SKRIPSI

ANALISIS SENTIMEN KANDIDAT CALON PRESIDEN 2024 MENGGUNAKAN *IndoBERT* DAN *SUPPORT VECTOR MACHINE*

Disusun dan diajukan oleh:

MUHAMMAD AKIB
D121 17 1304

PROGRAM STUDI SARJANA TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2024

LEMBAR PENGESAHAN SKRIPSI

ANALISIS SENTIMEN KANDIDAT CALON PRESIDEN 2024 MENGGUNAKAN *IndoBERT* DAN *SUPPORT VECTOR MACHINE*

Disusun dan diajukan oleh

Muhammad Akib
D121171304

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian
Studi Program Sarjana Program Studi Teknik Informatika
Fakultas Teknik Universitas Hasanuddin
Pada Tanggal 8 Mei 2024
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

A. Ais Prayogi Alimuddin, S.T., M.Eng.
NIP 198305102014041001

Pembimbing Pendamping,

Anugrayani Bustamin, S.T., M.T.
NIP 199012012018074001

Ketua Program Studi,

Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN. Eng.
NIP 197507162002121004

PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini ;

Nama	:	Muhammad Akib
NIM	:	D121171304
Program Studi	:	Teknik Informatika
Jenjang	:	S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

ANALISIS SENTIMEN KANDIDAT CALON PRESIDEN 2024
MENGGUNAKAN *IndoBERT* DAN *SUPPORT VECTOR MACHINE*

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Gowa, 8 Mei 2024

Yang Menyatakan

Muhammad Akib

ABSTRAK

MUHAMMAD AKIB. *Analisis Sentimen Kandidat Calon Presiden 2024 Menggunakan IndoBERT dan Support Vector Machine* (dibimbing oleh Ais Prayogi Alimuddin dan Anugrayani Bustamin)

Isu terkini yang tengah mengemuka dalam arena politik Indonesia adalah persiapan pemilihan calon presiden tahun 2024. Proses pemilihan presiden, yang menjadi salah satu pilar utama dalam demokrasi Indonesia, membawa dampak yang signifikan terhadap dinamika politik dan arus kebijakan di tanah air. Tantangan yang muncul dalam menanggapi isu ini adalah kompleksitas opini publik yang terwakili di platform Twitter.

Dengan beragamnya pandangan dan penilaian masyarakat terhadap calon-calon presiden, menentukan calon yang memiliki potensi untuk memimpin negara menjadi semakin kompleks dan menantang. Untuk menghadapi dinamika ini, penelitian ini bertujuan untuk mengembangkan sebuah sistem klasifikasi menggunakan metode *IndoBERT* (*Bidirectional Encoder Representations from Transformers*) dan *Support Vector Machine (SVM)* teknik One VS All (OvA). Sistem ini dirancang untuk melakukan analisis sentimen terhadap isu-isu yang berkaitan dengan calon-calon presiden 2024.

Penelitian ini melibatkan uji coba terhadap model yang dikembangkan, dan hasilnya menunjukkan bahwa model *IndoBERT* dengan jumlah epoch 10unggul dengan tingkat akurasi sebesar 84%, sedangkan *SVM* dengan parameter nilai c =10, gamma = scale dan menggunakan kernel linear mencapai akurasi sebesar 76%. Hasil ini menandakan kemampuan model dalam mengenali dan menganalisis sentimen masyarakat terkait isu-isu politik, yang dapat memberikan wawasan berharga untuk mendukung pengambilan keputusan di tingkat politik dan masyarakat.

Model yang telah dikembangkan mampu memberikan kontribusi nyata dalam pemahaman terhadap sentiment masyarakat terkait isu pemilihan calon presiden 2024. Keberhasilan model dalam memprediksi sentimen masyarakat menjadi landasan penting dalam mendukung keterlibatan aktif masyarakat dalam proses demokratis, dan sekaligus memberikan kontribusi dalam meningkatkan kualitas perdebatan dan diskusi publik.

Kata Kunci: Presiden, *IndoBERT*, *Support Vector Machine* (*SVM*)

.

ABSTRACT

MUHAMMAD AKIB. *Sentiment Analysis of 2024 Presidential Candidate Using IndoBERT and Support Vector Machine* (Supervised by Ais Prayogi Alimuddin and Anugrayani Bustamin)

The latest issue that is currently emerging in the Indonesian political arena is the preparation for the 2024 presidential candidate election. The presidential election process, which is one of the main pillars of Indonesian democracy, has a significant impact on political dynamics and policy flows in the country. The challenge that arises in responding to this issue is the complexity of public opinion represented on the Twitter platform.

With the diversity of public views and assessments of presidential candidates, determining which candidate has the potential to lead the country is becoming increasingly complex and challenging. To deal with this dynamic, this research aims to develop a classification system using the IndoBERT (Bidirectional Encoder Representations from Transformers) and Support Vector Machine (SVM) methods. This system is designed to conduct sentiment analysis on issues related to the 2024 presidential candidates.

This research involved testing the developed model, and the results showed that the IndoBERT model was superior with an accuracy rate of 84%, while SVM achieved an accuracy of 76%. These results indicate the model's ability to recognize and analyze public sentiment regarding political issues, which can provide valuable insights to support decision making at the political and societal levels.

The model that has been developed is able to make a real contribution to understanding public sentiment regarding the issue of the 2024 presidential candidate election. The success of the model in predicting public sentiment is an important basis for supporting active community involvement in the democratic process, and at the same time contributes to improving the quality of public debate and discussion.

Keywords: President, *IndoBERT*, Support Vector Machine (*SVM*).

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR SINGKATAN DAN ARTI SIMBOL

Lambang/Singkatan	Arti dan Keterangan
SVM	*Support Vector Machine*
BERT	*Bidirectional Encoder Representations from Transformers*
NLP	*Natural Language Processing*
CBOW	*Continuous Bag-of-Words*

DAFTAR LAMPIRAN

KATA PENGANTAR

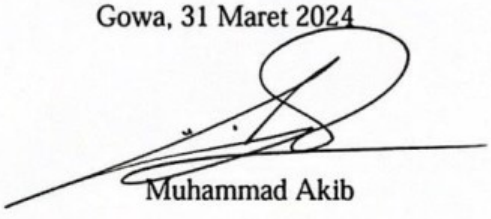
Dengan rasa syukur yang mendalam, penulis ingin mengungkapkan terima kasih kepada Allah SWT atas limpahan rahmat, petunjuk, dan karunia-Nya yang telah memberikan dukungan dalam menyelesaikan penyusunan tugas akhir. Tugas akhir ini berjudul "Analasis Sentimen Kandidat Calon Presiden 2024 Menggunakan BERT dan Support Vector Machine". Skripsi ini disusun sebagai bagian dari persyaratan untuk meraih gelar Sarjana (S1) di Program Studi Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin.

Penulis menyadari banyak kesulitan dan kendala yang dihadapi saat proses penyusunan tugas akhir ini. Perjalanan yang dilalui penulis dalam menyelesaikan skripsi ini tidak lepas dari tangan-tangan berbagai pihak yang senantiasa memberikan bantuan baik berupa materi maupun dorongan moril. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih kepada :

1. Orang Tua tercinta yang senantiasa mendoakan dan memberikan dukungan baik secara materil maupun spiritual selama menyelesaikan tugas akhir ini.
2. Bapak A. Ais Prayogi Alimuddin,S.T.,M.Eng selaku dosen pembimbing I, atas segala arahan, bimbingan, dan wawasan, serta waktu yang telah diluangkan dari awal hingga terselesainya tugas akhir ini.
3. Ibu Anugrayani Bustamin, S.T., M.T. selaku dosen pembimbing II, yang telah memberikan bimbingan, waktu, dan wawasan tambahan selama penelitian ini berlangsung.
4. Seluruh teman-teman Teknik Informatika Angkatan 2017 yang telah menjadi teman seperjuangan dan senantiasa memberikan dukungan dalam penyelesaian tugas akhir ini.
5. Serta semua pihak yang tidak dapat penulis sebut satu persatu atas semua bantuan dan dukungan yang telah diberikan hinggga terselesainya tugas akhir ini.

Akhir kata penulis menyadari bahwa tugas akhir ini masih jauh dari kata sempurna oleh karena itu penulis mengharapkan segala kritik dan saran yang membangun dari berbagai pihak. Semoga tugas akhir ini bermanfaat bagi penulis dan para pembaca.

Gowa, 31 Maret 2024



Muhammad Akib

BAB I
PENDAHULUAN

1.1 Latar Belakang

Media sosial telah menjadi bagian tak terpisahkan dalam kehidupan kita saat ini. Platform seperti Facebook, Twitter, Instagram, dan LinkedIn telah mengubah cara kita berinteraksi, berbagi informasi, dan menjalin hubungan dengan orang lain. Manfaat dari penggunaan media sosial sangat beragam dan signifikan. Dengan adanya sosial media digital, masyarakat dapat dengan mudah mengetahui berita terkini ataupun memberi opini tentang isu yang sedang hangat dalam berbagai bidang, seperti contoh dalam bidang politik. Salah satu isu hangat dalam bidang politik Indonesia pada saat ini yaitu pemilihan calon kandidat presiden 2024. Pemilihan presiden merupakan momentum penting dalam perjalanan sebuah negara. Seiring dengan perkembangan teknologi dan media sosial, opini publik dan pandangan masyarakat terhadap calon presiden semakin dapat terwakili dan terekam dalam berbagai platform online. Proses pemilihan calon presiden yang berlangsung setiap lima tahun merupakan salah satu aspek penting dalam sistem demokrasi, terutama di Indonesia. Seorang politisi yang berkeinginan untuk mencalonkan diri sebagai presiden tentu akan mengambil langkah untuk memeriksa atau mempertimbangkan popularitasnya berdasarkan pandangan publik. (Nardilasari et al., 2023).

Salah satu platform media sosial yang populer dan digunakan oleh masyarakat untuk menyampaikan pendapat mereka adalah Twitter. Twitter adalah platform media sosial yang memungkinkan pengguna untuk mengirim dan membaca pesan berbasis teks dengan batasan 140 karakter. Pada awal tahun 2013, pengguna Twitter menghasilkan lebih dari 500 juta kicauan setiap harinya (Putri et al., 2022). Keberadaan Twitter sebagai platform yang populer memudahkan ı internet untuk mengakses berbagai informasi, termasuk fakta, opini, dan ıng sedang hangat dibicarakan di masyarakat. Saat ini, di platform Twitter,

banyak beredar informasi mengenai beberapa calon yang diusung oleh partai politik dalam persiapan pemilihan presiden tahun 2024.

Dari banyaknya opini yang beredar di Twitter mengenai calon-calon presiden, menjadi semakin sulit untuk menentukan calon mana yang memiliki kekuatan dan potensi untuk menjadi presiden di masa depan. Karena itu, diperlukan sebuah analisis sentimen yang dapat memberikan pemahaman yang lebih mendalam tentang pandangan dan penilaian masyarakat terhadap masing-masing calon. Melalui analisis sentimen, dapat dilakukan evaluasi terhadap sentimen positif dan negatif yang terkait dengan calon presiden, sehingga dapat membantu dalam menilai kualitas kepemimpinan dan popularitas calon. Dengan demikian, analisis sentimen dapat menjadi alat penting dalam mendukung pengambilan keputusan yang lebih objektif dan akurat dalam konteks pemilihan presiden.

Analisis sentimen telah dibahas pada banyak literatur yang menggunakan metode yang bervariasi seperti *Naïve Bayes Classifier*, *Convolutional Neural Network* (CNN), *Long Short Term Memory* (*LSTM*) dan *LSTM* dua arah (bi LSTM). Salah satu metode yang banyak digunakan pada analisis sentimen adalah *Support Vector Machine* yang dimana pada awalnya prinsip kerja dari *Support Vector Machine* yaitu mengklasifikasi secara linier, kemudian *Support Vector Machine* dikembangkan sehingga dapat bekerja pada klasifikasi non *linear*. Metode lain yang dapat digunakan untuk melakukan analisis sentimen yaitu dengan menggunakan model bahasa pra pelatihan. Model bahasa pra pelatihan dan transformers akhir akhir ini menunjukkan peningkatan kemampuan dalam beberapa tugas pemrosesan bahasa alami dan sekarang dianggap sebagai *state of the art* pada representasi kata kontekstual (Ali et al., 2021). Selama beberapa tahun terakhir, *BERT* telah menjadi model representasi yang secara luas dan efisien digunakan, mencapai kinerja terdepan pada tugas-tugas tingkat kalimat dan token-level, mengungguli banyak arsitektur tugas khusus. *BERT* telah diusulkan dalam beberapa tahun terakhir untuk mengembangkan berbagai model canggih dalam berbagai tugas Pemrosesan Bahasa Alami. Penelitian ini akan mengembangkan sebuah sistem klasifikasi untuk ın sentimen analisis terhadap isu kandidat calon presiden 2024. Metode ınakan pada penelitian kali ini yaitu *BERT* dan *Support Vector Machine*

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan di atas maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana merancang sistem analisis sentimen menggunakan *BERT* dan *Support Vector Machine*?
2. Bagaimana performa sistem analisis sentimen pada kandidat calon presiden 2024 menggunakan *BERT* dan *Support Vector Machine* dengan parameter yang optimal?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menghasilkan perangkat lunak yang dapat melakukan analisis sentimen mengenai kandidat calon presiden 2024 menggunakan model *BERT* dan *Support Vector Machine*.
2. Mengetahui akurasi analisis sentimen mengenai kandidat calon presiden 2024 menggunakan model *BERT* dan *Support Vector Machine*.

1.4 Manfaat Penelitian

Dengan dilakukannya penelitian ini, diharapkan manfaat yang didapatkan antara lain:

1. Membantu menganalisis sentimen masyarakat terhadap kandidat calon presiden 2024.
2. Hasil penelitian dapat dijadikan sebagai rujukan penelitian terkait.

1.5 Ruang Lingkup

1. Data yang digunakan berupa cuitan berbahasa Indonesia dengan kata kunci ‘Ganjar Pranowo’, ‘Anies Baswedan’, dan ‘Prabowo Subianto’ yang berasal ari twitter.
   lasifikasi terdiri atas tiga kelas, yaitu positif, negatif dan netral

BAB 2
TINJAUAN PUSTAKA

2.1 Pemilu 2024

Pemilihan umum merupakan mekanisme untuk mengimplementasikan sistem demokrasi dan menerapkan nilai-nilai keempat Pancasila serta Pasal 1 (2) UUD Negara Republik Indonesia Tahun 1945 (Yusrin & Salpina, 2023). Pemilihan umum adalah aspek yang sangat signifikan dalam menjaga kedaulatan rakyat dan demokrasi di Indonesia. Pemilu yang efektif harus memperhatikan sistem yang digunakan dan konsekuensinya. Indonesia adalah salah satu negara yang menerapkan sistem pemilu proporsional (Pakaya et al., 2022).

Pemilihan umum dianggap sebagai tahap awal dalam rangkaian kehidupan negara yang demokratis. Oleh karena itu, pemilu menjadi penggerak utama dalam mekanisme sistem politik Indonesia. Hingga saat ini, pemilu tetap dianggap sebagai peristiwa kenegaraan yang sangat penting. Hal ini disebabkan oleh keterlibatan langsung seluruh rakyat dalam pemilu. Melalui pemilu, rakyat memiliki kesempatan untuk menyampaikan keinginan mereka dalam politik dan sistem negara. Di Indonesia, Pemilu telah diadakan sebanyak 12 kali. Pemilu pertama dilaksanakan pada tahun 1955, dan setelah itu diadakan secara berurutan pada tahun 1971, 1977, 1982, 1987, 1992, dan 1997. Setelah masa kepemimpinan Presiden Soeharto berakhir, Pemilu kembali diadakan pada tahun 1999, 2004, 2009, 2014, dan yang terakhir pada tahun 2019 dan kembali akan diadakan pada tahun 2024 (Pakaya et al., 2022). Masyarakat Indonesia akan menyelenggarakan perayaan demokrasi Pemilu 2024 yang mencakup Pemilihan Presiden 2024 (Pilpres 2024), Pemilihan Legislatif 2024 (Pileg 2024), dan Pemilihan Kepala Daerah 2024 (Pilkada 2024).

2.2 Twitter

itter adalah platform media sosial yang memungkinkan penggunanya rbagi pesan singkat yang disebut "*tweet.*" Didirikan pada tahun 2006,

Twitter memungkinkan pengguna untuk mengirim dan menerima pesan dengan batasan 280 karakter. Pengguna dapat membagikan pemikiran, berita, gambar, video, tautan, dan banyak lagi melalui tweet mereka. Salah satu fitur kunci dari Twitter adalah pengguna dapat mengikuti akun pengguna lain dan melihat tweet yang mereka bagikan di lini waktu mereka. Dengan mengikuti akun-akun yang menarik, pengguna dapat mendapatkan pembaruan langsung tentang topik yang diminati, berinteraksi dengan orang-orang yang memiliki minat yang sama, atau mengikuti berita dan acara terkini. Selain itu, Twitter juga digunakan sebagai alat untuk berpartisipasi dalam percakapan publik. Pengguna dapat menggunakan tanda pagar atau "hashtag" untuk menandai tweet mereka dengan topik tertentu dan bergabung dalam diskusi atau kampanye yang sedang berlangsung di Twitterr.

Twitter telah menjadi platform yang populer untuk berita, politik, hiburan, dan aktivisme. Banyak organisasi, selebriti, politisi, dan tokoh masyarakat menggunakan Twitter untuk berkomunikasi dengan pengikut mereka dan membagikan pandangan mereka dengan audiens yang lebih luas.Pada tahun 2022, Twitter menjadi salah satu platform media sosial yang populer di kalangan masyarakat Indonesia. Jumlah pengguna Twitter di Indonesia mencapai 18,45 juta atau sekitar 4,23% dari total pengguna Twitter di seluruh dunia yang mencapai 436 juta. Informasi ini dilaporkan oleh *We Are Social*. Tidak dapat disangkal bahwa Twitter merupakan media sosial yang nyaman dan mudah digunakan untuk berbagi informasi, terutama melalui penggunaan tagar atau hashtag terkait isu-isu tertentu (Jaafar et al., 2022).

Peningkatan jumlah pengguna Twitter telah menyebabkan peningkatan jumlah tweet yang diposting. Twitter juga merupakan salah satu platform media sosial yang efektif dalam menampung berbagai opini tentang produk, layanan, termasuk film. Hal ini dikarenakan Twitter memungkinkan penggunanya untuk memberikan komentar dengan cepat dan mengungkapkan sentimen, evaluasi, perilaku, dan emosi terkait dengan entitas tertentu seperti produk, layanan, organisasi, individu, permasalahan, topik, acara, dan atribut-atribut yang terkait
in et al., 2023). Masyarakat umum sering menggunakan Twitter sebagai
tuk mengekspresikan emosi yang terkait dengan suatu hal, termasuk emosi
ın positif (Tirtayasa & Wibowo, 2023).

2.3 *Natural Language Processing*

Natural Language Processing (NLP) adalah salah satu cabang kecerdasan buatan (AI) yang berkaitan dengan interaksi antara mesin dan manusia menggunakan bahasa alami untuk memecahkan masalah terkait data bahasa alami, seperti informasi, klasifikasi teks, dan lain-lain. NLP membantu komputer dalam berkomunikasi dengan manusia menggunakan bahasa manusia dan melakukan berbagai tugas yang terkait dengan bahasa tersebut. Selain itu, NLP juga dapat digunakan dalam pencarian dan seleksi dalam proses rekrutmen dengan mengidentifikasi kemampuan dari calon karyawan, bahkan melacak mereka sebelum mereka aktif mencari pekerjaan di pasar tenaga kerja (Wirdayanti et al., 2023).

2.4 *Text Mining*

Text mining merupakan suatu proses yang menggunakan statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengambil dan mengidentifikasi informasi yang bernilai. Definisi dari data mining adalah proses penemuan pola dalam data. Berdasarkan tujuannya, data mining dapat dikelompokkan menjadi deskripsi, estimasi, prediksi, klasifikasi, pengelompokan, dan asosiasi (Ikhromr et al., 2023). Data mining dapat juga diartikan sebagai suatu proses yang menggunakan satu atau lebih teknik pembelajaran mesin (*machine learning*) untuk menganalisis dan menghasilkan pengetahuan secara otomatis (Fauzan et al., 2023). Kehadiran data mining menjadi penting karena ada masalah ledakan data yang terjadi belakangan ini. Banyak organisasi telah mengumpulkan jumlah data yang sangat besar selama bertahun-tahun (Susanto & Sudiyatno, 2014).

2.5 Analisis Sentimen

Analisis sentimen adalah suatu bidang penelitian yang terus berkembang dan
gan dengan berbagai disiplin ilmu seperti Data Mining, *Natural Language*
ıg (NLP), dan *Machine Learning*. Fokus dari analisis sentimen adalah
engekstraksi sentimen dari sebuah kalimat berdasarkan kontennya.
t analysis, juga dikenal sebagai *opinion mining*, merupakan proses yang

dilakukan secara otomatis untuk memahami, mengekstrak, dan memproses data teks guna mendapatkan informasi yang terkandung dalam suatu kalimat opini. Tujuan dari analisis sentimen ini adalah untuk melihat pendapat atau kecenderungan opini seseorang terhadap suatu masalah atau objek, apakah memiliki kecenderungan positif, negatif, atau netral (Halim & Safuwan, 2023).

Analisis sentimen adalah studi komputasional tentang pendapat, sentimen, dan emosi yang dinyatakan dalam bentuk teks (Zulfa & Winarko, 2017). Analisis sentimen juga dapar diartikan sebagai suatu disiplin ilmu yang mengkaji pendapat, perasaan, penilaian, ulasan, dan sikap seseorang terhadap suatu hal, seperti organisasi, layanan, produk, isu, individu, dan peristiwa (Dinar et al., 2023). Manfaat dari analisis sentimen adalah sebagai sebuah konsep atau evaluasi yang berguna dalam berbagai bidang. Analisis sentimen ini memungkinkan pengelompokan polaritas teks dalam kalimat atau dokumen untuk memahami apakah opini yang terkandung dalam teks tersebut bersifat positif atau negatif (Insan et al., 2023).

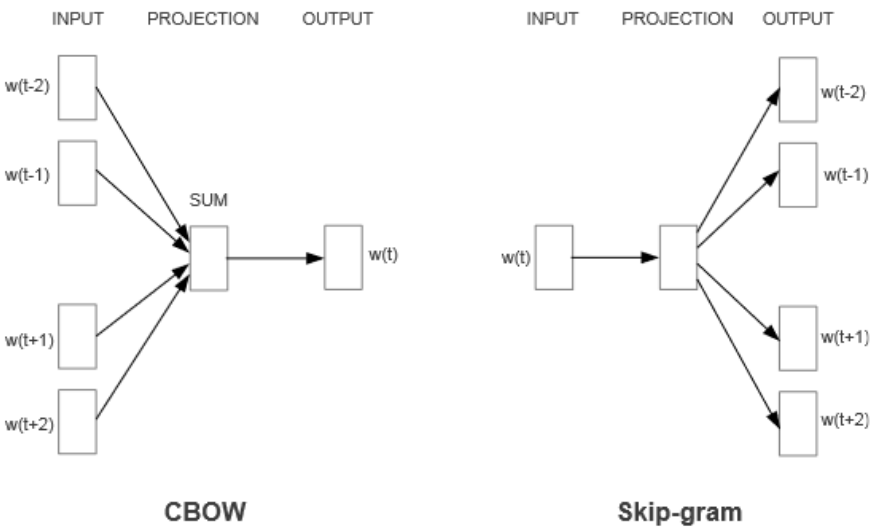
Analisis sentimen dapat memberikan manfaat dalam prediksi calon presiden dengan memahami dan mengukur pandangan, perasaan, dan opini masyarakat terhadap calon tersebut. Beberapa manfaat analisis sentimen untuk prediksi calon presiden meliputi:

1. Mengetahui persepsi publik: Analisis sentimen membantu dalam memahami bagaimana publik secara umum merespons dan memandang calon presiden. Hal ini dapat memberikan wawasan tentang popularitas, dukungan, dan pandangan masyarakat terhadap calon tertentu.
2. Mendeteksi tren dan perubahan opini: Melalui analisis sentimen, dapat diidentifikasi tren dan perubahan opini yang terjadi seiring berjalannya waktu. Informasi ini dapat membantu dalam memprediksi perubahan dukungan publik terhadap calon presiden dan mengantisipasi perubahan keadaan politik.
3. Memahami isu-isu penting: Analisis sentimen dapat membantu dalam engidentifikasi isu-isu yang penting bagi masyarakat dan bagaimana calon esiden meresponsnya. Informasi ini dapat digunakan untuk mengarahkan mpanye dan pesan politik calon.

4. Mengukur efektivitas kampanye: Dengan menganalisis sentimen, dapat dievaluasi sejauh mana kampanye calon presiden berhasil dalam mempengaruhi pandangan dan perasaan masyarakat. Informasi ini dapat digunakan untuk mengukur dampak kampanye dan memodifikasi strategi yang ada.

2.6 Word2Vec

Word2vec adalah metode yang populer dan banyak digunakan untuk mempelajari penyisipan kata dari teks yang belum diproses, yang diperkenalkan oleh Mikolov et al (Mikolov et al., 2013). Konsep word2vec, atau word embeddings, didasarkan pada gagasan merepresentasikan kata-kata secara terdistribusi. Word2Vec menggunakan jaringan saraf yang dangkal untuk mempelajari penyematan kata dengan memprediksi hubungan antara kata dan konteks sekitarnya. Pendekatan ini menetapkan hubungan antara kata-kata yang muncul dalam konteks yang serupa. Word2vec terdiri dari dua algoritma, yaitu *Skip-Gram* dan *Continuous Bag-of-Words* (CBOW), yang digunakan untuk menghasilkan vektor kata (Al-Saqqa & Awajan, 2019).



Gambar 1. Arsiketur model Word2Vec(Mikolov et al., 2013)

ambar 1 menampilkan arstiker dari masing masing Teknik yang ada pada ec. Dari gambar tersebut dapat kita lihat bahwa dalam arsitektur *CBOW*, libuat untuk kata dengan mempertimbangkan konteks sekitarnya. Di sisi

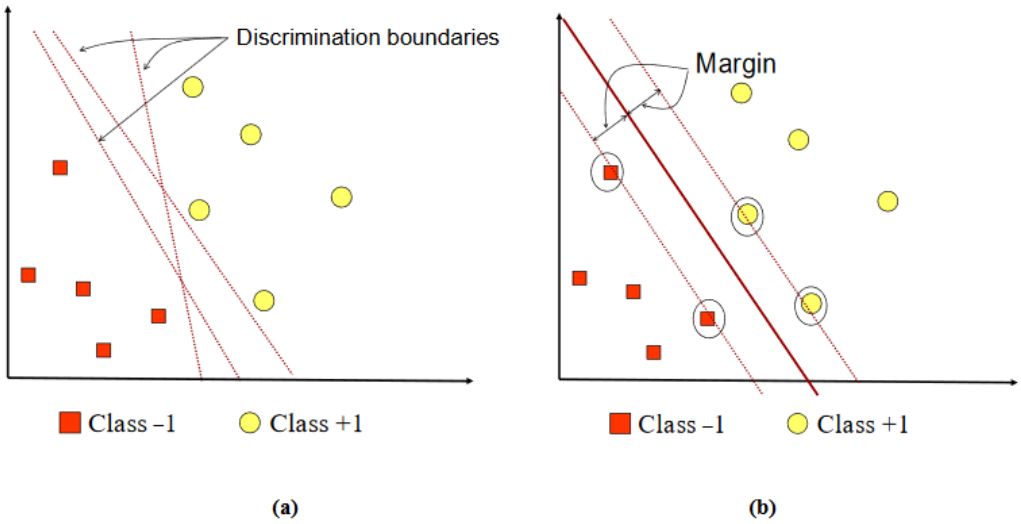
lain, model *Skip-Gram* memprediksi kata-kata di sekitarnya. *CBOW* memanfaatkan konteks untuk memprediksi kata target. *CBOW* memiliki keuntungan dalam waktu pelatihan yang lebih cepat dan memiliki akurasi yang sedikit lebih baik untuk kata-kata yang sering muncul. *Skip-Gram* menggunakan satu kata untuk memprediksi konteks target. *Skip-Gram* bekerja efektif dengan data pelatihan yang terbatas dan dapat mewakili kata-kata yang dianggap langka (Nurdin et al., 2020).

Word2Vec *CBOW (Continuous Bag-of-Words)* dan *Skip-Gram* adalah dua pendekatan yang berbeda dalam model Word2Vec untuk mempelajari representasi kata-kata dalam bentuk vektor. Dalam metode *CBOW*, model Word2Vec mencoba memprediksi kata target berdasarkan konteks kata sekitarnya. Ini berarti model menggunakan kata-kata konteks sebagai input dan mencoba memprediksi kata target. Metode ini cocok digunakan pada korpus teks dengan kata-kata yang sering muncul dan memiliki konteks yang lebih jelas. Sebaliknya, dalam metode *Skip-Gram*, model Word2Vec mencoba memprediksi kata-kata konteks berdasarkan kata target. Dalam hal ini, model menggunakan kata target sebagai input dan mencoba memprediksi kata-kata konteks yang mungkin muncul di sekitarnya. *Skip-Gram* lebih cocok untuk digunakan pada korpus teks dengan kata-kata yang jarang muncul atau memiliki konteks yang lebih kompleks. Walaupun *Skip-Gram* membutuhkan lebih banyak waktu untuk dilatih, ia cenderung menghasilkan vektor kata yang lebih baik dalam kasus-kasus di mana hubungan antara kata-kata yang jauh lebih penting. Pilihan antara *CBOW* dan *Skip-Gram* tergantung pada data teks yang digunakan dan tujuan dari pemodelan tersebut.

2.7 *Support Vector Machine*

Support Vector Machine (*SVM*) diperkenalkan pertama kali oleh Vapnik pada tahun 1992 sebagai kumpulan konsep unggulan dalam bidang pengenalan pola. Sebagai metode pengenalan pola, *SVM* masih tergolong relatif baru dalam usianya. Meskipun demikian, kemampuannya telah dievaluasi dalam berbagai aplikasi dan dianggap sebagai state of the art dalam pengenalan pola. Saat ini, *SVM* menjadi ı topik yang berkembang pesat. *SVM* adalah metode pembelajaran mesin roperasi berdasarkan prinsip Structural Risk Minimization (SRM). ·a adalah untuk menemukan *Hyperplane* terbaik yang dapat memisahkan

dua kelas pada ruang input. *SVM* memiliki kemampuan yang kuat dalam mempelajari batas keputusan yang optimal, yang memisahkan dengan jelas kelas-kelas yang berbeda dalam data. Hal ini membuat *SVM* menjadi salah satu metode yang populer dalam pengenalan pola. Dengan menggunakan konsep SRM, *SVM* berupaya mengurangi risiko struktural dan memperoleh model yang dapat memberikan generalisasi yang baik pada data yang belum pernah dilihat sebelumnya. Dengan cara ini, *SVM* dapat menghasilkan model yang dapat diterapkan pada berbagai masalah pengenalan pola dengan performa yang baik (Nugroho et al., 2003).



Gambar 2. Usaha *SVM* menemukan hyperline terbaik (Nugroho et al, 2003)

Gambar 2a menampilkan beberapa pola yang merupakan anggota dari dua kelas: +1 dan -1. Pola yang termasuk dalam kelas -1 ditandai dengan warna merah, sementara pola dalam kelas +1 ditandai dengan warna kuning. Klasifikasi masalah dapat diinterpretasikan sebagai usaha untuk menemukan garis (*Hyperplane*) yang dapat memisahkan kedua kelompok tersebut. Gambar 2a menunjukkan berbagai garis pemisah yang mungkin. *Hyperplane* pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur margin dari *Hyperplane* tersebut dan mencari titik maksimalnya. Margin adalah jarak antara *Hyperplane* dan pola terdekat dari masing-masing kelas. Pola-pola ini disebut sebagai support vector. Gambar 2b ‹kan *Hyperplane* terbaik yang ditandai dengan garis solid, yang terletak di ngah kedua kelas. Titik-titik merah dan kuning yang berada dalam

lingkaran hitam adalah support vector. Proses pembelajaran pada *SVM* berfokus pada mencari lokasi *Hyperplane* ini (Drajana, 2017).

Data dinotasikan sebagai $\overrightarrow{x_l} \epsilon R^2$ sedangkan label masing-masing dinotasikan $y_i \in$ {-1, +1} untuk i = 1, 2, …, l yang mana l adalah banyaknya data. Asumsi kedua kelas -1 dan +1 dapat terpisah secara sempurna oleh *Hyperplane* yang didefinisikan:

$$\vec{w} \cdot \vec{x} + b = 0 \quad (1)$$

Pola $\overrightarrow{x_l}$ yang termasuk kelas -1 (sampel negatif) dapat dirumuskan sebagai pola yang memenuhi pertidaksamaan pada persamaan 2

$$\vec{w} \cdot \vec{x} + b \leq -1 \quad (2)$$

Sedangkan pola $\overrightarrow{x_l}$ yang termasuk kelas +1 (sampel positif) memenuhi pertidaksamaan pada persamaan 3

$$\vec{w} \cdot \vec{x} + b \geq +1 \quad (3)$$

Dengan keterangan sebagai berikut:

$\vec{w}$ = vector pembobot

b = vector bias

Pemaksimalan jarak terdekat antara *Hyperplane* dengan pattern dilakukan dengan menghitung margin yang mana dirumuskan dengan $\frac{1}{||w||}$.

SVM menggunakan *kernel* untuk meningkatkan kemampuan pemisahan data yang kompleks. Dengan menggunakan *kernel, SVM* dapat mengklasifikasikan data yang tidak dapat dipisahkan secara *linear* di ruang asli. Berikut adalah persamaan *kernel linear* yang dapat digunakan pada *Support Vector Machine* (Bhavsar & Panchal, 2012):

Kernel Linear

$$K\,(xi, xj) = xi^T xj + c \quad (4)$$

mana xi dan xj dalah vektor fitur dari dua *instance* yang akan gkan, dan c adalah konstanta yang sering kali disebut sebagai parameter a *kernel linear*, perbandingan antara dua *instance* diukur dengan melihat

sejauh mana produk titik $xi^T xj$ dari vektor fitur ini mendekati satu sama lain. Jika produk titik besar, itu menunjukkan bahwa dua *instance* tersebut mirip dalam ruang fitur.

2.8 One vs All

Untuk membangun model SVM multiclass menggunakan metode One vs All, digunakan konsep classifier biner di mana kelas-kelas dibagi menjadi dua kelompok. Kelompok pertama terdiri dari satu kelas, sedangkan kelompok kedua terdiri dari kelas-kelas lainnya. *Classifier biner SVM* yang dihasilkan dilatih untuk menentukan apakah kelas tersebut termasuk dalam kelompok pertama atau kelompok kelas lainnya. Proses ini diulang untuk kelompok kedua yang terdiri dari lebih dari dua kelas, sehingga setiap kelompok memiliki satu kelas (Astriratma, 2020). Berikut adalah tabel yang menggambarkan cara kerja OvA pada SVM.

Tabel 1. Gambaraan cara kerja OvA SVM

yi = 1	yi = -1	Hipotesis
Kelas 1	Bukan Kelas 1	$f^1(x) = (w^1)x + b^1$
Kelas 2	Bukan Kelas 2	$f^2(x) = (w^2)x + b^2$
Kelas …	Bukan Kelas …	$f^{\cdots}(x) = (w^{\cdots})x + b^{\cdots}$
Kelas n	Bukan Kelas n	$f^n(x) = (w^n)x + b^n$

Dari Tabel 1, metode OvA akan membangun sejumlah model SVM sesuai dengan jumlah kelas (n). Setiap model klasifikasi dilatih dengan keseluruhan data untuk mencari kelas yang sesuai. Dengan demikian, menggunakan pendekatan ini, SVM multiclass diubah menjadi beberapa classifier SVM biner.

a) Kelas 1 vs Bukan Kelas 1

Ketika ingin memisahkan Kelas 1 dari kelas lainnya, langkah yang dilakukan membuat sebuah model SVM yang bertujuan untuk mengklasifikasikan tara Kelas 1 (label positif, yi = 1) dan bukan Kelas 1 (label negatif, yi = -otesis untuk Kelas 1 adalah $f^1(x) = (w^1)x + b^1$, di mana $w^1 + b^1$ adalah

parameter yang dipelajari oleh model SVM. Hipotesis untuk Bukan Kelas 1 adalah $f^1(x) = -(w^1)x + b^1$,, di mana $w^1 + b^1$ tetap digunakan, namun dengan pembalikan tanda (-) untuk membedakan kelas negatif.

b) Kelas 2 vs Bukan Kelas 2:

Proses ini diulang untuk Kelas 2, di mana kita membuat model SVM yang memisahkan antara Kelas 2 (yi = 1) dan bukan Kelas 2 (yi = -1). Hipotesis untuk Kelas 2 adalah $f^2(x) = (w^2) + b^2$, di mana $(w^2)x + b^2$, adalah parameter model SVM kelas 2. Hipotesis untuk Bukan Kelas 2 $-f^2(x) = -(w^2) + b^2$, dengan menggunakan parameter $(w^2)x + b^2$ yang sama, namun dengan tanda negatif.

c) Kelas ke-n vs Bukan Kelas ke-n

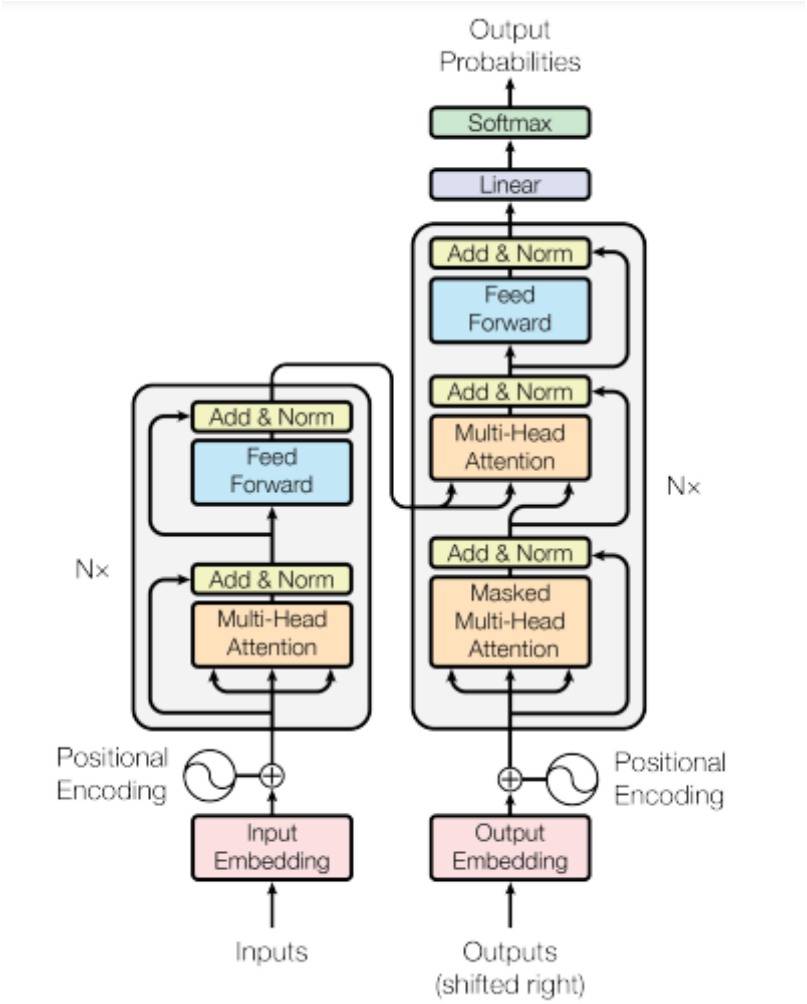
Proses ini juga diulang untuk Kelas ke-n, di mana kita membuat model SVM yang memisahkan antara Kelas ke-n (yi = 1) dan bukan Kelas ke-n (yi = -1). Hipotesis untuk Kelas ke-n adalah $f^n(x) = (w^n)x + b^n$, di mana $(w^n) + b^n$ adalah parameter model SVM kelas ke-n. Hipotesis untuk Bukan Kelas ke-n adalah $\mathrm{f}f^n(x) = -(w^n)x + b^n$, menggunakan parameter $(w^n) + b^n$ dengan tanda negatif.

Dengan demikian, metode OvA membangun sejumlah model SVM biner yang terpisah untuk setiap kelas, di mana setiap model SVM berusaha memisahkan kelas tersebut dari kelas-kelas lainnya. Proses ini memungkinkan klasifikasi multikelas dengan efektif menggunakan algoritma SVM.

2.9 *BERT*

BERT, yang merupakan singkatan dari "*Bidirectional Encoder Representations from Transformer*s" adalah sebuah model bahasa yang ıgkan oleh Google pada tahun 2018. Model ini menggunakan arsitektur ıer, yang merupakan jenis arsitektur jaringan saraf yang sangat sukses gas-tugas pemrosesan bahasa alami. *BERT* adalah model *deep learning*

yang telah menghasilkan kemajuan signifikan dalam berbagai tugas Pemrosesan Bahasa Alami (NLP). Model ini terdiri dari enam lapisan *Transformer* yang disusun di atas *encoder* dan *decoder* masing-masing. Kontribusi dari setiap lapisan Transformer ini menyebabkan proses pelatihan yang sangat kompleks, konfigurasi tinggi, waktu pelatihan yang besar, dan biaya yang signifikan (Atmaja & Yustanti, 2021). Arsitektur *Transformer* dapat dilihat pada Gambar 3 dibawah ini.



Gambar 3. Arsitektur Transformer (Vaswani et al., 2023)

Struktur *BERT* terdiri dari encoder Transformer yang bersifat multi-lapisan dan bersifat bidireksional (Vaswani et al., 2023). Secara umum, langkah-langkah dalam pelatihan *BERT* melibatkan:

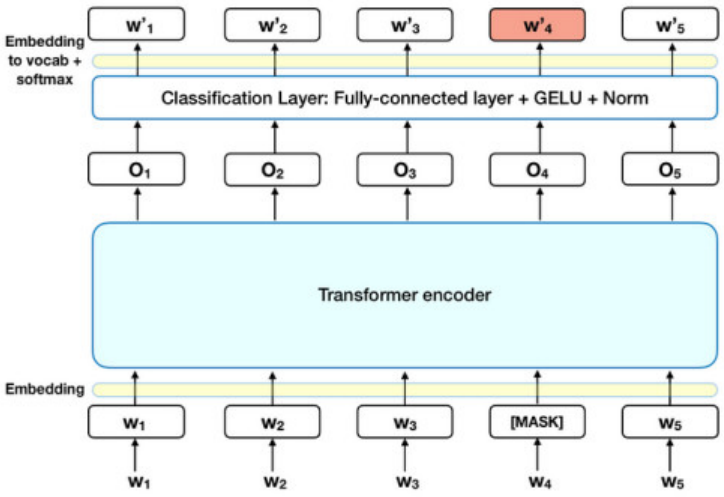
nbuatan model pretrained menggunakan dua skenario Pemrosesan Bahasa mi (NLP), yaitu *Masked Language Model* (MLM) yang ditunjukkan pada mbar 4 dan *Next Sentence Prediction* (NSP) yang ditunjukkan pada

Gambar 5, untuk menghasilkan embedding baru yang optimal untuk kedua kasus tersebut.

- *Masked Language Model*

Sebelum memberikan urutan kata ke *BERT*, sekitar 15% kata dalam setiap urutan digantikan dengan token [*MASK*]. Model kemudian berusaha memprediksi nilai asli dari kata-kata yang disamarkan, berdasarkan konteks yang diberikan oleh kata-kata lain yang tidak disamarkan dalam urutan tersebut. Secara teknis, prediksi kata keluaran memerlukan:

1. Penambahan lapisan klasifikasi di atas keluaran *encoder*.
2. Penggandaan vektor keluaran dengan matriks penyematan, mengubahnya ke dalam dimensi kosakata.
3. Penghitungan probabilitas setiap kata dalam kosakata dengan softmax.



Gambar 4. *Masked language model* (Devlin et al., 2019)

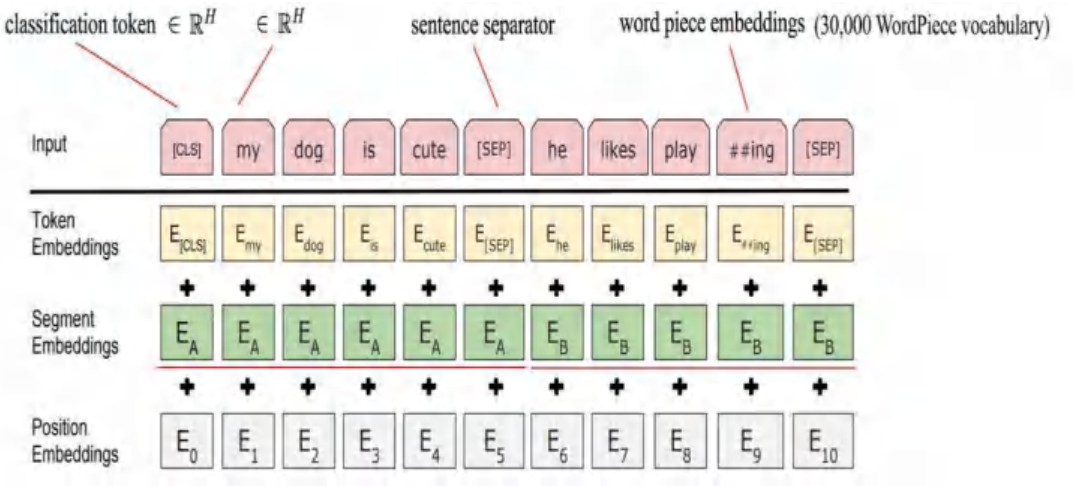
Fungsi kerugian *BERT* hanya mempertimbangkan prediksi nilai-nilai yang disamarkan dan mengabaikan prediksi kata-kata yang tidak disamarkan. Sebagai akibatnya, model ini konvergen lebih lambat dibandingkan dengan model-model searah, sebuah karakteristik yang diimbangi oleh peningkatan kesadaran konteksnya.

- *Next Sentence Prediction*

am proses pelatihan *BERT*, model menerima pasangan kalimat sebagai ut dan belajar untuk memprediksi apakah kalimat kedua dalam pasangan ebut merupakan kalimat berikutnya dalam dokumen asli. Selama

pelatihan, 50% dari input adalah pasangan di mana kalimat kedua adalah kalimat berikutnya dalam dokumen asli, sedangkan dalam 50% lainnya, kalimat acak dari korpus dipilih sebagai kalimat kedua. Asumsinya adalah kalimat acak tersebut akan tidak terhubung dengan kalimat pertama. Untuk membantu model membedakan antara dua kalimat tersebut selama pelatihan, input diproses dengan cara berikut sebelum memasuki model:

1. Token [CLS] dimasukkan di awal kalimat pertama dan token [SEP] dimasukkan di akhir setiap kalimat.
2. *Embedding* kalimat yang menunjukkan Kalimat A atau Kalimat B ditambahkan ke setiap token. *Embedding* kalimat serupa dalam konsepnya dengan embedding token dengan kosakata.
3. *Embedding* posisi ditambahkan ke setiap token untuk menunjukkan posisinya dalam urutan. Konsep dan implementasi *embedding* posisi dijelaskan dalam makalah Transformer.



Gambar 5. Next sentence prediction (NSP)(Devlin et al., 2019)

Untuk memprediksi apakah kalimat kedua benar-benar terhubung dengan yang pertama, langkah-langkah berikut dilakukan:

1. Seluruh urutan input melewati model Transformer.
2. Keluaran dari token [CLS] diubah menjadi vektor berbentuk 2×1, menggunakan lapisan klasifikasi sederhana (matris bobot dan bias yang dipelajari).

   Menghitung probabilitas *IsNextSequence* dengan *softmax*.

4. Saat melatih model *BERT*, Masked LM dan *Next Sentence Prediction* dilatih bersama-sama, dengan tujuan meminimalkan fungsi kerugian gabungan dari kedua strategi tersebut.

(2) Penggunaan *embedding* yang dihasilkan untuk dimasukkan kembali ke dalam jaringan saraf yang terhubung dengan output sesuai dengan kasus NLP yang ditargetkan.

2.10 *IndoBERT*

IndoBERT ialah model pra-pelatihan yang dipelajari dengan menggunakan Transformer, suatu algoritma yang diadaptasi dari prinsip kerja Convolutional Neural Network. Meskipun, terdapat perbedaan dalam proses ekstraksi fitur pada Transformer, dimana tidak ada konvolusi dengan kernel n x m seperti pada CNN, melainkan melibatkan encoder dan decoder. Sementara itu, IndoLEM adalah dataset yang dapat digunakan untuk tujuh tugas Pemrosesan Bahasa Alami (NLP) yang umumnya digeneralisasi menjadi tiga kategori, yakni morfo-sintaks, semantik, dan diskursif. IndoBERT merupakan sebuah model berbasis transformer yang mengadopsi gaya BERT. Model ini secara khusus dilatih *sebagai masked language model* (MLM) dengan menggunakan *framework Huggingface*. Model ini mengikuti konfigurasi standar untuk *BERT-Base*, yang mencakup 12 lapisan tersembunyi (*hidden layer*), 12 kepala perhatian (*attention head*), dan lapisan tersembunyi *feed-forward* ((Koto & Baldwin, 2020)

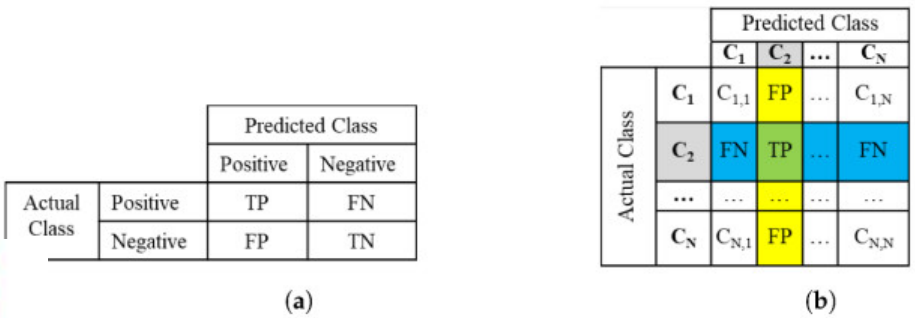
IndoBERT adalah sebuah model pre-trained yang dikembangkan khusus untuk bahasa Indonesia. Model ini didasarkan pada arsitektur BERT (Bidirectional Encoder Representations from Transformer) yang sangat sukses dalam penelitian Natural Language Processing (NLP). Dengan menggunakan dataset bahasa Indonesia yang besar, IndoBERT dilatih untuk memahami dan menganalisis teks dalam bahasa Indonesia dengan lebih baik, termasuk memahami makna kata-kata, sintaksis, dan konteks kalimat. Hal ini membuat IndoBERT menjadi salah satu alat yang penting dan efektif dalam pengolahan bahasa Indonesia di berbagai aplikasi

oBERT dapat melakukan beberapa tugas yang terkait dengan pemrosesan ami. Di antaranya adalah:

1. Klasifikasi Teks: IndoBERT dapat mengklasifikasikan teks berdasarkan topik atau kategori tertentu, seperti politik, bisnis, olahraga, atau hiburan. Ini berguna bagi analis data atau peneliti untuk mengorganisasi data dengan lebih baik.
2. Analisis Sentimen: IndoBERT dapat mengenali sentimen dalam suatu teks, seperti apakah ulasan restoran bersifat positif atau negatif. Hal ini membantu pemilik bisnis meningkatkan kualitas produk dan layanan mereka.
3. Ekstraksi Informasi: IndoBERT dapat mengekstraksi informasi penting dari teks, seperti nama, tanggal, dan lokasi dari sebuah artikel berita. Ini membantu dalam pengumpulan dan penyajian data secara efektif.
4. Penerjemahan Bahasa: IndoBERT dapat menerjemahkan teks dari bahasa Indonesia ke bahasa Inggris atau bahasa lainnya dengan akurasi tinggi, membantu para peneliti atau penerjemah dalam mengolah data dalam bahasa yang diinginkan.
5. Generasi Teks: IndoBERT dapat menghasilkan teks baru berdasarkan input yang diberikan, seperti ringkasan artikel berita atau jawaban atas pertanyaan yang diberikan. Ini mempermudah pengguna dalam mendapatkan informasi yang dibutuhkan dengan cepat.

2.11 *Confusion Matrix*

Confusion Matrix adalah metode yang sering digunakan untuk membandingkan hasil klasifikasi suatu sistem dengan hasil klasifikasi yang sebenarnya. *Confusion Matrix* ini berbentuk tabel yang menggambarkan kinerja model klasifikasi pada sejumlah data yang diuji, di mana nilai sebenarnya diketahui (Suparno, 2023).



Gambar 6. *Confusion Matrix* (a) Binary Classification (b) Multiclass Classification (Markoulidakis et al., 2021)

Gambar 6 menampilkan tabel *Confusion Matrix binary classification* dan *multiclass classification*. TP, FP, FN, dan TN adalah singkatan yang digunakan dalam evaluasi kinerja model klasifikasi seperti SVM (Support Vector Machine), dan mereka mewakili empat kemungkinan hasil yang bisa diperoleh saat melakukan prediksi terhadap data. Berikut adalah penjelasan singkat tentang masing-masing konsep:

a. True Positive (TP):

   TP terjadi ketika model klasifikasi benar-benar memprediksi kelas positif dengan benar.

   Contoh: Jika model memprediksi pasien memiliki penyakit (kelas positif), dan ternyata pasien tersebut benar-benar memiliki penyakit, maka prediksi tersebut merupakan True Positive (TP).

b. False Positive (FP):

   FP terjadi ketika model klasifikasi salah memprediksi kelas positif padahal sebenarnya kelasnya negatif.

   Contoh: Jika model memprediksi pasien memiliki penyakit (kelas positif), tetapi ternyata pasien tersebut sebenarnya tidak memiliki penyakit, maka prediksi tersebut merupakan False Positive (FP).

c. False Negative (FN):

   FN terjadi ketika model klasifikasi salah memprediksi kelas negatif padahal sebenarnya kelasnya positif.

   Contoh: Jika model memprediksi pasien tidak memiliki penyakit (kelas negatif), tetapi ternyata pasien tersebut sebenarnya memiliki penyakit, maka prediksi tersebut merupakan False Negative (FN).

d. True Negative (TN):

   TN terjadi ketika model klasifikasi benar-benar memprediksi kelas negatif dengan benar.

   Contoh: Jika model memprediksi pasien tidak memiliki penyakit (kelas negatif), dan ternyata pasien tersebut benar-benar tidak memiliki penyakit, maka prediksi tersebut merupakan True Negative (TN).

Tabel *Confusion Matrix* tersebut dapat digunakan untuk menghitung nilai-nilai metrik evaluation seperti nilai akurasi, presisi, *recall* dan *F1 Score*.

1) Akurasi, merupakan metrik yang mengukur seberapa tepat model dalam memprediksi keseluruhan jumlah data dengan benar. Akurasi dapat dihitung dengan menggunakan persamaan:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

2) Presisi, merupakan metrik yang mengukur seberapa tepat model dalam memprediksi data positif. Nilai presisi dapat dihitung dengan menggunakan persamaan berikut:

$$Presisi = \frac{TP}{TP + FP} \quad (6)$$

3) *Recall,* Merupakan metrik yang mengukur seberapa baik model dalam menemukan atau mengingat data positif. Nilai *recall* dapat dihitung dengan menggunakan persamaan berikut:

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

4) *F1 Score,* Merupakan penggabungan dari presisi dan *recall*. *F1 Score* adalah *harmonic mean* dari presisi dan *recall,* dan memberikan ukuran keseluruhan kinerja model dalam memprediksi data positif dengan baik. *F1 Score* dapat dinyatakan dengan persamaan berikut:

$$F1\ Score = \frac{2\ x\ Presisi \times Recall}{Presisi + Recall} \quad (8)$$

2.12 *Grid Search*

Grid Search adalah salah satu teknik yang digunakan dalam machine learning untuk menemukan kombinasi parameter terbaik untuk model yang sedang diuji. Teknik ini sering digunakan untuk mengoptimalkan hyperparameter dalam algoritma machine learning, seperti algoritma pembelajaran berbasis klasifikasi atau regresi. Cara kerja *Grid Search* adalah dengan melakukan pencarian melalui sejumlah kombinasi parameter yang telah ditentukan sebelumnya. Kombinasi parameter ini dibentuk dalam bentuk grid atau tabel, di mana setiap parameter beberapa nilai yang mungkin. *Grid Search* kemudian menguji model akan setiap kombinasi parameter ini dan mengukur kinerja model akan metrik evaluasi yang ditentukan, seperti akurasi, presisi, atau F1-

score. Hasil dari *Grid Search* adalah kombinasi parameter yang menghasilkan kinerja model terbaik berdasarkan metrik evaluasi yang dipilih.