PERBANDINGAN ALGORITMA *MULTINOMIAL NAIVE BAYES* DAN *DECISION TREE* PADA KLASIFIKASI EMOSI *TWEET*

SKRIPSI

Program Studi Sarjana Informatika
Jurusan Informatika

Oleh:
WINDARI OKTAPIA S
NIM D1041161032

FAKULTAS TEKNIK
UNIVERSITAS TANJUNGPURA
PONTIANAK
2023

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Windari Oktapia S

NIM : D1041161032

Menyatakan bahwa dalam skripsi yang berjudul "PERBANDINGAN ALGORITMA *MULTINOMIAL NAIVE BAYES* DAN *DECISION TREE* PADA KLASIFIKASI EMOSI *TWEET*" tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu perguruan tinggi manapun. Sepanjang pengetahuan Saya, tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis di acu dalam naskah ini dan disebutkan dalam DAFTAR PUSTAKA.

Demikian pernyataan ini dibuat dengan sebenar-benarnya. Saya danggup menerima konsekuensi akademis dan hukum di kemudian hari apabila pernyataan yang dibuat ini tidak benar.

Pontianak, 08 Juni 2023

Windari Oktapia S

NIM D1041161032

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,
RISET, DAN TEKNOLOGI
UNIVERSITAS TANJUNGPURA
FAKULTAS TEKNIK
Jalan Prof. Dr. H. Hadari Nawawi Pontianak 78124
Telepon (0561) 740186 Email : ft@untan.ac.id Website : http://teknik.untan.ac.id

HALAMAN PENGESAHAN

PERBANDINGAN ALGORITMA *MULTINOMIAL NAIVE BAYES* DAN *DECISION TREE* PADA KLASIFIKASI EMOSI *TWEET*

Program Studi Sarjana Informatika
Jurusan Informatika

Oleh :

Windari Oktapia S
NIM. D1041161032

Telah dipertahankan didepan Penguji Skripsi pada tanggal 08 Juni 2023 dalam sidang dan diterima sebagai salah satu persyaratan untuk memperoleh gelar sarjana.

Susunan Penguji Skripsi :

Ketua,

Dr. Arif Bijaksana P.N., S.T., M.T.
NIP 197208081998021002

Penguji Utama,

Tursina, S.T., M.Cs.
NIP 197801152002122003

Sekretaris,

Rina Septiriana, ST, M.Cs.
NIP 19870923202012200l

Penguji Pendamping,

Hafiz Muhardi, S.T., M.Kom.
NIP 0023079006

Pontianak, 08 Juni 2023
Dekan,

Dr. -Ing. Ir. Slamet Widodo, M.T., IPM
NIP 196712231992031002

HALAMAN PERSEMBAHAN

Puji Tuhan saya dapat menyelesaikan penulisan skripsi ini dengan penyertaan Tuhan Yesus Kristus. Skripsi ini saya dedikasi kepada keluarga (Bapak, Mama dan Egi), Op Mauliate Siregar dan Op Sotarduga Simanjuntak yang senantiasa mendukung aku disetiap langkah ku.

Terima kasih kepada teman teman saya Tina, wacen, 323, dan mak ti family yang selalu bersama ku saat pengerjaan skripsi ini.

I did it!~

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena dengan rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul ”PERBANDINGAN ALGORITMA *MULTINOMIAL NAIVE BAYES* DAN *DECISION TREE* PADA KLASIFIKASI EMOSI *TWEET*” sebagai syarat guna memperoleh gelar Sarjana Komputer (S.Kom.) pada Program Studi Informatika, Jurusan Informatika, Fakultas Teknik, Universitas Tanjungpura Pontianak.

Dalam penyelesaian penelitian ini penulis secara langsung maupun tidak langsung telah mendapatkan bimbingan, arahan dan masukan dari beberapa pihak. Untuk itu, penulis ingin mengucapkan terima kasih kepada: Bapak Dr. Arif Bijaksana Putra Negara, S.T., M.T. selaku Dosen Pembimbing Pertama dan Ibu Rina Septiriana, ST, M.Cs. selaku Dosen Pembimbing kedua. Kepada Ibu Tursina, S.T., M.Cs. selaku Penguji Utama dan Bapak Hafiz Muhardi, S.T., M.Kom. selaku Penguji Pendamping yang telah memberikan saran serta masukan untuk perbaikan dalam skripsi ini, serta seluruh dosen Jurusan Informatika, dan pegawai Fakultas Teknik Universitas Tanjungpura.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna karena adanya kerterbatasan ilmu dan pengalaman yang dimiliki. Oleh karena itu, penulis menerima seluruh kritik dan saran yang bersifat membangun. Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi semua pihak pembaca maupun penelitian selanjutnya.

Pontianak, 08 Juni 2023
Penulis,

Windari Oktapia S

ABSTRAK

Media sosial telah menjadi sebuah wadah “curhat" seseorang untuk dapat mengekspresikan emosi dan pikiran yang sedang dialami tanpa adanya batasan. Salah satu media sosial yang banyak digunakan adalah Twitter. Twitter memiliki pengguna aktif yang mencapai 238 juta dan memberi akses kepada pengguna agar dapat mencari informasi mengenai *tweet* secara spesifik. Sehingga Twitter dapat digunakan sebagai sumber data dalam menganalisis emosi seseorang berdasarkan tulisan/*tweet* yang dibuatnya. Dalam menganalisis emosi pada sebuah *tweet*, diperlukan suatu metode klasifikasi agar dapat mengklasifikasikan *tweet* ke dalam kelas emosi yang tepat. Klasifikasi emosi pada tweet bertujuan untuk mengelompokkan *tweet* ke dalam kelas emosi yang sudah ditentukan seperti, marah, senang, takut, cinta, dan sedih. Algoritma yang digunakan untuk membangun sebuah model machine learning klasifikasi emosi yaitu *Multinomial Naive Bayes* dan *Decision Tree*. Tujuan dari penelitian ini adalah untuk mengetahui algoritma klasifikasi *Multinomial Naive Bayes* dan *Decision Tree* manakah yang memiliki performa lebih baik dengan melakukan perbandingan hasil nilai accuracy antara algoritma klasifikasi. Penelitian ini juga menggunakan *feature extraction Bag of Word* dan *Bigram* serta metode SMOTE pada data dan penggunaan *Tuning Hyperparameter* pada algoritma. Hasil penelitian menunjukkan bahwa model klasifikasi *Multinomial Naive Bayes* menggunakan *feature extraction Bag of Word* dengan metode SMOTE dan Tuning Hyperparameter memiliki nilai accuracy yang paling tinggi yaitu 67.15%. Sedangkan model klasifikasi *Decision Tree* menggunakan *feature extraction Bag of Word* dengan metode SMOTE dan Tuning Hyperparameter yang memiliki nilai accuracy paling tinggi yaitu 53.90%.

Kata kunci : Klasifikasi Emosi, *Machine Learning Multinomial Naive Bayes , Decision Tree*

DAFTAR ISI

HALAMAN PERNYATAAN ..... **ii**
HALAMAN PENGESAHAN ..... **iii**
HALAMAN PERSEMBAHAN ..... **iv**
KATA PENGANTAR ..... **v**
ABSTRAK ..... **vi**
DAFTAR ISI ..... **vii**
DAFTAR GAMBAR ..... **x**
DAFTAR TABEL ..... **xii**
DAFTAR LAMPIRAN ..... **xiii**
KODE PROGRAM ..... **xiv**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Perumusan Masalah 4
1.3 Tujuan Penelitian 4
1.4 Batasan Masalah 4
1.5 Sistematika Penulisan 5
BAB II TINJAUAN PUSTAKA **6**
2.1 Kajian Terkait 6
2.2 Emosi 9
2.3 Data Mining 10
2.4 Text Preprocessing 10
2.5 Feature Extraction 11
2.5.1 *Bag of Word* 11
2.5.2 *Bigram* 12
2.6 Imbalanced Data 12
2.7 Pembagian Dataset 13
2.8 Multinomial Naive Bayes 14
2.9 *Decision Tree* 14
2.10 Tuning Hyperparameter 15
2.11 *10-Fold Cross Validation* 15
2.12 *Confusion Matrix* 16

BAB III METODOLOGI PENELITIAN 18
3.1 Alat dan Data Penelitian 18
3.1.1 Perangkat Keras 18
3.1.2 Perangkat Lunak 18
3.2 Data Penelitian 18
3.3 Langkah Penelitian 18
3.3.1 Pengumpulan Data 19
3.3.2 Text Preprocessing 21
3.3.3 Feature Extraction 22
3.3.4 SMOTE 23
3.3.5 Pembagian Dataset 23
3.3.6 Implementasi Algoritma 23
3.3.7 Tuning Hyperparameter 23
3.3.8 Skenario Pengujian 24
3.3.9 Evaluasi 31
BAB IV IMPLEMENTASI DAN HASIL 32
4.1 Implementasi 32
4.1.1 Text Preprocessing 32
4.1.2 Distribusi Frekuensi 35
4.1.3 Feature Extraction 38
4.1.4 SMOTE 39
4.1.5 Pembagian Dataset 41
4.1.6 Pemodelan 41
4.1.7 *Tuning Hyperparameter* 42
4.1.8 Hasil Skenario Pengujian 46
4.1.9 Evaluasi 51
4.2 Hasil Evaluasi 53
4.2.1 *Multinomial Naive Bayes* 54
4.2.2 *Decision Tree* 63
4.3 Perbandingan Evaluasi 71
4.4 Implementasi Aplikasi Generik 74
BAB V 76

5.1 Kesimpulan 76
5.2 Saran 76
DAFTAR PUSTAKA 78
LAMPIRAN ..... A
Lampiran A ..... A
Lampiran B (Classification Report) 1

DAFTAR GAMBAR

Gambar 3. 1 Langkah Penelitian 19
Gambar 3. 2 Jumlah Dataset 20
Gambar 3. 3 Ilustrasi *Case Folding* 21
Gambar 3. 4 Ilustrasi *Cleaning* 22
Gambar 3. 5 Ilustrasi normalisasi 22
Gambar 3. 6 Skenario Pengujian Pertama Multinomial Naive Bayes 25
Gambar 3. 7 Skenario Pengujian Pertama Decision Tree 25
Gambar 3. 8 Skenario Pengujian Kedua Multinomial Naive Bayes 25
Gambar 3. 9 Skenario Pengujian Kedua Decision Tree 26
Gambar 3. 10 Skenario Pengujian Ketiga Multinomial Naive Bayes 26
Gambar 3. 11 Skenario Pengujian Ketiga Decision Tree 27
Gambar 3. 12 Skenario Pengujian Keempat Multinomial Naive Bayes 27
Gambar 3. 13 Skenario Pengujian Keempat Decision Tree 27
Gambar 3. 14 Skenario Pengujian Kelima Multinomial Naive Bayes 28
Gambar 3. 15 Skenario Pengujian Kelima Decision Tree 28
Gambar 3. 16 Skenario Pengujian Keenam Multinomial Naive Bayes 29
Gambar 3. 17 Skenario Pengujian Keenam Decision Tree 29
Gambar 3. 18 Skenario Pengujian Ketujuh Multinomial Naive Bayes 30
Gambar 3. 19 Skenario Pengujian Ketujuh Decision Tree 30
Gambar 3. 20 Skenario Pengujian Kedelapan Multinomial Naive Bayes ... 31
Gambar 3. 21 Skenario Pengujian Kedelapan Decision Tree 31
Gambar 4. 1 Hasil *Case Folding* 32
Gambar 4. 2 Hasil Cleaning 33
Gambar 4. 3 Hasil Normalisasi 34
Gambar 4. 4 Distribusi Frekuensi 37
Gambar 4. 5 *Vocabulary Bag of Word* 38
Gambar 4. 6 Hasil *Bag of Word* 38
Gambar 4. 7 *Vocabulary Bigram* 39
Gambar 4. 8 Hasil *Bigram* 39
Gambar 4. 9 Data Sebelum SMOTE 40
Gambar 4. 10 Hasil SMOTE (*Bag of Word*) 40

Gambar 4. 11 Hasil SMOTE (*Bigram*) 41
Gambar 4. 12 Data Setelah SMOTE 41
Gambar 4. 13 *Confusion Matrix* Skenario Pertama 54
Gambar 4. 14 *Confusion Matrix* Skenario Kedua 54
Gambar 4. 15 *Confusion Matrix* Skenario Ketiga 55
Gambar 4. 16 *Confusion Matrix* Skenario Keempat 55
Gambar 4. 17 *Confusion Matrix* Skenario Kelima 56
Gambar 4. 18 *Confusion Matrix* Skenario Keenam 56
Gambar 4. 19 *Confusion Matrix* Skenario Ketujuh 57
Gambar 4. 20 *Confusion Matrix* Skenario Kedelapan 57
Gambar 4. 21 *Confusion Matrix* Skenario Pertama 63
Gambar 4. 22 *Confusion Matrix* Skenario Kedua 63
Gambar 4. 23 *Confusion Matrix* Skenario Ketiga 64
Gambar 4. 24 *Confusion Matrix* Skenario Keempat 64
Gambar 4. 25 *Confusion Matrix* Skenario Kelima 65
Gambar 4. 26 *Confusion Matrix* Skenario Keenam 65
Gambar 4. 27 *Confusion Matrix* Skenario Ketujuh 66
Gambar 4. 28 *Confusion Matrix* Skenario Kedelapan 66
Gambar 4. 29 Tampilan Aplikasi 74
Gambar 4. 30 Contoh *Tweet* 75

DAFTAR TABEL

Tabel 2. 1 Kajian Terkait 7
Tabel 2. 2 *Bag of Word* 11
Tabel 2. 3 *Bigram* 12
Tabel 2. 4 *Confusion Matrix* 16
Tabel 3. 1 Jumlah Dataset 20
Tabel 3. 2 Sampel Data *Tweet* 20
Tabel 3. 3 Skenario Pengujian 24
Tabel 4. 1 Kamus Normalisasi 34
Tabel 4. 2 Hasil *Text Preprocessing* 35
Tabel 4. 3 Distribusi Frekuensi 36
Tabel 4. 4 Hasil *Grid Search Multinomial Naive Bayes* 43
Tabel 4. 5 Hasil *Grid Search Decision Tree* 45
Tabel 4. 6 Desain Hasil Evaluasi Tiap Model 53
Tabel 4. 7 Desain Hasil Pengujian 53
Tabel 4. 8 Hasil Evaluasi Model *Multinomial Naive Bayes* 58
Tabel 4. 9 Hasil Evaluasi *Decision Tree* 67
Tabel 4. 10 Hasil Perbandingan Evaluasi 71
Tabel 4. 11 Hasil Klasifikasi Emosi Aplikasi Generik 75

DAFTAR LAMPIRAN

Lampiran A ..... A

Lampiran B (Classification Report) B-1

KODE PROGRAM

Kode Program 4. 1 *Case Folding* 32

Kode Program 4. 2 Cleaning 33

Kode Program 4. 3 Normalisasi 34

Kode Program 4. 4 35

Kode Program 4. 5 *Feature Extraction* (*Bag of Word*) 38

Kode Program 4. 6 *Feature Extraction* (*Bigram*) 39

Kode Program 4. 7 SMOTE (*Bag of Word*) 40

Kode Program 4. 8 SMOTE (*Bigram*) 40

Kode Program 4. 9 Pembagian Dataset 41

Kode Program 4. 10 *Grid Search Multinomial Naive Bayes* dengan *Bag of Word* 42

Kode Program 4. 11 *Grid Search Multinomial Naive Bayes* dengan *Bigram* 43

Kode Program 4. 12 *Grid Search Multinomial Naive Bayes* dengan *Bag of Word* menggunakan SMOTE 43

Kode Program 4. 13 *Grid Search Multinomial Naive Bayes* dengan *Bigram* menggunakan SMOTE 43

Kode Program 4. 14 *Decision Tree* dengan *Bag of Word* 44

Kode Program 4. 15 *Decision Tree* dengan *Bigram* 45

Kode Program 4. 16 *Decision Tree* dengan *Bag of Word* menggunakan SMOTE 45

Kode Program 4. 17 *Decision Tree* dengan *Bigram* menggunakan SMOTE 45

Kode Program 4. 18 Skenario Pertama 46

Kode Program 4. 19 Skenario Kedua 46

Kode Program 4. 20 Skenario Ketiga 47

Kode Program 4. 21 Skenario Keempat 47

Kode Program 4. 22 Skenario Kelima 47

Kode Program 4. 23 Skenario Keenam 48

Kode Program 4. 24 Skenario Ketujuh 48

Kode Program 4. 25 Skenario Kedelapan 48

Kode Program 4. 26 Skenario Pertama 49

Kode Program 4. 27 Skenario Kedua 49

Kode Program 4. 28 Skenario Ketiga 49
Kode Program 4. 29 Skenario Keempat 50
Kode Program 4. 30 Skenario Kelima 50
Kode Program 4. 31 Skenario Keenam 50
Kode Program 4. 32 Skenario Ketujuh 51
Kode Program 4. 33 Skenario Kedelapan 51
Kode Program 4. 34 Evaluasi data *training* tiap model 52
Kode Program 4. 35 Evaluasi data *test* tiap model 52

BAB I
PENDAHULUAN

1.1 Latar Belakang

Emosi merupakan salah satu aspek penting dalam kehidupan manusia. Emosi terjadi dikarenakan adanya reaksi terhadap rangsangan dari luar dan dalam diri sendiri, yang berkaitan dengan perubahan fisiologis dan berbagai pikiran (Prawitasllri, 2016). Sering kali manusia sulit mengekpresikan perasaan emosi secara lisan, namun cenderung mengungkapkan apa saja yang diinginkan dan dipikirannya kedalam media sosial (Radja & Citra, 2020).

Media sosial telah menjadi wadah "curhat" seseorang untuk mengekspresikan emosi dan pikiran yang mereka alami ke media sosial tanpa adanya (Radja & Citra, 2020). Salah satu media sosial yang banyak digunakan adalah Twitter. Menurut data Hootsuite pada bulan juli 2022 pengguna media sosial Twitter telah mencapai 238 juta pengguna aktif diseluruh dunia dan Indonesia memperoleh peringkat ke-4 didunia dengan jumlah pengguna Twitter mencapai 21.2 juta pengguna aktif (Kemp, 2022). Twitter merupakan media sosial yang memiliki pengguna dengan jangkauan yang luas dan pengguna dapat menulis pesan yang dibatasi tidak lebih dari 140 karakter serta dapat menyertakan gambar, video, maupun tautan blog yang disebut sebagai *tweet* (Maclean et al., 2013).

Twitter menyediakan *Application Programming Interface* (API) agar pengguna dapat mengakses dan mencari informasi mengenai *tweet* secara spesifik, misalnya topik tertentu, lokasi, trending terkini, dan lain sebagainya (Eka Sembodo et al., 2016). Sehingga Twitter dapat digunakan sebagai sumber data dalam menganalisis emosi seseorang berdasarkan tulisan/*tweet* yang dibuatnya.

Dalam menganalisis emosi pada sebuah *tweet*, diperlukan suatu metode klasifikasi agar dapat mengklasifikasikan *tweet* ke dalam kelas emosi yang tepat. Klasifikasi merupakan salah satu metode dalam data mining yang digunakan untuk mengelompokkan objek data ke dalam kelas yang telah

ditentukan berdasarkan karakteristik sebelumnya (Setio et al., 2020). Klasifikasi emosi pada *tweet* bertujuan untuk mengelompokkan *tweet* ke dalam kelas yang sudah ditentukan seperti, marah, senang, takut, cinta, dan sedih. Klasifikasi emosi memiliki manfaat seperti dapat mengetahui target kepuasan kustomer mengenai produk maupun layanan yang diberikan, dan pada bidang pemerintahan dapat mengetahui pendapat dan tanggapan/respon masyarakat mengenai kebijakan pemerintah. Beberapa algoritma klasifikasi yang umum digunakan diantaranya *K-Nearest Neighbors, Logistic Regression, Naive Bayes Classifier, Decision Tree,* dan *Support Vector Machine*. Setiap algoritma mempunyai cara bagaimana membangun model dan mempunyai kompleksitas model masing-masing (Müller & Guido, n.d.).

Berdasarkan penelitian (Saputri et al., 2019) terkait klasifikasi emosi *tweet* dengan judul "*Emotion Classification on Indonesian Twitter Dataset*" penelitian menggunakan algoritma klasifikasi *Logistic Regression*, *Support Vector Machine*, dan *Random Forest* dengan menggunakan kombinasi *feature extraction Bag of Word, Word Embeddings, Lexicon Based, Part-of-Speech* (POS) *tag* dan *Orthograpic Feature*. Dataset diambil dari teks *tweet Twitter* dan menggunakan lima kelas emosi (senang, marah, sedih, cinta, dan takut), menunjukkan bahwa algoritma klasifikasi *Logistic Regression* memperoleh *f1-score* terbaik sebesar 75,98%

Adapun penelitian terkait mengenai klasifikasi emosi *tweet* yang pernah dilakukan (Negara et al., 2021) yaitu membandingkan algoritma *Logistic Regression* dan *K-Nearest Neighbors*, dengan menggunakan *feature extraction* TF-IDF dan *Tuning Hyperparameter*. Penelitian ini menunjukkan bahwa algoritma klasifikasi *Logistic Regression* dengan menggunakan TF-IDF dan *Tuning Hyperparameter* memperoleh nilai akurasi 65% dan *F1-score* sebesar 66%.

Penelitian (Yang et al., 2016) melakukan analisis sentimen *Twitter* mengenai perbandingan metode *SVM, Multinomial Naive Bayes, Decision Tree C4.5, KNN dan Logistic Regression* dengan menggunakan unigram sebagai

feature extraction. Hasil penelitian menunjukkan bahwa metode *Multinomial Naive Bayes* menghasilkan nilai *accuracy* tertinggi sebesar 84.60% dibandingkan *accuracy SVM* sebesar 82.30%, *Logistic Regression* sebesar 79.50%, *Decision Tree C4.5* sebesar 65.30%, dan *KNN* sebesar 64.40%.

Penelitian (Maulidah et al., 2020) melakukan perbandingan algoritma klasifikasi *Decision Tree*, *Naive Bayes*, SVC, K-NN dan *Random Forest* untuk mengklasifikasikan kategori rekomendasi buku. Menghasilkan bahwa algoritma *Decision Tree* memiliki akurasi yang paling tinggi sebesar 99.95% daripada *Naive Bayes*, SVC, K-NN dan *Random Forest*.

Penelitian lain dilakukan (Permatasari, 2018) kinerja beberapa *Ensemble Features* dan *Bag of Word features*. Penelitian ini menunjukkan bahwa hasil akurasi *Bag of Word* memiliki pengujian tertinggi yaitu akurasi sebesar 94%, presisi sebesar 95.83%, *recall* sebesar 92% dan *f-measure* sebesar 93.88%.

(Trianto et al., 2020) juga melakukan perbandingan kinerja *feature extraction* N-GRAM (Unigram, *Bigram*, dan Trigram) pada kasus klasifikasi *rating* dengan menggunakan algoritma *Naive Bayes*. Hasil penelitian menunjukkan bahwa penggunaan *feature extraction Bigram* dengan menggunakan algoritma *Naive Bayes* memperoleh hasil yang lebih tinggi dari Unigram dan Trigram. Hasil akurasi yang didapat sebesar 94.06%, presisi sebesar 90.71%, dan *recall* sebesar 91.73%.

Salah satu permasalahan yang terjadi dalam membangun sebuah model *machine learning,* yaitu dataset yang digunakan tidak seimbang jumlah data antar kelas atau label. SMOTE (*Synthetic Minority Over-Sampling Technique*) digunakan untuk mengatasi permasalahan *imbalance data*, metode ini akan membuat replikasi dari data minoritas. Replikasi yang dimaksud ialah synthetic data dengan mencari *k-nearest neighbors* (ketetanggaan terdekat data sebanyak k pada data minoritas (Sofyan & Prasetyo, 2021). Pernyataan ini diperkuat dengan adanya penelitian yang telah dilakukan oleh (Saputra et al., 2020), yaitu melakukan optimasi pada algoritma *Naive Bayes* dengan menggunakan *feature extraction Bigram* dan menggunakan metode SMOTE

pada *tweet* myXLCare. Penelitian tersebut memperoleh hasil nilai akurasi 86.33%, presisi 82.58%, dan *recall* 92.38%, nilai tersebut lebih baik jika dibandingkan *Naive Bayes* dengan menggunakan *feature extraction Bigram* dan tanpa menggunakan SMOTE, yang memperoleh nilai akurasi 78.88%, presisi 74. 43% dan *recall* 76.90%.

Berdasarkan uraian penelitian (Saputri et al., 2019) dan (Negara et al., 2021), maka penelitian ini berfokus pada perbandingan algoritma klasifikasi emosi yaitu *Multinomial Naïve Bayes* dan *Decision Tree*. Kedua algoritma ini dibandingkan untuk mencari algoritma manakah yang memiliki performa yang lebih baik. Tidak hanya membandingkan algoritma klasifikasi, tetapi penelitian ini juga membandingkan hasil algoritma menggunakan *Tuning Hyperparameter* dan dataset menggunakan *feature extraction Bag of Word* dan *Bigram*, serta metode SMOTE untuk menyeimbangkan data yang digunakan.

1.2 Perumusan Masalah

Rumusan masalah pada penelitian ini antara lain sebagai berikut:

1. Bagaimana pengaruh Algoritma klasifikasi *Multinomial Naive Bayes* dan *Decision Tree* terhadap performa model *machine learning* klasifikasi emosi *tweet*?
2. Bagaimana pengaruh penggunaan *feature extraction Bag of Wod* dan *Bigram* serta metode SMOTE dalam data yang digunakan untuk membangun model *machine learning* klasifikasi emosi?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan dari penelitian ini adalah :

1. Mengetahui algoritma klasifikasi yang paling optimal antara Multinomial Naive Bayes dan Decision Tree dalam membangun model machine learning klasifikasi emosi pada data tweet, dengan melakukan perbandingan hasil performa model menggunakan metrik nilai akurasi.
2. Mengetahui pengaruh penggunaan data dari teknik *feature extraction Bag of Words* dan *Bigram s*erta metode SMOTE dalam meningkatkan

kinerja model machine learning klasifikasi emosi, serta menentukan teknik mana yang lebih efektif dalam membangun model.

1.4 Batasan Masalah

Batasan masalah dilakukan untuk menghindari meluasnya permasalahan yang ada, sebagai berikut :

1. Data awal yang digunakan merupakan data *tweet* dari penelitian (Saputri et al., 2019).
2. Lima kelas emosi yang digunakan, yaitu marah, senang, sedih, takut, dan cinta.

1.5 Sistematika Penulisan

Adapun sistematika dari penulisan tugas akhir ini disusun dalam lima bab yang terdiri dari Bab I Pendahuluan, Bab II Tinjauan Pustaka, Bab III Metodologi Penelitian dan Perancangan Sistem, Bab IV Hasil Perancangan dan Analisis Sistem, serta Bab V Penutup.

Bab I Pendahuluan adalah bab yang berisi latar belakang, perumusan masalah, tujuan penelitian, pembatasan masalah, dan sistematika penulisan.

Bab II Tinjauan Pustaka adalah bab yang berisi landasan teori berkaitan dengan penelitian yang akan dilakukan.

Bab III Metodologi Penelitian adalah bab yang berisi tentang Data dan Alat Penelitian, Preprocessing, *feature extraction*, Pemodelan , dan Evaluasi.

Bab IV Hasil dan Analisis adalah bab yang berisi penjelasan mengenai implementasi pada sistem, screenshoot tampilan antarmuka sistem yang sudah jadi, serta analisis hasil uji coba. Setiap bagian sistem yang ditampilkan akan dilakukan analisis terlebih dahulu untuk mengarah kepada suatu kesimpulan.

Bab V Penutup adalah bab yang berisi kesimpulan dari penelitian yang telah dilakukan dan saran/rekomendasi untuk perbaikan, pengembangan atau kesempurnaan / kelengkapan penelitian yang telah dilakukan.