**Penerapan SqueezeBERT dalam Klasifikasi Emosi Multi-label dengan Menggunakan PyTorch dan Tez**

Proposal Proyek Akhir

Oleh :

| 11S18015 | Yanada Sari Situmorang |
| --- | --- |
| 11S18002 | Putri Era Waty Bakara |
| 11S18010 | N. Priskila Napitupulu |
| 11S18012 | Suryani Vebhitry Siahaan |
| 11S18038 | Grace Winda Nissi Naibaho |
| 11S18063 | Elyta Junetta Siregar |
|  |  |



Institut Teknologi Del

2021/2022

# DAFTAR ISI

[**DAFTAR ISI**](#_heading=h.f57ul06or0bk) **1**

[**DAFTAR GAMBAR**](#_heading=h.30j0zll) **2**

[**PENDAHULUAN**](#_heading=h.3znysh7) **1**

[Latar Belakang](#_heading=h.2et92p0) 1

[Rumusan Masalah](#_heading=h.3dy6vkm) 2

[Tujuan Penelitian](#_heading=h.1t3h5sf) 3

[Ruang Lingkup Penelitian](#_heading=h.4d34og8) 3

[Hasil yang diharapkan](#_heading=h.2s8eyo1) 3

[Tahapan Penelitian](#_heading=h.vywbzj45pjhn) 3

[**DAFTAR REFERENSI**](#_heading=h.2jxsxqh) **5**

# DAFTAR GAMBAR

[Gambar 1.1 Hubungan Era Kelima Komputer dan Era Kelima Manajemen 2](#_heading=h.tyjcwt)

[Gambar 2.2 *Value Chain* Nike 4](#_heading=h.44sinio)

**DAFTAR TABEL**

[Tabel 2.1 Karakteristik A Berdasarkan B 4](#_heading=h.lnxbz9)

# PENDAHULUAN

Pada bab ini dijelaskan mengenai latar belakang topik penelitian, rumusan permasalahan penelitian, tujuan penelitian, ruang lingkup penelitian, metode penelitian dan sistematika penulisan dalam proyek mata kuliah Pemrosesan Bahasa Alami (PBA).

## Latar Belakang

Pemrosesan Bahasa Alami atau yang sering disebut sebagai *Natural Language Processing* (NLP) menjadi pilar penting dalam perkembangan AI saat ini. NLP sangat populer dan mempunyai dampak yang sangat besar di dalam bidang teknologi dan pemasaran, yang berhubungan dengan interaksi antara mesin dan manusia menggunakan bahasa natural untuk membantu komputer dalam memahami bahasa manusia.

Beberapa contoh pengaplikasian NLP diantaranya seperti *google translate, sentiment analysis, speech recognition, emotion recognition, chatbot,* dan lain sebagainya. Dari contoh tersebut adapun pengaplikasian yang akan dibahas pada penelitian ini adalah terkait *emotion recognition*. *Emotion recognition* merupakan interpretasi dan klasifikasi emosi (gembira, sedih, bahagia, marah, antisipasi, dan lain sebagainya) yang menggunakan data *sequence* mentah (audio, teks, atau video) yang bertujuan untuk membangun sistem yang dapat melakukan klasifikasi emosi, dengan terlebih dahulu memprediksi skor emosi pada data *sequence* di masa depan.

Emosi merupakan perasaan kuat yang berasal dari suasana hati atau interaksi seseorang satu sama lain. Banyak orang yang suka mengekspresikan emosi mereka atau berbagai peristiwa yang berarti seperti fakta, opini, emosi dan intensitas emosi pada berbagai jenis topik dalam data *sequence*, yang dalam penelitian ini berupa teks/kalimat pendek. Hal ini biasanya diekspresikan ke orang lain melalui media sosial, seperti twitter, facebook, *personal notes*, blog, novel, email, pesan obrolan, dan berita utama. Oleh karena itu, topik ini telah menarik perhatian para peneliti di bidang pemrosesan bahasa alami untuk melakukan analisis terhadap emosi yang diungkapkan dalam media sosial guna membangun suatu sistem yang dapat mendeteksi emosi tersebut. Analisis untuk mendeteksi emosi ini dapat menguntungkan banyak pihak di bidang perdagangan, kesehatan masyarakat, kesejahteraan sosial, dan lain-lain. Contoh penerapan nyata adalah suatu sistem dapat mendeteksi opini publik tentang politik, manajemen merek, dan pemantauan pasar saham.

Deteksi emosi merupakan masalah klasifikasi emosi multi-label yang membutuhkan prediksi beberapa skor emosi dari data *sequence* yang diberikan. Setiap data *sequence* yang diberikan dapat memiliki lebih dari satu emosi, sehingga masalahnya dapat diajukan sebagai masalah klasifikasi emosi multi-label. Klasifikasi emosi multi-label adalah sub-tugas dari klasifikasi emosi teks, yang bertujuan untuk mengidentifikasi emosi, seperti kegembiraan, kemarahan, kecemasan, dan lain-lain dari suatu teks pendek. Dalam proses klasifikasi emosi multi-label ini, dibutuhkan suatu model bahasa yang dapat membantu dalam proses pengklasifikasian emosi teks, salah satunya adalah model bahasa BERT.

BERT (*Bidirectional Encoder Representations from Transformer*)adalah kerangka kerja pembelajaran mesin yang bersifat *open source*, yang artinya siapa pun dapat menggunakannya. Saat ini, Google menggunakan model BERT tersebut untuk membantu dalam memahami bagaimana konteks kata dalam mengoptimalkan interpretasi permintaan pencarian pengguna. Hal ini pastinya menunjukkan bahwa pentingnya model tersebut sehingga banyak organisasi lain, kelompok penelitian, dan lain sebagainya yang menyempurnakan arsitektur model BERT dengan *supervised learning* untuk mengoptimalkannya demi efisiensi atau mengkhususkannya untuk tugas-tugas tertentu. Salah satu contoh penyempurnaan arsitektur model BERT ini adalah model SqueezeBERT. Salah satu tujuan penyempurnaan arsitektur tersebut adalah untuk menyempurnakan tugas pengklasifikasian emosi teks untuk data *sequence*. Selain itu, SqueezeBERT ini mampu berjalan 4,3x lebih cepat daripada basis BERT. Oleh karena itu, berdasarkan masalah pengklasifikasian emosi multi label dan model SqueezeBERT di atas, maka Penulis dalam penelitian ini akan melakukan penelitian terkait **“Penerapan SqueezeBERT dalam Klasifikasi Emosi Multi Label dengan Menggunakan PyTorch dan Tez”**.

## Rumusan Masalah

Adapun yang menjadi rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan model SqueezeBERT dalam proses pengklasifikasian emosi multi label berdasarkan skor prediksi emosi yang dihasilkan terhadap data *training* dan *testing* dalam skala data yang besar?
2. Bagaimana keakuratan klasifikasi emosi yang dihasilkan ketika dilakukan pengujian terhadap kalimat baru?

## Tujuan Penelitian

Adapun yang menjadi tujuan dalam penelitian ini adalah sebagai berikut:

1. Menerapkan model SqueezeBERT dalam proses pengklasifikasian emosi multi label berdasarkan skor prediksi emosi yang dihasilkan terhadap data *training* dan *testing* dalam skala data yang besar.
2. Mengetahui tingkat keakuratan klasifikasi emosi yang dihasilkan ketika dilakukan pengujian terhadap kalimat baru.

## Ruang Lingkup Penelitian

Adapun yang menjadi ruang lingkup dalam penelitian ini adalah:

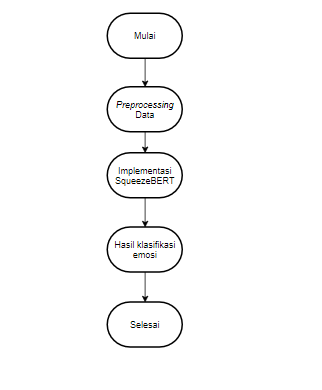
1. Penelitian ini menggunakan pendekatan SqueezeBERT.
2. Library yang digunakan pada proyek ini adalah PyTorch dan Tez.
3. Data set yang digunakan diambil dari *link* berikut <https://huggingface.co/datasets/go_emotions>.

## Hasil yang diharapkan

Hasil yang diharapkan dari pengerjaan proyek ini adalah mampu melakukan klasifikasi terkait emosi (bahagia, sedih, pengakuan, kaget, dan lain sebagainya) berdasarkan kalimat yang diberikan dengan terlebih dahulu melabeli setiap kelas emosi yang terdapat pada *dataset* yang digunakan dan kemudian menampilkan 5 label emosi berdasarkan emosi yang memiliki peluang terbanyak.

## Tahapan Penelitian

Berikut ini adalah tahapan yang akan dilakukan dalam menyelesaikan proyek pemrosesan bahasa alami :



Gambar 1. Tahapan Penelitian

1. Preprocessing Data

Pada tahap ini dilakukan data *preprocessing* yang merupakan proses untuk mengubah data berupa teks yang sebelumnya belum terstruktur akan menjadi sebuah data yang terstruktur. Preprocessing ini juga dilakukan untuk menghilangkan kata yang tidak diperlukan sampai menyeragamkan bentuk data.

1. Implementasi SqueezeBERT

Pada tahap ini peneliti akan melakukan penerapan pendekatan SqueezeBERT. Kemudian model yang dihasilkan dari penerapan pendekatan tersebut akan disimpan sebagai *knowledge* untuk mengklasifikasikan data yang akan di input.

1. Hasil Klasifikasi Emosi

Hasil akhir dari penelitian ini adalah model yang dapat mengklasifikasikan emosi (bahagia, sedih, pengakuan, kaget, dan lain sebagainya) berdasarkan kalimat yang diberikan dengan memberi label terlebih dahulu kepada dataset yang digunakan.

# DAFTAR REFERENSI

Jabreel, M., & Moreno, A. (2019). A Deep Learning-Based Approach for Multi-Label Emotion Classification in Tweets. *Applied Sciences*, *9*(6), 1123. https://doi.org/10.3390/app9061123

*SqueezeBERT*. Huggingface.co. (2021). Retrieved 11 November 2021, from<https://huggingface.co/transformers/model_doc/squeezebert.html>.

*squeezebert/squeezebert-uncased · Hugging Face*. Huggingface.co. (2021). Retrieved 11 November 2021, from<https://huggingface.co/squeezebert/squeezebert-uncased>.

Tyagi, N. (2021). *What is SqueezeBERT in NLP? | Analytics Steps*. Analyticssteps.com. Retrieved 11 November 2021, from<https://www.analyticssteps.com/blogs/what-squeezbert-nlp>.

*What is BERT (Language Model) and How Does It Work?*. SearchEnterpriseAI. (2021). Retrieved 11 November 2021, from<https://searchenterpriseai.techtarget.com/definition/BERT-language-model>.

*Penerapan Algoritma BERT dalam Search Engine Google*. MTI. (2021). Retrieved 11 November 2021, from<https://mti.binus.ac.id/2020/09/03/penerapan-algoritma-bert-dalam-search-engine-google/>.