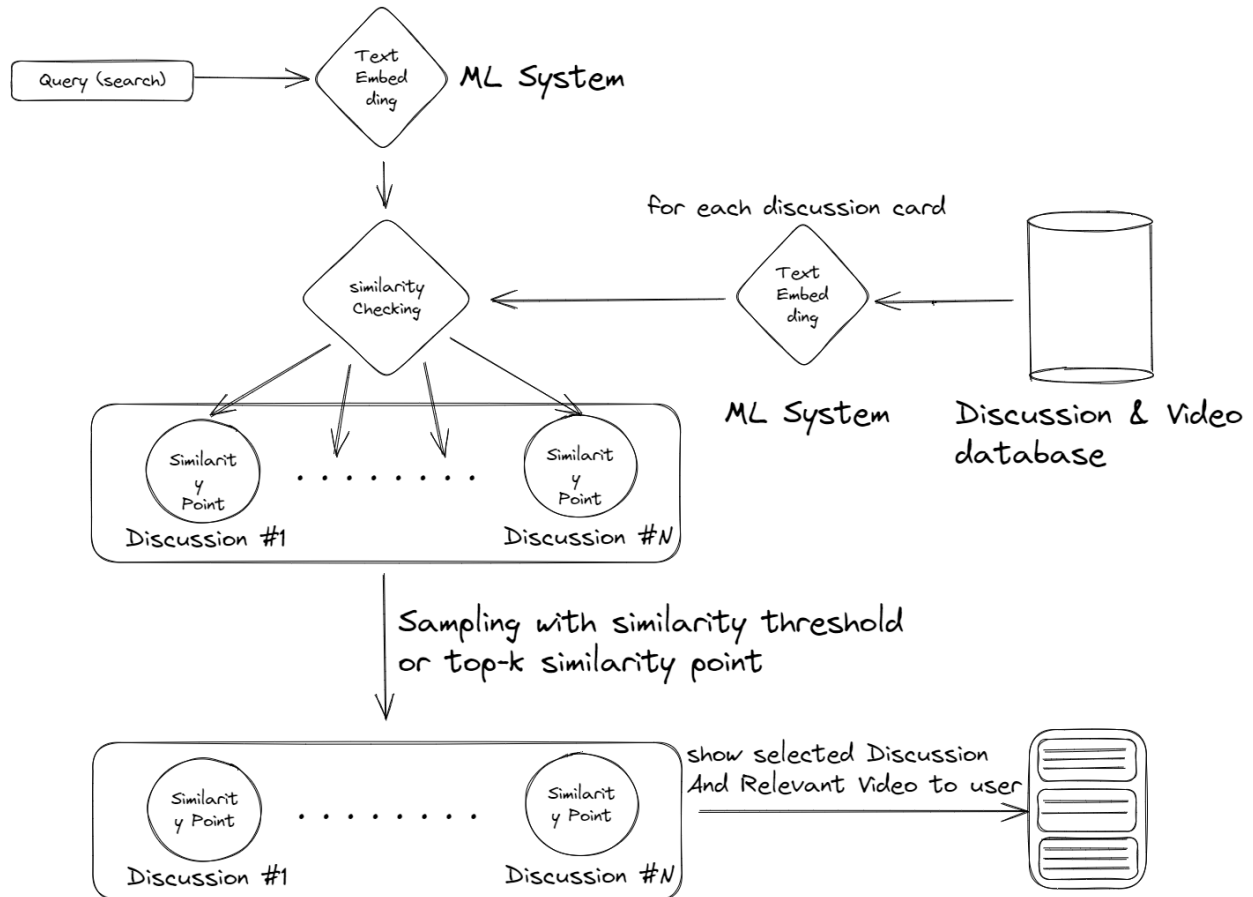


## Proposal - Bangkit 2023 Company-based Capstone



# Proposal - Bangkit 2023 Company-based Capstone

## Discussion and Relevant Video Semantic Search



# Proposal - Bangkit 2023 Company-based Capstone

## OVERALL PROJECT INFRASTRUCTURE

