**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) **1**

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

[**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) 8

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

[**ANALISIS DAN DESAIN**](#_heading=h.f7bqq3xgisjm) **12**

[Analisis Algoritma Naive Bayes](#_heading=h.a0z0jm4gcjr1) 14

[Desain](#_heading=h.x71w9e40iova) 15

[Desain Pre-Processing](#_heading=h.pmdcryji87vv) 15

[Desain TF-IDF](#_heading=h.oxuuwq3wm6gw) 15

[Desain Naive Bayes](#_heading=h.v6mo4pk70b77) 15

[**PEMBAGIAN TUGAS**](#_heading=h.edmpv1wmivtn) **16**

[**DAFTAR REFERENSI**](#_heading=h.a832xjvps9sp) **18**

# 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.

## Rumusan Masalah

Berdasarkan latar belakang masalah yang telah disampaikan sebelumnya, dapat ditarik rumusan masalah yang akan diselesaikan sebagai berikut.

1. Bagaimana mengklasifikasikan emosi berdasarkan data komentar menggunakan algoritma Naive Bayes.

## Tujuan

Tujuan penelitian dalam Tugas Pemrosesan Bahasa Alami ini adalah sebagai berikut:

1. Melakukan implementasi Naive Bayes dalam mengklasifikasikan emosi berdasarkan data komentar.

## Manfaat Penelitian

Manfaat penelitian dalam Tugas Pemrosesan Bahasa Alami ini adalah sebagai berikut:

1. Penelitian ini diharapkan mampu melakukan klasifikasi emosi berdasarkan kalimat yang diberikan dengan memberi label pada setiap kelas emosi sesuai dengan dataset yang diberikan.

## Ruang Lingkup Penelitian

Ruang Lingkup dari Tugas Pemrosesan Bahasa Alami ini adalah sebagai berikut:

1. Data yang diolah dikumpulkan dari *website* reddit.
2. Penelitian ini melakukan normalisasi hanya dalam bentuk huruf.
3. Penelitian ini menerima *inputan* dalam bentuk kalimat yang mempunyai kata singkatan bahasa Inggris.
4. Penelitian ini akan melakukan normalisasi kata singkatan yang di-*input* oleh *user*.

# TINJAUAN PUSTAKA

Pada bab ini diuraikan dasar teori yang didapatkan dari pustaka yang relevan dengan kajian Penelitian.

## Emosi

Emosi manusia adalah suatu keadaan jiwa kompleks yang terdiri dari tiga komponen berbeda yaitu pengalaman subjektif, respon psikologis dan respon ekspresif. Untuk lebih memahami emosi, peneliti melakukan identifikasi dan klasifikasi pada beberapa jenis emosi. Pada 1972, psikologis Paul Ekman menyatakan bahwa terdapat enam emosi dasar yang terdapat pada budaya manusia secara universal. Ekman melakukan penelitian di seluruh pelosok dunia dan menemukan enam emosi yang sama melalui ekspresi wajah. Enam emosi dasar adalah bahagia, sedih, terkejut, marah, jijik dan takut. Ekman menemukan bahwa enam ekspresi ini bersifat universal kepada seluruh manusia di dunia.

## Natural Language Processing

*Natural Language Processing* (NLP) atau pemrosesan bahasa alami merupakan bidang dari ilmu *computer science.* NLP adalah kajian tentang struktur yang berkaitan dengan memberi komputer untuk memahami teks dan kata-kata yang diucapkan dengan cara yang sama seperti yang dapat dilakukan manusia. NLP menggabungkan linguistik komputasi pemodelan bahasa manusia berbasis aturan dengan model statistik, pembelajaran mesin, dan pembelajaran mendalam. NLP menggerakkan program komputer yang menerjemahkan teks dari satu bahasa ke bahasa lain, menanggapi perintah lisan, dan meringkas teks dalam jumlah besar dengan cepat bahkan secara real time.

## Data Preprocessing

*Data preprocessing* adalah suatu proses yang digunakan untuk menghilangkan kata-kata atau teks yang tidak diperlukan dalam komputasi. Inti dari data *preprocessing* adalah mengubah data yang tidak terstruktur sesuai dengan kebutuhan dan proses *mining* yang dilakukan. Tahapan tahapan yang dilakukan dalam preprocessing ini ada 5 yaitu *text cleaning, case* *folding*, *tokenization, stopword removal, stemming.*

1. *Text cleaning*

Proses *text cleaning* (pembersihan teks) merupakan suatu proses dalam membersihkan data yang dimiliki dari suatu komponen yang tidak diperlukan dalam proses dan tahapan selanjutnya, seperti tanda baca, data tweet yang sama (retweet), link, dan sebagainya. Sehingga, jika telah dilakukan proses penghapusan link, tanda baca (punctuation), dan tweet yang berulang (retweet), maka akan semakin memudahkan peneliti dalam proses pembersihan data selanjutnya dan juga menambah keakuratan dalam melakukan proses pengklasifikasian sentimen pada tahapan berikutnya.

1. *Case Folding*

*Case folding* merupakan mengubah keseluruhan huruf menjadi huruf kecil (*lower* *case*). Hal ini dilakukan karena tidak semua teks konsisten dalam penggunaan huruf.

1. *Tokenizing*

*Tokenizing* adalah suatu tahap untuk memenggal setiap kata yang ada dalam kalimat. Proses ini dilakukan untuk menjadikan sebuah kalimat tersebut menjadi lebih bermakna. Dan suatu kata, tanda baca, angka, simbol, dan entitas penting lain dapat disebut sebagai token. Untuk proses ini diperlukan untuk mengenali setiap kata maupun huruf yang ada pada korpus normalisasi yang diperlukan untuk proses spelling normalization.

1. *Stopword removal*

Stopwords Removal merupakan suatu proses yang bertujuan untuk mengambil kata-kata penting dari tokenisasi. Stopwords merupakan kata-kata yang tidak mengandung sentimen atau bukan merupakan ciri dari sebuah sentimen. Misalnya ‘di’, ‘ke’, ‘dari’, ‘oleh’, dan sebagainya. Stoplist berisi kumpulan kata yang kurang relevan namun sering muncul pada sebuah dokumen atau tweet. Stoplist berisi sekumpulan stopwords.

1. *Stemming*

Proses stemming merupakan tahapan yang mengembalikan kata-kata ke bentuk dasarnya dengan cara menghapus semua awalan (prefix), imbuhan (affix) dan akhiran (suffix). Tahapan stemming digunakan agar memperkecil banyaknya indeks yang beda dari suatu dokumen dan mengelompokkan kata yang mempunyai arti yang sama untuk bentuk kata yang beda karena adanya imbuhan. Dan proses stemming diperlukan untuk kebutuhan pada proses selanjutnya yaitu proses penentuan sentimen dengan menggunakan kamus lexicon yang menggunakan kata-kata dasar.

## TF-IDF

Metode TF-IDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada *information retrieval*. Dimana *information retrieval* adalah bagian dari *computer science* yang berhubungan dengan pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri (Melita, et al., 2018). Metode *Term Frequency Inverse Document Frequency (TF­IDF)* adalah cara pemberian bobot hubungan suatu kata (*term*) terhadap dokumen. TF­IDF ini adalah sebuah ukuran statistik yang digunakan untuk mengevaluasi seberapa penting sebuah kata di dalam sebuah dokumen atau dalam sekelompok kata. Untuk dokumen tunggal tiap kalimat dianggap sebagai dokumen. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Bobot kata semakin besar jika sering muncul dalam suatu dokumen dan semakin kecil jika muncul dalam banyak dokumen. Pada TF-IDF digunakan rumus untuk menghitung bobot (W) masing-masing dokumen yang berikan:

dimana:

Wdt : bobot dokumen ke­d terhadap kata ke­t

tfdt : banyaknya kata yang dicari pada sebuah dokumen

Idft : Inversed Document Frequency (log (N/df))

N : total dokumen

df : banyak dokumen yang mengandung kata yang dicari.

## Klasifikasi

Klasifikasi adalah masalah pemodelan prediktif yang melibatkan keluaran label kelas yang diberikan beberapa masukan. Tugas klasifikasi umumnya adalah memprediksi satu label atau target kelas. Namun klasifikasi juga mungkin melibatkan prediksi terhadap dua atau lebih label kelas. Dua atau lebih label kelas dalam kasus klasifikasi disebut sebagai Multi-Label Classification. Adapun penjelasan terkait Multi-Label Classification disampaikan dalam subbab berikut ini.

### Multi-Label Classification

Multi-Label Classification adalah varian dari masalah klasifikasi dimana beberapa label atau target kelas dapat ditetapkan untuk setiap *instance*. Dalam multi-label classification, nol atau lebih label diperlukan sebagai output untuk setiap sampel input, dan output diperlukan secara bersamaan. Multi-Label Classification memerlukan algoritma pembelajaran mesin untuk mendukung prediksi beberapa kelas atau label. Dalam Multi-Label Classification, *training set* terdiri dari instance yang masing-masing terkait dengan satu set label, dan tugasnya adalah memprediksi set label dari *instance* yang tidak terlihat melalui analisis *instance training* dengan set label yang diketahui.

Multi-Label Classification dapat diterapkan pada kasus-kasus tertentu seperti penentuan *genre* dari sebuah film, penentuan jenis penyakit berdasarkan gejala tertentu, penentuan emosi berdasarkan text, dan lain sebagainya. Dalam penelitian ini Multi-Label Classification digunakan untuk mengklasifikasikan emosi dalam sebuah kalimat atau komentar dari sebuah media sosial.

### Naive Bayes Classification

Naïve Bayes adalah pendekatan statistic fundamental yang menggunakan probabilitas dan cost (Yasin, 2005). Bayes rule adalah suatu aturan dalam menentukan nilai probabilitas dari suatu kondisi tertentu menjadi sebuah kondisi yang didefinisikan. Kaitan antara Naïve Bayes dengan metode klasifikasi adalah bahwa hipotesis dalam teorema Bayes merupakan label kelas yang menjadi target mapping dalam melakukan klasifikasi, dan bukti klasifikasi merupakan fitur yang menjadi masukan dalam model klasifikasi. Selain dalam klasifikasi, metode ini dapat juga digunakan dalam mengekstraksi kata kunci pada dokumen, yang mampu memberikan hasil yang baik berupa fitur yang digunakan pada training set, di antaranya TF-IDF, dan posisi kata kunci pada dokumen terkait. Metode ini telah digunakan di berbagai penelitian dan sangat populer di bidangnya.

Di beberapa penelitian, dilakukan perbandingan antara metode-metode klasifikasi lainnya seperti pada penelitian yang dilakukan Syahmia Gusriani, yang membandingkan metode K-NN, Naïve Bayes, dan Decision Tree, untuk kasus Analisis Sentimen dan menghasilkan sebuah kesimpulan yaitu Naïve Bayes memiliki tingkat akurasi yang paling tinggi diantara dua metode lainnya. Tiga metode ini digunakan Syahmia karena metode-metode ini merupakan metode yang paling populer saat ini. Selain itu, pada penelitian yang dilakukan Xhemali, Daniela, Chris J. Hinde, dan Roger G. Stone yaitu meneliti perbandingan tingkat akurasi Naïve Bayes, Decision Tree, Neural Networks dalam klasifikasi data training web pages, dan menghasilkan kesimpulan bahwa Naïve Bayes Classifier merupakan metode yang mampu memberikan tingkat akurasi yang lebih baik dibanding metode classifier lainnya.

Persamaan teorema Bayes:

Keterangan:

*B* : Data class yang belum diketahui

*A* : Hipotesis data class spesifik

*P(A|B)* : Probabilitas hipotesis A berdasar kondisi B (posterior probability)

*P(A)* : Probabilitas hipótesis A (prior probability)

*P(B|A)* : Probabilitas Bberdasarkan kondisi pada hipotesis A

*P(B)* : Probabilitas B

Berikut merupakan model dari Teorema Naïve Bayes:

Persamaan tersebut merupakan model dari Teorema Naïve Bayes, yang akan digunakan dalam proses klasifikasi dengan data kontinu dengan rumus Densitas Gauss:

Keterangan:

*P* : Peluang

*Xi* : Aribut ke i

*xi* : Nilai atribut ke i

*Y* : Kelas yang di cari

*yj* : Sub kelas Y yang dicari

*u* : Mean (rata-rata dari seluruh atribut)

*o* : Deviasi standar (varian dari seluruh atribut)

Rumus untuk menghitung Mean:

Rumus untuk menghitung Deviasi Standar

# ANALISIS DAN DESAIN

Pada bab ini dijelaskan mengenai analisis dan desain penelitian yang menjadi acuan pada tahapan implementasi selama pelaksanaan Proyek Akhir PBA.

* 1. **Analisis**

Bagian ini akan menjelaskan proses pengolahan data sampai dengan proses pengklasifikasian emosi. Proses ini dilakukan secara manual menggunakan metode yang sebelumnya sudah ditentukan. Proses pengerjaan terdiri dari analisis data, pengolahan data (*cleaning data, stopword removal, case folding, tokenizing* dan *stemming*), metode *feature selection* (TF-IDF) dan terakhir proses klasifikasi menggunakan algoritma *Naïve Bayes*.

* + 1. **Analisis Data**

Proyek mata kuliah PBA ini menggunakan dataset yang bersumber dari tautan berikut: https://huggingface.co/datasets/go\_emotions. Dataset GoEmotions ini berisikan kumpulan data beranotasi manual terbesar yaitu terdiri dari 58.000 Reddit English comments, yang dikategorikan kedalam 27 label kategori emosi dan sebuah label netral. Data mentah disertakan serta versi dataset yang lebih kecil dan disederhanakan dengan pemisahan train/val/test yang telah ditentukan sebelumnya. Dataset GoEmotions ditujukan untuk keperluan pendalaman multi-class, multi-label emotion classification.

Dalam data mentah, label terdaftar sebagai kolomnya sendiri dengan entri biner 0/1 daripada daftar id seperti pada data yang disederhanakan.Selanjutnya pada proyek ini data akan dibagi menjadi data test dan data train.

* + 1. **Preprocessing Data**

Data preprocessing merupakan proses pengolahan data, sehingga data yang hendak di klasifikasi merupakan data yang baik. Pada penelitian ini data yang digunakan untuk preprocessing adalah data dari reddit, untuk data tahapan preprocessing yang dilakukan adalah *cleaning data, stopword removal, case folding, tokenizing* dan *stemming*. Berikut ini merupakan tahapan yang dilakukan pada proses preprocessing data.

1. Cleaning Data

Cleaning adalah proses yang diperlukan karena dapat menghasilkan data yang bersih dan mampu mengurangi noise secara efektif. Karakter-karakter yang dihapus pada tahapan cleaning merupakan karakter karakter yang tidak penting, sehingga harus dibuang karena dapat mengganggu kinerja sistem dalam melakukan ekstraksi kata kunci.

1. Stopword Removal

Pada tahap ini, semua kata-kata yang dianggap kurang memiliki makna yang penting dalam sebuah kalimat akan dihilangkan. Stop word merupakan kata-kata yang tidak penting yang tidak memiliki potensi sebagai kata kunci, seperti ialah, adalah, yaitu, ini, itu, dan sebagainya.

1. Case Folding

*Case Folding* merupakan proses yang dilakukan untuk menyamaratakan seluruh karakter pada teks menjadi huruf kecil. Pada tahap *case folding* semua huruf kapital, atau huruf besar akan diubah menjadi huruf kecil. Proses ini bertujuan untuk menyeragamkan data agar mudah di proses.

1. Tokenizing

Pada tahap *tokenizing*, data tweet yang telah melalui tahap *preprocessing non alphabet removal* dan *case folding* akan di pecah menjadi perkata.

1. Stemming

Proses stemming merupakan tahapan yang mengembalikan kata-kata ke bentuk dasarnya dengan cara menghapus semua awalan (prefix), imbuhan (affix) dan akhiran (suffix). Tahapan stemming digunakan agar memperkecil banyaknya indeks yang beda dari suatu dokumen dan mengelompokkan kata yang mempunyai arti yang sama untuk bentuk kata yang beda karena adanya imbuhan.

* + 1. **TF-IDF**

Metode TF-IDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada information retrieval (Ridwan, et al., 2019). Metode TF-IDF juga menggabungkan dua konsep penghitungan bobot, yaitu frekuensi kata dalam dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut. TF-IDF ini juga adalah sebuah ukuran statistik yang digunakan untuk mengevaluasi seberapa penting sebuah kata didalam sebuah dokumen atau dalam sekelompok kata. Frekuensi kata dalam dokumen tertentu menunjukkan pentingnya kata tersebut dalam dokumen. Berapa kali sebuah dokumen berisi kata ini menunjukkan seberapa umum kata itu. Oleh karena itu, jika frekuensi kata lebih tinggi dalam dokumen, dan frekuensi seluruh dokumen yang berisi kata lebih rendah dalam dokumen, bobot hubungan antara kata dan dokumen akan tinggi.

## Analisis Algoritma Naive Bayes

Pada metode pengklasifikasian mengenai *emotion* yang peneliti lakukan yaitu menggunakan metode *Naïve Bayes Classifier.* Pengklasifikasian ini dilakukan setelah melalui proses dan tahapan yang telah dilakukan sebelumnya.

## Desain

## Desain Pre-Processing

## Desain TF-IDF

## Desain Naive Bayes

# 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.