



### Lampiran 01. Kode *pre-processing* dataset

```
def remove_gelar(text):
    """
    Remove academic titles before and after names.
    """
    text = re.sub(r"\b((dr|prof|ir|hj|h)\.?\s*)+\b", "", text,
flags=re.IGNORECASE)
    text = re.sub(r"(\s+|\b)(sh|mh|msc|ph\.?d|mba|s\.?kom|m\.?kom|m\.?si|dr|ma|mm)\b\.\.??", r"\1", text, flags=re.IGNORECASE)
    return re.sub(r"\s+", '', text).strip()

# Comprehensive cleaning

def clean_text(text):
    text = str(text)
    # A) Remove leading metadata/tags
    metadata_patterns = [
        r'^tempo\.co,?\s*jakarta\s*[\---]\s*',
        r'^tempo\.co,?\s*',
        r'^info nasional[\---]?\s*',
        r'^[A-Z ]{3,}\s*'
    ]
    for pat in metadata_patterns:
        text = re.sub(pat, "", text, flags=re.IGNORECASE)
    # B) Remove any remaining occurrences
    text = re.sub(r'tempo\.co,?', "", text, flags=re.IGNORECASE)
    text = re.sub(r'info nasional', "", text, flags=re.IGNORECASE)
    # C) Lowercase
    text = text.lower()
    # D) Remove HTML tags and URLs
    text = re.sub(r'<[^>]+>', '', text)
    text = re.sub(r'http\S+|www\.\S+', '', text)
    # E) Remove academic titles
    text = remove_gelar(text)
    # F) Keep ASCII letters, numbers, basic punctuation and spaces
    text = "".join(c for c in text if (ord(c) < 128 and (c.isalnum() or c.isspace() or c in ".,!?:;-")))
    # G) Normalize whitespace
    text = re.sub(r"\s+", '', text).strip()
    return text
```

## Lampiran 02. *Setting random seed*

```
# -- Seed (comprehensive) --  
SEED = 42  
os.environ["PYTHONHASHSEED"] = str(SEED)  
random.seed(SEED)  
np.random.seed(SEED)  
torch.manual_seed(SEED)  
torch.cuda.manual_seed_all(SEED)  
torch.backends.cudnn.deterministic = True  
torch.backends.cudnn.benchmark = False  
set_seed(SEED)
```



### Lampiran 03. Source code post-processing

```
def postprocess_summary(text, max_sent=5, min_words=20):
    if not text:
        return ""

    # 1. Bersihkan token khusus & noise
    out = re.sub(r"<unk>|<pad>", "", text)

    # 2. Split kalimat
    sentences = re.split(r'(?<=[!?!])\s+', out)

    # 3. Deduplication kalimat
    seen, res = set(), []
    for s in sentences:
        s_clean = s.strip()
        if s_clean and s_clean not in seen:
            seen.add(s_clean)
            res.append(s_clean)

    # 4. Buang kalimat terakhir jika tidak lengkap (tidak diakhiri tanda baca)
    if res and not res[-1].endswith(('!', '!', '?')):
        res = res[:-1]

    # 5. Gabungkan kembali
    out = " ".join(res)

    # 6. Normalisasi spasi
    out = re.sub(r'\s+', ' ', out).strip()

    # 7. Truncate ke max_sent kalimat
    sent = re.split(r'(?<=[!?!])\s+', out)
    out = " ".join(sent[:max_sent])

    # 8. Kontrol panjang minimum (hindari ringkasan terlalu pendek)
    words = out.split()
    if len(words) < min_words and len(sent) > 1:
        # tambahkan kalimat berikutnya jika ada
        out = " ".join(sent[:min(max_sent, len(sent))])

    # 9. Hilangkan repetisi n-gram sederhana (3-gram)
    tokens = out.split()
    filtered = []
    for i in range(len(tokens)):
        if i >= 2 and tokens[i-2:i+1] == tokens[i-5:i-2]: continue
        filtered.append(tokens[i])
    out = " ".join(filtered)

    # 10. Normalisasi kapitalisasi & tanda baca akhir
    out = out[0].upper() + out[1:]
    if not out.endswith('.'):
        out += '.'

    return out
```

#### **Lampiran 04. Case-by-case demonstration**

##### **Top IndoBERT**

*Platform belanja online kesulitan melakukan pembersihan terhadap penjualan barang bekas impor seperti baju dan sepatu. ketua umum indonesian digital empowering community, m tesar sandikapura, mengatakan sebaiknya pemerintah fokus pada pembinaan masyarakat dan mencari tahu mengapa orang indonesia masih suka barang impor. penjualan pakaian bekas impor terutama pakaian dan sepatu sudah semakin berkembang, sehingga perlu adanya penanganan melalui kerjasama berbagai pihak. pemerintah melalui kementerian koperasi dan ukm telah bersepakat dengan pelaku e - commerce dan platform sosial media untuk melakukan takedown kepada penjual baju bekas impor.*

*Ahli psikologi forensik reza indragiri amriel membandingkan teori superior order defense dalam kasus peredaran narkoba mantan kapolda sumatera barat teddy minahasa dengan kasus richard eliezer alias bharada e yang menjadi eksekutor pembunuhan berencana terhadap brigadir j. reza menyebut tidak semua terdakwa yang mengklaim melakukan tindak pidana karena diperintah atasannya harus diterima, tetapi tetap harus diuji. namun, dia tidak bisa memastikan apakah dody prawiranegara yang mengaku mendapat perintah dari teddy minahasa untuk menukar 5 kilogram sabu dengan tawas sebagai tindakan tidak mampu melawan perintah atasan dalam perkara ini. demikian, tidak bisa dipastikan apakah teddy minahasa tidak mampu menghadapi konsekuensi yang sangat buruk jika tolak yang buruk.*

*Guru besar bidang keamanan pangan gizi di ipb, ahmad sulaeman, mengatakan bahwa pelabelan kemasan air minum berpotensi mengandung bpa belum perlu dilakukan. batas maksimum migrasi bpa di indonesia adalah 0,6 bpj dan ini masih sesuai dengan mayoritas batas parameter bpa negara - negara maju di dunia lainnya. direktur advokasi dan jaringan pshk, fajri nursyamsi, menyarankan agar permasalahan itu diselesaikan dalam pembahasan antar kementerian dan lembaga terkait sehubungan dengan adanya rencana perubahan peraturan bpom nomor 31 tahun 2018 tentang label pangan olahan harus diselesaikan.*

*Kejaksaan negeri cianjur menyatakan berkas perkara pidana kasus tabrak lari yang menewaskan selvi amalia nuraeni, mahasiswa fakultas hukum universitas suryakencana, dinyatakan lengkap atau p21. tersangka sugeng guruh gautama legiman dinyatakan melanggar pasal 310 ayat 4 dan 312 undang - undang nomor 22 tahun 2009 tentang lalu lintas dan angkutan jalan umum. kejar cianjur akan berkoordinasi dengan penyidik satuan lalu lintas kepolisian resor cianjur untuk melakukan tahap ii atau proses penyerahan tersangka dan barang bukti dari pihak penyidik kepolisian kepada pihak kejaksaan untuk selanjutnya disidangkan di pengadilan.*

### ***Divergent IndoBERT***

*Larangan buka puasa bersama bagi pejabat dan asn berdampak positif untuk menekan persepsi negatif terkait kinerja asn. mui bali tetap melaksanakan buka bersama selama ramadhan dan gelar griya atau open house hari raya idul fitri 1443 hijriah. beberapa minggu belakangan ini, kinerja asn dan pejabat mendapatkan sorotan luas dari masyarakat karena beberapa pejabat viral karena pamer harta kekayaan dan gaya hidup mewah di media sosial. presiden jokowi melarang seluruh pejabat dan pegawai pemerintah untuk melakukan protokol kesehatan. pemerintah memberikan kelonggaran kepada umat islam melaksanakan shalat tarawih di masjid dengan syarat telah melakukan vaksinasi covid - 19 dua dosis prasarana.*

*Gempa magnitudo 6,2 magnitudo mengguncang wilayah timur pada Selasa 143 pagi. pusat gempa berada di 448 km sebelah utara - barat laut port moresby dengan kedalaman sekitar 200 km dari permukaan air laut. belum ada informasi lebih lanjut terkait dampak gempa dan intensitas getaran yang dirasakan masyarakat setempat. informasi mengenai dampak gempa, dan tingkat intensitas getaran masyarakat setempat masih belum diketahui. informasi tentang dampak gempa.*

*Menteri koordinasi bidang politik, hukum dan HAM Mahfud MD menegaskan bahwa transaksi mencurigakan senilai Rp 300 triliun di Kementerian Keuangan telah mencapai Rp 349 triliun dan merupakan tindak pidana pencucian uang TPPU yang sangat besar. Ia berharap Kementerian Keuangan akan menyelesaikan laporan hasil analisis yang diduga TPPU dari PPAK. TPPU akan disidik oleh Kemenkeu sebagai penyidik, PNS penyidik di bidang pajak dan kepabeanan. akan ditindaklanjuti oleh TPPU atau mungkin diserahkan ke polisi, jaksa, atau KPK. akan diserahkan ke kepolisian, jaksa atau KPK dalam kasus ini.*



Kpk membuka pelaporan lhkpn hingga 31 maret 2023, menteri pariwisata dan ekonomi kreatif sandiaga uno menjadi pejabat terkaya dengan harta kekayaan tertinggi sebesar rp10, 9 triliun. namun, nilai kekayaannya yang dimiliki sandiaga belum mencapai rp300 miliar. sandiaga hanya terdata senilai rp10 miliar, sementara pada 2021, naik 178, 24 persen dalam setahun. demikian, harta kekayaannya dipicu oleh nilai surat berharga yang ia miliki yang naik 214, 1 persen menjadi rp9, 78 triliun pada 2021. naik nilai kekayaannya senilai rp3, 81 triliun.

#### **Bottom IndoBERT**

Badan pengatur jalan tol bpjt melakukan uji coba operasi di jalan tol bekasi - cawang - kampung melayu becakayu pada seksi 2a ujung jakasampurna - marga jaya. selama periode uji coba, pengguna jalan akan dikenakan tarif lama sebesar rp14. 000 untuk golongan i jenis kendaraan mobil, sedan, bus, pick up, dan truk kecil. jalan tol becakayu adalah jalan tol layang sepanjang 16, 78 km dan jongoendksi dengan jalan tol wiyoto wiyono via gt pisanan.

Seorang dukun spesialis penggandaan pembunuhan berantai oleh slamet tohari 45 di banjarnegara, jawa tengah yang menjadi korban pembunuhan tersebut ternyata menjanjikan kepada korban untuk menggandakan uang. polisi masih melakukan pendalaman terkait jumlah korban dari mbah slamet yang meninggal di desa balun, kecamatan wanayasa, banjarnegara. selain itu, polisi juga menemukan 10 mayat korban lain yang ditemukan di kebun milik tersangka. mayat korban lainnya masih dalam pendalaman oleh polisi. ditemukan 10 orang lain yang diduga telah diidentifikasi di jalan setapak di hutan.

Kerajaan majapahit memiliki lima jenis pajak yang biasanya ditarik oleh kerajaan majapahit, yaitu pajak tanah, pajak usaha, dan pajak usaha. namun, jika tanah ditelantarkan, pemiliknya akan dikenai denda oleh raja, sedangkan pajak usaha masih terbagi menjadi tiga jenis lagi. pemerintahan majapahit menarik pajak dari mereka karena adanya keahlian jenis pekerjaan sebagai mata pencaharian hidup. hanya ada pedagang, pengrajin, penjual jasa, pedagang, dan jasa. demikian, pajak dikenakan pada kelompok pengrajin atau juru dagang terhadap obyek pajak yang berupa komoditas perdagangan.

Harga emas antam naik menjadi rp1. 082. 000 per gram pada rabu 2932023. harga buyback juga naik rp5. 000 menjadi rp967.

### Top PEGASUS

*Calon penumpang melalui bank partner dengan tambahan potongan harga hingga rp 1 juta. diskon ini berlaku untuk periode pemesanan mulai senin, 10 april 2023 hingga ahad, 30 april 2023 dengan menggunakan kode promo thr23.*

*Atlet nomor lompat galah mengikuti pemusatan latihan nasional pelatnas sea games kamboja 2023 di lapangan madya, gelora bung karno. tim nasional atletik indonesia diperkirakan akan menurunkan 25 atlet untuk semua nomor di kamboja. para atlet atletik nomor berlatih dalam pemusatan latihan sea games 2023 dan sejumlah nama idan fauzan richsan, maria andriani melabesy, dan diva renata jayadi.*

*Pemerintah memastikan aturan devisa hasil ekspor dhe akan segera terbit pada maret 2023 dan dipastikan para eksportir tidak wajib mengkonversikan dolar mereka ke rupiah. menteri koordinator bidang perekonomian airlangga hartarto menjelaskan, lewat revisi aturan nomor 1 tahun 2019 tentang dhe sumber daya alam yang selama ini rajin diparkir di luar negeri akan kembali ke tanah air.*

*Pemilik nama asli park jinyoung akan mendaftar sebagai tentara aktif pada mei 2023. pihak bh entertainment akan menjalankan kewajibannya pada 8 mei mendatang. got7 resmi hengkang dari jyp entertainment pada 19 januari 2021, masing-masing anggota got7 memulai karier solo dengan agensi-agensi baru.*

### Divergent PEGASUS

*Komisi pemberantasan korupsi kpk menetapkan kembali hakim agung nonaktif gazalba saleh gs sebagai tersangka dalam kasus dugaan penerimaan gratifikasi dan tindak pidana pencucian uang tppu di mahkamah agung. kpk menerangkan bahwa gs sebelumnya telah menjadi perampasan aset hasil korupsi untuk negara.*

*Relawan jaringan nasional jarnas anies baswedan kota yogyakarta membagikan 500 paket takjil gratis yang juga berisi 500 buku saku anies dalam jalan kusumanegara pada kamis 64 sore. relawan juga menampilkan pertunjukan menyanyi di pinggir jalan untuk menghibur para pengguna jalan yang lewat.*



*Hakim ungkap awal mula penyebab terjadinya penganiayaan david ozora oleh mario dandy dalam persidangan perempuan ag 15 pada senin 104. dalam pertimbangannya, hakim menilai insiden itu dipicu rasa kesal mario terhadap david. hakim membacakan fakta-fakta dipertimbangan yang dilangsungkan di pengadilan negeri jakarta selatan dan simak video di as. newsupdate update update newsflash hakim.*

*Politikus partai golkar yang menjabat tenaga ahli kantor staf presiden ksp ali mochtar ngabalin tidak melihat kebersamaan presiden jokowi bersama prabowo subianto dan ganjar pranowo sebagai bentuk dukungan atau endorsement untuk pilpres 2024. presiden setiap orang menjadi menteri pasti orang yang dipercaya, apalagi menko keuangan, yang diberi kepercayaan kepada airlangga...*

### **Bottom PEGASUS**

*Jakarta cara daftar mudik gratis kemenhub melalui direktorat jenderal perhubungan darat, badan usaha milik negara bumh serta jasa raharja menyelenggarakan program mudik yang sama dengan harapan akan mengurangi pergerakan mudik dengan sepeda motor,ucap hendro dikutip dari antara.*

*Harga minyak memicu resesi dan melemahkan permintaan bahan bakar. minyak mentah berjangka west texas intermediate wti untuk pengiriman april terangkat 90 sen atau 1,35 menjadi usd67,64 per barel menjelang berakhirnya kontrak di new york mercantile exchange.*

*Pemerintah berencana memberikan bantuan sosial atau bansos pangan berupa beras, ayam, dan telur selama tiga bulan menjelang lebaran. bansos diperkirakan lebih dari rp 7,8 triliun.hal ini diungkap direktur jenderal anggaran kementerian keuangan atau kemenkeu isa rachmatarwata dalam konferensi pers apbn kita di jakarta pada Selasa, 14 maret 2023.*

*Daerah hijau perhutanan mendominasi kawasan inti pusat pemerintahan kipp di ibu kota baru yang semula merupakan bagian dari kecamatan sepaku, kabupaten penajampaserutara, kalimantan timur itu. menteri pekerjaan umum dan perumahan rakyat pupr basuki hadimuljono mengatakan kipp ikn nusantara akan berdiri di lahan 6.600 hektare.*

## Lampiran 05. Riwayat Hidup

### RIWAYAT HIDUP



**Ketut Tutur Diatmika**, anak keempat dari empat bersaudara ini adalah putra kandung dari pasangan Ketut Widia, S.Pd. dan Ni Made Menaka. Lahir di Singaraja pada tanggal 16 April 2002 dan saat ini penulis dan keluarga menetap di Kampung Baru, Singaraja. Penulis lahir dan dibesarkan ditengah keluarga yang memiliki ekonomi yang berkecukupan dan menomorsatukan pendidikan serta Agama. Penulis menyelesaikan pendidikan TK di TK Surapati Singaraja pada tahun 2006 hingga 2008. Kemudian menyelesaikan Pendidikan Sekolah Dasar di SD No 2 Kampung Baru pada tahun 2009 – 2014. Selanjutnya, penulis menyelesaikan pendidikan di SMP Negeri 1 Singaraja pada tahun 2014 – 2017 dan SMA Negeri 4 Singaraja pada tahun 2017 – 2020.