

Lampiran A- 1
Dataset Awal judul_tiktok.csv 1

Total baris: 99,065

Kolom: ['aweme_id', 'create_time', 'text', 'unique_id']

Video unik: 388

Pengguna unik: 76074

	aweme_id	create_time	text	unique_id
0	7376590288212479237	1717846551	Uda 9 bulanan main,,, \n-+10jt,, \nbelum bisa st...	user8149616194483
1	7376590288212479237	1717546502	bner saya juga ngerasain	hermawanaldsm
2	7376590288212479237	1717649591	untan	kuntan41
3	7376590288212479237	1717842605	aku sudah mengiklaskan 😊	rama.dony04
4	7376590288212479237	1717786502	entah sampai kapan.. 🤔	teguhtekek76

Lampiran B- 1

Contoh Dataset Setelah Preprocessing

Contoh *Dataset* Setelah *Preprocessing*

Dataset yang sudah diproses tersimpan di 'judol_tiktok_preprocessed.csv'

--- Data Hasil Preprocessing (Tabel) ---

	aweme_id	create_time	text	unique_id	text_clean
0	7376590288212479237	1717846551	Uda 9 bulanan main,,,n-+10jt,,\nbelum bisa st...	user8149616194483	uda bulanan main jt belum bisa stop
1	7376590288212479237	1717546502	bner saya juga ngerasain	hermawanaldsm	bner saya juga ngerasain
2	7376590288212479237	1717842605	aku sudah mengiklaskan 😊	rama.dony04	aku sudah mengiklaskan
3	7376590288212479237	1717786502	entah sampai kapan.. 🤔	teguhtekek76	entah sampai kapan
4	7376590288212479237	1717846195	Bang... ah udahlah pahit kalo diceritain 😭😭😭😭	orangpinggiran.76	bang ah udahlah pahit kalo diceritain
5	7376590288212479237	1717822167	sy sudah berhenti 2bln ini, terakhir win 60jt,...	penarilintassumut	saya sudah berhenti bln ini terakhir win jt da...
6	7376590288212479237	1718003766	saya hancur hidupku berantakan hampir ratusan ...	tokoonlinecom.www_	saya hancur hidupku berantakan hampir ratusan ...
7	7376590288212479237	1717847848	5 hari berturut turut di bantai trus	rudi.jkb	hari berturut turut di bantai trus
8	7376590288212479237	1717846998	awalnya udh berhenti, iseng depo 50 naik 300rb...	blender.spender29	awalnya sudah berhenti iseng depo naik rb eh b...
9	7376590288212479237	1717688229	modal trakhir ku jual ninja 27 jta,ku depositin...	ucilll_27	modal trakhir ku jual ninja jta ku depositin sm...
10	7288980752232975621	1697098662	Itu sejak awal. Gak semuanya saweran slot.	thisisrafljanuartatoo	itu sejak awal tidak semuanya saweran slot
11	7288980752232975621	1697108841	pertanyaan gw fee nya mau d kembaliin juga kag...	dhobret	pertanyaan gw fee nya mau d kembaliin juga kag...
12	7288980752232975621	1697098991	emang itu bukan cuma dari judi online tapi kur...	jealousguys	emang itu bukan cuma dari judi online tapi kur...
13	7288980752232975621	1697102368	lawak ujung"nya jdi drama	user891047271699	lawak ujung nya jdi drama
14	7288980752232975621	1697099631	lo mau nangkepin streamer juga percuma, bandar...	timnasindonesiainf	kamu mau nangkepin streamer juga percuma banda...
15	7288980752232975621	1697099482	ribet banget n drama,kalo emang ada indikasi k...	ocehancamelclub	ribet banget n drama kalo emang ada indikasi k...
16	7288980752232975621	1697098882	gw lebih setuju pembuat situs slot di tangkep ...	ruang_kesepian1	gw lebih setuju pembuat situs slot di tangkep ...

Lampiran C- 1

Hasil Perhitungan Leksikon Domain-Spesifik

Hasil Perhitungan Leksikon Domain-Spesifik

--- Contoh Perhitungan Sentimen Berbasis Leksikon (Detail) ---

	Text	Detected Positive Words	Positive Score	Detected Negative Words	Negative Score	Final Sentiment
0	uda bulanan main jt belum bisa stop	stop, bisa	2	-	0	Positif
1	bner saya juga ngerasain	-	0	-	0	Netral
2	aku sudah mengiklaskan	-	0	-	0	Netral
3	entah sampai kapan	-	0	-	0	Netral
4	bang ah udahlah pahit kalo diceritain	-	0	-	0	Netral
5	saya sudah berhenti bln ini terakhir win jt da...	berhenti, stop	2	-	0	Positif
6	saya hancur hidupku berantakan hampir ratusan ...	-	0	hancur	1	Negatif
7	hari berturut turut di bantai trus	-	0	-	0	Netral
8	awalnya sudah berhenti iseng depo naik rb eh b...	berhenti	1	-	0	Positif
9	modal trakhir ku jual ninja jta ku depositin sm...	-	0	-	0	Netral

Lampiran D- 1
Statistik & Distribusi Dataset

Statistik & Distribusi Dataset

Distribusi label:

label

Netral 52972

Positif 14683

Negatif 6127

Name: count, dtype: int64

Persentase:

label

Netral 71.8

Positif 19.9

Negatif 8.3

Name: count, dtype: float64

Kode Program Tahap *Preprocessing (Cleaning, Normalization, Filtering)*

```
# Kamus slang Indonesia (dapat diperluas sesuai kebutuhan)
SLANG_DICT = {
    'gpp': 'tidak apa-apa', 'bgt': 'banget', 'yg': 'yang', 'udh': 'sudah',
    'gua': 'saya', 'gue': 'saya', 'ga': 'tidak', 'gak': 'tidak',
    'ngga': 'tidak', 'nggak': 'tidak', 'aja': 'saja', 'krn': 'karena',
    'utk': 'untuk', 'dgn': 'dengan', 'jg': 'juga', 'sdh': 'sudah',
    'hrs': 'harus', 'blm': 'belum', 'dr': 'dari', 'sy': 'saya',
    'tp': 'tapi', 'jd': 'jadi', 'bs': 'bisa', 'skrg': 'sekarang',
    'judol': 'judi online', 'wkwk': '', 'wkwkwk': '', 'haha': '', 'hehe': '',
    'lo': 'kamu', 'lu': 'kamu', 'lbh': 'lebih', 'byk': 'banyak',
}

def preprocess_text(text):
    text = str(text).lower()
    # 1. Hapus URL
    text = re.sub(r'http\S+|www\S+', '', text)
    # 2. Hapus mention & hashtag
    text = re.sub(r'@\w+|#\w+', '', text)
    # 3. Hapus emoji
    emoji_pattern = re.compile(
        "[" u"\U0001F600-\U0001F64F" u"\U0001F300-\U0001F5FF"
        u"\U0001F680-\U0001F6FF" u"\U0001F1E0-\U0001F1FF"
        u"\U00002702-\U000027B0" u"\U000024C2-\U0001F251" "]"+",
        flags=re.UNICODE)
    text = emoji_pattern.sub('', text)
    # 4. Hapus karakter non-alfabet
    text = re.sub(r'^a-z\s', ' ', text)
    # 5. Normalisasi slang
    words = [SLANG_DICT.get(w, w) for w in text.split()]
    text = re.sub(r'\s+', ' ', ' '.join(words)).strip()
    return text

df['text_clean'] = df['text'].apply(preprocess_text)
print("Contoh hasil cleaning:")
for orig, clean in zip(df['text'].head(5), df['text_clean'].head(5)):
    print(f"  ASLI : {orig}")
    print(f"  BERSIH: {clean}\n")

def is_bot_like(text):
    """Filter komentar dengan pola repetisi karakter > 50% (indikasi bot)."""
    if len(text) == 0:
        return True
    chars = list(text.replace(' ', ''))
    if len(chars) == 0:
        return True
    most_common = max(set(chars), key=chars.count)
    return chars.count(most_common) / len(chars) > 0.5

n_before = len(df)
df = df[df['text_clean'].str.split().str.len() >= 3].reset_index(drop=True)
n_after_len = len(df)
df = df[~df['text_clean'].apply(is_bot_like)].reset_index(drop=True)
n_after_bot = len(df)

print(f"Sebelum filter          : {n_before:,}")
print(f"Setelah filter panjang : {n_after_len:,} (-{n_before-n_after_len:,})")
print(f"Setelah filter bot-like: {n_after_bot:,} (-{n_after_len-n_after_bot:,})")
print(f"\nTotal data siap untuk pelabelan: {len(df):,} ({len(df)/n_before*100:.1f}% dari data awal)")
```

Lampiran E- 2

Dokumentasi *Source Code* (Script Python)

Kode Program Pelabelan Leksikon

```
POSITIVE_WORDS = [
    'tangkap', 'berantas', 'hentikan', 'berhenti', 'stop', 'larang',
    'dukung', 'bagus', 'baik', 'setuju', 'mantap', 'luar biasa', 'hebat',
    'berhasil', 'sukses', 'selamat', 'alhamdulillah', 'syukur',
    'semangat', 'support', 'ayo', 'keren', 'tepat', 'benar',
    'pemberantasan', 'blokir', 'tutup', 'tindak', 'proses', 'hukum',
    'bisa', 'mampu', 'insaf', 'tobat', 'sadar', 'lepas', 'bebas',
    'edukasi', 'sosialisasi', 'pencegahan', 'lindungi'
]

NEGATIVE_WORDS = [
    'rugi', 'hancur', 'gagal', 'kalah', 'bangkrut', 'hutang', 'susah',
    'sedih', 'menderita', 'kecewa', 'parah', 'buruk', 'jelek', 'bodoh',
    'sia-sia', 'percuma', 'kecanduan', 'masalah', 'makin', 'bertambah',
    'korupsi', 'tidak bisa', 'gak bisa', 'lemah', 'lambat', 'malas',
    'lelet', 'bohong', 'tipu', 'palsu', 'salah', 'malu', 'marah',
    'kesal', 'frustrasi', 'jengkel', 'kacau', 'ribuan', 'miliaran',
    'candu', 'ketergantungan', 'korban', 'tidak mampu', 'tidak berdaya',
    'nangis', 'nagih'
]

def label_sentiment(text):
    text_lower = text.lower()
    pos_score = sum(1 for w in POSITIVE_WORDS if w in text_lower)
    neg_score = sum(1 for w in NEGATIVE_WORDS if w in text_lower)
    if pos_score > neg_score:
        return 'Positif'
    elif neg_score > pos_score:
        return 'Negatif'
    else:
        return 'Netral'

df['label'] = df['text_clean'].apply(label_sentiment)

print("Distribusi label:")
dist = df['label'].value_counts()
print(dist)
print(f"\nPersentase:")
print((dist / len(df) * 100).round(1))

# Contoh komentar per kelas
for label in ['Positif', 'Netral', 'Negatif']:
    print(f"=== Contoh sentimen {label} ===")
    samples = df[df['label'] == label]['text_clean'].sample(3, random_state=RANDOM_STATE).tolist()
    for s in samples:
        print(f" - {s[:100]}")
    print()
```

Lampiran E- 3
Dokumentasi *Source Code (Script Python)*

Kode Program *Custom WeightedTrainer* (Modifikasi *Hugging Face Trainer*)

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.svm import LinearSVC
from sklearn.pipeline import Pipeline
from sklearn.utils.class_weight import compute_class_weight

class_weights = compute_class_weight('balanced', classes=np.unique(y_train), y=y_train)
print("Class weights:", {le.classes_[k]: round(v, 3) for k, v in enumerate(class_weights)})

baseline_pipeline = Pipeline([
    ('tfidf', TfidfVectorizer(ngram_range=(1, 2), max_features=50000, sublinear_tf=True, min_df=2)),
    ('clf', LinearSVC(class_weight='balanced', C=1.0, max_iter=2000, random_state=RANDOM_STATE))
])

baseline_pipeline.fit(X_train, y_train)
print("Training baseline selesai.")
```

Lampiran E- 4

Dokumentasi *Source Code (Script Python)*

Kode Program *Fine-Tuning IndoBERT (Training loop & hyperparameter)*

```
from transformers import BertTokenizer, BertForSequenceClassification, Trainer, TrainingArguments
from torch.utils.data import Dataset
import torch.nn as nn

MODEL_NAME = "indobenchmark/indobert-base-p1"
MAX_LENGTH = 128

tokenizer = BertTokenizer.from_pretrained(MODEL_NAME)
model = BertForSequenceClassification.from_pretrained(MODEL_NAME, num_labels=3)

class SentimentDataset(Dataset):
    def __init__(self, texts, labels, tokenizer, max_length=128):
        self.texts = texts
        self.labels = labels
        self.tokenizer = tokenizer
        self.max_length = max_length

    def __len__(self):
        return len(self.texts)

    def __getitem__(self, idx):
        encoding = self.tokenizer(
            str(self.texts[idx]),
            truncation=True,
            padding='max_length',
            max_length=self.max_length,
            return_tensors='pt'
        )
        item = {k: v.squeeze(0) for k, v in encoding.items()}
        item['labels'] = torch.tensor(self.labels[idx], dtype=torch.long)
        return item

train_dataset = SentimentDataset(X_train, y_train, tokenizer, MAX_LENGTH)
val_dataset = SentimentDataset(X_val, y_val, tokenizer, MAX_LENGTH)
test_dataset = SentimentDataset(X_test, y_test, tokenizer, MAX_LENGTH)

print(f"Train dataset : {len(train_dataset):,}")
print(f"Val dataset   : {len(val_dataset):,}")
print(f"Test dataset    : {len(test_dataset):,}")
```

Lampiran E- 5

Dokumentasi *Source Code* (Script Python)

```
# Class weights untuk loss function (mengatasi imbalance)
class_weights_tensor = torch.tensor(class_weights, dtype=torch.float)

class WeightedTrainer(Trainer):
    def compute_loss(self, model, inputs, return_outputs=False, **kwargs):
        labels = inputs.pop("labels")
        outputs = model(**inputs)
        logits = outputs.logits
        loss_fct = nn.CrossEntropyLoss(weight=class_weights_tensor.to(logits.device))
        loss = loss_fct(logits, labels)
        return (loss, outputs) if return_outputs else loss

from sklearn.metrics import accuracy_score as acc_score, f1_score as f1_sc

def compute_metrics(eval_pred):
    predictions, labels = eval_pred
    predictions = np.argmax(predictions, axis=1)
    return {
        'accuracy': acc_score(labels, predictions),
        'f1_macro': f1_sc(labels, predictions, average='macro')
    }

# --- Hitung warmup_steps secara eksplisit (menggantikan warmup_ratio yang deprecated) ---
BATCH_SIZE = 16
NUM_EPOCHS = 4
WARMUP_FRACTION = 0.1 # 10% dari total steps, sesuai Tabel 2.5 proposal

steps_per_epoch = len(train_dataset) // BATCH_SIZE
total_steps = steps_per_epoch * NUM_EPOCHS
warmup_steps = int(WARMUP_FRACTION * total_steps)

print(f"Steps per epoch : {steps_per_epoch:,}")
print(f"Total steps      : {total_steps:,}")
print(f"Warmup steps      : {warmup_steps:,} ({WARMUP_FRACTION*100:.0f}% dari total steps)")

training_args = TrainingArguments(
    output_dir='./indobert_judol_results',
    num_train_epochs=NUM_EPOCHS,
    per_device_train_batch_size=BATCH_SIZE,
    per_device_eval_batch_size=32,
    learning_rate=2e-5,
    weight_decay=0.01,
    warmup_steps=warmup_steps, # diganti dari warmup_ratio -> warmup_steps (tidak deprecated)
    eval_strategy="epoch",
    save_strategy="epoch",
    load_best_model_at_end=True,
    metric_for_best_model="f1_macro",
    logging_steps=100,
    report_to="none"
)

trainer = WeightedTrainer(
    model=model,
    args=training_args,
    train_dataset=train_dataset,
    eval_dataset=val_dataset,
    compute_metrics=compute_metrics
)

print("\nTrainer siap. Jalankan sel berikutnya untuk memulai fine-tuning.")
print("Estimasi waktu: ~30-60 menit dengan GPU T4 untuk 51.700 data training, 4 epoch.")
```

Lampiran E- 6

Dokumentasi *Source Code (Script Python)*

Lampiran 5.e. Kode Program Evaluasi Model

```
# Pilih hasil yang ingin dianalisis (ganti sesuai bagian yang dijalankan)
y_pred = y_pred_baseline # atau: y_pred = y_pred_bert
classes = le.classes_

cm = confusion_matrix(y_test, y_pred)

fig, ax = plt.subplots(figsize=(7, 5.5))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', ax=ax,
            xticklabels=classes, yticklabels=classes,
            annot_kws={'size': 14, 'weight': 'bold'})

for i in range(len(classes)):
    for j in range(len(classes)):
        pct = cm[i, j] / cm[i].sum() * 100
        ax.text(j + 0.5, i + 0.72, f'({pct:.1f}%)', ha='center', va='center', fontsize=9, color='gray')
ax.set_xlabel('Label Prediksi', fontsize=13, fontweight='bold')
ax.set_ylabel('Label Aktual', fontsize=13, fontweight='bold')
ax.set_title('Confusion Matrix - Klasifikasi Sentimen', fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig_confusion_matrix.png', dpi=150, bbox_inches='tight')
plt.show()

report = classification_report(y_test, y_pred, target_names=classes, output_dict=True)

x = np.arange(len(classes))
width = 0.25
fig, ax = plt.subplots(figsize=(9, 5.5))
b1 = ax.bar(x - width, [report[c]['precision'] for c in classes], width, label='Precision', color='#185FA5')
b2 = ax.bar(x, [report[c]['recall'] for c in classes], width, label='Recall', color='#3B8C4B')
b3 = ax.bar(x + width, [report[c]['f1-score'] for c in classes], width, label='F1-Score', color='#E24B4A')
for bars in [b1, b2, b3]:
    for bar in bars:
        h = bar.get_height()
        ax.text(bar.get_x() + bar.get_width()/2, h + 0.003, f'({h:.3f})', ha='center', va='bottom', fontsize=9, fontweight='bold')
ax.axhline(report['macro avg']['f1-score'], color='gray', ls='--', alpha=0.7,
           label=f'Macro F1={report['macro avg']['f1-score']:.4f}')
ax.set_xticks(x)
ax.set_xticklabels(classes, fontsize=12)
ax.set_ylim(0.80, 1.03)
ax.set_title('Performa Model per Kelas Sentimen', fontsize=12, fontweight='bold')
ax.set_ylabel('Nilai Metrik', fontsize=12, fontweight='bold')
ax.legend(fontsize=11)
ax.grid(axis='y', alpha=0.3)
plt.tight_layout()
plt.savefig('fig_metrics_per_class.png', dpi=150, bbox_inches='tight')
plt.show()

STOPWORDS_ID = {
    'yang', 'dan', 'di', 'ke', 'dari', 'ini', 'itu', 'dengan', 'untuk',
    'pada', 'ada', 'tidak', 'saya', 'kamu', 'mereka', 'kami', 'kita',
    'sudah', 'juga', 'akan', 'jadi', 'nya', 'aja', 'lah', 'sih', 'deh',
    'dong', 'nih', 'sama', 'kalau', 'tapi', 'karena', 'mau', 'udah',
    'biar', 'emang', 'memang', 'ginana', 'kayak', 'terus', 'nggak',
    'kan', 'lagi', 'pak', 'bang', 'kalo', 'lebih', 'banget', 'bisa'
}

fig, axes = plt.subplots(1, 3, figsize=(16, 5))
for ax, lbl in zip(axes, ['Positif', 'Netral', 'Negatif']):
    txt = ' '.join(df[df['label'] == lbl]['text_clean'].astype(str).tolist())
    cmap = 'Greens' if lbl == 'Positif' else 'Blues' if lbl == 'Netral' else 'Reds'
    wc = WordCloud(width=500, height=380, background_color='white', max_words=70,
                  stopwords=STOPWORDS_ID, colormap=cmap, prefer_horizontal=0.85,
                  min_font_size=8).generate(txt)
    ax.imshow(wc, interpolation='bilinear')
    ax.axis('off')
    ax.set_title(f'Sentimen {lbl}\n({df[df["label"]==lbl].shape[0]:,} komentar)',
               fontsize=13, fontweight='bold', color=COLORS[lbl])
plt.suptitle('Word Cloud per Kelas Sentimen - Komentar TikTok Judi Online', fontsize=13, fontweight='bold', y=1.02)
plt.tight_layout()
plt.savefig('fig_wordcloud.png', dpi=150, bbox_inches='tight')
plt.show()
```

Lampiran E- 7

Dokumentasi *Source Code* (Script Python)

```
df['ym'] = df['dt'].dt.to_period('M')
monthly = df.groupby(['ym', 'label']).size().unstack(fill_value=0)
monthly = monthly[monthly.sum(axis=1) >= 30].tail(28)
monthly.index = [str(x) for x in monthly.index]

fig, ax = plt.subplots(figsize=(14, 5))
for lbl in ['Positif', 'Netral', 'Negatif']:
    if lbl in monthly.columns:
        ax.plot(range(len(monthly)), monthly[lbl], marker='o', ms=4, lw=2, label=lbl, color=COLORS[lbl])
ax.set_xticks(range(len(monthly)))
ax.set_xticklabels(monthly.index, rotation=45, ha='right', fontsize=8)
ax.set_xlabel('Periode', fontsize=12, fontweight='bold')
ax.set_ylabel('Jumlah Komentar', fontsize=12, fontweight='bold')
ax.set_title('Tren Temporal Sentimen Komentar TikTok - Isu Judi Online', fontsize=12, fontweight='bold')
ax.legend(fontsize=11)
ax.grid(alpha=0.3)
plt.tight_layout()
plt.savefig('fig_temporal_trend.png', dpi=150, bbox_inches='tight')
plt.show()
```

Lampiran F- 1 Dokumentasi Lingkungan Komputasi

Dokumentasi Lingkungan Komputasi

```

--- Informasi GPU ---
Thu Jul 9 12:03:45 2026

+-----+
| NVIDIA-SMI 580.82.07                  Driver Version: 580.82.07          CUDA Version: 13.0     |
+-----+-----+-----+-----+-----+-----+
| GPU  Name          Persistence-M   Bus-Id        Disp.A    Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap     Memory-Usage  GPU-Util  Compute M. |
|              |              |              |              |              |      MIG M. |
+-----+-----+-----+-----+-----+-----+
|   0   Tesla T4              Off      00000000:00:04:0  Off              0%            0 |
| N/A   53C    P8              10W / 70W      0MiB / 15360MiB             0%          Default |
|              |              |              |              |              |              |
+-----+-----+-----+-----+-----+-----+

+-----+
| Processes: |
| GPU   GI    CI          PID    Type   Process name                      GPU Memory |
|   ID   ID     ID              |              |           Usage |
+-----+-----+-----+-----+-----+
|  No running processes found |
+-----+

--- Informasi RAM ---
MemTotal:        13286944 kB

--- Versi Python ---
3.12.13 (main, Mar  4 2026, 09:23:07) [GCC 11.4.0]

--- Daftar Library Python yang Digunakan ---
matplotlib==3.10.0
matplotlib-inline==0.2.2
matplotlib-venn==1.1.2
numpy==2.0.2
pandas==2.2.2
pandas-datareader==0.11.1
pandas-gbq==0.30.0
pandas-stubs==2.2.2.240909
scikit-learn==1.6.1
seaborn==0.13.2
torch==2.11.0+cu128
torchao==0.10.0
torchaudio==2.11.0+cu128
torchcodec==0.11.0+cu128
torchdata==0.11.0
torchsummary==1.5.1
torchvision==0.26.0+cu128
transformers==5.12.1
wordcloud==1.9.6

```

Lampiran G- 1
Dokumentasi *Source Code* (Script Python)

Screenshot Output Preprocessing (Jumlah data sebelum & sesudah filtering)

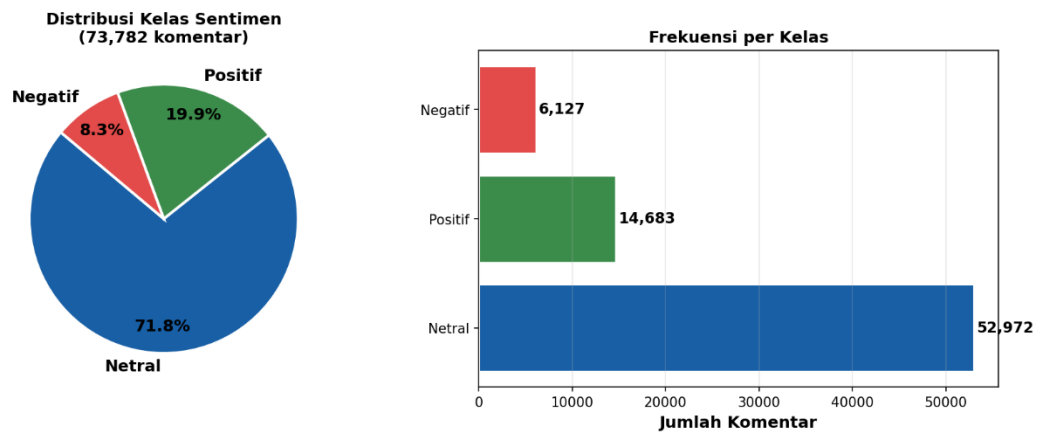
```
Sebelum filter          : 99,065  
Setelah filter panjang : 73,858 (-25,207)  
Setelah filter bot-like: 73,782 (-76)
```

Total data siap untuk pelabelan: 73,782 (74.5% dari data awal)

Lampiran G- 2
Dokumentasi *Source Code* (*Script Python*)

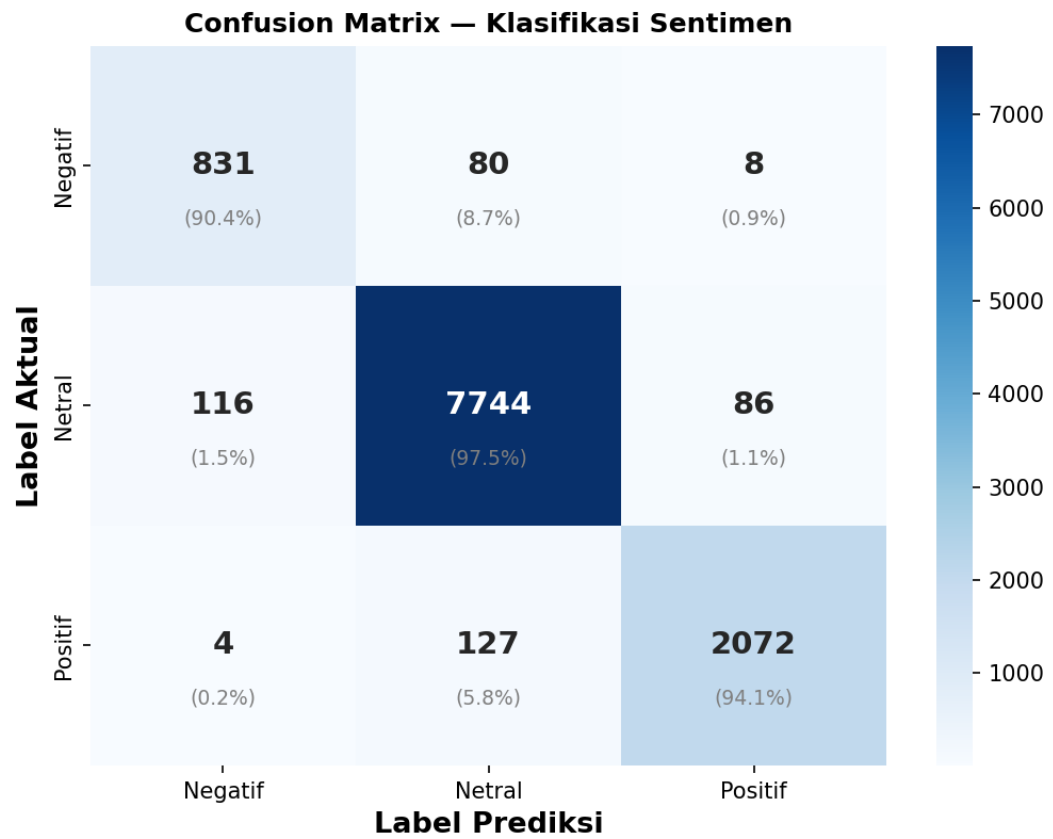
Screenshot Output Distribusi Kelas

Distribusi Sentimen Publik TikTok — Isu Judi Online



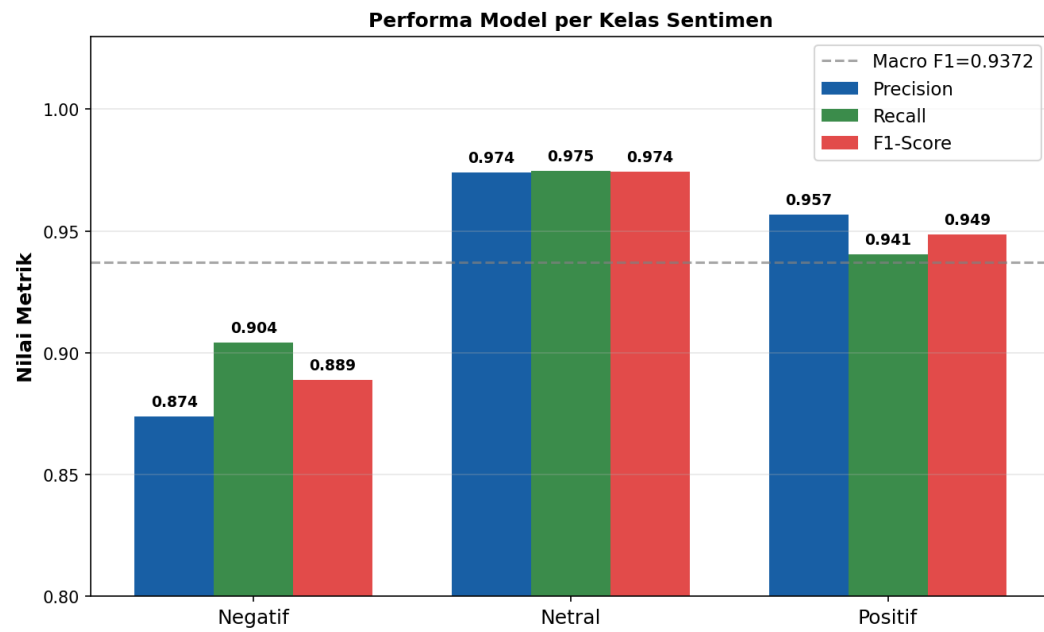
Lampiran G- 3
Dokumentasi *Source Code* (Script Python)

Screenshot Confusion Matrix



Lampiran G- 4
Dokumentasi *Source Code (Script Python)*

Lampiran 6.d. *Screenshot Classification Report*



Lampiran G- 5

Dokumentasi *Source Code* (Script Python)

Lampiran 6.d. *Screenshot Word Cloud* per Kelas Sentimen



Lampiran H- 1
Link Script Google Colab

<https://colab.research.google.com/drive/1qYz7MUe-2YzhRVpT4iqCcaW-Sa-91HTe?usp=sharing>

Lampiran I- 1
Jadwal Kegiatan

JADWAL KEGIATAN

No	Kegiatan	Maret				April				Mei				Juni			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Persiapan dan Finalisasi Proposal																
2	Validasi Dataset Kaggle																
3	<i>Preprocessing</i> Teks																
4	Pelabelan Sentimen dan Validasi IAA																
5	<i>Fine-Tuning</i> IndoBERT																
6	Evaluasi Model (Metrik dan <i>Confusion Matrix</i>)																
7	Analisis Distribusi Sentimen																
8	Penyusunan Laporan Final																
9	Revisi dan Persetujuan																

Lampiran J- 1
Lembar Bimbingan

HISTORI BIMBINGAN SKRIPSI

NIM : 22166013 Tahun Akademik : 2025/2026 Genap
 Nama : ANASTASYA KAROLIEN GUNAWAN Kelas : BA
 Judul Skripsi : Analisis Sentimen Komentar TikTok terhadap Judi Online Menggunakan Fine-Tuning IndoBERT
 Pembimbing 1 : Nugroho Adhi Santoso, S.Kom., M.Kom NIDN : 0629098701
 Pembimbing 2 : Bayu Aji Santoso, M.Kom NIDN : 06230797001
 Tanggal Cetak : 31 July 2025 18:38:27

Total Topik Bimbingan: 5 Total Sesi Bimbingan: 26 Jumlah ACC: 4 Jumlah Revisi: 13

No	Tanggal	Topik	Pengirim	Status	Keterangan
1	06/02/2026 21:13	Pengajuan Judul Skripsi	ANASTASYA KAROLIEN GUNAWAN	PENDING	Prediksi Penjualan Produk Fashion Menggunakan Metode Regresi Linier (Studi Kasus: PGE Official)
2	28/04/2026 17:32	Proposal Skripsi	ANASTASYA KAROLIEN GUNAWAN	-	Topik bimbingan baru
3	10/07/2026 12:26	ACC Judul Skripsi	Nugroho Adhi Santoso, S.Kom., M.Kom	ACC	Pembimbing 1 menyetujui judul: Analisis Sentimen Komentar TikTok terhadap Judi Online Menggunakan Fine-Tuning IndoBERT
4	10/07/2026 12:26	ACC Judul Skripsi	Bayu Aji Santoso, M.Kom	ACC	Pembimbing 2 menyetujui judul: Analisis Sentimen Komentar TikTok terhadap Judi Online Menggunakan Fine-Tuning IndoBERT
5	10/07/2026 13:18	Laporan Skripsi BAB 4 - 5	ANASTASYA KAROLIEN GUNAWAN	-	Topik bimbingan baru
6	12/07/2026 20:29	Laporan Skripsi BAB 4 - 5	Bayu Aji Santoso, M.Kom	REVISI	- Masih banyak yang perlu di perbaiki silahkan temui saya untuk berdiskusi di ruang prodi
7	14/07/2026 11:46	Proposal Skripsi	Nugroho Adhi Santoso, S.Kom., M.Kom	REVISI	1. Penulisan sitasi dalam 1 paragraf diberikan kelebihan dan kekurangan, hindari dalam 1 paragraf ada 1 atau 2 sitasi bahkan 3 sitasi yang tidak berhubungan dengan judul yang diambil. 2. Batasan masalah dalam pembuatan sistem yang dibuat harus disertakan dalam 1 paragraf, tidak terlalu panjang dibuat dengan jelas dan sesuai. 3. Rumusan masalah diperhatikan dalam pembuatan dan bagaimana kedepannya karena nantinya akan menjawab kesimpulan.
8	14/07/2026 11:46	Proposal Skripsi	Nugroho Adhi Santoso, S.Kom., M.Kom	REVISI	1. Landasan teori perlu ditambahkan tentang penjelasan gambar yang berhubungan dengan pengambilan teori berdasarkan judul yang diambil. 2. Perlu adanya penambahan teori yang diambil berdasarkan judul yang diambil pada landasan teori.
9	14/07/2026 11:47	Proposal Skripsi	Nugroho Adhi Santoso, S.Kom., M.Kom	REVISI	1. Pada metode penelitian untuk alur penelitian perlu adanya penjelasan dalam tahapan menyesuaikan penelitian yang diambil. 2. Gambar alur penelitian dibuat lebih

Lampiran J- 2
Lembar Bimbingan

No	Tanggal	Topik	Pengirim	Status	Keterangan
					sederhana sehingga mempermudah dijelaskan.
10	14/07/2026 11:47	Proposal Skripsi	Nugroho Adhi Santoso, S.Kom., M.Kom	REVISI	1. Jadwal penelitian menyesuaikan dengan kegiatan yang dilakukan pada saat penelitian berlangsung. 2. Jika kegiatan penelitian sudah dilaksanakan maka waktu yang sudah dilakukan akan menyesuaikan dengan bulan dan minggu yang sudah dilaksanakan.
11	14/07/2026 11:47	Proposal Skripsi	Nugroho Adhi Santoso, S.Kom., M.Kom	REVISI	1. Untuk keterangan gambar diperhatikan dan silahkan diberikan penjelasan tentang gambar tersebut. 2. Untuk keterangan tabel diperhatikan dan silahkan diberikan penjelasan tentang tabel tersebut. 3. Untuk kata asing dicetak miring. 4. Tanda baca dalam kalimat diperhatikan.
12	14/07/2026 11:48	Proposal Skripsi	Nugroho Adhi Santoso, S.Kom., M.Kom	ACC	ACC Proposal Lanjutkan ke Laporan Skripsi Bab 1 sampai Bab 3
13	14/07/2026 11:59	Laporan Skripsi BAB 1 - 3	ANASTASYA KAROLIEN GUNAWAN	-	Topik bimbingan baru
14	14/07/2026 12:29	Laporan Skripsi BAB 1 - 3	Nugroho Adhi Santoso, S.Kom., M.Kom	REVISI	1. Ditambahkan referensi kelebihan dan kekurangan yang akan disitasi 2. Ditambahkan referensi terdahulu 3. Rumusan Masalah dibuat rinci untuk menjawab kesimpulan
15	14/07/2026 12:30	Laporan Skripsi BAB 1 - 3	Nugroho Adhi Santoso, S.Kom., M.Kom	REVISI	1. Landasan teori ditambahkan definisi teori penjelasan gambar dan fungsi nya 2. Setiap penjelasan definisi teori ditambahkan sitasi 3. Penjelasan tentang metode yang diambil dan cara kerja nya
16	14/07/2026 12:30	Laporan Skripsi BAB 1 - 3	Nugroho Adhi Santoso, S.Kom., M.Kom	REVISI	1. Ditambahkan tetang penjelasan (Jenis, Sifat, dan Pendekatan Penelitian) 2. Pada alur penelitian penjelasan gambar diperjelas dan dari gambar alur penelitian tersebut diberikan penjelasan
17	14/07/2026 12:31	Laporan Skripsi BAB 1 - 3	Nugroho Adhi Santoso, S.Kom., M.Kom	REVISI	1. Setiap Keterangan gambar diberikan kalimat narasi menunjukkan gambar tersebut 2. Setiap Keterangan tabel diberikan kalimat narasi menunjukkan tabel tersebut
18	14/07/2026 12:31	Laporan Skripsi BAB 1 - 3	Nugroho Adhi Santoso, S.Kom., M.Kom	ACC	Laporan Skripsi Bab 1 sampai Bab 3 ACC Silahkan dilanjutkan ke Pembimbing 2 untuk melanjutkan laporan Bab 4 - Bab 5
19	26/07/2026 18:04	Laporan Skripsi BAB 4 - 5	ANASTASYA KAROLIEN GUNAWAN	-	Topik bimbingan baru
20	28/07/2026 05:18	Laporan Skripsi BAB 4 - 5	Bayu Aji Santoso, M.Kom	REVISI	- Pada 4.1 Deskripsi Dataset hilangkan redaksi Proses reduksi data dilakukan secara bertahap melalui tahapan preprocessing. Berdasarkan Tabel 4.1, terlihat bahwa sekitar 25,2% dari data awal (sebanyak 25.207 komentar) tereliminasi pada tahap length filtering. Hal ini mengonfirmasi bahwa seperempat interaksi di kolom komentar TikTok bersifat

Lampiran J- 3
Lembar Bimbingan

No	Tanggal	Topik	Pengirim	Status	Keterangan
					tidak informatif atau merupakan spam yang tidak memiliki nilai sentimen semantik. Setelah proses penyaringan akhir terhadap pola bot-like, diperoleh 73.782 komentar valid yang siap digunakan untuk pelabelan dan pemodelan. Jumlah ini merepresentasikan sampel opini publik yang autentik dan relevan untuk dianalisis lebih lanjut., pada Tabel 4. 1 Ringkasan Reduksi Data Preprocessing tampilkan 3 data teratas 3 data dibawah - hilangkan juga Hasil pelabelan menggunakan pendekatan Domain-Specific Lexicon-Based pada 73.782 data valid menghasilkan distribusi kelas yang sangat timpang (highly imbalanced), sebagaimana disajikan pada Tabel 4.2. Kelas Netral mendominasi sebesar 71,8%, diikuti oleh Positif (19,9%) dan Negatif (8,3%). Dominasi kelas Netral ini mencerminkan karakteristik percakapan di media sosial di mana mayoritas respons bersifat observasional, sinis, atau ambigu. Sebaliknya, minoritas ekstrem pada kelas Negatif (hanya 6.127 komentar) menjadi tantangan metodologis utama, karena model klasifikasi konvensional cenderung bias memprediksi seluruh data sebagai kelas mayoritas jika tidak ditangani dengan strategi penanganan ketidakseimbangan kelas yang tepat. Tabel 4. 2 Distribusi Kelas Sentimen Hasil Pelabelan Leksikon - pada 4.2.1 Preprocessing di buat langkah langkah 1. setiap langkah di Preprocessing Cleaning data , Normalisasi , filtering , tokenisasi , berikan penjelasan deskriptif di atasnya baru di bawahnya ada tabel menampilkan hasil dari Preprocessing Cleaning data , Normalisasi , filtering , tokenisasi 5 data teratas 5 data dibawah - pada 4.2.2 Pelabelan Data berikan sitasi rujukan jurnal untuk menghasilkan Tabel 4. 4 Hasil Proses Scoring Secara Granular - pada confusion matrix di jelaskan hasilnya - buat visualisasi hasil akurasi precision recall dan f1 score
21	29/07/2026 09:13	Laporan Skripsi BAB 4 - 5	ANASTASYA KAROLIEN OUNAWAN	-	Topik bimbingan baru
22	29/07/2026 09:17	Laporan Skripsi BAB 4 - 5	ANASTASYA KAROLIEN OUNAWAN	-	Revisi laporan
23	29/07/2026 09:19	Laporan Skripsi BAB 4 - 5	Bayu Aji Santoso, M.Kom	REVISI	kirim pdf

Lampiran J- 4
Lembar Bimbingan

No	Tanggal	Topik	Pengirim	Status	Keterangan
24	28/07/2026 09:53	Laporan Skripsi BAB 4 - 5	ANASTASYA KAROLIEN GUNAWAN	-	Revisi laporan
25	28/07/2026 10:59	Laporan Skripsi BAB 4 - 5	Bayu Aji Santoso, M.Kom	REVISI	Kenapa malah kirim file lama ?
26	28/07/2026 16:16	Laporan Skripsi BAB 4 - 5	ANASTASYA KAROLIEN GUNAWAN	-	Revisi laporan
27	31/07/2026 16:52	Laporan Skripsi BAB 4 - 5	ANASTASYA KAROLIEN GUNAWAN	-	Revisi laporan
28	31/07/2026 18:13	Laporan Skripsi BAB 4 - 5	Bayu Aji Santoso, M.Kom	ACC	-
29	31/07/2026 18:13	Laporan Skripsi BAB 4 - 5	Bayu Aji Santoso, M.Kom	ACC	-

Tanggal Cetak: 31 July 2026 18:38:27

Dosen Pembimbing 1



Nugroho Adhi Santoso, S.Kom., M.Kom

Dosen Pembimbing 2



Bayu Aji Santoso, M.Kom

Lampiran K- 1
Letter of Acceptance (LoA)



**LEMBAGA PENELITIAN DAN
PENGEMBANGAN PUSTAKA CENDEKIA**
Jurnal Cakrawala Akademika

HP : 0895404313613 / 081274887803 ; E-Mail : admin@jurnalpustakacendekia.com ; E-ISSN : 3062-9942



SURAT KETERANGAN PUBLIKASI ARTIKEL JURNAL

Letter Of Accepted (LoA)

Nomor : 2915/JCA/Vol3/No1/2026

Kepada YTH.

Anastasya Karolien Gunawan, Nugroho Adhi Santoso, Bayu Aji Santoso

Terimakasih telah mengirimkan artikel terbaik anda untuk diterbitkan pada **Jurnal Cakrawala Akademika** dengan judul:

**"ANALISIS SENTIMEN KOMENTAR TIKTOK TERHADAP JUDI
ONLINE DI INDONESIA MENGGUNAKAN FINE-TUNING INDOBERT"**

Berdasarkan hasil review dan keputusan tim editor, maka artikel tersebut dinyatakan **DITERIMA** untuk dipublikasikan pada **Jurnal Cakrawala Akademika** edisi **Volume 3 Nomor 1**.

Demikian surat keterangan ini kami sampaikan untuk dipergunakan sebagaimana mestinya, kami ucapkan terimakasih.

No Reg : LOA20260901220400



LOA20260901220400

Keaslian LOA Dapat
Diperiksa Dengan
Memindai QR Code
Disamping !

Palembang, 01 September 2026
Editor In Chief

Winda Eprilia, S.Pd.,Gr.,M.Pd.

Penerbit :

Lembaga Penelitian dan Pengembangan Pustaka Cendekia

Jl. Sei Sahang, NO. 5296, Kec. IB 1, Kel. Lorok Pakjo, Kota Palembang
0895404313613 / 081274887803
admin@jurnalpustakacendekia.com

Lampiran L- 1
Formulir Perubahan Judul

FORMULIR PERUBAHAN JUDUL SKRIPSI

Nama : Anastasya Karolien Gunawan
Nim : 22166013
Program Studi : Sistem Informasi
Tahun Akademik : 2025/2026
Pembimbing I : Nugroho Adhi Santoso,M.Kom

Judul Skripsi Sebelumnya :
PREDIKSI PENJUALAN PRODUK FASHION MENGGUNAKAN METODE
REGRESI LINIER(STUDI KASUS:PGE OFFICIAL)
Judul Skripsi Baru :
ANALISIS SENTIMEN KOMENTAR TIKTOK TERHADAP JUDI ONLINE
MENGGUNAKAN FINE-TUNING INDOBERT

Telah disetujui oleh Dosen Pembimbing SKRIPSI
pada tanggal 16 April 2026

Dosen Pembimbing I



Nugroho Adhi Santoso,M.Kom
NIPY. 08.025.560

Lampiran M- 1 Bukti Cek Plagiarisme



Analysis attestation
Compilatio Magister+ | Universitas Harkat Negeri

SKRIPSI TASYA ACC FINAL ALHAMDULILAH

ID : 0b7eb5d38269f8b4d2418bd750d601f2e56f81d0



37%
Suspicious texts

File name : SKRIPSI TASYA ACC FINAL ALHAMDULILAH.txt
Original file size : 6.12 MB
Number of words : 17,220

Submitter : S1 Sistem Informasi Harkat Negeri
Submission date : August 2, 2026
Upload type : interface
analysis end date : August 2, 2026

Summary (section 1/2)

Location of suspect texts in the document :



Included in the suspicious text score :



Similarities

12%

Syntactics 4% Semantics 7%

Passages with similarities to sources found in different collections.



AI detection

21%

Texts with stylistically similar formulations to AI-generated text.

This rate is an indicator, not proof. Check with the author that he/she has mastered the knowledge mentioned in the document.



Unrecognized languages

9%

Passages in which some of the vocabulary used is not part of the language dictionary. This may be an attempt by the author to modify the text to make detection impossible.



Not included in the percentage of suspicious texts :



Texts between quotes

3%

Passages between quotation marks, often revealing a quotation.

Lampiran M- 2

Bukti Cek Plagiarisme

Sources of similarities (section 2/2)



Similarities

12%

Passages with similarities to sources found in different collections.



Main source detected

No.	Description	Similarities	Locations
1	repository.nusaputra.ac.id Semantic detection repository.nusaputra.ac.id/eprint/1730/1/SINTA%20APRILIANI%20%20	3%	
2	myskripsi.ums.ac.id Semantic detection myskripsi.ums.ac.id/media/skripsi/laporan/2024/12/16/Skripsi-Nasywa...	3%	
3	SKRIPSI-22205013-Terbaru #f21d5b Semantic detection Comes from my group	3%	
4	Fine-Tuning IndoBERT Untuk Analisis Sentimen Beri... Semantic detection doi.org/10.31544/jtera.v11.i1.2026.51-60	2%	
5	archive.umsida.ac.id Semantic detection archive.umsida.ac.id/index.php/archive/preprint/download/10010/721...	2%	
6	6 #f4bedd Semantic detection Comes from my group	2%	
7	Evaluasi IndoBERTweet untuk Klasifikasi Sentimen... Semantic detection doi.org/10.55340/jiu.v15i1.2683	2%	
8	Analisis kepuasan pengguna aplikasi gramedia digit... Semantic detection etheses.uin-malang.ac.id/84650/1/220607110011.pdf	2%	
9	tesis.ums.ac.id Semantic detection tesis.ums.ac.id/public/bimb/1768966396_9576f708190b6d07b860.docx	2%	
10	Analisis Sentimen Publik Tagar #IndonesiaGelap ata... Semantic detection doi.org/10.53842/juki.v8i1.2300	<1%	

Referenced source (without similarities detected)

No.	Description
1	https://www.kaggle.com/datasets/aginanjar/dataset-judol-tiktok-komentar
2	https://datareportal.com/reports/digital-2025
3	https://nasional.kompas.com/read/2024/11/14/15382151/budi-gunawan
4	https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c
5	https://doi.org/10.54035/jem.v10i1



Curriculum Vitae

1. Data Pribadi

Nama Lengkap : Anastasya Karolien Gunawan
NIM : 22166013
Tempat, Tanggal Lahir : Cirebon, 23-12-2004
Jenis Kelamin : Perempuan
Alamat : Jl. Panggung Baru Rt. 03 Rw. 07 Kec.
Tegal Timur Kota. Tegal
No. HP / WhatsApp : 089525187477
Email : antasyacarolien@gmail.com
LinkedIn :
Github / Portofolio :

2. Riwayat Pendidikan

Tahun	Nama Intansi	Program Studi/Jurusan
2010-2016	SDN MKK 2 Tegal	-
2016-2019	SMPN 10 Tegal	-
2019-2022	SMAN 4 Tegal	-
2022-2026	Universitas Harkat Negeri Tegal	Sistem Informasi

3. Keahlian

Soft Skill	Hard Skill
Komunikasi Efektif & Pelayanan Pelanggan	Manajemen E-Commerce & Toko Online
Manajemen Waktu	Live Streaming & Penjualan
Teliti & Terorganisir	Administrasi Penjualan & Dokumen
Adaptif	Entri, Analisis, & Pengolahan Data
Conten Ideation	Pengelolaan Basis Data (Database) & Analisis entimen

FORMULIR PUBLIKASI JURNAL SKRIPSI

Nama : Anastasya Karolien Gunawan
NIM : 22166013
Program Studi : S1 Sistem Informasi
Tahun Akademik : 2025/2026
Pembimbing 1 : Nugroho Adhi Santoso, M.Kom
Pembimbing 2 : Bayu Aji Santoso, M.Kom

Judul Skripsi : ANALISIS SENTIMEN KOMENTAR TIKTOK
TERHADAP JUDI ONLINE DI INDONESIA
MENGUNAKAN FINE-TUNING INDOBERT

Judul Publikasi : ANALISIS SENTIMEN KOMENTAR TIKTOK
TERHADAP JUDI ONLINE DI INDONESIA
MENGUNAKAN FINE-TUNING INDOBERT

Nama Penerbit : Jurnal Cakrawala Akademika (JCA) PT. Pustaka
Cendekia Publisher

Link Abstrak : -
Halaman : -
Volume : 3
Sinta : Non Sinta

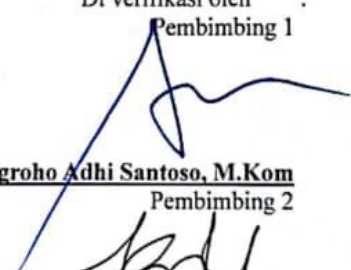
Saya yang bertanda tangan di bawah ini menyatakan bahwa artikel jurnal yang berasal dari skripsi saya telah dipublikasikan / disubmit pada jurnal sebagaimana tercantum di atas dan merupakan karya ilmiah saya sendiri.

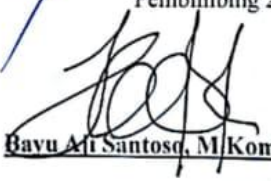
Apabila di kemudian hari terdapat pelanggaran akademik, saya bersedia menerima sanksi sesuai dengan ketentuan yang berlaku.

Tegal, 14 September 2026


Anastasya Karolien Gunawan

Di verifikasi oleh :
Pembimbing 1


Nugroho Adhi Santoso, M.Kom
Pembimbing 2


Bayu Aji Santoso, M.Kom