

**ANALISIS SENTIMEN KURIKULUM MERDEKA
MENGUNAKAN KLASIFIKASI *MULTINOMIAL
NAÏVE BAYES* DAN *BERNOULLI NAÏVE BAYES*
DENGAN SELEKSI FITUR *QUERY EXPANSION
RANKING***

SKRIPSI



MUHAMMAD YUSRAN

H051191078

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
FEBRUARI 2023**

**ANALISIS SENTIMEN KURIKULUM MERDEKA
MENGUNAKAN KLASIFIKASI *MULTINOMIAL
NAÏVE BAYES* DAN *BERNOULLI NAÏVE BAYES*
DENGAN SELEKSI FITUR *QUERY EXPANSION
RANKING***

SKRIPSI

**Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Sains pada
Program Studi Statistika Departemen Statistika Fakultas Matematika dan
Ilmu Pengetahuan Alam Universitas Hasanuddin**

MUHAMMAD YUSRAN

H051191078

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
FEBRUARI 2023**

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Analisis Sentimen Kurikulum Merdeka Menggunakan Klasifikasi *Multinomial Naïve Bayes* dan *Bernoulli Naïve Bayes* dengan Seleksi Fitur *Query Expansion Ranking*

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 22 Februari 2023



Muhammad Yusran

NIM H051191078

**ANALISIS SENTIMEN KURIKULUM MERDEKA
MENGUNAKAN KLASIFIKASI *MULTINOMIAL NAÏVE*
BAYES DAN *BERNOULLI NAÏVE BAYES* DENGAN SELEKSI
FITUR *QUERY EXPANSION RANKING***

Disetujui oleh:

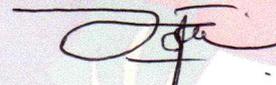
Pembimbing Utama



Siswanto, S.Si., M.Si.

NIP. 19920107 201903 1 012

Pembimbing Pertama



Dr. Anna Islamiyati, S.Si., M.Si.

NIP. 19770808 200501 2 002

Ketua Program Studi



Dr. Nurfitri Sunusi, S.Si., M.Si

NIP. 19720117 199703 2 002

Pada 22 Februari 2023

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Muhammad Yusran
NIM : H051191078
Program Studi : Statistika
Judul Skripsi : Analisis Sentimen Kurikulum Merdeka Menggunakan
Klasifikasi *Multinomial Naïve Bayes* dan *Bernoulli Naïve Bayes* dengan Seleksi Fitur *Query Expansion Ranking*

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Siswanto, S.Si., M.Si. (.....)
2. Sekretaris : Dr. Anna Islamiyati, S.Si., M.Si. (.....)
3. Anggota : Dr. Nurtiti Sunusi, S.Si., M.Si. (.....)
4. Anggota : Dr. Erna Tri Herdiani, S.Si., M.Si. (.....)

Ditetapkan di : Makassar

Tanggal : 22 Februari 2022

KATA PENGANTAR

Assalamu 'alaikum Warahmatullahi Wabaraktuh

Puji syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, hidayah, dan karunia-Nya yang telah diberikan kepada penulis sampai saat ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa Sallam* beserta para keluarga, sahabat dan pengikut beliau hingga akhir zaman. *Alhamdulillahirobbil'alamin*, berkat nikmat kemudahan dan pertolongan yang diberikan oleh Allah *Subhanahu Wa Ta'ala*, penulis dapat menyelesaikan skripsi yang berjudul “**Analisis Sentimen Kurikulum Merdeka Menggunakan Klasifikasi *Multinomial Naïve Bayes* dan *Bernoulli Naïve Bayes* dengan Seleksi Fitur *Query Expansion Ranking*” yang disusun sebagai salah satu syarat akademik untuk memperoleh gelar sarjana pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.**

Penulis menyadari bahwa dalam penyelesaian skripsi ini penulis telah melewati perjuangan panjang dan pengorbanan yang tidak sedikit. Namun berkat rahmat dan izin-Nya serta dukungan dari berbagai pihak yang turut membantu dalam bentuk moril maupun material sehingga dengan segala keterbatasan kemampuan dan pengetahuan, penulis dapat menyelesaikan skripsi ini. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya kepada Ayahanda **H. Abdullah** dan Ibunda **Hj. Maryam** yang telah membesarkan dan mendidik penulis, memberikan dukungan penuh, pengorbanan luar biasa, limpahan cinta dan kasih sayang, kesabaran hati, dan dengan ikhlas telah menemani setiap langkah penulis dengan doa dan restu mulianya. Ucapan terima kasih juga penulis haturkan kepada Kakak-kakak tersayang penulis **Marwah Asriati, Darma Ekawati, Muhammad Taufiq, Fausia Ferawati** dan **Ulfa Nur Ramadhani** yang senantiasa memberikan bantuan dan semangat kepada penulis dalam menyelesaikan masa studi penulis, serta keluarga besar penulis, terima kasih atas doa mulia dan dukungannya selama ini.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan dan ketulusan juga penulis ucapkan kepada:

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Nurtiti Sunusi S.Si., M.Si.**, selaku Ketua Departemen Statistika, sekaligus Penasihat Akademik penulis dan Tim Penguji yang dengan penuh kesabaran telah memberikan arahan, dorongan semangat dan motivasi kepada penulis selama menempuh pendidikan di Departemen Statistika.
4. **Bapak Siswanto, S.Si., M.Si.**, selaku Pembimbing Utama dan **Ibu Dr. Anna Islamiyati, S.Si., M.Si.** selaku Pembimbing Pertama yang dengan penuh kesabaran telah ikhlas meluangkan waktu dan pemikirannya di tengah berbagai kesibukannya untuk memberikan arahan, dorongan, dan motivasi kepada penulis mulai dari awal pengerjaan hingga selesainya penulisan tugas akhir ini.
5. **Ibu Dr. Erna Tri Herdiani, S.Si., M.Si.** selaku Tim Penguji yang telah meluangkan waktu dalam memberikan motivasi serta kritikan yang membangun kepada penulis dalam penyempurnaan tugas akhir ini.
6. Segenap Dosen Pengajar dan Staf yang telah memberikan ilmu dan kemudahan kepada penulis dalam berbagai hal selama menempuh pendidikan di Departemen Statistika.
7. Sahabat terbaik penulis sejak di bangku SMA, **Fauzi Asham Nasrul, M. Yusril Izah Mahendra Rastach, Taufik Kemal Thaha, Andi Muh. Arul Efansyah, Habib Yardan Wafa, Alief Fikri Nurham, Ahmad Mubaraq, Muwaffaq Ahnaf Shiddiq dan Muh. Nurfaizy**, yang senantiasa memberikan semangat dan menemani perjuangan pendidikan penulis.
8. Sahabat-sahabat terbaik semasa perkuliahan penulis, **Ummul Auliyah Syam, Alya Safira Irtiqa Miolo, Muh. Nur Iskandar Zulkarnain, Alfiyah Salsa Dila Sabir, Muhammad Rayhan Rifaldi, A. Ahmad Qeis Tenridapi, Wahyu Dwi Rahmawati, Fadhila Febriyanti Najamuddin, Diah Lestari, Vinaya Rifqi Anandari, Fatiyah Utami dan Mutmagfira**. Terima kasih atas kebersamaan, kebahagiaan, serta kebaikannya menjadi sosok guru dan keluarga bagi penulis. Terima kasih telah mengukir kenangan indah bersama penulis selama masa perkuliahan.

9. Kakak-kakak di Laboratorium Statistika. Terima kasih atas bimbingan, dorongan semangat, dan nasehatnya kepada penulis selama masa akhir perkuliahan penulis. Terima kasih untuk segala canda dan tawa yang menghibur penulis selama berada di lab.
10. Teman-teman **Statistika 2019**, terkhusus kepada **Agus Hermawan** selaku Ketua Angkatan 2019. Terima kasih atas ilmu, kebersamaan, suka dan duka selama menjalani perkuliahan di Departemen Statistika. Terima kasih telah menerima kehadiran penulis.
11. Teman-teman **Beasiswa Bakti BCA 2021-2022**, terkhusus kepada **Kelompok Tangguh**. Terima kasih telah memberikan pengalaman berharga dan wadah pengembangan diri bagi penulis.
12. **Keluarga Besar Ambalan Garuda Pancasila, Pasukan Rimbas, Kwartir Ranting Gerakan Pramuka Turikale dan Kwartir Cabang Gerakan Pramuka Maros**. Terima kasih telah menjadi rumah kedua bagi penulis untuk pulang dan memberikan kesempatan kepada penulis untuk mendapatkan pengetahuan dan pengalaman yang sangat berkesan.
13. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu. Terima kasih setinggi-tingginya untuk segala dukungan, partisipasi, dan apresiasi yang diberikan kepada penulis.

Penulis menyadari bahwa masih banyak kekurangan dalam penyusunan skripsi ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini dapat memberikan manfaat untuk berbagai pihak.

Makassar, 22 Februari 2023



Muhammad Yusran

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIK**

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Yusran
NIM : H051191078
Program Studi : Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty- Free Right*)** atas tugas akhir saya yang berjudul:

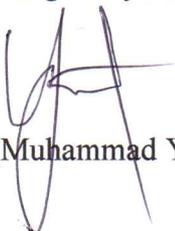
**“Analisis Sentimen Kurikulum Merdeka Menggunakan Klasifikasi
Multinomial Naïve Bayes dan *Bernoulli Naïve Bayes* dengan Seleksi Fitur
Query Expansion Ranking”**

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar tanggal 22 Februari 2023.

Yang menyatakan,


(Muhammad Yusran)

ABSTRAK

Analisis sentimen merupakan analisis data teks yang mengklasifikasikan data ke dalam sentimen positif dan negatif. Salah satu metode klasifikasi sentimen adalah *naïve bayes*, yaitu metode klasifikasi yang didasarkan pada teorema bayes. Tingginya dimensi fitur dalam proses klasifikasi menjadi masalah dalam analisis sentimen. Salah satu cara untuk mengatasi masalah tersebut adalah dengan menggunakan seleksi fitur. Penelitian ini bertujuan untuk memperoleh hasil klasifikasi sentimen terkait kurikulum merdeka pada *twitter* dan perbandingan kinerja klasifikasi dua metode klasifikasi pada pengklasifikasian sentimen terkait kurikulum merdeka. Metode klasifikasi yang digunakan adalah variasi dari *naïve bayes*, yaitu *multinomial naïve* dan *bernoulli naïve bayes* dengan *query expansion ranking* sebagai seleksi fitur. Data pada penelitian ini adalah 15000 *tweets* dengan kata kunci “Kurikulum Merdeka”. Klasifikasi sentimen menggunakan *multinomial naïve bayes* terdiri dari 106 *tweet* sentimen positif dan 164 *tweet* sentimen negatif dengan *accuracy* 98.88889%, *recall* 98.13084%, *precision* 99.05660% dan *f-measure* 98.59155%, sedangkan menggunakan *bernoulli naïve bayes* terdiri dari 95 *tweet* sentimen positif dan 175 *tweet* sentimen negatif dengan *accuracy* 94.81481%, *recall* 87.85047%, *precision* 98.94739% dan *f-measure* 93.06931%. Berdasarkan hal tersebut, diperoleh bahwa *multinomial naïve bayes* lebih baik dibandingkan *bernoulli naïve bayes* dalam mengklasifikasikan sentimen kurikulum merdeka.

Kata Kunci: Analisis Sentimen, *Multinomial Naïve Bayes*, *Bernoulli Naïve Bayes*, *Query Expansion Ranking*, Kurikulum Merdeka.

ABSTRACT

Sentiment analysis is an analysis of text data that classifies data into positive and negative sentiments. One of the sentiment classification methods is naïve bayes, which is a classification method based on bayes theorem. High dimension of features in the classification process becomes a problem in sentiment analysis. One way to overcome this problem is to use feature selection. This research aims to obtain the results of the classification of sentiments related to the kurikulum merdeka on Twitter and compare the classification performance of the two classification methods on the classification of sentiments of kurikulum merdeka. The classification method used is variation of naïve bayes, namely multinomial naïve and bernoulli naïve bayes with query expansion ranking as feature selection. Data in this research are 15,000 tweets with the keyword "Kurikulum Merdeka". Sentiment classification uses multinomial naïve bayes consisting of 106 tweets of positive sentiment and 164 tweets of negative sentiment with accuracy 98.88889%, recall 98.13084%, precision 99.05660% and f-measure 98.59155%, while using bernoulli naïve bayes consisting of 95 positive sentiment tweets and 175 negative sentiment tweets with 94.81481% accuracy, 87.85047% recall, 98.94739% precision and 93.06931% f-measure. Based on this, it was found that multinomial naïve bayes was better than bernoulli naïve bayes in classifying kurikulum merdeka sentiment.

Keywords: *Sentiment Analysis, Multinomial Naïve Bayes, Bernoulli Naïve Bayes, Query Expansion Ranking, Kurikulum Merdeka.*

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
LEMBAR PERNYATAAN KEOTENTIKAN	iii
HALAMAN PERSETUJUAN PEMBIMBING	iv
HALAMAN PENGESAHAN	v
KATA PENGANTAR	vi
ABSTRAK	x
ABSTRACT	xi
DAFTAR ISI	xii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvi
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA	5
2.1 Analisis Sentimen.....	5
2.2 Praproses Data Teks	5
2.3 Klasifikasi <i>Naïve Bayes</i>	6
2.3.1 Klasifikasi <i>Multinomial Naïve Bayes</i>	7
2.3.2 Klasifikasi <i>Bernoulli Naïve Bayes</i>	9
2.4 Seleksi Fitur <i>Query Expansion Ranking</i>	10
2.5 <i>Confusion Matrix</i>	11
2.6 <i>Text Visualization</i>	13
2.7 Kurikulum Merdeka	13
2.8 <i>Twitter</i>	14
BAB III METODOLOGI PENELITIAN	15

3.1 Sumber Data	15
3.2 Struktur Data	15
3.3 Tahapan Analisis	15
BAB IV HASIL DAN PEMBAHASAN.....	18
4.1 Deskripsi Data	18
4.2 Praproses Data Teks	19
4.2.1 <i>Data Cleaning</i>	20
4.2.2 <i>Case Folding</i>	26
4.2.3 <i>Spelling Normalization</i>	27
4.2.4 <i>Stemming</i>	28
4.2.5 <i>Stopword Removal</i>	29
4.2.6 <i>Tokenizing</i>	30
4.3 <i>Text Visualization</i>	32
4.3.1 <i>Word Cloud</i> untuk Keseluruhan Data <i>Tweets</i>	32
4.3.2 <i>Word Cloud</i> untuk Data <i>Tweets</i> dengan Sentimen Positif.....	33
4.3.3 <i>Word Cloud</i> untuk Data <i>Tweets</i> dengan Sentimen Negatif	33
4.4 Seleksi Fitur dengan <i>Query Expansion Ranking</i>	34
4.5 Pembagian Data Menjadi Data Latih dan Data Uji.....	35
4.6 Klasifikasi Sentimen Menggunakan <i>Naïve Bayes</i>	36
4.6.1 Klasifikasi Sentimen Menggunakan <i>Multinomial Naïve Bayes</i>	36
4.6.2 Klasifikasi Sentimen Menggunakan <i>Bernoulli Naïve Bayes</i>	37
4.7 Pengklasifikasian Data Uji	39
4.7.1 Pengklasifikasian Data Uji Menggunakan <i>Multinomial Naïve Bayes</i>	39
4.7.2 Pengklasifikasian Data Uji Menggunakan <i>Bernoulli Naïve Bayes</i>	42
4.8 Evaluasi Kinerja Klasifikasi.....	45
4.8.1 Evaluasi Kinerja Klasifikasi <i>Multinomial Naïve Bayes</i>	45
4.8.2 Evaluasi Kinerja Klasifikasi <i>Bernoulli Naïve Bayes</i>	47
4.9 Perbandingan Kinerja Klasifikasi <i>Multinomial Naïve Bayes</i> dengan <i>Bernoulli Naïve Bayes</i>	48
BAB V KESIMPULAN DAN SARAN	50
5.1 Kesimpulan.....	50
5.2 Saran.....	50

DAFTAR PUSTAKA.....	51
LAMPIRAN.....	57

DAFTAR GAMBAR

Gambar 4.1 Diagram Batang Kelas Sentimen Data *Tweets* Kurikulum Merdeka . 19

Gambar 4.2 *Word Cloud* Keseluruhan Data 32

Gambar 4.3 *Word Cloud* Sentimen Positif 33

Gambar 4.4 *Word Cloud* Sentimen Negatif 33

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	12
Tabel 3.1 Struktur Data Penelitian	15
Tabel 4.1 Hasil Penambangan Data.....	18
Tabel 4.2 Struktur Data <i>Tweets</i> Sebelum Praproses Data Teks	19
Tabel 4.3 Struktur Data <i>Tweets</i> Sebelum dan Setelah Penghapusan <i>Username</i>	20
Tabel 4.4 Struktur Data <i>Tweets</i> Sebelum dan Setelah Penghapusan URL.....	21
Tabel 4.5 Struktur Data <i>Tweets</i> Sebelum dan Setelah Penghapusan <i>Hashtag</i>	23
Tabel 4.6 Struktur Data <i>Tweets</i> Sebelum dan Setelah Penghapusan <i>Emoticon</i>	24
Tabel 4.7 Struktur Data <i>Tweets</i> Sebelum dan Setelah Penghapusan Nomor	25
Tabel 4.8 Struktur Data <i>Tweets</i> Sebelum dan Setelah Penghapusan Tanda Baca...25	
Tabel 4.9 Struktur Data <i>Tweets</i> Sebelum dan Setelah <i>Case Folding</i>	26
Tabel 4.10 Struktur Data <i>Tweets</i> Sebelum dan Setelah <i>Spelling Normalization</i>	27
Tabel 4.11 Struktur Data <i>Tweets</i> Sebelum dan Setelah <i>Stemming</i>	28
Tabel 4.12 Struktur Data <i>Tweets</i> Sebelum dan Setelah <i>Stopword Removal</i>	29
Tabel 4.13 Struktur Data <i>Tweets</i> Sebelum dan Setelah <i>Tokenizing</i>	30
Tabel 4.14 Struktur Data <i>Tweets</i> Sebelum dan Setelah Praproses Data Teks.....	31
Tabel 4.15 Hasil Perhitungan Nilai Skor QER.....	35
Tabel 4.16 Proporsi Data Latih dan Data Uji	36
Tabel 4.17 Hasil Perhitungan Peluang Tiap Fitur <i>Multinomial Naïve Bayes</i>	37
Tabel 4.18 Hasil Perhitungan Peluang Tiap Fitur <i>Bernoulli Naïve Bayes</i>	39
Tabel 4.19 Peluang Fitur-fitur Data Uji <i>Multinomial Naïve Bayes</i>	41
Tabel 4.20 Kelas Prediksi Data Uji Menggunakan <i>Multinomial Naïve Bayes</i>	42
Tabel 4.21 Peluang Fitur-fitur Data Uji <i>Bernoulli Naïve Bayes</i>	44
Tabel 4.22 Kelas Prediksi Data Uji Menggunakan <i>Bernoulli Naïve Bayes</i>	45
Tabel 4.23 <i>Confusion Matrix Multinomial Naïve Bayes</i>	46
Tabel 4.24 <i>Confusion Matrix Bernoulli Naïve Bayes</i>	47
Tabel 4.25 Perbandingan Kinerja Klasifikasi.....	49

DAFTAR LAMPIRAN

Lampiran 1 Hasil Perhitungan Nilai Skor QER	58
Lampiran 2 Hasil Perhitungan Peluang Tiap Fitur <i>Multinomial Naïve Bayes</i>	59
Lampiran 3 Hasil Perhitungan Peluang Tiap Fitur <i>Bernoulli Naïve Bayes</i>	60
Lampiran 4 Kelas Prediksi Data Uji Menggunakan <i>Multinomial Naïve Bayes</i>	61
Lampiran 5 Kelas Prediksi Data Uji Menggunakan <i>Bernoulli Naïve Bayes</i>	62

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis sentimen adalah suatu bidang studi untuk menganalisis opini orang terhadap entitas seperti produk, layanan, instansi, individu, masalah, peristiwa, dan topik. Tugas dasar dalam analisis sentimen adalah mengolah berbagai macam tanggapan yang telah diberikan oleh masyarakat atau para pakar melalui berbagai media yang ada untuk mendapatkan informasi sentimen yang terkandung dalam suatu opini (Salam dkk., 2018). Analisis sentimen ini berfokus pada opini seseorang yang mengekspresikan sentimen positif atau negatif (Septian dkk., 2019). Oleh karena itu, dalam analisis sentimen diperlukan adanya klasifikasi yang digunakan untuk mengelompokkan tanggapan ke dalam sentimen positif atau sentimen negatif (Pamungkas dan Kharisudin, 2021). Beberapa metode yang umum digunakan pada proses klasifikasi dalam analisis sentimen antara lain *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* (Fithriasari dkk., 2020).

Metode *naïve bayes* merupakan metode pengklasifikasian dengan bentuk model peluang dan statistik yang disederhanakan dengan berdasar pada teorema bayes dengan asumsi bahwa setiap fitur bersifat bebas (*independent*). Fikri dkk. (2020) telah melakukan penelitian terkait perbandingan metode *naïve bayes* dan SVM pada analisis sentimen *twitter*. Hasil penelitian yang dilakukan menunjukkan bahwa akurasi *naïve bayes* sebesar 73.65% lebih baik dibandingkan SVM yang memiliki akurasi sebesar 70.20%. Penelitian terkait lainnya juga telah dilakukan oleh Surohman dkk., (2020) mengenai analisis sentimen terhadap *review fintech* dengan metode *naïve bayes* yang menghasilkan nilai akurasi sebesar 84.76% dan metode KNN yang menghasilkan akurasi sebesar 82,92%. Berdasarkan nilai akurasi tersebut dapat bahwa metode *naïve bayes* lebih baik daripada metode KNN dalam mengklasifikasikan sentimen.

Beberapa klasifikasi *naïve bayes* yang digunakan dalam analisis sentimen adalah *Multinomial Naïve Bayes* dan *Bernoulli Naïve Bayes*. *Multinomial naïve bayes* merupakan metode yang memperhatikan jumlah kemunculan kata dalam dokumen, sedangkan *bernoulli naïve bayes* merupakan metode yang hanya memperhatikan kemunculan kata pada dokumen dan mengabaikan jumlah

kemunculannya (Manning dkk., 2009). Metode *multinomial naïve bayes* digunakan oleh Hamzah (2021) pada penelitiannya dengan akurasi sebesar 87.74% dan metode *bernoulli naïve bayes* digunakan pada penelitian yang dilakukan oleh Pavitha dkk. (2022) dengan akurasi sebesar 87.5%. Penelitian terkait lainnya dilakukan oleh Wardani dkk. (2020) mengenai analisis sentimen pemindahan ibu kota negara. Klasifikasi menggunakan *bernoulli naïve bayes* menghasilkan performa dengan tingkat sensitifitas 90.19%, sementara *multinomial naïve bayes* menghasilkan performa dengan tingkat sensitifitas 93.45%. Sehingga, baik metode *bernoulli* maupun *multinomial* memiliki performa yang baik dalam melakukan klasifikasi sentimen.

Analisis sentimen membagi dua kelas yaitu positif dan negatif yang dilakukan dengan menghitung nilai setiap fiturnya. Hal tersebut mengakibatkan tingginya dimensi fitur dalam proses klasifikasi menjadi masalah dalam analisis sentimen karena menyebabkan pengklasifikasian menjadi tidak efisien (Fanissa dkk., 2018). Berdasarkan hal tersebut, maka seleksi fitur dibutuhkan untuk mengurangi jumlah fitur dalam proses klasifikasi. Selain itu, seleksi fitur juga dapat meningkatkan akurasi klasifikasi dari metode yang digunakan (Mongkareng dkk., 2017). Salah satu metode seleksi fitur dalam analisis sentimen adalah *Query Expansion Ranking* (QER). QER diajukan oleh Parlar dkk. (2018) sebagai sebuah seleksi fitur baru untuk menilai kata-kata yang diperlukan dan memberikan bobot pada kata untuk mencari informasi pada dokumen. Hasil dari penelitian tersebut menunjukkan bahwa QER meningkatkan kinerja analisis sentimen dalam hal akurasi dibandingkan dengan seleksi fitur lain, yakni *Information Gain*, *Chi-Square*, *Document Frequency Difference*, dan *Optimal Orthogonal Centroid*. Penelitian yang dilakukan Fanissa dkk. (2018) menunjukkan bahwa metode *naïve bayes* dengan seleksi fitur QER menghasilkan nilai akurasi 86.6% untuk analisis sentimen pariwisata di Kota Malang. Hal ini menunjukkan bahwa seleksi fitur QER bekerja dengan baik dengan metode *naïve bayes*.

Pertumbuhan teknologi informasi khususnya media sosial telah mengubah cara manusia berkomunikasi satu sama lain (Bourequat dan Mourad, 2021). Media sosial menjadi salah satu layanan publik untuk menyampaikan ataupun memperoleh berita, opini maupun komentar terhadap suatu isu. Salah satu media sosial yang

banyak diminati oleh masyarakat Indonesia adalah *twitter* (Agustina dkk., 2020). Pengguna *twitter* dapat mengemukakan opininya melalui postingan yang disebut *tweet*. Opini pengguna yang diposting melalui *tweet* dapat berupa pendapat, saran ataupun kritikan tentang topik-topik tertentu (Prasetya dkk., 2021). Kurikulum merdeka menjadi salah satu isu yang saat ini ramai dibicarakan di media sosial, termasuk pada *twitter*. Kurikulum merdeka adalah kurikulum dengan pembelajaran intrakurikuler yang beragam dengan konten yang lebih optimal agar peserta didik memiliki cukup waktu untuk mendalami konsep dan menguatkan kompetensi (Kemdikbudristek, 2022). Hingga kini, kurikulum merdeka sebagai kurikulum baru yang diluncurkan oleh Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi untuk mengatasi krisis pembelajaran masih menuai pro dan kontra.

Berdasarkan uraian tersebut, maka penelitian ini akan difokuskan dalam melakukan analisis sentimen terkait penerapan kurikulum merdeka pada media sosial *twitter* menggunakan klasifikasi *multinomial naïve bayes* dan *bernoulli naïve bayes* dengan seleksi fitur *query expansion ranking*. Hasil dari penelitian ini dapat memberikan gambaran opini masyarakat mengenai kurikulum merdeka dan menjadi acuan bagi pemerintah, khususnya Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi dalam mengambil arah kebijakan pendidikan di Indonesia.

1.2 Rumusan Masalah

Rumusan masalah yang dibahas pada penelitian ini adalah sebagai berikut:

1. Bagaimana hasil klasifikasi sentimen kurikulum merdeka pada media sosial *twitter* menggunakan klasifikasi *multinomial naïve bayes* dan *bernoulli naïve bayes* dengan seleksi fitur *query expansion ranking*?
2. Bagaimana perbandingan kinerja klasifikasi *multinomial naïve bayes* dan *bernoulli naïve bayes* dengan seleksi fitur *query expansion ranking* dalam klasifikasi sentimen kurikulum merdeka?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan pada penelitian ini adalah *tweet* berbahasa Indonesia yang ditambah menggunakan *library snsrape* pada bahasa pemrograman *python* dari tanggal 21 Juli 2022 hingga 17 November 2022.

2. *Tweet* yang diklasifikasikan sebagai sentimen positif merupakan *tweet* yang berisi dukungan dan keunggulan kurikulum merdeka, sedangkan *tweet* yang diklasifikasikan sebagai sentimen negatif merupakan *tweet* yang berisi tentang penolakan dan ketidakpuasan terhadap kurikulum merdeka.
3. Pembagian data menjadi data latih dan data uji dilakukan dengan perbandingan 80:20.
4. Pengujian kinerja klasifikasi menggunakan *confusion matrix* dengan nilai *accuracy*, *precision*, *recall (sensitivity)* dan *f-measure*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Memperoleh hasil klasifikasi sentimen kurikulum merdeka pada media sosial *twitter* menggunakan klasifikasi *multinomial naïve bayes* dan *bernoulli naïve bayes* dengan seleksi fitur *query expansion ranking*.
2. Memperoleh hasil perbandingan kinerja klasifikasi *multinomial naïve bayes* dan *bernoulli naïve bayes* dengan seleksi fitur *query expansion ranking* dalam klasifikasi sentimen kurikulum merdeka.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut:

1. Menambah wawasan dan pengetahuan mengenai pengklasifikasian dengan metode *multinomial naïve bayes* dan *bernoulli naïve bayes* serta seleksi fitur *query expansion ranking*.
2. Memberikan gambaran opini masyarakat mengenai kurikulum merdeka dan menjadi acuan bagi pemerintah, khususnya Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi dalam mengambil arah kebijakan pendidikan di Indonesia.

BAB II

TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Analisis sentimen adalah bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya (Liu, 2012). Analisis sentimen akan mengelompokkan polaritas dari teks yang terdapat pada kalimat untuk mengetahui pendapat yang dikemukakan dalam kalimat tersebut apakah bersifat positif atau negatif. Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau pendapat. Polaritas mempunyai arti apakah teks yang ada dalam dokumen, kalimat, atau pendapat memiliki aspek positif atau negatif (Amrullah dkk., 2020).

2.2 Praproses Data Teks

Praproses atau *preprocessing* merupakan proses untuk mempersiapkan data menjadi format yang lebih mudah untuk diproses. Praproses data teks mengubah suatu teks yang tidak terstruktur menjadi lebih terstruktur sehingga lebih mudah untuk diproses (Ferryawan dkk., 2020). Praproses merupakan tahapan yang sangat penting dalam analisis sentimen (Rosid dkk., 2020). Praproses juga dilakukan untuk mendapatkan hasil yang lebih akurat (Nasution dkk., 2019).

Berikut merupakan tahapan dari praproses data teks:

1. *Data cleaning*, yaitu pembersihan data dari *noise* seperti *username*, URL, *hashtag*, *emoticon*, nomor dan tanda baca (Zuhdi dkk., 2019).
2. *Case folding*, yaitu proses penyamaan bentuk huruf dalam sebuah dokumen. Hal ini dilakukan karena tidak semua teks konsisten dalam penggunaan huruf. *Case folding* dibutuhkan untuk mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar atau dalam hal ini menjadi *lowercase* (Alita dan Rahman, 2020).
3. *Spelling normalization*, yaitu perbaikan dan substitusi kata-kata yang salah eja, kata tidak baku, atau singkatan dengan bentuk tertentu (Simatupang dan Utomo, 2019).

4. *Stemming*, yaitu proses mengubah suatu kata menjadi kata dasar dengan menghilangkan imbuhan-imbuhan (awalan, akhiran, maupun perpaduan keduanya) pada kata dalam teks (Saputra dkk., 2015).
5. *Stopword removal*, yaitu menghapus kata-kata yang cukup umum dan sering muncul namun tidak mempunyai makna yang signifikan pada teks atau kalimat (Hidayatullah, 2016).
6. *Tokenizing*, yaitu proses memecah dokumen atau teks menjadi kumpulan kata dengan memisahkannya menggunakan spasi (Wahyuni dkk., 2017).

2.3 Klasifikasi *Naïve Bayes*

Klasifikasi digunakan untuk mengkategorikan setiap data yang ada ke dalam suatu kelas. Klasifikasi secara khusus pada teks atau dokumen memiliki tujuan untuk mengelompokkan suatu dokumen ke dalam suatu kategori tertentu (Manning dkk., 2009). Dokumen yang dimaksud disini adalah sebuah teks atau kumpulan kata-kata yang dianggap sebagai satu kesatuan informasi, dalam penelitian ini setiap *tweet* dianggap sebagai satu dokumen yang akan diklasifikasikan. Pengklasifikasian secara umum terdiri dari dua tahapan, yakni tahapan pelatihan (*training*) dan pengujian (*testing*). Tahapan pelatihan dilakukan menggunakan data latih yang telah diketahui kelasnya, sedangkan tahapan pengujian dilakukan dengan menggunakan data uji untuk mengukur keakuratan pengklasifikasian dari metode yang digunakan (Wijaya dan Santoso, 2016).

Klasifikasi *naïve bayes* adalah metode klasifikasi berdasarkan peluang dan statistik yang didasarkan pada teorema bayes dengan asumsi bahwa setiap fitur bersifat bebas (*independence*) (Buntoro, 2017). Teorema bayes adalah perhitungan statistik dengan menghitung peluang kemiripan kasus lama yang ada dibasis kasus dengan kasus baru (Sartika dan Sensuse, 2017). Klasifikasi *naïve bayes* dikatakan sebagai metode klasifikasi yang sederhana karena kemudahan penggunaannya serta waktu pemrosesan yang cepat bila diterapkan pada *data* yang besar, mudah diimplementasikan dengan strukturnya yang cukup sederhana dan tingkat efektivitas yang tinggi (Hadna dkk., 2016). Teorema bayes pada umumnya memiliki persamaan sebagai berikut (Raschka, 2014):

$$P(c|d) = \frac{P(c)P(d|c)}{P(d)} \quad (2.1)$$

dengan:

d : dokumen/*tweet* yang belum diklasifikasikan

c : notasi kelas (positif atau negatif)

$P(c|d)$: peluang dokumen d diklasifikasikan ke dalam kelas c

$P(c)$: peluang kelas c

$P(d|c)$: peluang dokumen/*tweet* d di dalam kelas c

$P(d)$: peluang dokumen/*tweet* d

Klasifikasi *Naïve Bayes* memiliki beberapa variasi, termasuk diantaranya adalah *Multinomial Naïve Bayes* dan *Bernoulli Naïve Bayes*.

2.3.1 Klasifikasi *Multinomial Naïve Bayes*

Multinomial naïve bayes merupakan variasi dari metode *naïve bayes* yang bekerja dengan cara menghitung frekuensi setiap kata pada dokumen (Fanissa dkk., 2018). Metode *multinomial* dirancang untuk menentukan frekuensi dari kata (*term frequency*), yaitu berapa kali suatu kata muncul dalam suatu dokumen. Suatu istilah/kata mungkin sangat penting dalam menentukan sentimen, menjadikan metode ini sebagai salah satu pilihan klasifikasi pada analisis sentimen (Singh dkk., 2019).

Pada metode *multinomial naïve bayes* dokumen diwakili oleh serangkaian kemunculan n buah kata (*term*) t dari dokumen seperti yang terdapat pada Persamaan (2.2) (Sholehah, 2018).

$$d = \langle t_1, t_2, \dots, t_k, \dots, t_{n_d} \rangle; t_k \in V \quad (2.2)$$

dengan:

V : himpunan kata (*vocabulary*) dalam data latih

n_d : jumlah kata pada dokumen d

Metode ini memperhitungkan jumlah kemunculan dari setiap kata. Peluang dokumen d terdapat dalam kelas c dihitung menggunakan Persamaan (2.3) (Sholehah, 2018).

$$P(c|d) \propto P(c) \prod_{1 \leq k \leq n_d} P(t_k|c) \quad (2.3)$$

dengan:

$P(t_k|c)$: peluang dari kata (*term*) t_k terdapat dalam dokumen kelas c

$P(c)$: peluang sebuah dokumen muncul di kelas c

Penduga dari $P(c)$ dihitung dengan menggunakan Persamaan (2.4) (Susanti, 2016).

$$\hat{P}(c) = \frac{N_c}{N} \quad (2.4)$$

dengan:

N_c : jumlah dokumen dengan kelas c pada data latih

N : jumlah total dokumen data latih

Penduga dari $P(t_k|c)$ dihitung dengan menggunakan Persamaan (2.5) (Susanti, 2016).

$$\hat{P}(t_k|c) = \frac{T_{ct}}{\sum_{t' \in V} T_{ct'}} \quad (2.5)$$

dengan:

T_{ct} : jumlah kemunculan kata (*term*) t dalam data latih pada kelas c

$\sum_{t' \in V} T_{ct'}$: jumlah seluruh istilah yang terdapat pada seluruh dokumen data latih di kelas c

Perhitungan peluang pada Persamaan (2.5) dapat bernilai nol jika terdapat kata pada data uji yang tidak terdapat pada data latih, hal ini tentunya dapat menjadi masalah karena akan membuat dokumen tidak dapat diklasifikasikan ke dalam sebuah kelas. Untuk menangani hal tersebut, maka digunakan *laplace smoothing*, yaitu menambahkan angka satu ke setiap hitungan. Peluang tersebut dihitung menggunakan Persamaan (2.6) (Manning dkk., 2009).

$$\hat{P}(t_k|c) = \frac{T_{ct} + 1}{\sum_{t' \in V} (T_{ct'} + 1)} = \frac{T_{ct} + 1}{(\sum_{t' \in V} T_{ct'}) + B} \quad (2.6)$$

dengan:

B : jumlah kata (*term*) pada *vocabulary*

Tujuan dari klasifikasi teks adalah menemukan kelas terbaik untuk dokumen. Kelas terbaik dalam klasifikasi *multinomial naïve bayes* adalah kelas yang paling mungkin, disebut juga dengan *maximum a posteriori* (MAP) disebut dengan c_{map} yang didefinisikan pada Persamaan (2.7) dengan C adalah himpunan $\{positif, negatif\}$ (Manning dkk., 2009).

$$c_{map} = \underset{c \in C}{\operatorname{argmax}} \hat{P}(c) \prod_{1 \leq k \leq n_d} \hat{P}(t_k|c) \quad (2.7)$$

2.3.2 Klasifikasi *Bernoulli Naïve Bayes*

Bernoulli naïve bayes merupakan variasi dari metode *naïve bayes* yang menetapkan bahwa dokumen diwakili oleh vektor atribut biner yang menunjukkan kata yang muncul dan tidak muncul dalam dokumen. Metode ini hanya memperhatikan kemunculan kata dalam dokumen dan mengabaikan frekuensi dari data yang muncul (Wardani dkk., 2020).

Pada metode *bernoulli naïve bayes*, kata (*term*) akan diberi nilai 1 apabila terdapat pada dokumen dan bernilai 0 apabila tidak terdapat pada dokumen seperti terdapat pada Persamaan (2.8) (Sholehah, 2018).

$$d = \langle e_1, e_2, \dots, e_i, \dots, e_M \rangle; e_i \in \{0,1\} \quad (2.8)$$

Dengan :

d : kumpulan nilai biner yang menunjukkan untuk setiap *term* itu muncul atau tidak dalam dokumen

M : banyaknya fitur

Peluang dokumen d terdapat dalam kelas c dihitung menggunakan Persamaan (2.9) (Manning dkk., 2009).

$$P(c|d) \propto P(c) \prod_{\substack{1 \leq i \leq M \\ e_i=1}} P(t_i|c) \prod_{\substack{1 \leq i \leq M \\ e_i=0}} (1 - P(t_i|c)) \quad (2.9)$$

dengan:

$P(t_i|c)$: peluang kemunculan dari kata (*term*) ke- i (t_i) terdapat dalam dokumen kelas c —dalam posisi apapun dan kemungkinan berkali-kali

$P(c)$: peluang sebuah dokumen muncul di kelas c

Penduga dari $P(c)$ dihitung dengan menggunakan Persamaan (2.10).

$$\hat{P}(c) = \frac{N_c}{N} \quad (2.10)$$

dan penduga dari $P(t_i|c)$ dihitung dengan menggunakan Persamaan (2.11) (Sholehah, 2018).

$$\hat{P}(t_i|c) = \frac{N_{ct}}{N_c} \quad (2.11)$$

dengan:

N_c : jumlah dokumen dengan kelas c pada data latih

N_{ct} : banyaknya dokumen yang memuat *term* t dalam dokumen kelas c pada data latih

N : jumlah total dokumen data latih

Metode *bernoulli* ini juga memerlukan *laplace smoothing* seperti metode *multinomial* untuk mengatasi perhitungan peluang bernilai 0 yang diakibatkan oleh adanya kata pada data uji yang tidak terdapat pada data latih. Peluang tersebut dihitung menggunakan Persamaan (2.12) (Manning dkk., 2009).

$$\hat{P}(t_i|c) = \frac{N_{ct} + 1}{N_c + 2} \quad (2.12)$$

Kelas yang paling mungkin atau *maximum a posteriori* (MAP) untuk *bernoulli naïve bayes* didefinisikan pada Persamaan (2.13) dengan C adalah himpunan $\{positif, negatif\}$ (Manning dkk., 2009).

$$c_{map} = \operatorname{argmax}_{c \in C} \hat{P}(c) \prod_{\substack{1 \leq i \leq M \\ e_i=1}} P(t_i|c) \prod_{\substack{1 \leq i \leq M \\ e_i=0}} 1 - P(t_i|c) \quad (2.13)$$

2.4 Seleksi Fitur *Query Expansion Ranking*

Seleksi fitur merupakan tahapan penting dalam klasifikasi sentimen karena dapat menyelesaikan masalah tingginya dimensi fitur dan *noisy feature* dengan memilih fitur-fitur penting. Seleksi fitur ini dapat meningkatkan kinerja klasifikasi sentimen (Prastyo dkk., 2022). Seleksi fitur adalah proses pengoptimalan untuk mengurangi sekumpulan besar fitur asli menjadi *subset* fitur yang relatif lebih kecil yang signifikan untuk akurasi klasifikasi yang cepat dan efektif (Atsqalani dkk., 2022). Meskipun beberapa fitur dihapus, seleksi fitur ini tidak mengurangi kualitas dari analisis sentimen (Siniwi dkk., 2021).

Query Expansion Ranking (QER) adalah salah satu seleksi fitur yang secara khusus dikembangkan untuk analisis sentimen. QER didasarkan pada metode *query expansion* yang berguna untuk meningkatkan kualitas *query* dan digabungkan dengan metode *probabilistic weighting model* kemudian digunakan untuk memberikan skor pada fitur dalam analisis sentimen (Siniwi dkk., 2021). Persamaan (2.14) digunakan untuk menghitung nilai skor QER untuk fitur (Parlar dkk., 2018).

$$Score_f = \frac{|p_f + q_f|}{|p_f - q_f|} \quad (2.14)$$

dengan:

$Score_f$: nilai skor QER untuk fitur f

p_f : rasio dokumen dengan kelas positif yang memuat fitur f

q_f : rasio dokumen dengan kelas negatif yang memuat fitur f

Nilai dari p_f dan q_f berturut-turut dihitung menggunakan Persamaan 2.15 dan Persamaan 2.16 (Parlar dkk., 2018).

$$p_f = \frac{DF_+^f + 0.5}{N^+ + 1.0} \quad (2.15)$$

$$q_f = \frac{DF_-^f + 0.5}{N^- + 1.0} \quad (2.16)$$

dengan:

DF_+^f : jumlah dokumen yang mengandung fitur f dalam kelas positif

DF_-^f : jumlah dokumen yang mengandung fitur f dalam kelas negatif

N^+ : jumlah dokumen dalam kelas positif

N^- : jumlah dokumen dalam kelas negatif

Perhitungan pada Persamaan 2.15 dan 2.16 dilakukan penambahan konstanta kecil ke pembilang dan penyebut untuk menghindari peluang nol (Parlar dkk., 2018).

Metode seleksi fitur QER menghitung terlebih dahulu nilai skor QER untuk tiap fitur sebelum fitur yang memiliki skor terendah dipilih dan digunakan dalam proses klasifikasi. Ketika fitur memiliki skor rendah, perbedaan antara peluang untuk kelas positif dan negatif tinggi. Oleh karena itu fitur tersebut lebih memiliki kelas yang spesifik dan lebih bernilai untuk proses klasifikasi (Parlar dkk., 2018).

2.5 Confusion Matrix

Hal penting dalam klasifikasi adalah mengukur kinerja dari klasifikasi yang dilakukan. *Confusion matrix* adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah (Normawati dan Prayogi, 2021). *Confusion matrix* dapat memberikan informasi perbandingan hasil klasifikasi yang dilakukan dengan klasifikasi sebenarnya (Wibowo dkk., 2021). *Confusion matrix* ditunjukkan pada Tabel 2.1.

Tabel 2.1 *Confusion Matrix*

Kelas Aktual	Kelas Prediksi	
	Positif	Negatif
Positif	<i>TP</i>	<i>FN</i>
Negatif	<i>FP</i>	<i>TN</i>

Berikut merupakan penjelasan dari masing-masing nilai pada *confusion matrix*:

1. *TP (True Positive)* adalah jumlah data positif yang terklasifikasi dengan benar sebagai positif.
2. *FN (False Negative)* adalah jumlah data positif yang terklasifikasi salah sebagai negatif.
3. *FP (False Positive)* adalah jumlah data negatif yang terklasifikasi salah sebagai positif.
4. *TN (True Negative)* adalah jumlah data negatif yang terklasifikasi dengan benar sebagai negatif.

Berdasarkan pada nilai-nilai yang merupakan representasi kinerja klasifikasi tersebut dapat dihitung (Dellia dan Tjahyanto, 2017):

1. *Accuracy*, perbandingan antara data yang terklasifikasi benar dengan keseluruhan data.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (2.17)$$

2. *Recall (Sensitivity)*, perbandingan antara data positif yang terklasifikasi benar dengan keseluruhan data positif.

$$Recall = \frac{TP}{TP + FN} \quad (2.18)$$

3. *Precision*, perbandingan antara data positif yang terklasifikasi benar dengan keseluruhan data yang terklasifikasi positif.

$$Precision = \frac{TP}{TP + FP} \quad (2.19)$$

4. *F-measure*, digunakan untuk pengukuran *precision* dan *recall* secara bersamaan.

$$F - measure = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (2.20)$$

Menurut Tarkus dkk. (2020), nilai-nilai kinerja klasifikasi dikatakan sangat baik jika telah menunjukkan angka diatas 85%.

2.6 *Text Visualization*

Text visualization merupakan teknik yang digunakan untuk menampilkan visualisasi data dalam bentuk teks (Kurniasari dkk., 2021). Salah satu metode dari *text visualization* yang umum digunakan adalah *word cloud*. *Word cloud* memberikan visualisasi frekuensi kata dalam teks sambil memberi lebih banyak penekanan pada kata-kata yang lebih sering muncul (Shahid dkk., 2017). Pada visualisasi menggunakan *word cloud*, ukuran huruf dari kata-kata akan berbanding lurus dengan frekuensi kemunculannya, sehingga semakin sering topik tersebut muncul, maka ukurannya akan tampak semakin besar (Ma'arif, 2018).

2.7 **Kurikulum Merdeka**

Kurikulum Merdeka adalah kurikulum dengan pembelajaran intrakurikuler yang beragam di mana konten akan lebih optimal agar peserta didik memiliki cukup waktu untuk mendalami konsep dan menguatkan kompetensi. Guru memiliki keleluasaan untuk memilih berbagai perangkat ajar sehingga pembelajaran dapat disesuaikan dengan kebutuhan belajar dan minat peserta didik. Proyek untuk menguatkan pencapaian profil pelajar pancasila dikembangkan berdasarkan tema tertentu yang ditetapkan oleh pemerintah (Kemdikbudristek, 2022).

Kurikulum Merdeka menjadi program yang diharapkan dapat melakukan pemulihan dalam pembelajaran, dimana menawarkan 3 karakteristik diantaranya pembelajaran berbasis proyek, pengembangan *soft skill* dan karakter sesuai dengan profil pelajar pancasila, pembelajaran pada materi esensial dan struktur kurikulum yang lebih fleksibel. Kurikulum merdeka juga ingin melakukan terobosan yang menjadi jurang penghalang diantara bidang-bidang keilmuan. Implementasi kurikulum merdeka di beberapa sekolah penerak dilaksanakan di tahun pertama dengan cukup baik, kemudian dikembangkan di banyak sekolah tahun sekarang sehingga dalam implementasinya kurikulum merdeka setelah dianalisis lebih baik

dan sesuai dengan kultur Indonesia daripada kurikulum 2013 (Jojor dan Sihotang, 2022).

2.8 *Twitter*

Twitter merupakan satu dari beberapa media sosial yang memiliki popularitas cukup tinggi, termasuk di Indonesia. Media sosial berlogo *larry bird* berwarna biru ini didirikan Jack Dorsey pada tahun 2006. Fungsi utama dari *twitter* adalah mengirim pesan yang disebut dengan *tweet* (Yusran dkk., 2022). *Twitter* menempati posisi kelima pada kategori media sosial yang sering digunakan setelah *Youtube*, *Whatsapp*, *Facebook* dan *Instagram* (Fairuz dkk., 2021). *Twitter API (Application Programming Interface)* adalah sebuah aplikasi yang dikembangkan oleh pihak *twitter* agar mempermudah pihak pengembang lain untuk mengakses informasi dari *twitter* (Anjani dan Fauzan, 2021). API merupakan metode yang digunakan oleh program pada komputer untuk berkomunikasi antara satu dengan yang lain untuk meminta dan menyajikan informasi yang diperlukan (Putri dkk., 2022). *Twitter API* ini dapat digunakan untuk melakukan penambangan data pada *twitter*.