KLASIFIKASI EMOSI PADA CERITA ANAK DENGAN PENDEKATAN *TEXT MINING* MENGGUNAKAN METODE *NAÏVE BAYES*

SKRIPSI

Oleh :
ADELIA TASYA VANESSA
NIM. 200605110152

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024

KLASIFIKASI EMOSI PADA CERITA ANAK DENGAN PENDEKATAN *TEXT MINING* MENGGUNAKAN METODE *NAÏVE BAYES*

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
ADELIA TASYA VANESSA
NIM. 200605110152

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024

HALAMAN PERSETUJUAN

KLASIFIKASI EMOSI PADA CERITA ANAK DENGAN PENDEKATAN *TEXT MINING* MENGGUNAKAN METODE *NAÏVE BAYES*

SKRIPSI

Oleh :
ADELIA TASYA VANESSA
NIM. 200605110152

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 24 September 2024

Pembimbing I,	Pembimbing II,
Dr. M Ainul Yaqin, M. Kom	Dr. Zainal Abidin, M. Kom
NIP. 19761013 200604 1 004	NIP. 19760613 200501 1 004

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Tr. Fachrul Kurniawan, M.MT, IPU
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

KLASIFIKASI EMOSI PADA CERITA ANAK DENGAN PENDEKATAN *TEXT MINING* MENGGUNAKAN METODE *NAÏVE BAYES*

SKRIPSI

Oleh :
ADELIA TASYA VANESSA
NIM. 200605110152

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 29 November 2024

Susunan Dewan Penguji

Ketua Penguji	: Syahiduz Zaman, M.Kom NIP. 19700502 200501 1 005	( )
Anggota Penguji I	: Dr. Agung Teguh Wibowo Almais, M.T NIP. 19860301 202321 1 016	( )
Anggota Penguji II	: Dr. M. Ainul Yaqin, M.Kom NIP. 19761013 200604 1 004	( )
Anggota Penguji III	: Dr. Zainal Abidin, M.Kom NIP. 19760613 200501 1 004	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Ir. Fachrul Kurniawan, M.MT, IPU
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: Adelia Tasya Vanessa
NIM	: 200605110152
Fakultas / Program Studi	: Sains dan Teknologi / Teknik Informatika
Judul Skripsi	: Klasifikasi Emosi Pada Cerita Anak Dengan Pendekatan *Text Minning* Menggunakan Metode *Naïve Bayes*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 29 November 2024
Yang membuat pernyataan,

METERAI TEMPEL
9AAKX455534500

Adelia Tasya Vanessa
NIM.200605110152

MOTTO

“Selalu berpikir positif dan melihat peluang disetiap kesulitan”

HALAMAN PERSEMBAHAN

Puja dan puji Syukur atas kehadirat Allah subhanahu wa ta'ala, serta shalawat dan salam bagi Rasul-Nya Penulis mempersembahkan hasil karya ini kepada:

Orang tua penulis yang sangat dicintai dan disayangi, Ayah Suharyono dan Ibu Sri Ratri Amalia serta seluruh keluarga, yang tak pernah berhenti memberi dukungan, do'a, semangat serta motivasi kepada penulis. Terimakasih atas segala bentuk pengorbanan dan dukungan sehingga penulis dapat menyelesaikan perkuliahan selama ini.

KATA PENGANTAR

Assalamu'alaikum warahmatullahi wabarakatuh.

Alhamdulillah, segala puji dan syukur senantiasa penulis panjatkan pada Allah subhanahu wa ta'ala yang berkat Rahmat, serta hidayah-Nya, sehingga penulis dapat menyempurnakan skripsi ini dengan tepat. Sholawat serta salam tetap tercurahkan kepada Nabi Muhammad SAW yang telah menuntun umat manusia menuju jalan yang lebih baik.

Penulis mengucapkan rasa terima kasih yang begitu besar kepada seluruh pihak yang memberikan dukungan dan membantu rampungnya skripsi ini. Ucapan terima kasih ini penulis disampaikan kepada:

1. Prof. Dr. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Harini, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Ir. Fachrul Kurniawan, M.MT, IPU., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. M. Ainul Yaqin, M.Kom, dan Dr. Zainal Abidin, M.Kom selaku dosen pembimbing I dan II, yang telah membimbing dan memberikan arahan kepada peneliti sehingga dapat menyelesaikan skripsi ini.
5. Syahiduz Zaman, M.Kom dan Dr. Agung Teguh Wibowo Almais, M.T selaku Dosen Penguji I dan II, yang telah memberikan ilmu, saran dan arahan dalam menyelesaikan skripsi.

6. Seluruh civitas akademik Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang, terutama seluruh dosen, yang telah memberikan bimbingan dan ilmu yang bermanfaat.
7. Kedua orang tua tercinta yang selalu memberikan semangat dan do'a pada penulis hingga rampungnya skripsi ini.
8. Ardi Firman, Victor Satrio, Kezia Varda, Azmia Almas dan Putri Aisyah selaku sahabat penulis yang telah memberikan semangat yang tiada henti kepada penulis hingga rampungnya skripsi.
9. Teman-teman (INTEGER) Angkatan 2020 senantiasa selalu memberikan semangat dan dukungan dalam berjuang bersama dalam mengejar gelar S.Kom dan pengalaman di Universitas yang sama.

Penulis menyadari bahwa dalam penyusunana skripsi ini masih jauh dari kata sempurna. Maka dari itu penulis menerima saran, kritik dan masukan yang bersifat membangun sehingga dapat menjadi lebih baik kedepannya. Penulis berharap semoga skripsi ini dapat memberikan manfaat untuk kedepannya..

Wassalamu'alaikum warahmatullahi wabarakatuh.

Malang, 29 November 2024

Penulis

DAFTAR ISI

HALAMAN PENGAJUAN ii
HALAMAN PERSETUJUAN iii
HALAMAN PENGESAHAN iv
PERNYATAAN KEASLIAN TULISAN v
MOTTO vi
HALAMAN PERSEMBAHAN vii
KATA PENGANTAR viii
DAFTAR ISI x
DAFTAR GAMBAR xii
DAFTAR TABEL xiii
ABSTRAK xiv
ABSTRACT xv
البحث مستخلص xvi
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 7
1.3 Batasan Masalah 7
1.4 Tujuan Penelitian 8
1.5 Manfaat Penelitian 8
BAB II STUDI PUSTAKA 9
2.1 Peneliatian Terdahulu 9
2.2 Emosi 12
2.3 Cerita Anak 14
2.4 Text Mining 15
2.5 Metode *Naïve Bayes* 16
2.6 Pembobotan Kata 17
2.7 *Confusion Matrix* 18
BAB III DESAIN DAN IMPLEMENTASI 20
3.1 Pengumpulan Data 20
3.2 Desain Sistem 22
3.3 *Prepocessing* Data 23
3.3.1 *Case Folding* 24
3.3.2 *Stop Word Removal* 26
3.3.3 *Lemmatization* 27
3.4 Proses Metode *Naïve Bayes* 29
3.5 Desain Eksperimen 31
BAB IV UJI COBA DAN PEMBAHASAN 33
4.1 Uji Coba 33
4.1.1 Proses pengujian dengan model 90:10 33
4.1.2 Proses Pengujian dengan model 80:20 38
4.1.3 Proses Pengujian dengan model 70:30 44
4.2 Pembahasan 49
4.2.1 Perbandingan Algoritma 50

4.2.2 Visualisasi Hasil.....51
4.3 Integrasi Islam.....53
BAB V KESIMPULAN DAN SARAN 57
5.1 Kesimpulan.....57
5.2 Saran.....57
DAFTAR PUSTAKA
LAMPIRAN

DAFTAR GAMBAR

Gambar 3.1 Alur Diagram.....23
Gambar 3.2 Hasil Pengecekan Tipe Data.....24
Gambar 3.3 Label Teks Cerita 24
Gambar 4.1 *Confusion Matrix* Data Train *Multinomial Naïve Bayes* 90:10.....34
Gambar 4.2 *Confusion Matrix* Data Test *Multinomial Naïve Bayes* 90:10.....35
Gambar 4.3 *Confusion Matrix* Data Train *Complement Naïve Bayes* 90:10 37
Gambar 4.4 *Confusion Matrix* Data Test *Complement Naïve Bayes* 90:10 38
Gambar 4.5 *Confusion Matrix* Data Train *Multinomial Naïve Bayes* 80:20.....39
Gambar 4.6 *Confusion Matrix* Data Test *Multinomial Naïve Bayes* 80:20.....41
Gambar 4.7 *Confusion Matrix* Data Train *Complement Naïve Bayes* 80:20 42
Gambar 4.8 *Confusion Matrix* Data Test *Complement Naïve Bayes* 80:20 44
Gambar 4.9 Confussion Matrix Data Train *Multinomial Naïve Bayes* 70:30 45
Gambar 4.10 *Confusion Matrix* Data Test *Multinomial Naïve Bayes* 70:30.....47
Gambar 4.11 *Confusion Matrix* Data Train *Complement Naïve Bayes* 70:30 48
Gambar 4.12 Confussion Matrix Data Test Complement *Naïve Bayes* 70:3.....49
Gambar 4.13 Visualisasi Hasil 53

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu.....11
Tabel 3.1 Tampilan Input *Dataset*.....20
Tabel 3.2 *Split Dataset*.....21
Tabel 3.3 Jumlah Data.....22
Tabel 3.4 Hasil *Case Folding*.....25
Tabel 3.5 Hasil *Stopwords Removal*.....26
Tabel 3.6 Hasil *Lemmatization*.....28
Tabel 3.7 *Preprocessing Pipeline* 28
Tabel 3.8 Kombinasi Parameter Terbaik.....32
Tabel 3.9 Skenario Uji.....32
Tabel 3.10 Nilai *Hyperparameter* 32
Tabel 4.1 *Classification Report Data Train Multinomial Naïve Bayes 90:10*.....33
Tabel 4.2 *Classification Report Data Test Multinomial Naïve Bayes 90:10*.....35
Tabel 4.3 *Classification Report Data Train Complement Naïve Bayes 90:10*.....36
Tabel 4.4 *Classification Report Data Test Complement Naïve Bayes 90:10*.....37
Tabel 4.5 *Classification Report Data Train Multinomial Naïve Bayes 80:20*.....39
Tabel 4.6 *Classification Report Data Test Multinomial Naïve Bayes 80:20*.....40
Tabel 4.7 *Classification Report Data Train Complement Naïve Bayes 80:20*.....42
Tabel 4.8 *Classification Report Data Test Complement Naïve Bayes 80:20*.....43
Tabel 4.9 *Classification Report* Data *Train Multinomial Naïve Bayes 70:30* 44
Tabel 4.10 *Classification Report* Data *Test Multinomial Naïve Bayes 70:30*.....46
Tabel 4.11 *Confussion Report* Data *Train Complement Naïve Bayes 70:30*.....47
Tabel 4.12 *Confussion Report Data Test Complement Naïve Bayes 70:30*.....48
Tabel 4.13 Hasil Evaluasi Model Klasifikasi 50
Tabel 4.14 Visualisasi Hasil.....51

ABSTRAK

Vanessa, Adelia Tasya. 2024. **Klasifikasi Emosi Pada Cerita Anak Dengan Pendekatan *Text Mining* Menggunakan Metode *Naïve Bayes*.** Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. M. Ainul Yaqin, M.Kom (II) Dr. Zainal Abidin, M.Kom.

Kata kunci*:* Klasifikasi, Emosi, Cerita Anak, *Naïve Bayes.*

Cerita anak digunakan untuk mengembangkan kemampuan membaca dan pemahaman bahasa. Tidak hanya itu cerita anak juga dapat memperkaya imajinasi, memperluas wawasan, dan menambah pengetahuan tentang berbagai emosi pada anak-anak melalui sebuah cerita. Penelitian tentang klasifikasi emosi pada cerita anak sedikit banyak dilakukan. Namun, ketidakseimbangan jumlah antar kelas sering terjadi masalah. Maka, penelitian ini bertujuan untuk mengetahui hasil performa dari metode *Naïve Bayes* dengan perbandingan algoritma antara *Multinomial Naïve Bayes* dan *Complement Naiave Bayes* untuk klasifikasi emosi pada cerita anak. Setiap cerita anak melalui *preprocessing* data pada penelitian ini meliputi *case folding, stopword removal* dan *lemmatization*. Untuk ekstraksi fitur dan pembobotan fitur digunakan metode *TF-IDF*. Pengujian data dilakukan dengan menggunakan *split data* 90:10, 80:20, dan 70:30. Berdasarkan hasil pengujian, dengan menggunakan algoritma *Multinomial Naïve Bayes* diperoleh rata-rata akurasi sebesar 0.67 atau 67% dan mengalami kesulitan dalam mengklasifikasi emosi. Sedangkan menggunakan Complement Naive Bayes diperoleh rata-rata 0.89 atau 89%. Dalam penelitian ini dapat ditunjukan hasil dengan menggunakan Complement Naive Bayes lebih baik dalam mengklasifikasi emosi. Metode *Naïve Bayes* dalam mengklasifikasikan emosi pada teks cerita anak dengan cukup baik. Tetapi hasil yang diperoleh masih belum cukup maksimal, masih perlu dikaji lagi untuk metode penyeimbangan data. Melalui proses *preprocessing* dan representasi numerik menggunakan *TF-IDF*, model juga mampu menghitung probalibilitas kemunculan emosi pada teks cerita anak.

ABSTRACT

Vanessa, Adelia Tasya. 2024. **Klasifikasi Emosi Pada Cerita Anak Dengan Pendekatan *Text Mining* Menggunakan *Metode Naïve Bayes.*** Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. M. Ainul Yaqin, M.Kom (II) Dr. Zainal Abidin, M.Kom

Children's stories are used to develop reading skills and language comprehension. Not only that, children's stories can also enrich imagination, broaden horizons, and increase knowledge about various emotions in children through a story. More research has been conducted on the classification of emotions in children's stories. However, an imbalance in numbers between classes is often a problem. So, this research aims to determine the performance results of the *Naïve Bayes* method by comparing the algorithms between *Multinomial Naïve Bayes* and *Complement Naïve Bayes* for classifying emotions in children's stories. Each child's story goes through data preprocessing in this research including case folding, stopword removal, and lemmatization. For feature extraction and feature weighting, the *TF-IDF* method is used. Data testing was carried out using split data of 90:10, 80:20, and 70:30. Based on the test results, using the *Multinomial Naïve Bayes* algorithm obtained an average accuracy of 0.67 or 67% and experienced difficulty in classifying emotions. Meanwhile, using *Complement Naïve Bayes* obtained an average of 0.89 or 89%. In this research, it can be shown that the results using *Complement Naïve Bayes* are better at classifying emotions. The *Naïve Bayes* method in classifying emotions in children's story texts is quite good. However, the results are still not optimal, and the data balancing method must be studied further. Through preprocessing and numerical representation using *TF-IDF*, the model can also calculate the probability of the appearance of emotions in children's story texts.

Key words: Classification, Emotion, Children’s Story, *Naïve Bayes*

البحث مستخلص

فانيسا، أديليا تاسيا. 2024. تصنيف الانفعالات في قصص الأطفال باستخدام منهج التنقيب في النص باستخدام طريقة ساذج بايز. أُطرُوحَة. برنامج دراسة الهندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم
الإسلامية الحكومية، مالانج. المشرف: (أنا) د. م. عين اليقين، م.كوم (الثاني) د. زين العابدين، م.كوم

الكلمات المفتا :حيةالتصنيف، العاطفة، قصة الأطفال، ساذج بايز

تستخدم قصص الأطفال لتنمية مهارات القراءة واستيعاب اللغة. وليس هذا فحسب، فقصص الأطفال يمكنها أيضًا إثراء الخيال وتوسيع الآفاق وزيادة المعرفة بالمشاعر المختلفة لدى الأطفال من خلال القصة. تم إجراء المزيد من الأبحاث حول تصنيف العواطف في قصص الأطفال. ومع ذلك، غالبًا ما يمثل عدم التوازن في الأعداد بين الفئات مشكلة. لذا، يهدف هذا البحث إلى تحديد
نتائج أداء طريقة *Naïve Bayes* من خلال مقارنة الخوارزميات بين *Multinomial Naïve Bayes* و *Complement Naïve Bayes* لتصنيف العواطف في قصص الأطفال. تمر قصةكل طفل بالمعالجة المسبقة للبيانات في هذا البحث بما في ذلك طي الحالة وإزالةكلمة التوقف والإضفاء الطابع الشخصي. لاستخراج الميزة ووزن الميزة، يتم استخدام طريقة TF-IDF. تم إجراء اختبار البيانات باستخدام بيانات مقسمة 90:10، 80:20، و70:30. بناءً على نتائج الاختبار، باستخدام خوارزمية *Naïve Bayes* المتعددة الحدود، تم الحصول على دقة متوسطة تبلغ 0.67 أو 67% وواجهت صعوبة في تصنيف المشاعر. وفي الوقت نفسه، باستخدام *Complement Naïve Bayes* حصل على متوسط 0.89 أو 89%. في هذا البحث، يمكن إثبات أن النتائج باستخدام *Complement Naïve Bayes* هي الأفضل في تصنيف العواطف. إن طريقة *Naïve Bayes* في تصنيف العواطف في نصوص قصص الأطفال جيدة جدًا. ومع ذلك، فإن النتائج لا تزال غير مثالية، ويجب دراسة طريقة موازنة البيانات بشكل أكبر. ومن خلال المعالجة المسبقة والتمثيل العددي باستخدام TF-IDF يمكن للنموذج أيضًا حساب احتمالية ظهور العواطف في نصوص قصص الأطفال.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Cerita anak adalah satu dari sekian banyak bentuk narasi yang secara khusus digunakan untuk memenuhi kebutuhan dan minat anak-anak dalam proses belajar dan hiburan. Cerita anak memiliki ciri khas tersendiri yang membedakannya dari jenis cerita lainnya, yaitu fokus pada tema, karakter, dan konflik yang relevan dengan dunia dan pengalaman anak-anak. Dalam literatur, cerita anak sering kali disajikan dalam bentuk naratif yang sederhana, dengan plot yang mudah dipahami dan karakter yang dapat diidentifikasi oleh anak-anak. Pentingnya cerita anak dalam pembentukan literasi anak telah diakui secara luas, karena cerita-cerita ini tidak hanya mengembangkan kemampuan membaca dan pemahaman bahasa, tetapi juga memperkaya imajinasi, memperluas wawasan, dan membentuk nilai-nilai sosial serta emosional pada anak-anak. Menurut (Masruroh & Ramiati, 2022) mengatakan bahwa cerita anak merupakan tipe buku yang dipilih oleh banyak orang dewasa untuk dibaca bersama anak-anak dan cerita anak juga dapat memberikan kesempatan untuk menambah kosa kata sehingga anak-anak dapat mengembangkan kemapuan bahasanya.

Cerita anak juga memainkan peran penting dalam perkembangan emosional anak dengan memberikan informasi yang kaya tentang emosi dan berfungsi sebagai alat penting untuk sosisalisasi emosi. Selain itu, cerita anak dapat membantu kesadaran emosional, empati, dan pemahaman tentang keadaan emosi yang berbeda

dengan menyajikan berbagai emosi (Ding et al., 2021). Analisis emosi melibatkan identifikasi keadaan emosional atau psikologis tertentu dari sebuah tokoh berdasarkan teks cerita anak, dan analisis emosi ini menggali lebih dalam emosi yang spesifik seperti kegembiraan, ketakutan, kesedihan, dan keberanian, yang mempengaruhi respons emosional pembaca (Nandwani & Verma, 2021). Di samping itu, cerita anak juga mengundang pembaca untuk melakukan analisis terhadap konflik, karakter, dan tema yang disajikan. Dengan demikian, melalui cerita anak, anak-anak tidak diajak untuk merasakan berbagai emosi, tetapi juga diajak untuk berpikir kritis, mengasah keterampilan analitis, dan mengembangkan pemahaman mereka tentang nilai-nilai kehidupan.

Salah satu langkah sederhana yang dapat kita ambil adalah memberikan edukasi melalui cerita-cerita anak. Ini adalah cara yang efektif untuk membantu anak-anak dalam memahami dunia di sekitar lingkungan mereka dan mengembangkan keterampilan serta pemahaman yang mereka butuhkan untuk bertahan dan tumbuh. Banyak orang tua yang masih memberikan cerita anak tanpa mengetahui layak tidaknya cerita anak tersebut dibacakan kepada anak di usia dini, seperti genre atau alur cerita yang sesuai dengan umur anak tersebut. Perkembangan emosi anak merupakan aspek krusial dalam psikologi perkembangan, yang dapat dipengaruhi oleh berbagai faktor, termasuk cerita anak.

Perkembangan adalah transformasi psikologis yang muncul dari kemajuan alami kemampuan mental dan fisik anak, yang dilengkapi oleh pengaruh eksternal dan pengalaman pendidikan selama jangka waktu tertentu menuju kedewasaan. Kemajuan ini menandakan pertumbuhan fungsi dan struktur tubuh, pengembangan

kompetensi, peningkatan kapasitas untuk menghadapi tantangan, kemampuan untuk bertanggung jawab besar, dan pencapaian kebebasan dalam mengutarakan ekspresi kreatif mereka. Evolusi ini secara signifikan dibentuk oleh lingkungan tempat anak dibesarkan, mengarahkan mereka menuju kedewasaan (S. Sukatin et al., 2020).

Orang tua memiliki peran yang sangat penting dalam kehidupan anak-anak terutama untuk membantu dan mendampingi anak agar bisa mengelola emosinya. Hal tersebut tergantung sebagaimana mereka akan berbicara dengan anak-anak tentang pengelolahan emosi, orangtua dapat mengambil dengan pendekatan melatih emosi atau mengabaikan emosi (Q. Y. H. Sukatin, 2020). Cerita anak tidak hanya berfungsi sebagai alat hiburan dan pendidikan, tetapi juga sebagai media yang efektif untuk mengembangkan kesadaran emosional, empati, dan kemampuan sosial anak-anak. Dengan menyajikan berbagai keadaan emosi melalui plot yang sederhana dan karakter yang mudah diidentifikasi, cerita anak membantu anak-anak mengenali dan memahami emosi mereka sendiri serta emosi orang lain.

Oleh karena itu, penelitian yang mendalam mengenai analisis emosi dalam cerita anak menjadi relevan dalam memahami dampak cerita anak terhadap perkembangan emosional dan kognitif anak-anak. Dengan merumuskan pendekatan yang tepat dalam mengklasifikasikan sebuah emosi dalam teks cerita anak, diharapkan penelitian ini dapat memberikan kontribusi yang signifikan dalam bidang literasi anak dan pengembangan pendidikan.

Analisis emosi dalam sebuah cerita anak menjadi relevan karena cerita anak memiliki peran penting dalam pengembangan literasi, pemahaman bahasa, dan

nilai-nilai sosial serta emosional pada anak-anak. Dengan fokus pada tema, karakter, dan konflik yang relevan dengan dunia anak-anak, cerita anak membantu mereka mengembangkan imajinasi, memperluas wawasan, dan membentuk pemahaman tentang emosi. Analisis emosi dalam cerita anak tidak hanya membantu anak-anak mengenali dan memahami emosi mereka sendiri serta emosi orang lain, tetapi juga memfasilitasi pengembangan kesadaran emosional dan empati. Selain itu, penelitian ini dapat memberikan kontribusi yang signifikan dalam bidang literasi anak dan pengembangan pendidikan dengan merumuskan pendekatan yang tepat untuk meningkatkan literasi dan pendidikan anak-anak melalui pemahaman yang lebih baik tentang dampak cerita anak terhadap perkembangan emosional dan kognitif mereka

Analisis emosi sangat penting untuk perkembangan emosional anak, karena membantu dalam mengenali dan memahami berbagai emosi yang mereka alami. Dengan menggunakan metode Naïve Bayes, kita dapat mengklasifikasikan emosi seperti senang, sedih, marah, dan takut dalam cerita anak. Pendekatan ini memungkinkan kita mempelajari pola-pola linguistik yang berkaitan dengan emosi tersebut, sehingga dapat mendukung perkembangan emosional anak secara lebih efektif. Dengan menggunakan bantuan algoritma *Naïve Bayes*, bisa mempelajari pola-pola kata dan frasa-frasa yang terkait dengan berbagai emosi

Dalam penelitian ini, metode *Naïve Bayes* dapat mengembangkan pemahaman yang lebih dalam tentang klasifikasi emosi yang terkandung dalam teks cerita anak. Selain berfungsi sebagai alat untuk mengklasifikasikan emosi dalam teks cerita, metode ini juga dapat digunakan untuk memahami dampak emosional yang

mungkin dimiliki cerita tersebut pada anak-anak. Dengan demikian, klasifikasi emosi pada teks cerita menggunakan metode *Naïve Bayes* membuka pintu bagi eksplorasi yang lebih mendalam tentang kekuatan naratif sehingga anak-anak bisa menganalisis sebuah emosi sebagai pengaruh emosional yang terdapat dalam sebuah teks cerita yang dibacakan kepada anak-anak.

Dalam islam, penting untuk memberikan perhatian dan perlindungan kepada anak usia dini, karena mereka adalah Amanah dari Allah SWT. Sebagai kutipan dari Al-Qur'an yang menekankan pentingnya memperlakukan anak-anak dengan kasih sayang dan keadlian. Salah satunya adalah Surah Al-Isra Ayat 24 yang berbunyi sebagai berikut :

وَاخْفِضْ لَهُمَا جَنَاحَ الذُّلِّ مِنَ الرَّحْمَةِ وَقُلْ رَّبِّ ارْحَمْهُمَا كَمَا رَبَّيٰنِيْ صَغِيْرًا

"Dan rendahkanlah dirimu terhadap keduanya dengan penuh kasih sayang dan ucapkanlah, "Wahai Tuhanku! Sayangilah keduanya sebagaimana mereka berdua telah mendidik aku pada waktu kecil." (Q.S Al-Isra Ayat 24).

Menurut Surah Al-isra ayat 24 menyatakan menyoroti sikap hormat dan kasih sayang yang harus dimiliki anak terhadap kedua orang tua, serta memberi pengajaran terhadap anak untuk memohon rahmat dan kasih sayang Allah bagi orang tua mereka. Ayat ini menekankan pentingnya relasi spiritual antara anak-anak dan orang tua, serta pengakuan atas pengorbanan dan orang tua memberikan kasih sayang terhadap anak.

Ayat Surah Al-isra ayat 24 tersebut mencerminkan nilai-nilai kasih sayang, perawatan, dan orang tua harus bertanggung jawab terhadap anak-anak mereka. Kasih sayang, perhatian, dan kepedulian pembaca adalah beberapa perasaan yang

baik yang dapat dipicu oleh cerita yang berpusat pada hubungan antara orang tua dan anak serta menekakan pentingnya mendidik dan melindungi mereka. Sebaliknya, cerita anak yang menyoroti pengabaian atau penelantaran anak oleh orang tua dapat memicu emosi negative seperti kekhawatiran, kesedihan, atau kemarahan. Oleh karena itu, narasi untuk anak-anak yang terinspirasi dari prinsip moral yang mirip dengan yang terkandung dalam ayat dapat menghasilkan pengalaman emosional yang kuat bagi para pendengar atau pembaca, dan dapat memainkan peran penting dalam membentuk pemahaman dan perilaku orang tua terkain dengan perlindungan dan perawatan terhadap anak-anak.

Penelitian yang telah dilakukan oleh (Ayomi & Dewi, 2023) menunjukan bahwa analisis emosi yang dilakukan dan metode yang digunakan adalah *Naïve Bayes* diperoleh performansi yang dihasilkan lebih baik dengan menggunakan algoritma *Multinomial Naïve Bayes.* Dan pada penelitiant ersebut, peneliti menyarankan untuk memperbaiki akurasi yang akan dilakukan dalam penelitian selanjutnya dengan menggunakan metode koreksi kesalahan ejaan sebelum dilakukannya konversi pada penelitian.

Dari klasifikasi emosi terhadap cerita anak, penulis menggunakan metode *Naïve Bayes* yang mana dalam pengklasifikasiannya dilakukan pemilihan kata yang akan menghasilkan hasil emosi. Penulis sangat berhati-hati dalam pengerjaain klasifikasi ini mengingat hadits Rasulullah yang berbunyi:

عَنْ أَبِي هُرَيْرَةَ رَضِيَ اللَّهُ عَنْهُ أَنَّ رَسُولَ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ قَالَ
مَنْ سَلَكَ طَرِيقًا يَلْتَمِسُ فِيهِ عِلْمًا، سَهَّلَ اللَّهُ لَهُ بِهِ طَرِيقًا إِلَى الْجَنَّةِ

“Barang siapa menempuh jalan untuk mencari ilmu, Allah akan memudahkan baginya jalan menuju surga” (HR. Muslim).

Hadist ini mendorong pentingnya menuntut ilmu, termasuk ilmu yang berkaitan dengan teknologi dan pengembangan secara modern. Dalam hal ini, penelitian klasifikasi emosi dengan metode *Naïve Bayes* adalah contoh penerapan ilmu modern. Upaya ini bisa dianggap sebagai bentuk menuntuut ilmu yang dianjurkan dalam hadist.

Beberapa kelebihan yang dimiliki oleh metode *Naïve Bayes* membuat penulis memilih metode tersebut sebaagai metode untuk melakukan klasifikasi emosi. Oleh karena itu, dilakukan penelitian dengan judul "Klasisfikasi Emosi Pada Cerita Anak dengan Pendekatan *Text Mining* menggunakan Metode *Naïve Bayes*" untuk membantu pembaca dalam mengetahui sebuah emosi pada teks cerita anak.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah diatas, rumusan masalah sebagai berikut bagaimana mengetahui emosi yang terkandung dalam teks cerita anak untuk perkembangan emosional anak?

1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini sebagai berikut:

1. Jenis cerita anak yang menjadi objek analisis, fokus pada cerita anak dalam bentuk teks, seperti cerita dongeng fabel yang ditujukan untuk anak-anak dalam rentang usia tertentu dan teks dalam berbahasa inggris.
2. Analisis emosi pada teks cerita anak dalam empat emosi primer yaitu senang, sedih, marah, dan takut.

3. Perbandingan algoritma yang dilakukan pada model *Naïve Bayes* ini menggunakan kedua algoritma yaitu algoritma *Multinomial Naïve Bayes* dan *Complement Naïve Bayes*

1.4 Tujuan Penelitian

Tujuan utama dalam penelitian bertujuan untuk melakukan klasifikaasi emosi yang dapat memengaruhi perkembangan emosional anak dengan mengetahui beberapa emosi yang terkandung dalam teks cerita.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dapat di peroleh dari penelitian ini adalah

1. Memberikan beberapa panduan bagi orang tua, pendidik, dan penulis untuk memilih cerita anak dengan tujuan mendukung perkembangan emosional anak.
2. Menyediakan pemahaman lebih dalam tentang bagaimana klasifikasi emosi tercermin dalam teks, khususnya dalam cerita anak dengan menerapkan metode *Naïve Bayes.*

BAB II

STUDI PUSTAKA

2.1 Peneliatian Terdahulu

Pada penelitian yang akan dilakukan ini, penulis menemukan beberapa penelitian yang telah dilakukan sebagai sebuah rujukan pada penelitian ini. Ada beberapa rujukan yang diantara adalah penelitian (Jannah & Prasetyo, 2022) menggunakan sentimen analisis dan emosi dengan pendekatan berbasis leksikon. Leksikon ini merupakan daftar kata-kata dalam Bahasa inggris terkait dengan emosi dan sentimen yang terdiri dari 14.182 kata (*unigram*). Dari hasil *scraping*, data *tweet* yang dihasilkan berjumlah 31.859 *tweet* dari 29.122 *user*. Dari data hasil tersebut menunjukkan banyaknya tren *tweet* diunggah selama Maret sampai Juli 2020. Dari gambaran sentimen public selama pandemi periode Maret hingga Juli 2020, jumlah *tweet* dengan kategori sentimen negative sebanyak 15.303. Sementara itu, sentimen positif berada diurutan kedua dengan jumlah *tweet* 12.231. Dan grafik sentimen netral sebanyak 4.325 *tweet.* Dari hasil diatas, menandakan bahwa banyak publik dengan berpendapat negative, tetapi tidak jarang juga yang memiliki pandangan positif selama masi pandemi.

Penelitian lain (Arifin & Al-Idrus, 2024) melakukan klasifikasi emosi pada twitter terhadap bakal calon presiden pada pemilu 2024. Tujuan dari penelitian ini untuk mengetahui pandangan Masyarakat terhadap para bakal calon presiden tersebut berdasarkan klasifikasi emosinya. Dengan menggunakan pembagian data dengan metode 80:20, yaitu data latih dibagi menjadi 80% dan data uji menjadi

20%. Model algoritma *Naïve Bayes* tanpa menggunakan label netral menunjukkan performa yang lebih baik, dengan nilai akurasi model sebesar 58% pada data Anies Baswedan, 58% pada data Prabowo Subianto, dan 76% pada data Ganjar Pranowo, serta gabungan 69%. Klasifikasi pada skenario pengujian menggunakan label netral menunjukkan akurasi sebesar 55% pada data Anies Baswedan, 60% pada data Ganjar Pranowo, dan 53% pada data Prabowo Subianto, sedangkan untuk gabungan semuanya nilai akurasi sebesar 51%.

Penelitian lain dilakukan oleh (Fahrudin et al., 2022) yang dalam penelitiannya menganalisis persepsi Masyarakat terhadap isu-isu pelecehan seksual yang sedang banyak dibicarakan di platform twitter. Dari tujuan penelitian tersebut, menggunakan sentimen analisis dan leksikon emosi. Dengan menggunakan metode *Naïve Bayes*, hasil skurasi model klasifikasi yang dibangun sebesar 74% dengan presentase 72% tweet negative, 18.8% tweet positif, dan 9.2% tweet netral. Berdasarkan hasil penelitian tersebut menunjukan bahwa opini negative terkait dengan isu kekerasan pelecehan seksual masih lebih banyak dibandingkan dengan opini positif dan opini netral. Pada hasil analisis emosi menggunakan emotion lexicon muncul berbagai emosi seperti ketakutan, kewaspadaan, kemarahan, kepercayaan dan lainnya.

Penelitian tentang analisis emosi pada media sosial twitter (Ayomi & Dewi, 2023) dengan menggunakan metode *Multinomial Naïve Bayes* dan *symthetic minority oversampling technique* yang tujuannya untuk mengetahui hasil performa dari gabungan metode untuk analisis emosi cuitan dari media sosial Twitter. Pengujian ini dilakukan dengan menggunakan K-Fold Cross Validation, dengan

hasl pengujian metode SMOTE memperoleh nilai rata-rata akurasi sebesar 65% dan rata-rata nilai f1-score sebesar 66%. Sedangkan tanpa menggunakan SMOTE memperoleh nilai rata-rata akurasi sebesar 64% dan rata-rata f1-score sebesar 65%. Hasil yang diperoleh 1% lebih baik dengan menggunakan SMOTE dalam pengkategorian emosi.

Dalam penelitian lain juga dilakukan oleh (Baharuddin & Muyassar, 2023) tentang klasifikasi Emosi berbasis *sentimen analysis*. Media sosial "X" menjadi media sosial yang digunakan oleh banyak instansi untuk mempublikasikan sebuah informasi dan juga sebagai sarana komunikasi dengan Masyarakat. Pada penelitian ini, peneliti menggunakan dataset sebnyak 536 post. Metode yang digunakan menggunakan metode *Naïve Bayes* karena dapat mendeteksi beberapa emosi di beberapa topik pada platform "X". Dari penelitian tersebut, dapat disimpulkan bahwa metode *Naïve Bayes* ini baik dalam proses klasifikasi emosi di platform "X".

Penelitian terbaru penulis adalah klasifikasi emosi pada cerita anak dengan menggunakan metode *Naïve Bayes* ini bertujuan untuk mengetahui emosi apa saja yang terkandung dalam sebuah cerita anak dengan genre fabel. Pada klasifikasi ini peneliti menggunakan beberapa perbandingan dari data latih dan data uji, melakukan perbandingan algoritman metode *Naïve Bayes* yaitu *Multinomial Naïve Bayes* dan *Complement Naïve Bayes*, dan menampilkan visualisasi hasil presentase emosi pada klasifikasi emosi cerita anak.

Tabel 2.1 Penelitian Terdahulu

No	Penulis, tahun, judul	Input	Metode	Output
1.	(Jannah & Prasetyo, 2022), Analisis Sentimen dan Emosi Publik pada Awal	Pengumpulan data twitter dengan menggunakan *web scraping*.	Analisis sentimen dan emosi berdasarkan	Presentasi tweet dengan sentimen positif, negatif, dan netral

Lanjutan Tabel Penelitian Terdahulu

No	Penulis, tahun, judul	Input	Metode	Output
	Pandemi COVID-19 Berdasarkan Data Twitter dengan Pendekatan Berbasis Leksikon		*sentimen leksikon*	
2.	(Arifin & Al-Idrus, 2024), Klasifikasi Emosi Pengguna Twitter Terhadap Bakal Calon Presiden Pada Pemilu 2024 Menggunakan Algortima *Naïve Bayes*	Tweet di platform media sosial *Twitter*	*Naïve Bayes*	Hasil klasifkasi evalusi model dari tiap tweet calon presiden 2024
3.	(Fahrudin et al., 2022), Pemodelan Teks Tweet pada Isu Pelecehan Seksual Berbasis Analisis Sentimen dan Leksikon Emosi	Kumpulan teks *tweet* yang relevel dengan iso pelecehan seksual	*Naïve Bayes* dan *Lexicon-based sentimen analysis*	Klasifikasi *tweet* sebagai positif, negative, atau netral
4.	(Ayomi & Dewi, 2023), Analisis Emosi Pada Media Sosial Twitter Menggunakan Metode *Multinomial Naïve Bayes* dan Synthetic Minority Oversampling Technique	Dataset *Twitter*	Metode *Naïve Bayes*	Hasil perbandingan performannsi multinomial dan SMOTE
5.	(Baharuddin & Muyassar, 2023), Klasifikasi Emosi pada Platform X Menggunakan Metode Machine Learning Berbasis Sentimen Analysis	Kumpulan dataset yang dilakukan secara otomatis dengan memanfaatkan tool python dan disimpan dalam format csv.	*Metode Naïve Bayes Classfier*	Hasil analisis berupa presentase sentimen negative dan positif beserta hasil presentase klasifikasi topik.

2.2 Emosi

Analisis emosi merupakan proses mengidentifikasikan, mengkategorikan, atau mengklasifikasikan sebuah emosi berdasarkan pada ciri-ciri atau fitur-fitur yang ditemukan dalam data. Klasifikasi emosi melibatkan penggunaan teknik pembelajaran mesin untuk mennetukan emosi yang terkandung dalam sebuah teks cerita anak seperti senang, sedih, marah, takut.

Menurut penelitian (Ramadhan et al., 2021) mengatakan bahwa emosi adalah sebuah fenomenal dalam kehidupan. Dengan emosi, dapat mengkomunikasikan apa saja yang dirasakan untuk ditanggapi secara serius agar mendekat pada sikap dan tindakan cinta kasih yang tertuang dalam orang tua kepada anak-anak dengan membacakan sebuah teks cerita anak dengan mengetahui emosi yang terkandung.

Emosi didefinisikan sebagai pola perubahan yang kompleks, termasuk gairah fisiologis, perasaan, proses kognitif, dan reaksi perilaku yang dibuat sebagai respons terhadap situasi yang dianggap signifikan secara pribadi. Tujuan utama dari analisis emosi ini adalah untuk memperkaya pemahaman kita tentang pengaruh narasi anak terhadap pembacanya, mengidentifikasi pola emosional dalam literatur anak, dan memahami bagaimana narasi tersebut dapat memengaruhi perkembangan emosional dan kesejahteraan anak. Emosi dasar yang sering digunakan pada deteksi emosi berdasarkan basis emosi oleh Plutchick yang dikonseptualisasikan secara berpasangan yang berlawanan, yaitu *Joy-Sadness* (Senang-Sedih), *Trust-Disgust* (Percaya-Jijik), *Fear-Anger* (Takut-Marah) dan *Surprise-Anticipation* (Terkejut-Antisipasi).

Emosi merupakan salah satu aspek dalam kecerdasan manusia dan berperan penting untuk mengambil keputusan dalam kehidupan sehari-hari. Dalam interaksi manusia dengan komputer banyak mengabaikan hubungan emosional yang terjadi padahal hal tersebut sangat penting dalam komunikasi antar manusia pada kehidupan kita sehari-hari.

2.3 Cerita Anak

Cerita anak adalah narasi atau teks yang dirancang khusus untuk anak-anak, dengan tujuan memberikan hiburan, pendidikan, atau pesan moral yang sesuai dengan tahap perkembangan mereka. Bahasa yang digunakan dalam cerita anak biasanya sederhana dan mudah dipahami, dan cerita-cerita tersebut seringkali memiliki alur dan karakter yang menarik. Dongeng, fabel, cerita rakyat, dan cerita kreatif lainnya yang ditujukan untuk pembaca muda semuanya dianggap sebagai cerita anak. Pembaca muda diberi informasi, hiburan, dan imajinasi serta potensi kreatif mereka dibangkitkan saat mereka membaca literatur anak. Ini juga membantu mereka dalam memahami dunia luar.

Membacakan buku cerita anak merupakan salah pendekatan yang dapat digunakan dalam mengembangkan karakter anak. Selain itu, juga cara penyampaian cerita anak usia dini dimana anak menjadi pendengar yang baik yang tidak menggurui tetapi bersifat menyenangkan (Mayar et al., 2022). Cerita anak yang bersifat menyenangkan dapat mebghibur anak-anak, cerita anak yang dismapaikan mengandung nilai moral, anak usia dini menjadi aktif dari cerita anak yang disampaikan bersifat menarik yang memicu rasa penasaran terhadap anak dan mengembangkan karakter yang baik untuk anak sesuai dengan tahap perkembangannya.

Menurut pendapat (Mayar et al., 2022) dari cerita anak yang dibacakan anak mampu memahami tentang Masyarakat dan kehidupan secara umum. Sehingga mampu mengembangkan kreativitas dalam berimajinasi anak. Imajinasi anak dapat dibangun ketika anak usia dini dibacakan sebuah cerita dan mereka mampu

memvisualisasikan daya pikirnya masing-masing denga napa yang anak-anak dengar dan rasakan.

Dari segi emosi, membacakan cerita anak membuat anak usia dini memahami kejadian atau alur cerita yang runtut sehingga pendengar bisa merasakan alur cerita tersebut secara emosi. Setelah itu ingatannya akan mengingat sebuah peristiwa dan menghubungkannya dengan kondisi didalam sebuah cerita. Semakin sering anak dibacakan sebuah cerita anak, maka semakin banyak juga hal yang timbun dibawah alam sadar anak-anak (Gusmayanti & Dimyati, 2021).

2.4 Text Mining

Text mining mengacu pada proses mengekstraksi data yang signifikan dan relevan dari teks yang tidak terorganisir, termasuk dokumen, artikel, email, dan pesan. Tujuan utama dari *text mining* adalah untuk memeriksa dan memahami pola, tren, dan hubungan yang tersembunyi dalam teks tersebut. Pendekatan yang digunakan dalam text mining meliputi penggunaan algoritma komputasi untuk melakukan aktivitas seperti pengelompokan, klasifikasi, pengambilan informasi, analisis sentimen, dan pengembangan model prediktif. *Teks mining* atau juga disebut dengan analisis teks digunakan untuk guna mengubah yang semula tidak terstruktur menjadi format yang terstruktur untuk mengidentifikasikan pola, topik, kata kunci, dan atribut lainnya dalam sebuah data (Agung & Santara, 2024).

Menurut (Azzahra & Wibowo, 2020) mengatakan bahwa syarat umum *text mining* merupakan informasi yang diambil dapat menjadi data-data yang berguna, yang diaman data tersebut bersifat jelas dan eksplisit. *Text mining* adalah penemuan dari beberapa pengetahuan yang menarik pada dokemen teks. Hal tersebut

merupakan tantangan untuk menemukan pengetahuan yang akurat pada teks dokumen yang diinginkan, Dimana pengetahuan dapat berubah menjadi lebih efektif dan juga daoat memperbarui pola penemuan pada proses penerapan ke *text mining*.

2.5 Metode *Naïve Bayes*

Naïve Bayes adalah sebuah metode klasifikasi dalam statistika dan machine learning yang berdasarkan pada teorema Bayes dengan asumsi independensi anatar fitur. Meskipun sederhana, pendekatan ini seringkali cukup berhasil, terutama ketika digunakan dengan dataset yang besar dan kompleks.

Dalam penelitian (Ramadhan et al., 2021) mengatakan bahwa metode *Naïve Bayes* merupakan metode yang tidak memiliki aturan, *Naïve Bayes* ini menggunakan cabang matematika yang dikenal sebagai teori probabilitas untuk menemukan kemungkinan klasifikasi yang paling mungkin, dengan cara melihat frekuensi tiap klasifikasi pada data. Algortima *Naïve Bayes* memberikan suatu cara untuk mengkombinasikan peluang terdahulu dengan syarat kemungkinan menjadi sebuah formula yang dapat digunakan untuk menghitung peluan dari tiap kemungkinan yang akan terjadi. Dalam penelitian ini algortima klasifikasi *Naïve Bayes* yang akan digunakan adalah *Multinomial Naïve Bayes* dan *Complement Naïve Bayes.*

1. *Multinomial Naïve Bayes*

   Model ini termasuk pengembangan dari algortima klasifikasi *Naïve Bayes* dimana kelas tidak hanya ditentukan dengan kata yang muncul, tetapi juga

dengan jumlah kemunculan sehingga cocok untuk klasifikasi teks atau dokumen, terlebih lagi untuk ukuran dokumen yang besar.

2. *Complement Naïve Bayes*

Model pengembangan ini cenderung memiliki peforma yang lebih baik dalam klasifikasi ketika kelas dalam *dataset* tidak seimbang. Perbedaan yang signifikan dari *Multinomial Naïve Bayes* adalah dengan menggunakan data *Train* yang seimbang dari tiap kela ssehingga dapat mengurangi bias yang ada karena *dataset* tidak seimbang (Husada & Toba, 2020).

Naïve Bayes mengasumsikan bahwa keberadaan fitur dalam sebuah fitur dalam sebuah kelas tidak ada kaitannya dengan keberadaan fitur lainnya. Walaupun fitur tersebut berkaitan antara satu dengan yang lainnya, *Naïve Bayes* tetap menganggap fitur tersebut *independent* (Sastypratiwi et al., 2022). Dalam metode tersebut terdapat dua perhitungan utama yang akan dilakukan yaitu, perhitungan probabilitas kelas (*prior probability*) dan perhitungan probabilitas kondisional (*conditional probability*).

2.6 Pembobotan Kata

Pembobotan kata dilakukan dengan perhitungan Metode *TF-IDF* (*Term Frequency Inverse Document Frequency*). *TF-IDF* ini merupakan suatu metode yang digunakan untuk menentukan seberapa jauh keterhubungan kata terhadap dokumen dengan memberikan bobot di setiap kata. Metode ini menggabungkan 2 konsep yaitu frekuensi dari kemunculan suatu kata yang terdapat pada dokumen dan invers frekuensi dokumen yang mengandung kata tersebut (Ramadhan et al., 2021).

TF (*Term Frequency*) merupakan besarnya frekuensi kemunculan suatu kata dalam dokumen tertentu yang dapat menunjukkan seberapa penting kata dalam dokumen. Sedangkan IDF (*Invers Document Frequency*) menunjukkan tingkat kepentingan suatu kata dalam suatu Kumpulan dokumen, semakin besar nilai TF akan semakin kecil nilai IDF dari suatu kata. Berikut persamaan umum yang digunakan untuk menghitung pembobotan kata.

$$TF\ (d,t) = f(d,t) \tag{2.3}$$

Dimana :
t : Term (istilah
TF : Jumlah Term
$t\ (d,t)$: Frekuensi munculnya term t pada dokumen d

2.7 *Confusion Matrix*

Confusion Matrix adalah salah satu cara yang dapat digunakan untuk menilai hasil metodologi kategorisasi. Sistem klasifikasi informasi saat ini dan yang diantisipasi dimasukkan dalam matriks kebingungan. *Precision, recall, accuracy* dan *F1-Socre* merupakan salah satu metode untuk evaluasi klasifikasi yang bersifat biner. Pada penelitian ini, metode *Naïve Bayes* digunakan sebagai fitur klasifikasi teks pada cerita anak, sehingga metrik *precision*, *recall*, *accurancy* dan *F1-Score* dapat digunakan untuk mengevaluasi kinerja pada model yang menggunakan *Naïve Bayes.* Adanya metrik-metrik tersebut untuk membantu dalam mengukur seberapa baik model memprediksi kelas-kelas yang diinginkan (Azizah et al., 2023).

a. *Precission*

Precission merupakan nilai perbandingan dengan prediksi *True Positive* (TP) dengan keseluruhan data yang diprediksi positif. *Precission* digunakan untuk menetukan seberapa sering data benar ketika memprediksi positif.

$$Precision = \frac{TP}{(TP + FP)} \times 100 \quad (2.4)$$

b. *Recall*

Recall merupakan nilai perbandingan prediksi *True Positive* (TP) dengan keseluruhan data yang benar positif. *Recall* mendefinisikan bagaimana data benar untuk semua prediksi.

$$Recall = \frac{TP}{(TP + FN)} \quad (2.5)$$

c. *Accuracy*

Accuracy merupakan rasio prediksi benar (positif dan negative) dengan keseluruhan data.

$$Acxuracy = \frac{TP + FN}{Total\ Sampel} \quad (2.6)$$

d. *F1-Score*

F1-Score merupakan nilai rata-rata *precision* dan *recall* yang dibobotkan.

$$F1 - Score\ = \frac{2TP}{2TP + FN + FP} \quad (2.7)$$

Atau jika nilai *precision* dan *recall* sudah diketahui bisa menggunakan

$$F1 - Score\ = \frac{2(Precision\ \times Recall)}{Precision + Recall} \quad (2.8)$$

Dimana :
TP merupakan *True Positive*
FP merupakan *False Positive*
TN merupakan *True Negative*
FN merupakan *False Negative*

BAB III

DESAIN DAN IMPLEMENTASI

Penelitian ini akan dilakukan dalam beberapa tahap, menggunakan pendekatan yang metodis dan terstruktur untuk memungkinkan penulis untuk mendapatkan hasil yang diinginkan. Sebuah desain atau rencana sistem yang sesuai dengan kondisi penelitian diperlukan untuk mencapai tujuan penelitian.

3.1 Pengumpulan Data

Pada tahap ini penulis melakukan pengumpulan data yang akan digunakan pada penelitian tugas akhir. Pencarian data ini memanfaatkan platform *Kaggle* dari rentang Pebruari hingga Maret 2023. Dalam proses pencarian data, digunakan kata kunci berupa "Cerita Anak", dan "Fabel's Story" untuk mendapat dataset terkait dengan subjek penelitian ini. Pengumpulan data diambil dari platform kaggle https://www.kaggle.com/datasets/muhammadardiputra/aesop-fables-dataset. Data tersebut akan disimpan dalam format *Comma Separated Value* (CSV).

3.1.1 Input Dataset

Dataset yang digunakan dari *kaggle* sebanyak 57 cerita anak yang disimpan dalam file berformat csv. Dataset terdiri dari 25 cerita dengan label *anger*, 12 cerita dengan label *Fear*, 11 cerita dengan label *Happy*, dan 9 cerita dengan label *Sad*. Berikut adalah *dataset* yang akan digunakan dalam penelitian.

Tabel 3.1 Tampilan Input *Dataset*

id	title	label	cerita	emosi
1	The Wolf and The Kid	Fear	There was once a little Kid whose growing horns made him think he was a grown-up Billy Goat and	Happy, Sad,

Lanjutan Tabel Tampilan *Input Dataset*

id	title	label	cerita	emosi
			able to take care of himself. So one evening when the flock started home from the pasture and his mo...	Anger, Fear.
2	The Tortoise And The Ducks	Sad	The Tortoise. you know. carries his house on his back. No matter how hard he tries he cannot leave home. They say that Jupiter punished him so because he was such a lazy stay-at-home that he would...	Sad, Fear
3	Ants and Cocoons	Happy	In a shady forest. many wild and tame animals live. There were rabbits. birds. cats. dragonflies. butterflies and others. One day. the forest was hit by a terrible storm. The wind blew very hard. ...	Fear, Sad, Happy
4	Kera Became King	Anger	The King of the jungle "Lion" was shot by hunters. and the jungle dwellers became restless. They no longer had a king. Soon all the jungle dwellers gathered to elect a new King. The first person n...	Sad, Anger, Fear
5	Deer and Mouse	Sad	In the forest lived two deer. Their names were Kanca and Manggut. The two deer were brothers. Manggut was the older brother of Kanca. Kanca. on the other hand. was the younger brother of Manggut. ...	Sad, Anger
6	Salt Carrying Donkey	Sad	One summer day. there was a donkey walking in the mountains. The donkey was carrying several sacks of salt on its back. The sacks were very heavy. and the sun was shining brightly. "It's so hot. I...	Anger, Happy
7	Princess Jasmine Wangi	Sad	In a kingdom. there was a princess named Melati Wangi. She was a beautiful and clever princess. At home she always sang. But unfortunately she was arrogant and looked down on others…	Sad, Fear, Happy
8	The Deer has Been Hit by a Stone	Sad	The breeze made the forest dwellers sleepy. So did the deer. To get rid of his sleepiness he took a walk in the forest while puffing out his chest. While walking he said. "Who doesn't know Kancil....	Sad, Anger
9	Cowardly Bats	Anger	On a grassland in Africa. a Lion was eating a meal. Suddenly an eagle flew low and snatched the Lion's food. "Damn it". said the lion. The King of the forest was so angry that he ordered all the a...	Anger, Fear, Happy, Sad
10	Bende Wasiat	Happy	The tiger was reflecting in the river while washing his face. "Hmm. how dashing I am. my body is strong and muscular and the color of my skull is very beautiful." said the tiger to himself. The ti...	Happy, Anger, Fear

3.1.2 Pembagian Dataset

Sebelum masuk dalam pengujian, pertama kali data akan dibagi menjadi 2 yaitu data latih dan data uji dengan presentasi pada Tabel 3.2 sebagai berikut:

Tabel 3.2 *Split Dataset*

Data Latih	Data Uji
90%	10%
80%	20%

*0*Lanjutan Tabel *Split Dataset*

Data Latih	Data Uji
70%	30%

Berikut jumlah data yang akan dilakukan *Training* dan *Testing* pada Tabel 3.3 sebagai berikut :

Tabel 3.3 Jumlah Data

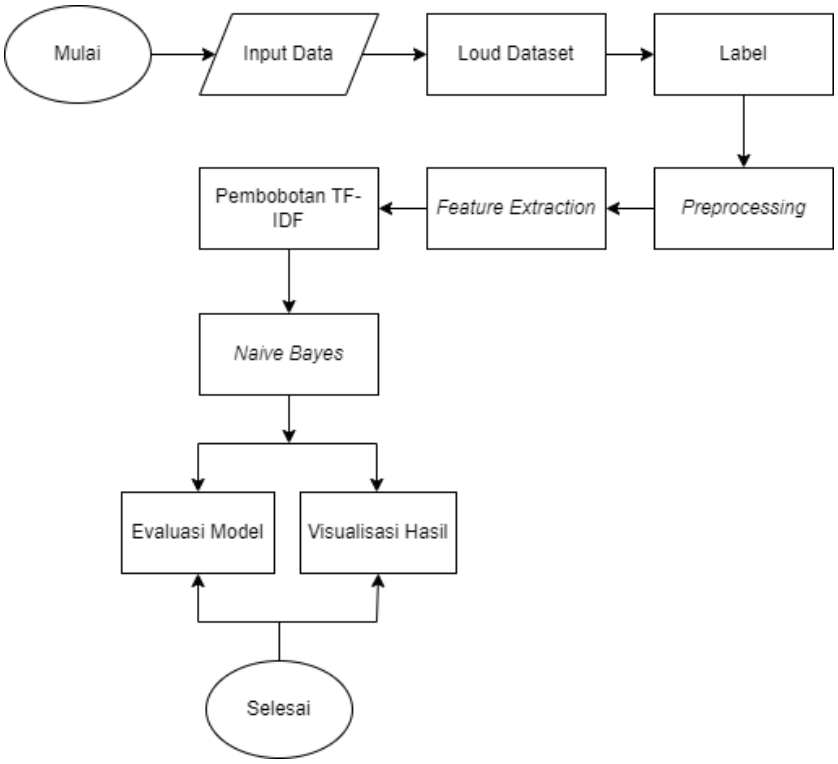
Split Data	Jumlah Data	
90 : 10	Jumlah X *Train*ing Data	51
	Jumlah X *Test*ing Data	6
	Jumlah Y *Train*ing Data	51
	Jumlah Y *Test*ing Data	6
80 : 20	Jumlah X *Train*ing Data	45
	Jumlah X *Test*ing Data	12
	Jumlah Y *Train*ing Data	45
	Jumlah Y *Test*ing Data	12
70 : 30	Jumlah X *Train*ing Data	39
	Jumlah X *Test*ing Data	18
	Jumlah Y *Train*ing Data	39
	Jumlah Y *Test*ing Data	18

Pembagian dataset pada Tabel 3.2 diatas diambil dengan tujuan untuk menganalisis manakah dari pembagian dataset tersebut yang memiliki performa terbaik merujuk pada penelitian (Saputra & Waluyo, 2022) yang memiliki performa terbaik dengan menggunakan beberapa uji pada data latih sebesar 90%, 80% dan 70% dan data uji sebesar 10%, 20% dan 30%. Dalam tahap evaluasi, model 70/30 tersebut memiliki kelebihan representasi yang baik terutama dalam mengolah data yang digunakan untuk mempelajari pola dari suatu data (Saputra & Waluyo, 2022).

3.2 Desain Sistem

Desain sistem merupakan gambaran alur dari sebuah sistem yang akan dibentuk. Sistem dimulai dari proses awal yaitu dengan input data berupa file *Comma Separated Value* (CSV) berisi kumpulan cerita anak. Kemudian,

melakukan langkah-langkah pra-pemrosesan teks. Selanjutnya, data dibagi menjadi data *Train* dan data *Test* untuk melatih dan menguji model. Setelah itu, data dilatih untuk mempelajari probabilitas kelas emosi berdasarkan fitur-fitur yang relevan. Model dilatih dan dievaluasi menggunakan data latih dan data uji untuk dihitung berapa akurasi disetiap teks cerita. Dengan urutan ini, skema menyajikan langkah-langkah dari pemrosesan awal hingga evaluasi model dalam analisis data teks.



Gambar 3.1 Alur Diagram

3.3 *Prepocessing* Data

Preprocessing memiliki tujuan untuk membersihkan data dan mengubah data yang tidak diperlukan dalam melakukan penelitian ini agar memiliki hasil yang lebih optimal. *Preprocessing* data yang dilakukan diantaranya yaitu, yang pertama

adalah menyesuaikan tipe data, karena saat melakukan pengecekan *dataset* terdapat tipe data yang tidak sesuai.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 57 entries, 0 to 56
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   id      57 non-null     int64
 1   title   57 non-null     object
 2   label   57 non-null     object
 3   cerita  57 non-null     object
 4   emosi   57 non-null     object
dtypes: int64(1), object(4)
memory usage: 2.4+ KB
```

Gambar 3.2 Hasil Pengecekan Tipe Data

Setelah dilakukan pengecekan tipe data, melakukan tahap menampilkan label dalam teks cerita anak sebagimana pada gambar berikut

```
df['label'].value_counts()
✓ 0.0s

label
Anger    25
Fear     12
Happy    11
Sad       9
Name: count, dtype: int64
```

Gambar 3.3 Label Teks Cerita

Setelah menampilkan label, lanjut tahapan *preprocessing* dalam penelitian ini terdapat beberapa tahap, yaitu:

3.3.1 *Case Folding*

Case Folding adalah proses mengubah perbedaan penggunaan bentuk huruf, pada bagian ini dilakukan perubahan diseluruh huruf kapital (*uppercase*) kemudian di ubah menjadi huruf kecil (*lowercase*) agar seragam.

```
import re
def casefolding(text):
  text = text.lower()
  text = re.sub(r'[^\w\s]','',text)
  text = re.sub(r'[-+]?[0-9]+','', text)
  text = text.strip()
  return text

raw_sample = df['cerita'].iloc[1]
case_folding = casefolding(raw_sample)

data_case_folding = pd.DataFrame({
    'Raw Data': [raw_sample],
    'Case Folding': [case_folding]
})

display(data_case_folding)
```

Dari pseudocode *case folding* diatas, menampilkan hasil *case folding* pada Tabel 3.4 sebagai berikut :

Tabel 3.4 Hasil *Case Folding*

Raw Data	Case Folding
The Tortoise. you know. carries his house on his back. No matter how hard he tries he cannot leave home. They say that Jupiter punished him so because he was such a lazy stay-at-home that he would not go to Jupiters wedding. even when especially invited. After many years. Tortoise began to wish he had gone to that wedding. When he saw how gaily the birds flew about and how the Hare and the Chipmunk and all the other animals ran nimbly by. always eager to see everything there was to be seen. the ortoise felt very sad and discontented. He wanted to see the world too. and there he was with a house on his back and little short legs that could hardly drag him along. One day he met a pair of Ducks and told them all his trouble. We can help you to see the world. said the Ducks. Take hold of this stick with your teeth and we will carry you far up in the air where you can see the whole countryside. But keep quiet or you will be sorry. The Tortoise was very glad indeed. He seized the stick firmly with his teeth. the two Ducks took hold of it one at each end. and away they sailed up toward the clouds. Just then a Crow flew by. He was very much astonished at the strange sight and cried: This must surely be the King of Tortoises! Why certainly began the Tortoise. But as he opened his mouth to say these foolish words he lost his hold on the stick. and down he fell to the ground. where he was dashed to pieces on a rock.	the tortoise you know carries his house on his back no matter how hard he tries he cannot leave home they say that jupiter punished him so because he was such a lazy stayathome that he would not go to jupiters wedding even when especially invited after many years tortoise began to wish he had gone to that wedding when he saw how gaily the birds flew about and how the hare and the chipmunk and all the other animals ran nimbly by always eager to see everything there was to be seen the ortoise felt very sad and discontented he wanted to see the world too and there he was with a house on his back and little short legs that could hardly drag him along one day he met a pair of ducks and told them all his trouble we can help you to see the world said the ducks take hold of this stick with your teeth and we will carry you far up in the air where you can see the whole countryside but keep quiet or you will be sorry the tortoise was very glad indeed he seized the stick firmly with his teeth the two ducks took hold of it one at each end and away they sailed up toward the clouds just then a crow flew by he was very much astonished at the strange sight and cried this must surely be the king of tortoises why certainly began the tortoise but as he opened his mouth to say these foolish words he lost his hold on the stick and down he fell to the ground where he was dashed to pieces on a rock

3.3.2 *Stop Word Removal*

Stop Word Removal adalah proses menghilangkan kata yang tidak diperlukan. Dengan menghapus *stop word removal*, dapat menyederhanakan teks dan fokus pada kata-kata yang lebih penting dalam analisis teks.

```
stopwords_en = stopwords_en

def remove_stop_words(text):
    stop_words = set(stopwords.words('english'))
    words = text.split()
    filtered_words = [word for word in words if word not in
stop_words]
    return ' '.join(filtered_words)
```

Dari pseudocode diatas, menampilkan hasil *stopwords removal* pada Tabel 3.5 sebagai berikut :

Tabel 3.5 Hasil *Stopwords Removal*

Case Folding	Stopwords Removal
the tortoise you know carries his house on his back no matter how hard he tries he cannot leave home they say that jupiter punished him so because he was such a lazy stayathome that he would not go to jupiters wedding even when especially invited after many years tortoise began to wish he had gone to that wedding when he saw how gaily the birds flew about and how the hare and the chipmunk and all the other animals ran nimbly by always eager to see everything there was to be seen the ortoise felt very sad and discontented he wanted to see the world too and there he was with a house on his back and little short legs that could hardly drag him along one day he met a pair of ducks and told them all his trouble we can help you to see the world said the ducks take hold of this stick with your teeth and we will carry you far up in the air where you can see the whole countryside but keep quiet or you will be sorry the tortoise was very glad indeed he seized the stick firmly with his teeth the two ducks took hold of it one at each end and away they sailed up toward the clouds just then a crow flew by he was very much astonished at the strange sight and cried this must surely be the king of tortoises why certainly began the tortoise but as	tortoise know carries house back matter hard tries cannot leave home say jupiter punished lazy stayathome would go jupiters wedding even especially invited many years tortoise began wish gone wedding saw gaily birds flew hare chipmunk animals ran nimbly always eager see everything seen ortoise felt sad discontented wanted see world house back little short legs could hardly drag along one day met pair ducks told trouble help see world said ducks take hold stick teeth carry far air see whole countryside keep quiet sorry tortoise glad indeed seized stick firmly teeth two ducks took hold one end away sailed toward clouds crow flew much astonished strange sight cried must surely king tortoises certainly began tortoise opened mouth say foolish words lost hold stick fell ground dashed pieces rock

Lanjutan Tabel 3.5 Hasil *Stopword Removal*

Case Folding	Stopwords Removal
he opened his mouth to say these foolish words he lost his hold on the stick and down he fell to the ground where he was dashed to pieces on a rock	

3.3.3 *Lemmatization*

Lemmatization merupakan proses untuk menghilangkan imbuhan kata agar kata tersebut dapat diubah ke bentuk dasarnya. Proses ini awalnya dilakukan dalam *text mining* dengan kosakata terbuka dengan tujuan untuk mempersingkat waktu komputasi dan meningkatkan kemampuan pengambilan informasi (Hickman et al., 2022).

Sebagai contoh, kata-kata seperti "running", "ran", dan "runs" semua dapat diubah menjadi bentuk dasar "run". Begitu pula, kata "better" dapat diubah menjadi "good". Proses ini lebih kompleks daripada stemming, yang hanya memotong akhiran kata tanpa memperhatikan konteks atau makna kata tersebut.

```
# Definisikan fungsi untuk lemmatization
def lemmatization(text):
    lemmatizer = WordNetLemmatizer()
    words = nltk.word_tokenize(text)
    lemmatized_words = [lemmatizer.lemmatize(word,
get_wordnet_pos(word)) for word in words]
    return ' '.join(lemmatized_words)

# Fungsi untuk mendapatkan part of speech (POS) tag
def get_wordnet_pos(word):
    """Map POS tag to first character lemmatize() accepts"""
    tag = nltk.pos_tag([word])[0][1][0].upper()
    tag_dict = {"J": wordnet.ADJ,
                "N": wordnet.NOUN,
                "V": wordnet.VERB,
                "R": wordnet.ADV}
    return tag_dict.get(tag, wordnet.NOUN)
```

Dari pseudocode diatas, menampilkan hasil *Lemmatzation* pada Tabel 3.6 sebagai berikut :

Tabel 3.6 Hasil *Lemmatization*

Stop Word Removal	Lemmatization
tortoise know carries house back matter hard tries cannot leave home say jupiter punished lazy stayathome would go jupiters wedding even especially invited many years tortoise began wish gone wedding saw gaily birds flew hare chipmunk animals ran nimbly always eager see everything seen ortoise felt sad discontented wanted see world house back little short legs could hardly drag along one day met pair ducks told trouble help see world said ducks take hold stick teeth carry far air see whole countryside keep quiet sorry tortoise glad indeed seized stick firmly teeth two ducks took hold one end away sailed toward clouds crow flew much astonished strange sight cried must surely king tortoises certainly began tortoise opened mouth say foolish words lost hold stick fell ground dashed pieces rock	tortoise know carry house back matter hard try leave home say jupiter punish lazy stayathome would go jupiter wedding even especially invite many year tortoise begin wish go wedding saw gaily bird flew hare chipmunk animal ran nimbly always eager see everything see ortoise felt sad discontent want see world house back little short leg could hardly drag along one day met pair duck told trouble help see world say duck take hold stick teeth carry far air see whole countryside keep quiet sorry tortoise glad indeed seize stick firmly teeth two duck take hold one end away sail toward cloud crow flew much astonish strange sight cry must surely king tortoise certainly begin tortoise open mouth say foolish word lose hold stick fell ground dash piece rock

Setelah melakukan proses *Preprocessing* Data, selanjutnya melakukan proses *Text Preprocessing Pipeline* yang sangat penting dalam analisis teks karena memastikan bahwa data yang telah melalui *preprocessing* tersebut siap digunakan oleh algoritma pemodelan. Hasil dari *Text Preprocessing Pipeline* diberi nama *clean_text* yang dapat meningkatkan akurasi dan kinerja model. Hasil dari *Text Preprocessing Pipeline* seperti tabel berikut

Tabel 3.7 *Preprocessing Pipeline*

title	label	cerita	emosi	label_id	clean_text
The Wolf and The Kid	Fear	There was once a little Kid whose growing horns made him think he was a grown-up Billy Goat and able to take care of himself. So one evening when the flock started home from the pasture and his mo...	Happy, Sad, Anger, Fear	1	little kid whose grow horn make think grownup billy goat able take care one even flock start home pasture mother call kid paid heed kept right nibble tender grass little later lift head flock go a...

Lanjutan Tabel 3.7 *Preprocessing Pipeline*

title	label	cerita	emosi	label_id	clean_text
The Tortoise and The Ducks	Sad	The Tortoise. you know. carries his house on his back. No matter how hard he tries he cannot leave home. They say that Jupiter punished him so because he was such a lazy stay-at-home that he would...	Sad, Fear	3	tortoise know carry house back matter hard try leave home say jupiter punish lazy stayathome would go jupiter wedding even especially invite many year tortoise begin wish go wedding saw gaily bird..
Ants and Cocoons	Happy	In a shady forest. many wild and tame animals live. There were rabbits. birds. cats. dragonflies. butterflies and others. One day. the forest was hit by a terrible storm. The wind blew very hard. ...	Fear, Sad, Happy	2	shady forest many wild tame animal live rabbit bird cat dragonfly butterfly others one day forest hit terrible storm wind blew hard hit tree leaf kraak sound branch break many animal unable save e...
Kera Became King	Anger	The King of the jungle "Lion" was shot by hunters. and the jungle dwellers became restless. They no longer had a king. Soon all the jungle dwellers gathered to elect a new King. The first person n...	Sad, Anger, Fear	0	king jungle lion shot hunter jungle dweller become restless longer king soon jungle dweller gather elect new king first person nominate leopard leopard refuse look human run away say rhino strong ...
Deer and Mouse	Sad	In the forest lived two deer. Their names were Kanca and Manggut. The two deer were brothers. Manggut was the older brother of Kanca. Kanca. on the other hand. was the younger brother of Manggut. ...	Sad, Anger	3	forest live two deer name kanca manggut two deer brother manggut old brother kanca kanca hand young brother manggut although brother character different kanca diligent kind whereas manggut lazy li...

3.4 Proses Metode *Naïve Bayes*

Metode *Naïve Bayes* untuk klasifikasi emosi dalam teks cerita anak menggunakan pendekatan text mining melibatkan beberapa langkah penting yang

harus diikuti secara sistematis. Pada awalnya, fase pengumpulan data melibatkan pengadaan bahan naratif untuk anak-anak dari berbagai sumber seperti karya sastra dan platform online. Selanjutnya, proses ini mencakup anotasi untuk mengidentifikasi dan mengkarakterisasi teks atau frasa yang menyampaikan keadaan emosional tertentu seperti kegembiraan, kesedihan, frustrasi, dan ketakutan. Setelah fase ini, dokumen yang dikumpulkan menjalani langkah-langkah pemrosesan awal. *Case Folding* digunakan untuk mengubah perbedaan bentuk huruf kapital (*uppercase*) menjadi huruf kecil (*lowercase*). Lalu *Stop Word Removal* merupakan proses menghilangkan kata-kata yang tidak digunakan. Kemudian dilanjutkan dengan *Lemmatization* adalah proses menghilangkan imbuhan kata agar kata tersebut dapat diubah ke bentuk dasarnya. Proses-proses ini secara kolektif merupakan langkah-langkah penting dalam pra-pemrosesan teks.

Selanjutnya, teks dikonversi menjadi format numerik yang dapat dianalisis oleh algoritma *Naïve Bayes* pada tahap ekstraksi fitur. Untuk ini, representasi *TF-IDF* (*Term Frequency-Inverse Document Frequency*) biasanya digunakan. Data kemudian dibagi menjadi data latih dan data uji seperti pada pseudocode dibawah ini.

```
# import library
import random
from sklearn.model_selection import Train_Test_split

# Pembagian data Training dan Testing
X = selected_x
y = df.label
X_Train, X_Test, y_Train, y_Test = Train_Test_split(X, y,
Test size=0.2, random state=0)
```

Menggunakan data latih, model *Naïve Bayes* dilatih. Model ini kemudian menentukan kemungkinan munculnya sebuah kata dalam setiap kategori emosi.

Setelah itu, pengujian model dilakukan menggunakan data uji untuk menilai kinerja model dalam memprediksi emosi dari teks cerita anak berdasarkan probabilitas yang dihitung selama pelatihan.

Kinerja model dinilai menggunakan ukuran seperti *F1-score*, *accuracy*, *precision*, dan *recall*. Matriks kebingungan (*Confusion Matrix*) juga digunakan untuk memvisualisasikan distribusi prediksi yang benar dan salah. Setelah model dievaluasi dan mencapai kinerja yang memuaskan, model tersebut dapat diimplementasikan untuk menganalisis emosi dalam teks cerita anak yang baru. Berikut merupakan pseudocode untuk mengetahui *Confusion Matrix*.

```
from sklearn.metrics import confusion_matrix,
classification_report

# Melakukan prediksi
predicted = model.predict(X_Test)

# Menghitung Confusion Matrix
CM = confusion_matrix(y_Test, predicted)

# Menampilkan classification report
print(classification_report(y_Test, predicted, zero_division=1))
```

Setiap tahap dalam proses ini sangat penting untuk menjamin bahwa model dapat dengan tepat dan berhasil mengkategorikan emosi dalam teks cerita anak serta menawarkan analisis yang mendalam tentang konten emosional dalam literatur anak-anak.

3.5 Desain Eksperimen

Dalam desain eksperimen ini dilakukan untuk mengevaluasi mode yang telah dibuat dan bertujuan untuk mengetahui kombinasi parameter mana yang

menghasilkan akurasi terbaik. Untuk kombinasi parameter terbaik yang digunakan dalam analisis emosi pada teks cerita anak tersebut terdapat pada tabel berikut

Tabel 3.8 Kombinasi Parameter Terbaik

No	Jenis Parameter	Teknik Pra-Pemrosesan	Teknik Ekstraksi Fitur
1	*Multinomial Naïve Bayes* (MNB)	*-Case Folding* *-Stopwords Removal* *- Lemmatization*	*TF-IDF*
2	*Complement Naïve Bayes* (CNB)		

Adapun Skenario Uji coba untuk mengetahui kombinasi parameter yang menghasilkan akurasi terbaik. Pembagian scenario uji coba terdapat pada tabel berikut :

Tabel 3.9 Skenario Uji

Train-Test Split	Ratio Split Data	Tuning Hyperparameter
	90 : 10 80 : 20 70 : 30	

Berdasarkan scenario di atas, pengujian akan dilakukan dengan melibatkan eksplorasi kinerja model dengan variasi jumlah fitur dan penyesuaian nilai hyperparameter. Detail nilai dari tiap parameter akan digunakan dalam skenario pengujian adalah sebagai berikut:

Tabel 3.10 Nilai *Hyperparameter*

Parameter	Nilai
MultinomialNB	*[90,10], [80,20], [70,30]*
ComplementNB	*[90,10], [80,20], [70,30]*

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Uji Coba

Proses pengklasifikasian dan pengujian model dalam penelitian menggunakan bahasa pemograman python dengan menggunakan fungsi *MultinomialNB* dan *ComplementNB* juga proses evaluasi *classification report* dan *Confusion Matrix* untuk menampilkan nilai akurasi pada model.

4.1.1 Proses pengujian dengan model 90:10

1. Skenario menggunakan data *Train MultinomialNB*

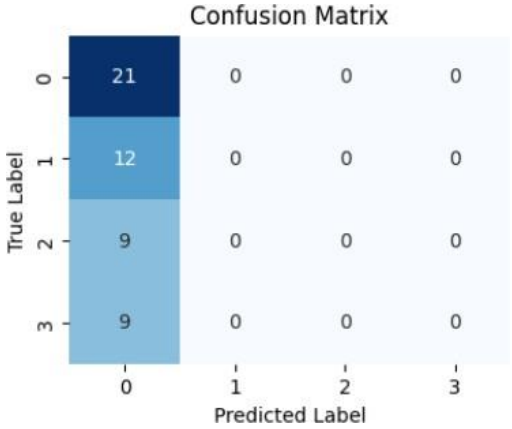
Hasil *accuracy* yang ditampilkan tiap label pada data *Train* dengan menggunakan algoritma *MultinomialNB* memiliki jumlah yang berbeda, Hasil klasifikasi tiap label dapat dilihat pada Tabel 4.1 sebagai berikut :

Tabel 4.1 *Classification Report Data Train Multinomial Naïve Bayes 90:10*

Label	*Precision*	*Recal*	*F1-Score*
Anger **[0]**	0.41	1.00	0.56
Fear **[1]**	0.00	0.00	0.00
Happy **[2]**	0.00	0.00	0.00
Sad **[3]**	0.00	0.00	0.00
Accuracy	0.41		

Berdasarkan tabel 4.1 diatas, kinerja yang ditunjukan model bekerja cukup baik dalam mengklasifikasikan emosi kelas 0 (*Anger*), namun kurang baik dalam mengklasifikasikan emosi lainnya seperti kelas 1 (*Fear*), kelas 2 (*Happy*), dan kelas 3 (*Sad*). Hal ini mengindikasikan adanya ketidakseimbangan dalam data pelatihan atau model yang belum cukup

kompleks untuk menangkap nuansa emosi yang berbeda. *Accuracy* keseluruhan model juga masih tergolong rendah, menunjukkan adanya ruang untuk perbaikan.



Gambar 4.1 *Confusion Matrix Data Train Multinomial Naïve Bayes 90:10*

Pada Gambar 4.1 diatas, menunjukkan seberapa baik model dalam memprediksi kelas sebenarnya dari suatu data. Model sangat baik dalam memprediksi kelas 0 (*Anger*), namun dalam memprediksi kelas 1 (*Fear*), 2 (*Happy*) dan 3 (*Sad*) sama sekali tidak dapat memprediksi dengan benar. Hal ini mengindikasikan adanya masalah dalam model, seperti data pelatihan yang tidak seimbang atau fitur yang tidak relevan untuk kelas 1, 2, dan 3. Untuk meningkatkan kinerja model, perlu dilakukan analisis lebih lanjut terhadap data dan model, serta dilakukan penyesuaian seperti penambahan data, pemilihan fitur yang lebih baik, atau penggunaan algoritma yang lebih kompleks.

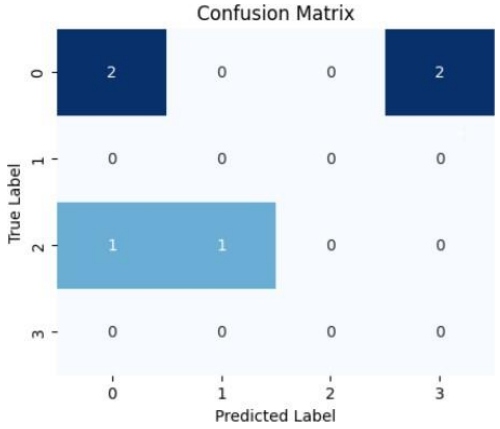
2. Skenario menggunakan data *Test MultinomialNB*

Hasil *accuracy* yang ditampilkan tiap label pada data *Test* dengan menggunakan algoritma *MultinomialNB* memiliki jumlah yang berbeda. Hasil klasifikasi tiap label dapat dilihat pada Tabel 4.2 sebagai berikut :

Tabel 4.2 *Classification Report Data Test Multinomial Naïve Bayes 90:10*

Label	***Precision***	***Recal***	***F1-Score***
Anger **[0]**	0.67	0.50	0.57
Fear **[1]**	0.00	1.00	0.00
Happy **[2]**	1.00	0.00	0.00
Sad **[3]**	0.00	1.00	0.00
Accuracy	0.67		

Berdasarkan Tabel 4.2 diatas, model menunjukkan kinerja yang baik dalam mengklasifikasikan kelas 0 (*Anger*), dengan nilai *precision*, *recall*, dan *F1-score* yang tinggi. Namun, model kurang baik dalam mengklasifikasikan emosi pada kelas 2 (*Happy*). *Accuracy* keseluruhan model masih tergolong cukup baik, tetapi dominan dipengaruhi oleh kinerja yang sangat baik pada kelas 0 (*Anger*).



Gambar 4.2 *Confusion Matrix Data Test Multinomial Naïve Bayes 90:10*

Pada Gambar 4.2 diatas, model memiliki kinerja yang cukup baik dalam memprediksi kelas 0 (*Anger*). Dari 4 data yang sebenarnya berlabel 0, model berhasil memprediksi semuanya dengan benar. Namun, model mengalami kesulitan dalam memprediksi kelas 1 (*Fear*). Dari 2 data yang sebenarnya berlabel 2 (*Happy*), model salah memprediksi semuanya sebagai kelas 0 (*Anger*).

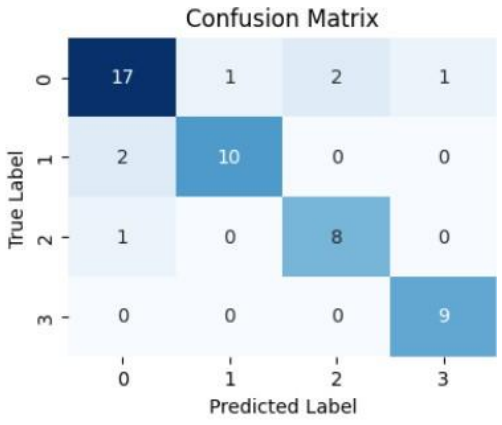
3. Skenario menggunakan data *Train ComplementNB*

Hasil *accuracy* yang ditampilkan tiap label pada data *Train* dengan menggunakan algoritma *ComplementNB* memiliki jumlah yang berbeda. Hasil klasifikasi tiap label dapat dilihat pada Tabel 4.3 sebagai berikut :

Tabel 4.3 *Classification Report Data Train Complement Naïve Bayes 90:10*

Label	***Precision***	***Recal***	***F1-Score***
Anger **[0]**	0.85	0.81	0.83
Fear **[1]**	0.91	0.83	0.87
Happy **[2]**	0.80	0.89	0.84
Sad **[3]**	0.90	1.00	0.95
Accuracy	0.86		

Berdasarkan Tabel 4.3 diatas, model menunjukkan kinerja yang cukup baik dalam mengklasifikasikan semua kategori emosi. Nilai *precision, recall,* dan *F1-score* yang tinggi untuk setiap kelas mengindikasikan bahwa model dapat mengidentifikasi dengan baik dan lengkap contoh-contoh dari setiap emosi. *Accuracy* keseluruhan model juga tinggi, menunjukkan bahwa model secara umum dapat melakukan klasifikasi dengan baik.



Gambar 4.3 *Confusion Matrix Data Train Complement Naïve Bayes 90:10*

Pada Gambar 4.3 diatas, model umumnya memiliki kinerja yang cukup baik dalam memprediksi kelas yang benar. Dari 17 data yang sebenarnya berlabel 0 (*Anger*), model berhasil memprediksi 17 data dengan benar. Namun, ada beberapa model yang melakukan kesalahan dalam memprediksi. Ada 2 data yang berlabel 1 (*Fear*), namun model salah memprediksi sebagai kelas 0 (*Anger*).

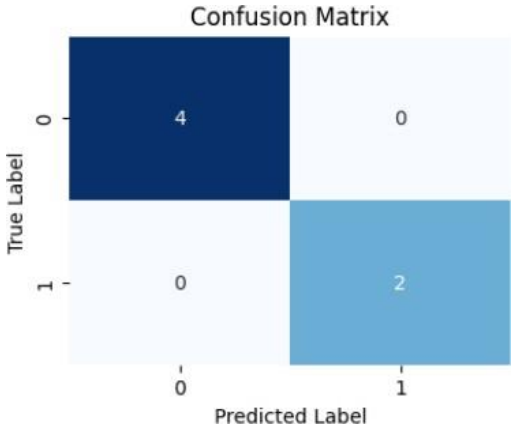
4. Skenario menggunakan data *Test ComplementNB*

Hasil *accuracy* yang ditampilkan tiap label pada data *Test* dengan menggunakan algoritma *ComplementNB* memiliki jumlah yang berbeda. Hasil klasifikasi tiap label dapat dilihat pada Tabel 4.4 sebagai berikut :

Tabel 4.4 *Classification Report Data Test Complement Naïve Bayes 90:10*

Label	***Precision***	***Recal***	***F1-Score***
Anger **[0]**	1.00	1.00	1.00
Happy **[2]**	1.00	1.00	1.00
Accuracy	1.00		

Berdasarkan Tabel 4.4 diatas, model menunjukkan kinerja yang baik dalam mengklasifikasikan kedua kategori emosi, yaitu kelas 0 (*Anger*) dan kelas 2 (*Happy*). Nilai *precision*, *recall*, dan *F1-score* yang semuanya bernilai 1.00 untuk setiap kelas mengindikasikan bahwa model dapat mengidentifikasi dengan baik dan lengkap contoh-contoh dari setiap emosi. *Accuracy* keseluruhan model juga mencapai 100%, menunjukkan bahwa model dapat melakukan klasifikasi dengan baik.



Gambar 4.4 *Confusion Matrix Data Test Complement Naïve Bayes 90:10*

Pada Gambar 4.4 diatas, model memiliki kinerja yang baik dalam memprediksi kelas. Model berhasil memprediksi semuanya dengan benar dari 4 data yang berlabel 0 (*Anger*). Begitu juga dengan 2 data yang berlabel 1 (*Fear*), model juga berhasil memprediksi semuanya dengan benar. Hal ini menunjukkan bahwa model telah belajar pola dalam data dengan sangat baik dan mampu mengklasifikasikan data dengan *accuracy* yang tinggi.

4.1.2 Proses Pengujian dengan model 80:20

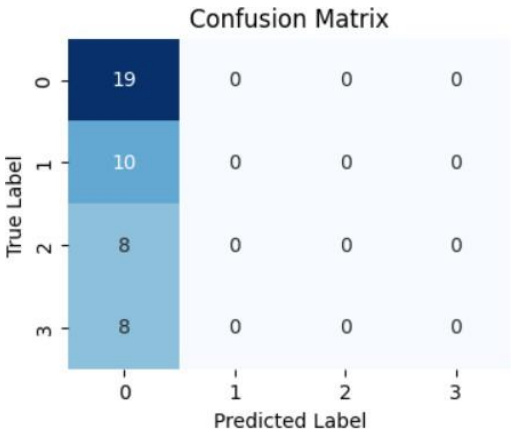
1. Skenario menggunakan data *Train MultinomialNB*

Hasil *accuracy* yang ditampilkan tiap label pada data *Train* dengan menggunakan algoritma *MultinomialNB* memiliki jumlah yang berbeda, Hasil klasifikasi tiap label dapat dilihat pada Tabel 4.5 sebagai berikut :

Tabel 4.5 *Classification Report Data Train Multinomial Naïve Bayes 80:20*

Label	*Precision*	*Recal*	*F1-Score*
Anger **[0]**	0.42	1.00	0.59
Fear **[1]**	0.00	0.00	0.00
Happy **[2]**	0.00	0.00	0.00
Sad **[3]**	0.00	0.00	0.00
Accuracy	0.42		

Berdasarkan Tabel 4.5 diatas, model menunjukkan kinerja yang buruk dalam mengklasifikasikan semua kategori emosi, kecuali pada kelas 0 (*Anger*). Nilai *precision*, *recall*, dan *F1-score* yang rendah untuk kelas 1 (*Fear*), 2 (*Happy*), dan 3 (*Sad*) bahwa emosi tersebut sama sekali tidak dapat diidentifikasikan. *Accuracy* keseluruhan model juga rendah, menunjukkan bahwa model secara umum tidak dapat melakukan klasifikasi dengan baik.



Gambar 4.5 *Confusion Matrix Data Train Multinomial Naïve Bayes 80:20*

Pada Gambar 4.5 diatas, model memiliki kinerja yang baik dalam memprediksi tiga kelas yaitu kelas 0 (*Anger*), 1 (*Fear*) dan 2 (*Sad*). Dari 19 data yang berlabel 0 (*Anger*), model berhasil memprediksi semuanya dengan benar. Model juga berhasil memprediksi 10 data yang berlabel 1 (*Fear*) dan 8 data yang berlabel 2 (*Happy*) dengan benar. Namun, model mengalami kesulitan dalam memprediksi kelas 3 (*Sad*). Dari 8 data yang berlabel 3 (*Sad*), model salah memprediksi semuanya sebagai kelas lain.

2. Skenario menggunakan data Test MultinomialNB

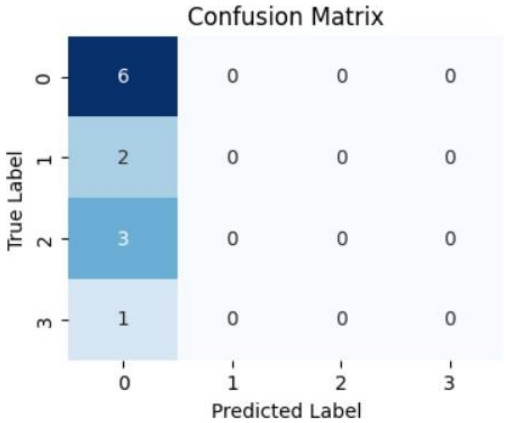
Hasil *accuracy* yang ditampilkan tiap label pada data *Test* dengan menggunakan algoritma *MultinomialNB* memiliki jumlah yang berbeda. Hasil klasifikasi tiap label dapat dilihat pada Tabel 4.6 sebagai berikut :

Tabel 4.6 *Classification Report Data Test Multinomial Naïve Bayes 80:20*

Label	***Precision***	***Recal***	***F1-Score***
Anger **[0]**	0.50	1.00	0.67
Fear **[1]**	0.00	0.00	0.00
Happy **[2]**	0.00	0.00	0.00
Sad **[3]**	0.00	0.00	0.00
Accuracy	0.50		

Berdasarkan Tabel 4.6 diatas, model menunjukkan kinerja yang tidak konsisten dalam mengklasifikasikan emosi. Model hanya mampu mengidentifikasi semua contoh emosi pada kelas 0 (*Anger*) dengan benar, tetapi *precision* yang didapat hanya 0.50, berarti tidak semua prediksi kelas 0 (*Anger*) benar-benar merupakan kelas 0 (*Anger*). Model sama sekali tidak dapat mengidentifikasi untuk kelas emosi 1 (*Fear*), *2* (*Happy*) dan 3 (*Sad*). Hal ini

ditunjukkan oleh nilai *precision*, *recall*, dan *F1-score* yang semuanya 0.00. *Accuracy* keseluruhan model juga rendah (0.50), menunjukkan bahwa model secara umum tidak dapat melakukan klasifikasi dengan baik. Hasil ini mengindikasikan bahwa model masih perlu banyak perbaikan, terutama dalam mengidentifikasi emosi selain 0 (*Anger*).



Gambar 4.6 *Confusion Matrix Data Test Multinomial Naïve Bayes 80:20*

Pada Gambar 4.6 diatas, model memiliki kinerja yang baik dalam memprediksi kelas. Dari 6 data yang berlabel 0 (*Anger*), model berhasil memprediksi semuanya dengan benar. Begitu juga dengan 2 data yang berlabel 1 (*Fear*), 3 data yang berlabel 2 (*Happy*), dan 1 data yang berlabel 3 (*Sad*), model juga berhasil memprediksi semuanya dengan benar.

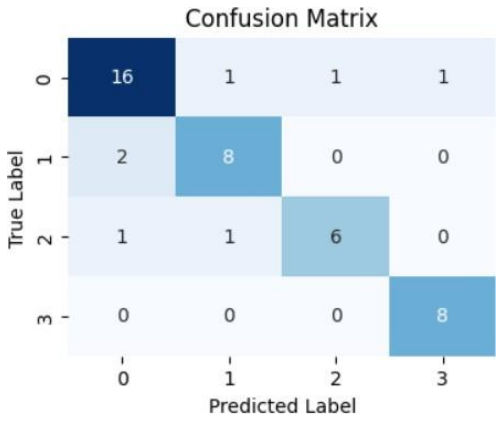
3. Skenario menggunakan data *Train ComplementNB*

Hasil *accuracy* yang ditampilkan tiap label pada data *Train* dengan menggunakan algoritma *ComplementNB* memiliki jumlah yang berbeda. Hasil klasifikasi tiap label dapat dilihat pada Tabel 4.7 sebagai berikut :

Tabel 4.7 *Classification Report Data Train Complement Naïve Bayes 80:20*

Label	*Precision*	*Recal*	*F1-Score*
Anger **[0]**	0.84	0.84	0.84
Fear **[1]**	0.80	0.80	0.80
Happy **[2]**	0.86	0.75	0.80
Sad **[3]**	0.89	1.00	0.94
Accuracy	0.84		

Berdasarkan tabel 4.7 diatas, model menunjukkan kinerja yang sangat baik dalam mengklasifikasikan semua kategori emosi. *Nilai precision, recall,* dan *F1-score* yang tinggi untuk semua kelas bahwa model dapat mengidentifikasi dengan baik dan lengkap contoh-contoh dari setiap emosi. *Accuracy* keseluruhan model juga sangat tinggi (0.84), menunjukkan bahwa model secara umum dapat melakukan klasifikasi dengan baik.



Gambar 4.7 *Confusion Matrix Data Train Complement Naïve Bayes 80:20*

Pada Gambar 4.7 diatas, model memiliki kinerja yang cukup baik dalam memprediksi kelas, terutama untuk kelas 0 (*Anger*), 1 (*Fear*) dan 3 (*Sad*). Namun, model masih mengalami sedikit kesulitan dalam membedakan kelas 0

(*Anger*) dan 2 (*Happy*). Hal ini terlihat dari adanya beberapa data kelas 2 (*Happy*) yang salah diprediksi sebagai kelas 0 (*Anger*).

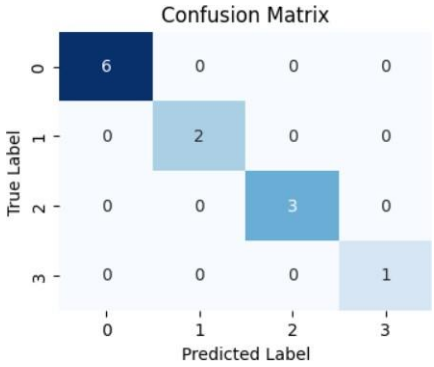
4. Skenario menggunakan data *Test ComplementNB*

Hasil *accuracy* yang ditampilkan tiap label pada data *Test* dengan menggunakan algoritma *Complement*NB memiliki jumlah yang berbeda. Hasil klasifikasi tiap label dapat dilihat pada Tabel 4.8 sebagai berikut :

Tabel 4.8 *Classification Report Data Test Complement Naïve Bayes 80:20*

Label	***Precision***	***Recal***	***F1-Score***
Anger **[0]**	1.00	1.00	1.00
Fear **[1]**	1.00	1.00	1.00
Happy **[2]**	1.00	1.00	1.00
Sad **[3]**	1.00	1.00	1.00
Accuracy	1.00		

Berdasarkan tabel 4.8 diatas, model menunjukkan kinerja yang sangat baik dalam mengklasifikasikan semua kategori emosi. Nilai *precision, recall,* dan *F1-score* yang semuanya 1.00 untuk semua kelas mengindikasikan bahwa model dapat mengidentifikasi dengan sangat baik dan lengkap contoh-contoh dari setiap emosi. Akurasi keseluruhan model juga 1.00, menunjukkan bahwa model secara sempurna dapat melakukan klasifikasi.



Gambar 4.8 *Confusion Matrix Data Test Complement Naïve Bayes 80:20*

Pada Gambar 4.8 diatas, model memiliki kinerja yang sangat baik dalam memprediksi kelas sebenarnya untuk semua kelas. Model berhasil memprediksi semuanya dengan benar dari 6 data yang berlabel 0 (*Anger*). Begitu juga dengan 2 data yang berlabel 1 (*Fear*), 3 data yang sebenarnya berlabel 2 (*Happy*), dan 1 data yang sebenarnya berlabel 3 (*Sad*), model juga berhasil memprediksi semuanya dengan benar.

4.1.3 Proses Pengujian dengan model 70:30

1. Skenario menggunakan data *Train MultinomialNB*

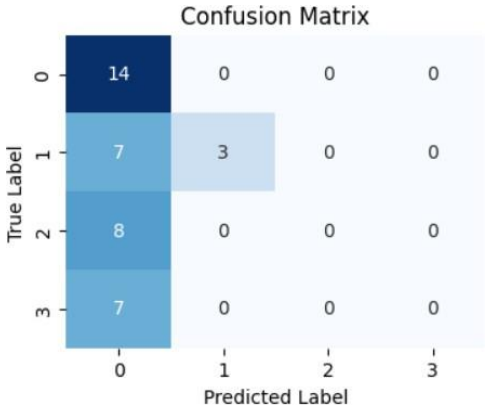
Hasil *accuracy* yang ditampilkan tiap label pada data *Train* dengan menggunakan algoritma *MultinomialNB* memiliki jumlah yang berbeda. Hasil klasifikasi tiap label dapat dilihat pada Tabel 4.9 sebagai berikut :

Tabel 4.9 *Classification Report* Data *Train Multinomial Naïve Bayes 70:30*

Label	***Precision***	***Recal***	***F1-Score***
Anger **[0]**	0.39	1.00	0.56
Fear **[1]**	1.00	0.30	0.46
Happy **[2]**	0.00	0.00	0.00
Sad **[3]**	0.00	0.00	0.00
Accuracy	0.44		

Pada Tabel 4.9 diatas, nilai dalam tabel menunjukkan kinerja model untuk setiap kelas. Contohnya, untuk kelas 0 (*Anger*), *precision* model adalah 0.39, *recall* model adalah 1.00, dan *F1-score* model adalah 0.56. Artinya, model mengklasifikasikan 39% contoh pada kelas 0 (*Anger*) dengan benar, dan 100% contoh "*Anger*" yang sebenarnya diidentifikasi dengan benar oleh model. *F1-score* 0.56 menunjukkan keseimbangan yang baik antara *precision* dan *recall*.

Accuracy keseluruhan model adalah 0.35, yang berarti model mengklasifikasikan 35% dari semua contoh dengan benar. Rata-rata terbobot menghitung rata-rata metrik ini dengan mempertimbangkan jumlah contoh dalam setiap kelas.



Gambar 4.9 *Confussion Matrix Data Train Multinomial Naïve Bayes 70:30*

Pada Gambar 4.9 diatas, model klasifikasi untuk data *Train*. Dalam model pada gambar, model bekerja sangat baik pada kelas 1 (*Fear*), tetapi dapat ditingkatkan pada kelas 2 (*Happy*) dan 3 (*Sad*). Hal ini dapat dilakukan dengan mengumpulkan lebih banyak data untuk kelas 2 (*Happy*) dan 3 (*Sad*), atau dengan menggunakan algoritma klasifikasi yang berbeda.

2. Skenario menggunakan data *Test MultinomialNB*

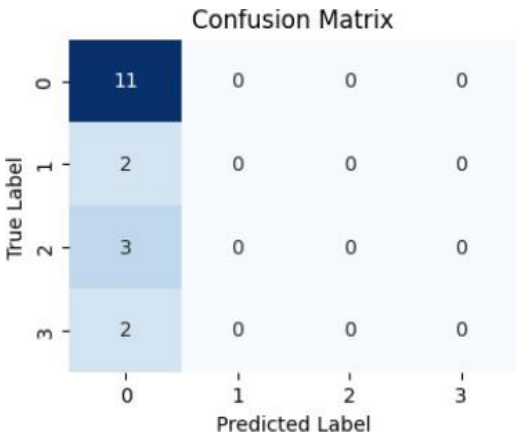
Hasil *accuracy* yang ditampilkan tiap label pada data *Test* dengan menggunakan algoritma *MultinomialNB* memiliki jumlah yang berbeda. Hasil klasifikasi tiap label dapat dilihat pada Tabel 4.10 sebagai berikut :

Tabel 4.10 *Classification Report* Data *Test Multinomial Naïve Bayes 70:30*

Label	***Precision***	***Recal***	***F1-Score***
Anger **[0]**	0.61	1.00	0.76
Fear **[1]**	0.00	0.00	0.00
Happy **[2]**	0.00	0.00	0.00
Sad **[3]**	0.00	0.00	0.00
Accuracy	0.61		

Pada Tabel 4.10 diatas, akurasi klasifikasi emosi adalah 0.15. Hal ini berarti bahwa hanya 15% dari prediksi yang benar. *Precision* menunjukkan proporsi prediksi positif yang benar-benar positif. Pada model klasifikasi emosi paling *precision* atau model lebih akurat dalam mengidentifikasikan kelas 0 (*Anger*). Pada *recall* proporsi pada model klasifikasi emosi untuk kelas 0 (*Anger*), karena model tidak melewatkan sampelnya. *F1-score* rata-rata harmonis *precision* dan *recall*. *F1-score* yang tinggi menunjukkan kinerja klasifikasi yang baik. Dalam hal ini, model klasifikasi emosi memiliki *F1-score* terbaik untuk kelas 0 (*Anger*).

Kelas 0 (*Anger*) memiliki performa terbaik dengan *precision, recall* dan *F1-score* tertinggi, serta dukungan sampel terbanyak. Sedangkan kelas 1 (*Fear*), 2 (*Happy*) dan 3 (*Sad*) memiliki *precision, recall* dan *F1-score* yang rendah, menunjukkan model kesulitan mengklasifikasikan emosi pada kelas tersebut.



Gambar 4.10 *Confusion Matrix Data Test Multinomial Naïve Bayes 70:30*

Pada Gambar 4.10 diatas, performa model secara keseluruhan dapat dilihat dari rata-rata *accuracy*. Dalam penjelasan *classification report* pada data *Test* tersebut, dalam kasus model pada gambar, model bekerja sangat baik pada kelas 1 (*Fear*), tetapi dapat ditingkatkan pada kelas 2 (*Happy*) dan 3 (*Sad*).

3. Skenario menggunakan data *Train ComplementNB*

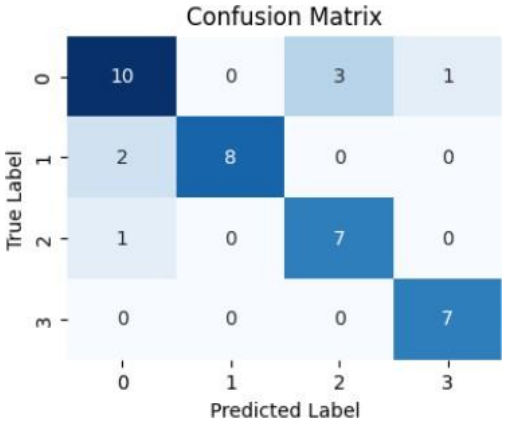
Hasil *accuracy* yang ditampilkan tiap label pada data *Train* dengan menggunakan algoritma *ComplementNB* memiliki jumlah yang berbeda. Hasil klasifikasi tiap label dapat dilihat pada Tabel 4.11 sebagai berikut :

Tabel 4.11 *Confussion Report* Data *Train Complement Naïve Bayes 70:30*

Label	***Precision***	***Recal***	***F1-Score***
Anger **[0]**	0.77	0.71	0.74
Fear **[1]**	1.00	0.80	0.89
Happy **[2]**	0.70	0.88	0.78
Sad **[3]**	0.88	1.00	0.93
Accuracy	0.82		

Pada Tabel 4.11 diatas, model bekerja paling baik pada Kelas 1 (*Fear*), dengan akurasi 90%. Ini berarti 90% sampel yang sebenarnya berada di Kelas 1 (*Fear*) diprediksi dengan benar. Klasifikasi bekerja paling buruk pada Kelas 3 (*Sad*), dengan akurasi 60%. Ini berarti hanya 60% sampel yang sebenarnya berada di Kelas 3 (*Sad*) diprediksi dengan benar.

Pada Gambar 4.11, terdapat empat kelas yaitu kelas 0 (*Anger*), kelas 1 (*Fear*), kelas 2 (*Happy*) dan kelas 3 (*Sad*). *Accuracy* keseluruhan klasifikasi adalah 82,5%, sampel diklasifikasikan dengan benar.



Gambar 4.11 *Confusion Matrix Data Train Complement Naïve Bayes 70:30*

4. Skenario menggunakan data *Test ComplementNB*

Hasil *accuracy* yang ditampilkan tiap label pada data *Test* dengan menggunakan algoritma *ComplementNB* memiliki jumlah yang berbeda. Hasil klasifikasi tiap label dapat dilihat pada Tabel 4.12 sebagai berikut :

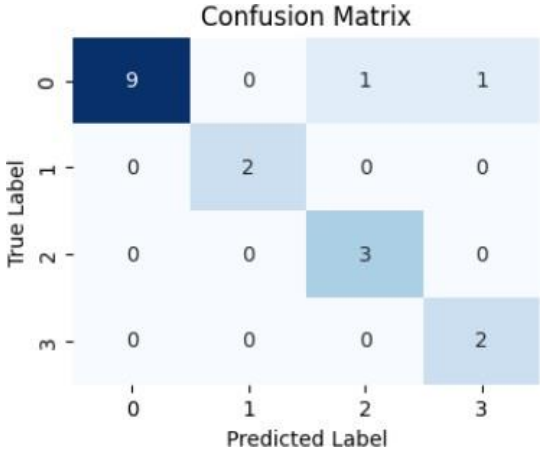
Tabel 4.12 *Confussion Report Data Test Complement Naïve Bayes 70:30*

Label	*Precision*	*Recal*	*F1-Score*
Anger **[0]**	1.00	0.82	0.90
Fear **[1]**	1.00	1.00	1.00

*0*Lanjutan Tabel *Confussion Report Data Test Complement Naïve Bayes 70:30*

Label	*Precision*	*Recal*	*F1-Score*
Happy **[2]**	0.75	1.00	0.86
Sad **[3]**	0.67	1.00	0.80
Accuracy	0.89		

Pada tabel 4.12 diatas, klasifikasi memiliki nilai *Accuracy* 89%. Ini berarti bahwa klasifikasi mengklasifikasikan 89% data teks dengan benar. Performa terbaik pada kelas 1 (*Fear*) dengan *precission, recall* dan *F1-score* 1.00. Performa kategori lain lebih rendah, dengan akurasi keseluruhan 89%. Klasifikasi dapat ditingkatkan dengan mengumpulkan lebih banyak data untuk kelas 1 (*Fear*).



Gambar 4.12 *Confussion Matrix Data Test Complement Naive Bayes 70:3*

4.2 Pembahasan

Dalam pembahasan berikut, terdapat 2 hasil dalam penelitian ini yaitu dengan perbandingana algoritma *Naïve Bayes* seperti *Multinomial Naïve Bayes* dan *Complement Naïve Bayes*, dan Visualisasi Hasil.

4.2.1 Perbandingan Algoritma

Dari hasil percobaan menggunakan *Multinomial Naïve Bayes* (*MultinomialNB*) dan *Complement Naïve Bayes* (*ComplementNB*) pada data *Train* dan *Test*, terlihat perbedaan kinerja yang signifikan antara kedua model.

Tabel 4.13 Hasil Evaluasi Model Klasifikasi

Algortima	**Data Latih**	**Data Uji**	**Performa**			
			Accuracy	***Precision***	***Recall***	***F-1 Score***
MultinomialNB	90%	10%	0.67	0.44	0.67	0.53
	80%	20%	0.50	0.25	0.50	0.33
	70%	30%	0.61	0.37	0.61	0.46
ComplementNB	90%	10%	1.00	1.00	1.00	1.00
	80%	20%	1.00	1.00	1.00	1.00
	70%	30%	0.89	0.92	0.95	0.89

Pada data pelatihan, *MultinomialNB* merupakan algoritma ini menunjukkan kinerja yang bervariasi tergantung pada proporsi data latih dan data uji. Pada umumnya, kinerja algoritma ini membaik seiring dengan bertambahnya proporsi data latih. Namun, pada beberapa kasus, algoritma ini mengalami penurunan kinerja.

Sebaliknya, *ComplementNB* menunjukkan kinerja yang jauh lebih baik. Pada data pelatihan, *ComplementNB* mencapai akurasi 1.00, dengan semua kategori memiliki *precision* dan *recall* yang tinggi. *ComplementNB* secara konsisten mengungguli *MultinomialNB* dalam hal *presisi*, *recall*, dan *accuracy* untuk kedua *dataset*. *ComplementNB* lebih efektif dalam mengklasifikasikan emosi dari data teks, terutama dalam mendeteksi kategori yang beragam dengan lebih akurat dibandingkan *MultinomialNB*.

4.2.2 Visualisasi Hasil

Setelah melalui proses *preprocessing data* dan melatih model dengan metode *Naïve Bayes*, penulis menampilkan visualisasi hasil klasifikiasi data teks cerita anak untuk mengetahui emosi yang terkandung dalam sebuah teks cerita anak tersebut. Hasil visualisasi yang dihasilkan dapat dilihat pada Tabel 4.14 sebagai berikut :

Tabel 4.14 Visualisasi Hasil

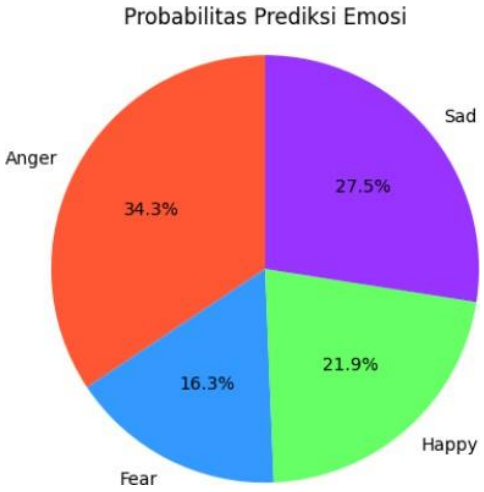
Teks yang Dimasukkan	Teks setelah Preprocessing	Prediksi Emosi	Keterangan Probabilitas
A Giant Oak stood near a brook in which grew some slender Reeds. When the wind blew. the great Oak stood proudly upright with its hundred arms uplifted to the sky. But the Reeds bowed low in the w...	giant oak stood near brook grew slender reeds. wind blew. great oak stood proudly upright hundred arms uplifted sky. reeds bowed low wind sang sad mournful song. reason complain. said oak. slighte...	Anger	Anger: 34.34%, Fear: 16.31%, Happy: 21.87%, Sad: 27.48%

Pada Tabel 4.14 diatas, dapat dilihat bahwa proses yang dilakukan merupakan hasil dari input data sebuah teks cerita yang ingin diklasifikasikan. Probabilitas emosi dihasilkan melalui metode *predict_proba* dari model *Multinomial Naïve Bayes* yang telah dilatih. Setelah teks cerita diproses dan diubah menjadi representasi numerik menggunakan *TF-IDF*, model dilatih untuk memprediksi emosi berdasarkan data latih yang ada. Saat melakukan prediksi, metode *predict_proba* digunakan untuk menghitung probabilitas dari setiap kelas emosi yang mungkin. Probabilitas ini berupa *array* dengan nilai yang menunjukkan seberapa besar kemungkinan teks tersebut termasuk dalam setiap kategori emosi (misalnya marah, takut, senang, sedih). Probabilitas ini kemudian dipetakan dengan label emosi menggunakan *LabelEncoder*, dan

hasilnya disajikan dalam bentuk persentase. Selain ditampilkan dalam teks, probabilitas juga divisualisasikan dalam grafik batang untuk menunjukkan distribusi prediksi model terhadap berbagai emosi.

Hasil prediksi emosi yang diperoleh tidak hanya menunjukkan jenis emosi yang teridentifikasi, tetapi juga disertai dengan probabilitas atau persentase prediksi untuk setiap emosi yang terkandung. Persentase ini memberikan gambaran tentang seberapa besar kemungkinan suatu teks mengandung emosi tertentu, memberikan informasi yang lebih mendalam dan terperinci. Dengan adanya probabilitas ini, pengguna dapat memahami lebih jelas nuansa emosi dalam teks, serta tingkat kepercayaan dari model prediksi terhadap klasifikasi yang dihasilkan.

Hasil klasifikasi teks tersebut disajikan dalam bentuk *pie chart* untuk memvisualisasikan distribusi berbagai emosi yang teridentifikasi dalam teks. Dengan cara ini, dapat dengan mudah melihat perbandingan antara emosi-emosi yang ada dan memahami mana yang paling dominan atau signifikan dalam teks. Penyajian data dalam bentuk visual seperti ini sangat membantu dalam menyederhanakan interpretasi hasil analisis dan membuat informasi yang kompleks lebih mudah dicerna seperti pada Gambar 4.13 sebagai berikut:



Gambar 4.13 Visualisasi Hasil

Pada gambar 4.13 diatas, mempresentasikan probabilitas prediksi berbagai macam emosi. Gambar didasarkan pada beberapa faktor, seperti kemungkinan cerita dalam suatu teks yang telah diklasifikasikan *anger, fear, happy* dan *sad.* Berdasarkan gambar 4.13 tersebut, emosi yang paling mungkin diprediksi adalah "*Anger*", sedangkan emosi yang paling kecil kemungkinannya adalah "*Fear*". Secara keseluruhan, gambar ini menunjukkan bahwa model prediksi emosi mampu memprediksi emosi dengan tingkat akurasi yang cukup baik.

4.3 Integrasi Islam

Dari hasil penelitian yang telah dilakukan, sistem yang dibangun pada dasarnya tentang mengklasifikasikan suatu emosi pada teks cerita anak yang digunakan sebagai tolak ukur dalam perkembangan emosi pada anak -anak. Hal ini berkaitan dengan surah Luqman ayat 13, sebagai berikut:

وَاِذْ قَالَ لُقْمٰنُ لِابْنِهٖ وَهُوَ يَعِظُهٗ يٰبُنَيَّ لَا تُشْرِكْ بِاللّٰهِ ۗاِنَّ الشِّرْكَ لَظُلْمٌ عَظِيْمٌ

"*(Ingatlah) ketika Luqman berkata kepada anaknya, yaitu dia menasihatinya} memberi perlajaran dan menumbuhkan minat anaknya "Wahai anakku, janganlah menyekutukan Allah. Sesungguhnya menyekutukan (Allah) itu benar-benar kezaliman yang besar*" (QS. Luqman : 13)

Berdasarkan tafsir ibnu Katsir bahwa ayat ini menunjukkan pentingnya pendidikan dan pembinaan anak sejak dini, termasuk pembinaan emosional dan spiritual. Sedangkan Tafsir Ash-Shaghir menambahkan bahwa nasihat ini juga menunjukkan tanggung jawab orang tua dalam menambahkan ilmu yang benar kepada anak dengan cara lembut dan bijaksana (*Tafsir Surah Luqman Ayat 13, 2024*). Dengan mengutamakan tauhid, ayat ini memberikan pedoman abadi tentang pentingnya membangun hubungan manusia dengan Allah SWT di atas keesaan-Nya serta menjadikan tauhid sebagai pondasi utama dalam Pendidikan.

Dalam hal mengklasifikasi suatu emosi juga berkaitan dengan surah Az-Zumar 9, sebagai berikut:

أَمَّنْ هُوَ قَٰنِتٌ ءَانَاءَ ٱلَّيْلِ سَاجِدًا وَقَآئِمًا يَحْذَرُ ٱلْءَاخِرَةَ وَيَرْجُوا رَحْمَةَ رَبِّهِۦ قُلْ هَلْ يَسْتَوِى ٱلَّذِينَ يَعْلَمُونَ وَٱلَّذِينَ لَا يَعْلَمُونَ
إِنَّمَا يَتَذَكَّرُ أُوْلُوا ٱلْأَلْبَٰبِ

"(Apakah kamu hai orang musyrik yang lebih beruntung) ataukah orang yang beribadat di waktu-waktu malam dengan sujud dan berdiri, sedang ia takut kepada (azab) akhirat dan mengharapkan rahmat Tuhannya? Katakanlah: "Adakah sama orang-orang yang mengetahui dengan orang-orang yang tidak mengetahui?" Sesungguhnya orang yang berakallah yang dapat menerima pelajaran." (QS. Az-Zumar 39:9).

Dan berrdasarkan Tafsir Ibnu Katsir, Pada ayat ini, Allah mengingatkan manusia tentang perbedaan besar antara orang-orang yang berilmu dan yang tidak berilmu. Mereka yang mengetahui, yakni orang-orang yang memiliki pengetahuan tentang Allah, wahyu-Nya, dan syariat-Nya, dipandang lebih tinggi derajatnya

dibandingkan mereka yang tidak mengetahui. Ini karena ilmu membimbing seseorang kepada iman, ketakwaan, dan amal yang benar (*Surat Az-Zumar Ayat 9*, 2024).

Dan untuk mengetahui perbedaan emosi pada suatu cerita juga berkaitan dengan surah Al-Hujurat ayat 13, sebagai berikut:

يَٰٓأَيُّهَا ٱلنَّاسُ إِنَّا خَلَقۡنَٰكُم مِّن ذَكَرٖ وَأُنثَىٰ وَجَعَلۡنَٰكُمۡ شُعُوبٗا وَقَبَآئِلَ لِتَعَارَفُوٓاْۚ إِنَّ أَكۡرَمَكُمۡ عِندَ ٱللَّهِ أَتۡقَىٰكُمۡۚ إِنَّ ٱللَّهَ عَلِيمٌ خَبِيرٞ ١٣

"Wahai manusia! Sesungguhnya Kami menciptakan kamu dari seorang laki-laki dan seorang perempuan, dan Kami menjadikan kamu berbangsa-bangsa dan bersuku-suku agar kamu saling mengenal. Sesungguhnya yang paling mulia di antara kamu di sisi Allah ialah orang yang paling bertakwa. Sungguh, Allah Maha Mengetahui, Maha Teliti." (QS. Al-Hujurat 49:13).

Berdasarkan Tafsir Ibnu Katsir dan Tafsir Al-Mukhtashar, ayat ini menjelaskan bahwa manusia diciptakan dari satu asal, yaitu Nabi Adam dan Hawa, dan menjadikan manusia berbangsa-bangsa serta bersuku-suku agar saling mengenal, memajami satu sama lain, bukan untuk saling membanggakan diri sendiri, merendahkan dan juga berselisih (*Tafsir Surat Al-Hujurat Ayat 13, 2024*). Keragaman ini merupakan Rahmat Allah SWT yang bertujuan untuk memperkaya hubungan kemanusiaan, bukan menjadi alas an perpecahan. Ukuran kemuliaan manusia di sisi Allah SWT tidak hanya ras, kekayaan, atau status sosial, melainkan takwa-ketaatan dan kesalehan kepada Allah SWT. Ayat ini menanamkan pesan universal tentang persaudaraan, kesetaraan, dan pentingnya membangun harmoni dalam kehidupan dengan menjadikan takwa sebagai tolok ukur utama.

Berdasarkan tafsir QS. Luqman (31):13 dan QS. Al-Hujurat (49):13 keduanya menekankan pentingnya mendidik anak mencakup aspek spiritual, emosional, dan sosial. Dan juga mengajarkan agar anak-anak diajarkan dasar agama sejak dini sebagai fondasi utama kehidupan, dengan pendekatan yang lembut dan penuh kasih, seperti nasihat Luqman kepada anaknya yang menanamkan nilai-nilai keimanan dan membentuk emosi yang positif. Sementar dengan QS. Al-Hujurat menyoroti bahwa keragaman manusia dalam bangsa dan suku adalah Rahmat Allah SWT yang bertujuan untuk memperkaya hubungan sosial, bukan menciptakan perpecahan, dengan takwa sebagai tolok ukur utama kemuliaan. Kedua ayat tersebut memberikan pedoman penting untuk menanamkan nilai-nilai kasih sayang, rasa hormat, dan penghargaan terhadap perbedaan, sekaligus mendidik anak mengelola emosi mereka sesuai dengan prinsip. Dengan membangun karakter yang berlandaskan spiritual, emosi yang sehat, dan kesadaran sosial, Pendidikan ini menciptakan generasi yang beriman, bertakwa, dan harmonis dalam kehidupan bermasyarakat.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Pada penelitian ini berfokus pada penggunaan metode *Naïve Bayes* untuk mengklasifikasi teks cerita anak dengan mempertimbangkan berbagai rasio pembagian data yaitu 90:10, 80:20, dan 70:30. Hasil eksperimen yang telah dilakukan, dapat disimpulkan bahwa algoritma *Complement Naïve Bayes* menunjukkan kinerja lebih baik dibandingkan *Multinomial Naïve Bayes* dalam mengklasifikasikan teks. *Complement Naïve Bayes* mencapai *accuracy* sebesar 0.89 pada data uji dengan proporsi data latih 70:30, dengan nilai *presisi*, *recall* dan *f1-score* yang konsisten tinggi pada semua kelas emosi. Sebaliknya, *Multinomial Naïve Bayes* hanya mampu mencapai *accuracy* tertinggi sebesar 0.67 pada data uji dengan proporsi data latih 90:10, dan mengalami kesulitan dalam mengklasifikasikan emosi. Hasil ini mengindikasikan bahwa *Complement Naïve Bayes* lebih cocok untuk menangani *dataset* yang digunakan dalam eksperimen ini dibandingkan dengan *Multinomial Naïve Bayes.* Model *Naïve Bayes* juga berhasil mengklasifikasikan emosi pada teks cerita dengan cukup baik. Melalui proses *preprocessing* dan representasi numerik menggunakan *TF-IDF*, model mampu menghitung probabilitas kemunculan setiap emosi pada teks tersebut.

5.2 Saran

Peneliti menyadari bahwa penelitian ini masih memiliki kekurangan, maka dari itu diharapkan pada penelitian selanjutnya ada peningkatan dan perbaikan agar hasil

yang di dapat menjadi lebih baik. Berikut beberapa saran dari peneliti yang diharapkan dapat memberikan kontribusi bagi penelitian selanjutnya:

1. Melakukan evaluasi dengan dataset yang lebih banyak dan beragam. Hal ini akan membantu mengukur kemampuan model dalam menghadapi variasi data yang lebih luas.
2. Penelitian selanjutnya dapat mencoba membandingkan denga metode lain yang berbeda.
3. Menambahkan fitur tambahan selain menggunakan *TF-IDF* dapat membantu menangkap informasi yang lebih kompleks dalam teks cerita.

DAFTAR PUSTAKA

Agung, S., & Santara, A. (2024). Implementasi Text Mining untuk Analisis Review pada Aplikasi Crowdfunding LX dan ST Menggunakan Metode Sentiment Analysis. *LANCAH: Jurnal Inovasi dan Tren*, 2(1), 124–130. https://doi.org/ 10.35870/ljit.v2i1.2245

Arifin, K., & Al-Idrus, S. I. (2024). Klasifikasi Emosi Pengguna Twitter Terhadap Bakal Calon Presiden Pada Pemilu 2024 Menggunakan Algoritma Naïve Bayes. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer)*, 23(1), 37. https://doi.org/10.53513/jis.v23i1.9558

Ayomi, F. A. J., & Dewi, K. E. (2023). Analisis Emosi pada Media Sosial Twitter Menggunakan Metode Multinomial Naive Bayes dan Synthetic Minority Oversampling Technique. *Komputa : Jurnal Ilmiah Komputer Dan Informatika, 12(2), 9–19*. https://doi.org/10.34010/komputa.v12i2.9454

Azizah, I., Cholissodin, I., & Yudistira, N. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi Shopee di Google Play menggunakan Metode Word Embedding dan Long Short Term Memory (LSTM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(5), 2453–2459. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12748

Azzahra, S. A., & Wibowo, A. (2020). Analisis Sentimen Multi-Aspek Berbasis Konversi Ikon Emosi dengan Algoritme *Naïve Bayes* untuk Ulasan Wisata Kuliner Pada Web Tripadvisor. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(4), 737. https://doi.org/10.25126/jtiik.2020731907

Baharuddin, R. R., & Muyassar, A. (2023). Klasifikasi Emosi pada Platform X Menggunakan Metode Machine Learning Berbasis Sentiment Analysis. *SEMANTIK Prosiiding*, 174–181. https://jurnal.ftkom.uncp.ac.id/index.php/semantik/article/view/38

Ding, R., He, W., & Wang, Q. (2021). A Comparative Analysis of Emotion-Related Cultural Norms in Popular American and Chinese Storybooks. *Journal of Cross-Cultural Psychology*, 52(2), 209–226. https://doi.org/10.1177/0022022120988900

Fahrudin, T. M., Sari, A. R. F., Iffadah, A. S., Windyadari, C. C., & ... (2022). Pemodelan Teks Tweet pada Isu Pelecehan Seksual Berbasis Analisis Sentimen dan Leksikon Emosi. *PROSIDING SEMINAR NASIONAL SAINS DATA*, 12–23. https://doi.org/ 10.33005/senada.v2i1.33

Gusmayanti, E., & Dimyati, D. (2021). Analisis Kegiatan Mendongeng dalam Meningkatkan Perkembangan Nilai Moral Anak Usia Dini. *Jurnal Obsesi :*

Jurnal Pendidikan Anak Usia Dini, 6(2), 903–917. https://doi.org/10.31004/obsesi.v6i2.1062

Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations. *Organizational Research Methods*, 25(1), 114–146. https://doi.org/10.1177/1094428120971683

Husada, I. N., & Toba, H. (2020). Pengaruh Metode Penyeimbangan Kelas Terhadap Tingkat Akurasi Analisis Sentimen pada Tweets Berbahasa Indonesia. *Jurnal Teknik Informatika Dan Sistem Informasi*, 6(2), 400–413. https://doi.org/10.28932/jutisi.v6i2.2743

Jannah, Y. A. N., & Prasetyo, R. B. (2022). Analisis Sentimen dan Emosi Publik pada Awal Pandemi COVID-19 Berdasarkan Data Twitter dengan Pendekatan Berbasis Leksikon (Analysis of Public Sentiment and Emotion at the Beginning of COVID-19 Pandemic Based on Twitter Data with Lexicon Based Approach). *Seminar Nasional Official Statistics 2022*, 597–607. https://doi.org/ 10.34123/semnasoffstat.v2022i1.1483

Masruroh, F., & Ramiati, E. (2022). Pembentukan Karakter Gemar Membaca Pada Anak Usia Dini Melalui Media Buku Cerita Bergambar. *INCARE, International Journal of Educational Resources*, 2(6), Article 6. https://doi.org/10.59689/incare.v2i6.353

Mayar, F., Natari, R., Cendana, H., Hutasuhut, B. R. S., Aprilia, S., & Nurhikmah, N. (2022). Peran Dongeng dalam Meningkatkan Kreativitas Anak Usia Dini. *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, 6(5), 4600–4607. https://doi.org/10.31004/obsesi.v6i5.2615

Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1), 1–19. https://doi.org/10.1007/s13278-021-00776-6

Ramadhan, G. S., Irawan, B., Setianingsih, C., & Telkom, U. (2021). Klasifikasi emosi pada lirik lagu menngunakan algoritma *Naïve Bayes* dan particle swarm optimization classification of emotions on song lyrics using *Naïve Bayes* algorithm and particle swarm optimization. *E-Proceeding of Engineering*, 8(5), 6293–6306. https://repository.telkomuniversity.ac.id/pustaka/172288/klasifikasi-emosi-pada-lirik-lagu-menggunakan-algoritma-na-ve-bayes-dan-particle-swarm-optimization.html

Saputra, R. A., & Waluyo, S. (2022). Penerapan Algoritma Naive Bayes Dalam Analisis Kenaikan Bahan Bakar Minyak Pada Twitter. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta-Indonesia*,

1(September), 569–575. https://senafti.budiluhur.ac.id/index.php/senafti/article/view/305

Sastypratiwi, H., Muhardi, H., & Noveanto, M. (2022). Klasifikasi Emosi Pada Lirik Lagu Menggunakan Algoritma Multiclass SVM dengan Tuning Hyperparameter PSO. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(4), 2279. https://doi.org/10.30865/mib.v6i4.4609

Sukatin, Q. Y. H. (2020). Analisis Psikologi Perkembangan Sosial Emosional Anak Usia Dini. *Bunayya : Jurnal Pendidikan Anak*, 6(2), Article 2. https://doi.org/10.22373/bunayya.v6i2.7311

Sukatin, S., Chofifah, N., Turiyana, T., Paradise, M. R., Azkia, M., & Ummah, S. N. (2020). Analisis Perkembangan Emosi Anak Usia Dini. *Golden Age: Jurnal Ilmiah Tumbuh Kembang Anak Usia Dini*, 5(2), 77–90. https://doi.org/10.14421/jga.2020.52-05

Surat Al-Hujurat Ayat 13 Arab, Latin, Terjemah dan Tafsir | Baca di TafsirWeb. (2024). Retrieved December 8, 2024, from https://tafsirweb.com/9783-surat-al-hujurat-ayat-13.html

Surat Az-Zumar Ayat 9 Arab, Latin, Terjemah dan Tafsir | Baca di TafsirWeb. (2024). Retrieved December 8, 2024, from https://tafsirweb.com/8671-surat-az-zumar-ayat-9.html

Surat Luqman Ayat 13 Arab, Latin, Terjemah dan Tafsir | Baca di TafsirWeb. (2024). Retrieved December 8, 2024, from https://tafsirweb.com/7497-surat-luqman-ayat-13.html

LAMPIRAN