PROJECT
PROPOSAL

ANALISIS ULASAN PELANGGAN E-COMMERCE MENGGUNAKAN KLASIFIKASI SENTIMEN BERBASIS TRANSFORMER DAN PEMODELAN TOPIK

Novia Natasya (203012510034)
Erlangga Dewa Sakti (203012510035)

❯ Start Slide
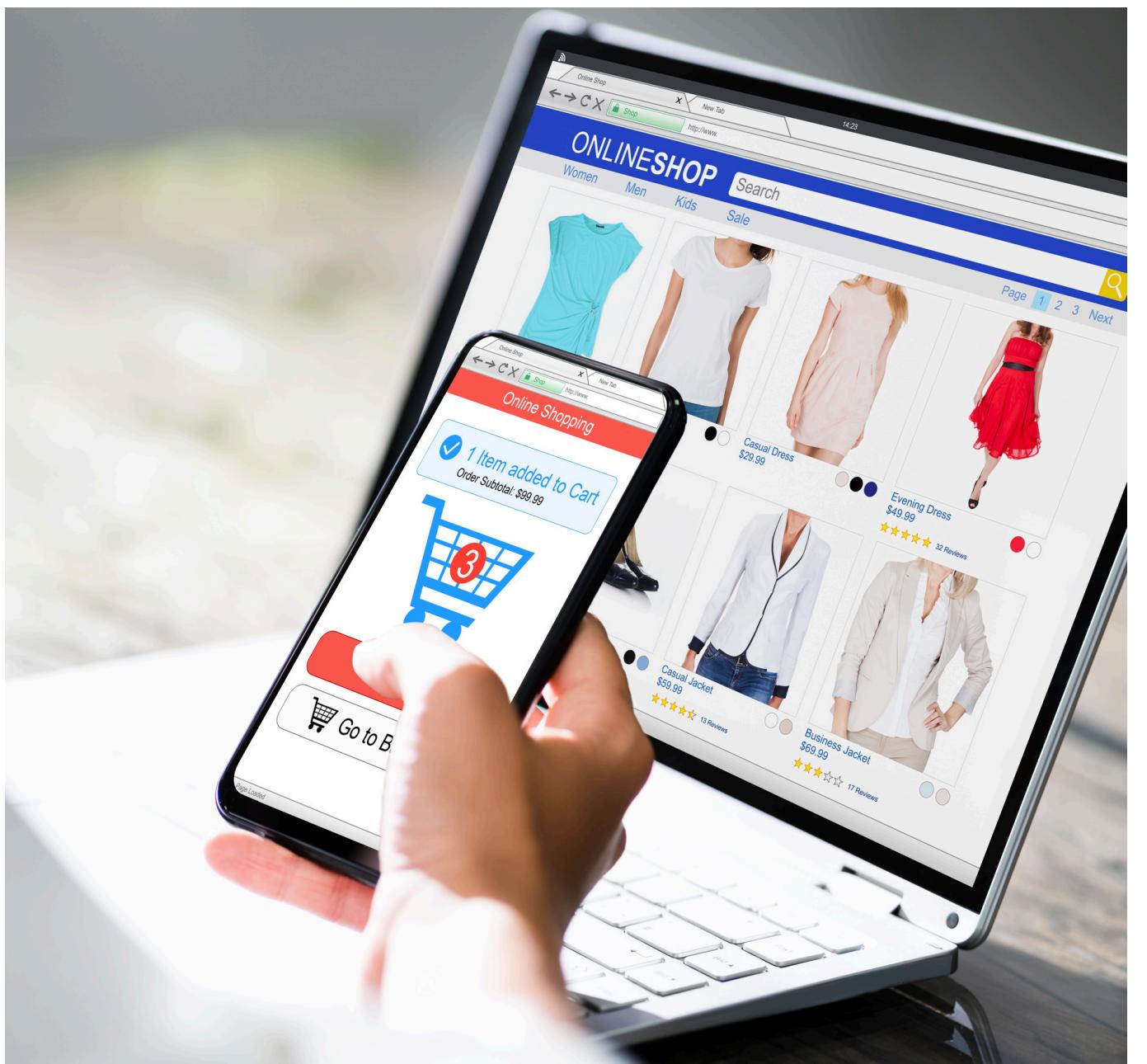
We Bring Your
Ideas To Life

Ulasan adalah sumber feedback vital di industri e-commerce untuk memahami kepuasan pelanggan dan performa produk.

Tantangan Utama: Volume data ulasan sangat besar dan tidak terstruktur, sehingga analisis manual tidak efisien.
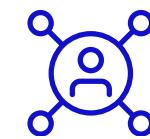
The Ideas

Menggunakan **Natural Language Processing (NLP)** untuk otomatisasi analisis dan mendapatkan insight seperti memprioritaskan perbaikan fitur produk.

Tujuan Proyek
Our Goals

Pertama

Bagaimana distribusi sentimen (positif, negatif, netral) dari ulasan pelanggan secara keseluruhan?

Kedua

Apa saja topik utama yang paling sering dibicarakan pelanggan dalam ulasan mereka?

Ketiga

Bagaimana distribusi sentimen untuk setiap topik? (Contoh: Apakah sentimen untuk topik "baterai" positif, sementara untuk "pengiriman" negatif?)

Dataset
ECInstruct

ECInstruct dibagi menjadi training sets dan validation sets.

Sumber:
https://huggingface.co/datasets/NingLab/ECInstruct

Data Split Distributions

Split	Size
Train	92,022
Validation	9,253
Total	116,528

Overall **116,528**
SA only **30,000**

output string	input string	options string	task string	split string
E: very negative	Clock time would not set. Returned and received replacement. New one was able to set time initially. After short time the…	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train
B: positive	This is a very good mp3 player. I found it very workable, and the sound quality is great. It's just as good as a cd, or mayb…	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train
A: very positive	Another Benchmade, another love story. GREAT PRICE on Amazon, I got a really good deal!	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train
B: positive	This molding is paintable, holds well and easy to install. I ran it on my wall and ceiling to hide our wiring and it looks…	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train
A: very positive	The XF times adapter is well made as I've compared with other similar adapters, this one was one of the best I came across,…	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train
A: very positive	Since this type of exercise is coming back in style, I was glad I was able to find one easily on line.	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train
B: positive	I use the USB reader to transfer photos from my Olympus D-400 Zoom digital camera to my iMac. Before buying my iMac, I had…	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train
A: very positive	Works great, haven't had any issues with this tray and have cooked with out about 5 times.	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train
E: very negative	The bean and coffee holder would stick in the coffeemaker from the beginning and I had to use a knife to get it out. Finally …	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train
B: positive	I use this item with P90x and I get a great workout with it. I'm pretty fit and this seems to be just right. I do wonder if…	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train
A: very positive	Bounce capability of the 420ex is worth the money over the 220ex. I use this with a Canon G5 and while the 420ex seems to…	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train
A: very positive	I was a bit skeptical about this product before I bought it. Having used it a couple of times I would say I made a very goo…	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train
A: very positive	GREAT! best price you can find on these. don't pay 15-20$ for one sock. .	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train
A: very positive	A little pricey for what they do and how basic they are, but they have a reputable name behind them, which in this case…	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train
A: very positive	Great product! Adjustable waist, reinforced knees, nice color and style!	["A: very positive", "B: positive", "C: neutral", "D: negative", "E: very negative"]	Sentiment_Analysis	train

```
Descriptive statistics of the DataFrame:
              output                                          input                              options              task  split  consolidated_output
count          30000                                          30000                                28000             30000  30000                30000
unique             5                                          12999                                    1                 1      3                    3
top    A: very positive   Ditch your drip and percolated coffee makers. ...   ["A: very positive", "B: positive", "C: neutra...   Sentiment_Analysis  train             Positive
freq           18426                                              4                                28000             30000  20000                23614
```

Value counts for 'output' column:

	count
output	
A: very positive	18426
B: positive	5188
E: very negative	2700
C: neutral	2276
D: negative	1410

	count
consolidated_output	
Positive	23614
Negative	4110
Neutral	2276

dtype: int64

Value counts for 'task' column:
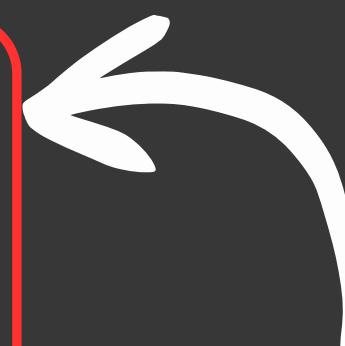
	count
task	
Sentiment_Analysis	30000

dtype: int64

Value counts for 'split' column:

	count
split	
train	20000
test	8000
val	2000

dtype: int64

Menggabungkan dari 5 label menjadi 3 label:
- "Very Positive" & "Positive" --> Positive.
- "Very" negative" & "Negative" --> Negative.
- "Neutral" == Neutral.

Tools & Libraries
## For The Project

Bahasa Pemrograman: Python
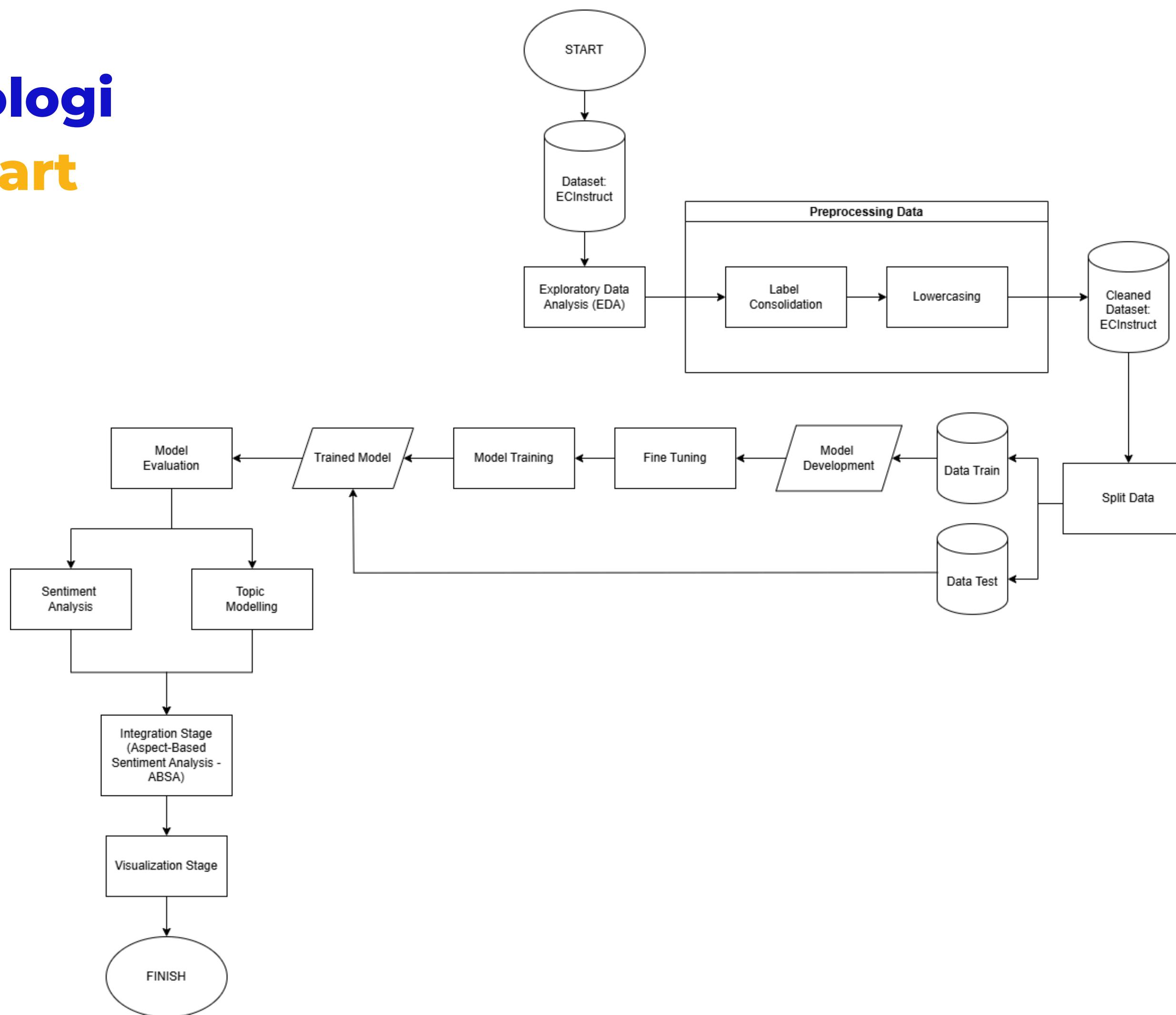
Library Utama:
- Hugging Face Transformers
- BerTopic
- Pandas & Scikit-learn
- Matplotlib & Seaborn

Metodologi
Flowchart

Metodologi
Gambaran Umum

Exploratory Data Analysis (EDA)

Proses awal untuk memahami karakteristik, pola, dan hubungan dalam dataset.

Pra-pemrosesan Data

Membersihkan dan menyiapkan data ulasan.

Analisis Sentimen

Melatih model Transformer (seperti BERT) untuk mengklasifikasikan sentimen.

Pemodelan Topik

Menggunakan BerTopic untuk mengidentifikasi topik-topik utama.

Analisis Sentimen Berbasis Aspek (ABSA)

Menggabungkan hasil sentimen dan topik.

Visualisasi & Interpretasi

- Visualisasi hasil sentimen dan topik.
- Analisis Sentimen Berbasis Aspek (ABSA)

Metodologi
Detail Teknis

Sumber Data

Dataset ECInstruct, diambil 30.000 (SA task) dari 116.528 sampel data untuk tugas NLP analisis sentiment e-commerce.

Pra-pemrosesan

- Menyederhanakan 5 label sentimen menjadi 3 (Positif, Netral, Negatif).
- Hanya lowercasing, karena stopwords dan tanda baca sangat penting untuk konteks model Transformer (Alzahrani et. al.)

Model yang Digunakan

- Analisis Sentimen: Model Transformer (BERT).
- Pemodelan Topik: Library BerTopic.

Metrik Evaluasi

Akurasi, Presisi, Recall, dan F1-Score untuk mengukur performa model

[1] E. Alzahrani and L. Jololian, "How Different Text-Preprocessing Techniques using the Bert Model Affect the Gender Profiling of Authors," Academy and Industry Research Collaboration Center (AIRCC), Sep. 2021, pp. 01–08. doi: 10.5121/csit.2021.111501.

Hasil yang Diharapkan
Output Research

Laporan Distribusi Sentimen

Grafik yang menunjukkan persentase ulasan di setiap label.

Analisis Sentimen Berbasis Aspek (ABSA)

Melakukan pemetaan sentimen untuk setiap aspek topic produk.

Daftar Topik Utama

Identifikasi 5-10 topik teratas beserta kata kuncinya.

About The
Analysis

Conclusions
The Project

Proyek ini menunjukkan bagaimana NLP dengan model berbasis transformer dapat mengubah ulasan pelanggan menjadi aset strategis yang lebih akurat bagi bisnis e-commerce untuk meningkatkan performa penjualan.

TERIMA KASIH

FOR YOUR ATTENTION!

Prinsip Sains Data

❯ End Slide