**Multi-Label Emotion Classification**

**using Naïve Bayes**

**(study case: Reddit Comments)**

Proposal Tugas Pemrosesan Bahasa Alami

Oleh :

| 11S18001 | Efrica C. Situmeang |
| --- | --- |
| 11S18007 | Aldi Irvan Siagian |
| 11S18013 | Grace Noelia Simorangkir |
| 11S18040 | Anjel Riska Pardede |
| 11S18041 | Grace Widya Simanjuntak |
| 11S18048 | Romual Naibaho |



11S4037 - Pemrosesan Bahasa Alami

Fakultas Informatika dan Teknik Elektro

Institut Teknologi Del

2021

# DAFTAR ISI

[**DAFTAR ISI**](#_heading=h.1fob9te) **i**

[**DAFTAR TABEL**](#_heading=h.2et92p0) **ii**

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

[Latar Belakang](#_heading=h.3dy6vkm) 1

[Rumusan Masalah](#_heading=h.4d34og8) 3

[Tujuan](#_heading=h.1fej77yahbmn) 3

[Manfaat Penelitian](#_heading=h.2s8eyo1) 3

[Ruang Lingkup Penelitian](#_heading=h.17dp8vu) 4

[**TINJAUAN PUSTAKA**](#_heading=h.3rdcrjn) **5**

[Emosi](#_heading=h.26in1rg) 5

[Natural Language Processing](#_heading=h.bcb8n0ptcr) 5

[Data Preprocessing](#_heading=h.mp47h7gcu7md) 6

[TF-IDF](#_heading=h.wwmcqcfrccpr) 7

[Klasifikasi](#_heading=h.nmjucfc41k8p) 8

[Multi-Label Classification](#_heading=h.j383wr3syyjj) 9

[Naive Bayes Classification](#_heading=h.ywl16mnco3si) 9

[**PEMBAGIAN TUGAS**](#_heading=h.edmpv1wmivtn) **13**

[**DAFTAR REFERENSI**](#_heading=h.a832xjvps9sp) **15**

# DAFTAR TABEL

[Tabel 1 Daftar Tugas](#_heading=h.2bn6wsx) 12

[Tabel 2 Pembagian Tugas](#_heading=h.qsh70q) 13

# PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang topik penelitian, tujuan penelitian, manfaat penelitian, ruang lingkup penelitian.

## Latar Belakang

Emosi adalah pola reaksi yang kompleks, yang melibatkan pengalaman, perilaku, dan fisiologis, di mana seorang individu mencoba untuk menangani masalah atau peristiwa yang signifikan secara pribadi. Dalam kehidupan sehari-hari, penyampaian emosi dapat disampaikan secara non-verbal menggunakan ekspresi wajah maupun verbal berupa tingkah laku. Emosi bersifat subjektif dan temporer yang muncul atau dipicu oleh stimulus seperti perlakuan dari orang sekitar atau lingkungannya.

Deteksi emosi adalah masalah umum yang berpotensi pada peningkatan seperti interaksi manusia atau komputer yang lebih baik. Setiap orang tentu pernah mengalami berbagai emosi. Analisis emosi adalah tindakan untuk menentukan sikap terhadap target atau topik. Sikap dapat berupa polaritas (positif atau negatif) atau keadaan emosional seperti senang, marah, atau sedih. Deteksi emosi pada teks dapat dianggap sebagai masalah klasifikasi pada konsep pemrosesan bahasa alami (NLP) dan Machine learning. Menurut hasil survey yang telah dilakukan pada tahun 2014 - 2018 terdapat 6 dari 10 penelitian yang melakukan penelitian tentang deteksi emosi pada teks (Al-Saaqa et al., 2018). Deteksi emosi pada teks telah menarik minat yang cukup besar karena penerapannya ke berbagai domain, termasuk klasifikasi teks, klasifikasi adegan dan video, dan bioinformatika terutama untuk kepentingan analisis emosi.

Pada penelitian ini akan menggunakan Reddit Comments (komentar dari media sosial Reddit) yang terdiri dari 27 kategori label emosi. Reddit merupakan situs web hiburan dan juga berita tempat user dapat berkontribusi dalam bentuk postingan pranala atau teks. Reddit menjadi salah satu media sosial yang populer di benua Amerika dan Eropa. Para pengguna Reddit biasanya akan membuat postingan yang mengusung tema seperti politik, ekonomi, budaya, agama, organisasi, atau tema lainnya dan akan menjadi topik perbincangan banyak orang. Semakin hari, ulasan pada Reddit yang disampaikan oleh pengguna Reddit semakin bertambah dan semakin beragam seperti komentar, kritik, opini yang positif, negatif atau netral. Namun tidak semua pengguna dapat memahami apa bentuk emosi yang ada pada postingan maupun ulasan tersebut. Karena bentuk emosi sangat sulit diidentifikasi jika hanya melihat tulisan saja. Oleh sebab itu penelitian ini dilakukan untuk mengklasifikasikan emosi dari berbagai komentar atau ulasan yang ada sehingga pengguna dapat mengetahui bagaimana bentuk emosi yang ada dalam ulasan tersebut.

Deteksi emosi dilakukan dengan pendekatan klasifikasi emosi. Klasifikasi adalah proses dengan menggabungkan atau mengelompokkan dua atau lebih data yang memiliki kesamaan pada suatu kriterianya. Terdapat beberapa metode yang dapat digunakan untuk melakukan klasifikasi pada teks, diantaranya Support Vector Machine, Naïve Bayes, K-Nearest Neighbor, Multinomial Naïve Bayes, Multi Task Regression, Algoritma Incremental Regression, dan lain-lain. Sedangkan untuk metode fitur ekstraksi juga terdapat beberapa yang sering digunakan, diantaranya TF-IDF, N-Gram, Word Embedding, dan lain-lain. Namun pada penelitian ini beberapa pendekatan atau metode yang akan digunakan dalam melakukan fitur ekstraksi yaitu TF-IDF yang kemudian diklasifikasikan dengan metode Naïve Bayes. Berdasarkan penelitian sebelumnya fitur TF-IDF memberikan hasil yang lebih baik sekitar 3% - 4% jika dibandingkan dengan fitur N-Gram (Ahuja et al., 2019). TF-IDF merupakan teknik yang paling sederhana sehingga mudah untuk melakukan pengecekan corpus yang besar dan juga dapat menghitung banyak istilah yang ada dalam sebuah dokumen. Algoritma pengklasifikasian Naive Bayes adalah salah satu pengklasifikasian statistik, pengklasifikasian ini dapat memprediksi probabilitas anggota kelas data yang akan masuk ke kelas tertentu, menurut perhitungan probabilitas. Klasifikasi ini menunjukkan akurasi dan kecepatan yang tinggi bila diterapkan kedalam database yang besar. Metode ini sering digunakan untuk memecahkan masalah di bidang Machine Learning karena metode ini memiliki tingkat akurasi yang tinggi dengan perhitungan sederhana.

## 

# PEMBAGIAN TUGAS

##### Table 1. Daftar Tugas

| **No** | **Pekerjaan** | **Keterangan** |
| --- | --- | --- |
| 1 | Studi Literatur | Melakukan studi literatur terkait penelitian yang berasal dari jurnal dan *paper*. |
| 2 | Pengumpulan Data | Melakukan pengumpulan data berupa kalimat-kalimat berbahasa Indonesia dengan teknik *scraping* dari beberapa buku, jurnal dan *website*. |
| 3 | Analisis Pengolahan Data | Melakukan analisis mengenai pengolahan data yang berupa *preprocessing* yang dilakukan pada data |
| 4 | Analisis Algoritma | Melakukan analisis mengenai algoritma yang digunakan untuk melakukan normalisasi. |
| 5 | Implementasi | Melakukan implementasi. |
| 6 | Pembahasan, Kesimpulan dan Saran | Membuat penjelasan dari pengujian model yang telah dilakukan serta membuat kesimpulan akhir dari tugas akhir dan saran untuk penelitian selanjutnya. |

##### Table 2. Pembagian Tugas

| **No** | **Nama** | **Pekerjaan** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **1** | Efrica Cahyani Situmeang |  |  |  |  |  |  |  |
| **2** | Aldi Irvan Siagian |  |  |  |  |  |  |  |
| **3** | Grace Noelia Simorangkir |  |  |  |  |  |  |  |
| **4** | Anjel Riska Pardede |  |  |  |  |  |  |  |
| **5** | Grace Widya Simanjuntak |  |  |  |  |  |  |  |
| **6** | Romual Naibaho |  |  |  |  |  |  |  |

# DAFTAR REFERENSI

Al-Saaqa, S., Abdel-Nabi, H., & Awajan, A. (2018). A Survey of Textual Emotion Detection. 2018 8th International Conference on Computer Science and Information Technology, CSIT 2018, 136–142. <https://doi.org/10.1109/CSIT.2018.8486405>

Xhemali, D., Hinde C.J., And Stone R.G., 2009.” Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages”, Journal of Computer Science Issues, Vol. 4, No. 1.

Yasin, U., 2005. Keyword Extraction Using Naive Bayes. pp. 1-5.

Maulana, M. 2016. Emotion Detection of Indonesian Tweets Using Naive Bayes Classification. repository its.

Melita, R., Amrizal, V., Suseno, H. B. & Dirjan, T. 2018. Penerapan Metode Term Frequency Inverse Document Frequency. *Jurnal Teknik,* Vol. 11 No. 2.

P. Ekman, Universals and cultural differences in facial expressions of emotion, vol. 19, Nebraska: University of Nebraska Press, 1972.