PATRICK GILBERT
R I O
SHABYAKSA HERMAWAN

BINUS SUPPORT AI

Project Pitching 1.

THE GROUP LEADER

THE PROGRAMMER

THE RESEARCHER

PROBLEM BACKGROUND

Cause & Reason

The purpose of this project is to create a Binus Support chatbot that can actually help us with our college-related problems of which that a Live-Agent's assistance is not necessary.

Our very own struggle, learned the hard way

When I was an oblivious freshmen, as I waited faithfully for the indeterminate response of a Binus Support Live-Agent :(

LITERATURE REVIEW

Text Summarization for Big Data Analytics:

A Comprehensive Review of GPT 2 and BERT Approaches

Bharathi Mohan Gurusamy, Prasanna Kunnar Rangarajan

As a contextual embedding model, BERT excells at examining context and summarising texts by indentifying connections between words.

According to the data from the experiments done by the authors, BERT shows superiority over GPT 2 in extracting key information.

LITERATURE REVIEW

Language Models are Unsupervised Multitask Learners (GPT 2)

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever

GPT-2 and its performance in next word prediction, achieving state-of-the-art results on various language modeling benchmarks.

This capability allows it to generate coherent and contextually relevant text and capture complex language patterns, excelling in reading comprehension and question answering.

LITERATURE REVIEW

IndoLEM and IndoBERT:

A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP

Fajri Koto, Afshin Rahimi, Jey Han Lau, Timothy Baldwin

IndoBert is a BERT-style language model specialised in Indonesian, pre-trained with IndoLEM, a new comprehensive benchmark dataset for Indonesian natural language processing (NLP).

PURPOSE & BENEFIT

(Minor) Quality of life improvement for Binusians (& to be)

Get your questions answered right away

No more unfounded queues or waitlists for answers to your questions that don't require a Live-Agent's assistance.

Eg. , “Apa saja peraturan dan larangan di area kampus?”

METHODS

BERT
GPT 2.0
INDOBERT

Tools that we use

And this is how we do it

*We decided to combine all three methods, the explaination is as follows

FIRST METHOD

Python & BERT

BERT uses deep bidirectional representations and effective pre-training strategies to achieve results across multiple tasks with a unified architecture which demonstrates its versatility and power in understanding language.

Better at discerning context for better answers :)

Is designed to be an input for another system):

SECOND METHOD

Python & GPT-2

Capable of generating human like texts :)

Struggles with gaining context from long texts):

This is why we need BERT

- LANGUAGE MODELING: USES A PROBABILISTIC FRAMEWORK OF LANGUAGE MODELING, WHERE THE MODEL PREDICTS THE NEXT WORD BASED ON PREVIOUS WORDS.

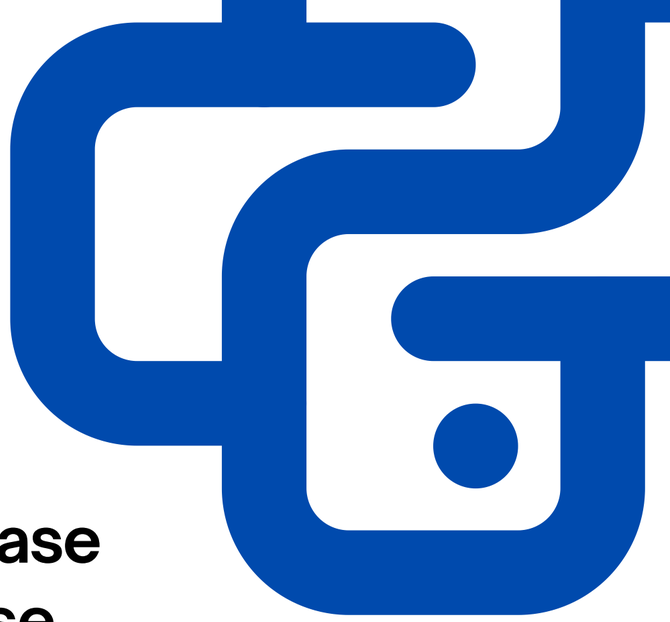
- TASK CONDITIONING: THE MODEL CAN CONDITION ITS OUTPUTS BASED ON THE INPUT AND THE TASK, ALLOWING IT TO PERFORM VARIOUS NLP TASKS BY INTERPRETING NATURAL LANGUAGE INSTRUCTIONS.

THIRD METHOD

Python & IndoBERT

A new variation of a BERT-style monolingual language model for Indonesian that has been shown to excel in seven tasks covering morpho-syntax, semantics, and discourse analysis.

BASED?
BASED ON WHAT?



We chose Python as our base of operation mainly because of its ease of use and extensive external support.

Python

3.12.0

Python 3.12.0 is not the latest version, but it is the “latest” version that almost every libraries support.

BASED?
BASED ON WHAT?

HTML parsing and text extraction

PDF downloads and text extraction

BERT & IndoBERT for text reading.
GPT-2 for generating output

Python

And the power of friendship

COULD THIS BE... A PROTOTYPE? 🦋

```
8   urls = [
9       "https://binus.ac.id/program/", #nanti masukin satu-satu link masing-masing jurusan    "nanti": Unknown
10      "https://sevima.com/jurusan/mobile-application-technology/",
11      "https://binus.ac.id/2021/10/11-jurusan-di-school-of-computer-science-basis-lulusan-canggih/"
12  ]
13
```

PROBLEMS 11 | OUTPUT | DEBUG CONSOLE | TERMINAL | PORTS | COMMENTS | Code

```
About UsAbout BINUSHistory
Educational Philosophy
Vision and Mission
BINUS Higher Education & BINUS University Officers Directory
Brochure Collection

QualityISO-9001 Certification
IT Based Business Process
Delivered Ontime Graduation
International Quality Graduates
Independent Auditors Report2015
2014
2013
2012
```

WHERE HAVE YOU BEEN?

imperfections

Pardon our inability to showcase our progress, due to our current model's schizophrenic nature of talking nonsense.

Where we are right now

Our program has shown promises of being able to fetch informations from websites and PDF texts. Though still challenged by contextual fetching, its definitely got the spirit!!

PLANS

Where we wanna be next

We yearn to improve our methods and are open to trying other QnA models & NLP services that are still within our budget of Rp 0,-

“Its gotta be at least as good if not better than the Live-Support team" -Researcher, Shabyaksa

PROGRESS SUMMARY

What we have done	What we are currently on	Our next steps
Early research	Looking and trying on available NLP alternatives	Application of found alternatives
Early method applications	Keep improving and working on our code	Implementation of said improvements
First prototyping		Keep it going

NOTABLE LINKS

Literature 1 : **Text Summarization for Big Data Analytics**

- https://www.researchgate.net/publication/374044569_Text_Summarization_for_Big_Data_Analytics_A_Comprehensive_Review_of_GPT_2_and_BERT_Approaches

Literature 2: **Language Models are Unsupervised Multitask Learners**

- https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Literature 3: **IndoLEM and IndoBERT**

- https://arxiv.org/pdf/2011.00677
- https://scholar.google.co.id/citations?view_op=view_citation&hl=en&user=RA9l3s4AAAAJ&citation_for_view=RA9l3s4AAAAJ:8k81kl-MbHgC

GIT inspo: https://github.com/OShaikhDevOps/bertBot.git

END OF SLIDE

Thank You

Please feel free to ask any questions or provide feedback.

From group 9