



Seminar Nasional Ilmu Teknik dan Aplikasi Industri (SINTA)

Alamat Prosiding: sinta.eng.unila.ac.id



Identifikasi kelas emosi majemuk pada kalimat majemuk Bahasa Indonesia menggunakan model *multinomial naïve bayes*

Aripin*, W Agastya dan H Haryanto

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang, Jl. Imam Bonjol 207 Semarang 50131

INFORMASI ARTIKEL

Riwayat artikel:

Diterima: 1 Oktober 2020

Direvisi: 19 November 2020

Kata kunci:

Kelas Emosi Majemuk
Kalimat Majemuk
Bahasa Indonesia
Multinomial Naïve Bayes
Persamaan Batas Dominan

ABSTRAK

Dalam proses komunikasi tatap muka, ekspresi wajah dapat mendukung agar proses komunikasi dapat lebih efektif. Ekspresi wajah seseorang dapat menunjukkan kondisi emosi sehingga dalam berkomunikasi orang lain dapat menyesuaikan. Dalam pengembangan film animasi wajah dan produksi *game*, ekspresi wajah juga diperlukan dalam adegan dialog sehingga film animasi dan *game* tersebut dapat terlihat lebih alami dan realistis. Dalam komunikasi yang menggunakan media teks, maka setiap kalimat yang digunakan dalam proses komunikasi tersebut juga merepresentasikan kondisi emosi seseorang. Secara umum, struktur kalimat yang digunakan dalam berkomunikasi adalah kalimat yang memiliki struktur majemuk. Emosi dari suatu kalimat dapat dikenali dengan melakukan proses klasifikasi terhadap kalimat tersebut. Hasil klasifikasi emosi terhadap kalimat dengan struktur majemuk biasanya menghasilkan kelas-kelas emosi majemuk (beberapa gabungan dari kelas-kelas emosi dasar). Penelitian ini bertujuan untuk melakukan klasifikasi terhadap kalimat majemuk Bahasa Indonesia menjadi beberapa kelas emosi dominan dengan menggunakan model *Multinomial Naïve Bayes*. Persamaan batas dominan digunakan untuk menentukan kelas-kelas yang dominan berdasarkan ambang batas tertentu. Hasil penelitian ini dapat diimplementasikan dalam pengembangan model karakter *virtual* berekspresi sehingga dapat mendukung pengembangan animasi wajah yang ekspresif untuk mendorong otomatisasi dalam proses produksi film animasi dan *game* khususnya animasi wajah. Hasil penelitian ini menunjukkan bahwa tingkat akurasi dalam proses klasifikasi dengan menggunakan *Multinomial Naïve Bayes* dan persamaan batas dominan untuk kalimat-kalimat majemuk Bahasa Indonesia adalah 75,47%.

1. Pendahuluan

Proses komunikasi tatap muka yang melibatkan manusia, umumnya dilakukan secara verbal dan non-verbal. Komunikasi secara verbal dapat diartikan dengan komunikasi langsung yang menggunakan bahasa lisan maupun tulisan sebagai media komunikasi. Sedangkan komunikasi non-verbal merupakan komunikasi yang umumnya menggunakan bahasa tubuh seperti ekspresi wajah, gelengan kepala, gerakan tangan tindakan dan sebagainya. Komunikasi verbal dan non-verbal sangat memerlukan ekspresi wajah untuk menegaskan kondisi emosi yang sedang dialami oleh seseorang sehingga komunikasi dapat berlangsung lebih efektif (Prawitasllri, 2016). Pada era teknologi seperti sekarang ini, komunikasi yang melibatkan ekspresi wajah juga digunakan dalam produksi film animasi, game dan penutur virtual lainnya. Karakter virtual dituntut dapat memvisualisasikan adegan komunikasi dalam bentuk adegan dialog maupun sebagai penutur secara realistis dan alami layaknya manusia. Banyak aplikasi yang melibatkan karakter virtual yang realistis dan alami yang banyak beredar dipasaran.

Namun bahasa yang digunakan dalam aplikasi tersebut adalah bahasa asing terutama bahasa Inggris.

Bahasa merupakan salah satu alat komunikasi yang digunakan manusia untuk bertukar informasi dan berdialog. Bahasa Indonesia merupakan salah satu bahasa di dunia yang telah menjadi bahasa resmi dan alat komunikasi bagi orang Indonesia. Indonesia merupakan negara dengan keragaman suku dan bahasa yang sangat tinggi dibandingkan dengan negara lain di dunia. Bahasa Indonesia memiliki peran penting bagi negara ini untuk menjadi pemersatu keanekaragaman bahasa yang ada dan sebagai sarana komunikasi nasional. Indonesia memiliki kedudukan penting di dunia internasional (Widodo, 2020). Bahasa Indonesia dipelajari di lebih dari 45 negara dan banyak digunakan di Wordpress. Fakta lain menunjukkan bahwa kunjungan orang asing ke Indonesia semakin meningkat sehingga kebutuhan penggunaan Bahasa Indonesia juga semakin meningkat (Badan Pusat Statistik, 2019). Hal-hal ini menunjukkan bahwa Bahasa Indonesia memiliki posisi penting diantara bahasa-bahasa lain di dunia dan sangat berpotensi menjadi bahasa asing kedua setelah Bahasa Inggris.

* Aripin.

E-mail: arifin@dsn.dinus.ac.id

Struktur kalimat Bahasa Indonesia umumnya terdiri dari subyek, predikat, obyek, pelengkap dan keterangan. Secara umum, struktur kalimat Bahasa Indonesia dapat digolongkan berdasarkan jumlah klausa (gagasan yang ada dalam suatu kalimat) menjadi struktur kalimat tunggal dan struktur kalimat majemuk. Kalimat tunggal merupakan kalimat Bahasa Indonesia yang hanya memiliki satu klausa (gagasan) saja (Widodo, 2020). Sedangkan kalimat Bahasa Indonesia yang majemuk merupakan kalimat Bahasa Indonesia yang memiliki lebih dari satu klausa (gagasan) dan setiap klausa tersebut dipisahkan oleh kata penghubung (konjungsi).

Emosi merupakan keadaan kondisi perasaan seseorang yang dapat berpengaruh pada perilaku, perubahan fisiologi dan berpikir. Emosi seseorang dipengaruhi oleh rangsangan dari luar maupun dari dalam diri seseorang. Emosi ini dapat ditularkan pada orang lain yang memperhatikan kondisi emosi seseorang tersebut. Misalnya, seseorang yang sedang mengalami kondisi emosi yang positif seperti senang, gembira dan hangat, maka akan membuat orang lain yang ikut memperhatikan kondisi emosi seseorang tersebut dengan ikut merasakan senang, gembira dan kehangatan. Hal demikian juga terjadi pada emosi negatif seperti sedih, marah dan takut (Ekman dan Cordaro, 2011).

Emosi dari suatu kalimat Bahasa Indonesia dapat dikenali melalui kata-kata sifat yang terdapat dalam kalimat tersebut (Sasangka, 2014). Kondisi emosi seperti marah, sedih, senang, takut dan terkejut dapat dikenali dari jumlah kata-kata sifat yang lebih banyak yang mencerminkan kondisi emosi tersebut. Jumlah kata-kata sifat untuk kondisi emosi marah, sedih, senang dan takut memiliki lebih banyak kata-kata sifat dibandingkan dengan kata-kata sifat untuk kondisi emosi seperti jijik, malu dan terkejut (Sasangka, 2014). Dalam penelitian kondisi emosi yang digunakan sebagai acuan adalah enam emosi dasar yang digolongkan oleh Paul Ekman. Enam emosi dasar tersebut adalah senang, sedih, marah, takut, jijik dan terkejut. Emosi senang menggambarkan kondisi keberhasilan ataupun kepuasan seseorang terhadap sesuatu. Sedih menggambarkan kegagalan atau penyesalan seseorang. Marah menggambarkan kondisi emosi frustrasi dan adanya ganjalan (kendala) terhadap sesuatu. Takut menggambarkan kondisi emosi khawatir terhadap suatu hal yang mengancam. Sedangkan jijik menggambarkan kondisi penolakan terhadap sesuatu dan rasa rishi. Dan terkejut menggambarkan perasaan terhadap sesuatu yang datang secara tiba-tiba (mendadak). Pemetaan emosi-emosi tersebut dari suatu kalimat Bahasa Indonesia dapat dilakukan dengan menggunakan salah satu aplikasi dari bidang ilmu *Natural Language Processing (NLP)*.

Pemetaan emosi berdasarkan sebuah kalimat Bahasa Indonesia juga dapat dilakukan dengan menggunakan proses klasifikasi dengan algoritma tertentu. Beberapa algoritma tersebut antara lain adalah K-NN (*k-Nearest Neighbor*), SVM (*support Vector Machine*), Naïve Bayes dan masih banyak lagi. Salah satu algoritma yang dikembangkan dari model Naïve Bayes adalah *Multinomial Naïve Bayes (MNB)*. Model MNB merupakan model probabilistik untuk menentukan kelas (kategori). MNB mengolah data dengan cara distribusi multinomial yaitu dengan mengukur tingkat intensitas dari suatu kategori (dalam penelitian ini adalah emosi). Dengan cara ini dapat dihitung frekuensi kemunculan kata-kata sifat tertentu yang merepresentasikan sebuah emosi dalam sebuah kalimat berbahasa Indonesia. Proses klasifikasi pada kalimat Bahasa Indonesia yang memiliki struktur majemuk dapat menghasilkan lebih dari satu kelas emosi. Nilai probabilitas dari tiap kelas emosi yang dihasilkan dari proses klasifikasi dihitung nilai ambang batasnya untuk menentukan kelas-kelas emosi yang dominan.

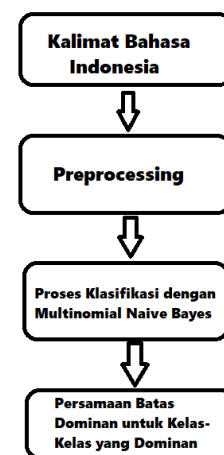
Persamaan batas dominan merupakan salah satu metode yang digunakan untuk menentukan kelas-kelas emosi mana saja yang dominan berdasarkan nilai ambang batasnya. Persamaan batas dominan bekerja dengan cara mengolah nilai probabilitas yang dihasilkan dari proses klasifikasi yang dibandingkan dengan nilai ambang batasnya, Apabila nilai probabilitas lebih besar dari nilai ambang batasnya, maka kelas emosi tersebut dikategorikan sebagai kelas emosi yang dominan.

Penelitian-penelitian sebelumnya yang menggunakan model Naïve Bayes dalam melakukan proses klasifikasi diuraikan dalam paragraf ini. Sebuah penelitian yang menggunakan kombinasi dari algoritma *Naïve Bayes* dan *Fuzzy Logic* yang digunakan untuk memvisualisasikan ekspresi wajah pada karakter virtual berdasarkan masukan sebuah kalimat (Arief dkk., 2010). Proses klasifikasi terhadap kalimat Bahasa Indonesia pernah dilakukan dengan menggunakan algoritma k-NN (Aripin, 2012; Sofiyana dkk., 2012). Dalam penelitian ini nilai akurasi yang dihasilkan dari proses klasifikasi masih tergolong rendah. Terdapat sebuah penelitian yang membahas proses klasifikasi emosi dengan menggunakan algoritma VSM (Sitorus dan Dachlan, 2017). Algoritma Naïve Bayes juga dapat digunakan untuk melakukan klasifikasi emosi terhadap lirik lagu berbahasa Indonesia (Noviardi dkk., 2018). Sebuah penelitian membahas mengenai klasifikasi dengan menggunakan model Naïve Bayes terhadap teks chatting untuk menghasilkan respon emosi berdasarkan makna teks tersebut (Martiana dkk., 2011). Berdasarkan uraian beberapa penelitian sebelumnya yang menggunakan algoritma *Naïve Bayes*, maka algoritma ini dipilih sebagai sebuah model yang diyakini dapat memberikan hasil akurasi yang lebih baik.

2. Metodologi

Beberapa tahap penelitian yang perlu dilakukan agar mampu mengidentifikasi kelas emosi majemuk dari sebuah kalimat berbahasa Indonesia secara garis besar digambarkan seperti Gambar 1.

Preprocessing terhadap kalimat Bahasa Indonesia yang dimasukan kedalam sistem adalah *case folding, filtering number & punctuation, tokenizing, filtering anstem, stemming, set idiom, set data-not, dan filtering stopword*. Tahap selanjutnya adalah proses klasifikasi terhadap dataset yang dihasilkan dari preprocessing. Nilai probabilitas dari proses klasifikasi difilter berdasarkan nilai ambang batasnya. Nilai probabilitas yang lebih besar dari nilai ambang batasnya dipilih sebagai kelas emosi yang dominan.



Gambar 1. Tahapan Penelitian

2.1 Akuisisi data dan oversampling

Dalam tahap ini, dikumpulkan data kalimat berbahasa Indonesia dari berbagai sumber, seperti buku, majalah, koran dan media digital. Kalimat yang dikumpulkan merupakan kalimat dengan struktur kalimat tunggal maupun kalimat majemuk. Selanjutnya, seluruh data kalimat tersebut dikelompokkan berdasarkan kelas-kelas emosinya dengan pengecakan dan validasi oleh ahli bahasa dan ahli bidang psikologi untuk memastikan kevalidan data kalimat agar sesuai dengan kelompok emosinya. Selain itu, jumlah kata sifat yang merepresentasikan kondisi emosi tertentu juga diidentifikasi untuk memastikan keseimbangan jumlah kata sifat yang digunakan dalam kalimat-kalimat Bahasa Indonesia. Hasil identifikasi jumlah kata yang merepresentasikan kondisi emosi tertentu disajikan di Tabel 1.

Tabel 1. Jumlah kata sifat dalam setiap kelas-kelas emosi

Kelas Emosi	Jumlah Kata
Senang	276
Sedih	367
Marah	243
Takut	137
Jijik	91
Terkejut	29

Dari Tabel 1 dapat dilihat bahwa jumlah kata sifat untuk masing-masing kelas emosi memiliki jumlah yang tidak sama (terjadi ketidakseimbangan). Selisih jumlah data yang terlalu besar dapat menyebabkan permasalahan pada saat proses klasifikasi. Pada saat sebuah kalimat dimasukkan ke dalam system dapat saja terjadi bahwa kalimat tersebut tidak masuk dalam kategori kelas emosi manapun. Kelas yang minoritas menjadi prioritas saat sebuah model sedang memproses kalimat tersebut. Langkah yang dapat dilakukan untuk mengatasi permasalahan diatas adalah dengan menambahkan metode oversampling untuk menggandakan setiap data yang ada dikelas minoritas sampai mendapatkan kondisi jumlah data yang seimbang. Hasil dari penerapan metode oversampling ditunjukkan di Tabel 2.

Tabel 2. Jumlah data setelah penerapan metode oversampling

Kelas Emosi	Jumlah Kata
Senang	365
Sedih	367
Marah	365
Takut	362
Jijik	364
Terkejut	364

Tahap selanjutnya setelah jumlah data seimbang untuk masing-masing kelas emosi adalah pemrosesan data awal (*preprocessing*).

2.2 Pemrosesan data awal (*preprocessing*)

Kalimat-kalimat Bahasa Indonesia yang telah dikumpulkan tidak dapat langsung dilakukan proses klasifikasi. Data kalimat harus melalui pemrosesan data awal agar data dapat dilakukan proses klasifikasi. Beberapa kegiatan pemrosesan data awal yang dilakukan dalam penelitian ini terdiri dari : *case folding, filtering number & punctuation, tokenizing, filtering unstem, stemming, set idiom, set data-not, dan filtering stopword*.

Tahap awal *preprocessing* adalah *case folding* yang merubah seluruh huruf dalam kalimat Bahasa Indonesia menjadi huruf kecil semua. Proses ini bertujuan untuk mengurangi perbedaan yang disebabkan perbedaan penulisan huruf besar dan huruf

kecil. Langkah selanjutnya adalah menghilangkan (menghapus) seluruh tanda baca dan tulisan berupa nomor-nomor yang ada pada kalimat tersebut. Proses ini dinamakan *filtering number & punctuation*. Hal ini perlu dilakukan agar data yang tidak memiliki informasi yang berkaitan dengan pemilihan kelas ditiadakan.

Tahap *tokenizing* bertujuan untuk memisahkan setiap kata dalam kalimat Bahasa Indonesia sehingga dapat mempermudah proses perhitungan dalam penerapan model. Setelah itu, tahap pemilahan kata-kata yang tidak boleh masuk pada tahap *stemming*. Tahap ini disebut dengan tahap *filtering unstem*. Kata-kata yang dipilih dalam tahap ini tidak diubah menjadi kata dasar karena akan mengubah maknanya. Misalnya kata ‘meninggal’ tidak akan diubah ke dalam kata dasar ‘tinggal’ karena ‘meninggal’ dan ‘tinggal’ berbeda makna. Tahap selanjutnya adalah *stemming* yaitu mengubah setiap kata menjadi kata dasar dengan menghilangkan imbuhan yang ada pada kata tersebut. Proses *stemming* tidak akan berlaku pada kata-kata yang telah dipilah dalam tahap *filtering unstem*. Proses *stemming* merupakan proses yang sangat penting dalam model klasifikasi yang sangat berpengaruh pada tingkat akurasi.

Tahap selanjutnya adalah *set idiom* yaitu menggabungkan kata-kata dasar yang termasuk dalam daftar kata idiom. Misalnya kata ‘tanggung’ digabungkan dengan kata ‘jawab’ menjadi ‘tanggung jawab’. Proses set idiom bertujuan untuk menghindari adanya kata-kata yang termasuk dalam idiom tetapi masih terpisah. Langkah selanjutnya adalah proses *set data-not* yaitu menggabungkan kata-kata pengecualian dengan kata yang ada untuk mempertahankan makna dari setiap kata tersebut. Misalnya kata ‘tidak suka’ memiliki makna yang berbeda dengan kata ‘suka’. Proses selanjutnya menghilangkan kata sambung yang dinamakan dengan *filtering stopword*. Proses menghilangkan kata sambung dilakukan dengan cara mencocokkan kata-kata dengan mengacu pada daftar kata kata sambung yang akan dihapus (*stopword*). Misalnya kata sambung ‘yang’, ‘di’, ‘dan’, ‘pada’ dan lain sebagainya. Proses ini dilakukan ditahap akhir untuk menghilangkan sisa-sisa kata dari tahap *filtering unstem, set idiom, dan set data-not*.

2.3 Klasifikasi dengan multinomial Naïve Bayes

Multinomial Naïve Bayes (MNB) merupakan serapan dari teorema Bayes denominatornya diasumsikan konstan. Metode ini mengambil keputusan berdasarkan perhitungan probabilitas. Keuntungan MNB adalah memiliki performansi yang lebih baik dan metode perhitungan yang tidak terlalu rumit (Shivam Bachhety dkk., 2018).

$$P(C_k|X) \propto P(C_k).P(X|C_k) \tag{1}$$

dimana $P(C_k|X)$ adalah *posterior probability* pada dokumen X terhadap kelas C_k , $P(C_k)$ adalah prior probability atas kemunculan kelas C_k di seluruh C pada tiap dokumen X. $P(X|C_k)$ adalah *likelihood probability* dari kemunculan term fitur X pada kelas C_k .

Dalam penelitian ini MNB diaplikasikan langsung untuk menghitung jumlah frekuensi kemunculan kata-kata tertentu pada setiap kalimat Bahasa Indonesia. Model MNB yang digunakan ditunjukkan seperti persamaan 2.

$$P(w_i|c_j) = \frac{\text{count}(w_i,c_j)+\alpha}{\text{count}(c_j)+|V|} \tag{2}$$

dimana $P(w_i|c_j)$ merupakan nilai probabilitas kata ke i terhadap kelas j , $\text{count}(w_i, c_j)$ adalah nilai frekuensi kemunculan kata

ke i pada kelas j dan α adalah nilai Laplacian smoothing (dimana $\alpha = 1$). Nilai $count(c_j)$ merupakan jumlah anggota kelas j . $|V|$ adalah jumlah anggota dari seluruh kelas tanpa pengandaan.

Model MNB melakukan proses klasifikasi terhadap kalimat-kalimat Bahasa Indonesia dilakukan dengan cara memilih kelas yang didasarkan pada hasil nilai probabilitas yang paling tinggi. Selanjutnya, untuk menentukan kelas emosi dominan dilakukan dengan menggunakan persamaan batas dominan berdasarkan nilai probabilitas yang paling tinggi tersebut. Dengan persamaan batas dominan dapat ditentukan kelas emosi yang memiliki nilai intensitas yang dominan. Nilai probabilitas yang telah dihasilkan dari model MNB selanjutnya diubah terlebih dahulu menjadi nilai prosentase untuk mendapatkan nilai yang universal. Perubahan ke nilai prosentase juga dimaksudkan untuk menghindari nilai probabilitas yang berbeda-beda setiap kalimat Bahasa Indonesia. Perubahan ke nilai prosentase dilakukan dengan menggunakan persamaan 3.

$$Persentase_i = \frac{Probabilitas_i}{\sum Probabilitas} \times 100\% \quad (3)$$

$Persentase_i$ merupakan nilai prosentase dari kelas emosi ke i dan $Probabilitas_i$ merupakan nilai probabilitas kelas emosi ke i . Sedangkan $\sum Probabilitas$ adalah nilai total dari seluruh nilai probabilitas kelas emosi.

Nilai prosentase dari setiap kelas emosi selanjutnya ditentukan nilai batas dominan dengan menggunakan persamaan batas dominan seperti disajikan di persamaan 4.

$$dominan = \min(p) + \beta \quad (4)$$

Dominan adalah nilai batas dominan. Nilai batas dominan yang diperoleh digunakan untuk menentukan kelas-kelas dominan. Nilai $\min(p)$ merupakan nilai prosentase terkecil dan β adalah nilai yang ditambahkan untuk membedakan kalimat netral dan kalimat selain netral (memiliki kelas emosi dominan tertentu). Nilai default β adalah 3%. Dalam penelitian nilai β ditentukan berdasarkan hasil eksperimen dari nilai probabilitas yang dihasilkan pada setiap kalimat Bahasa Indonesia yang dimasukkan ke sistem. Kelas emosi dominan ditentukan berdasarkan nilai prosentase yang lebih besar dibandingkan dari nilai batas dominan yang dihasilkan dari persamaan 3.

3. Hasil dan Pembahasan

Pada tahap ini, kami melakukan 9 kali eksperimen dengan menggunakan beberapa variasi pada jumlah data pelatihan dan jumlah data uji. Setiap eksperimen dihitung nilai *precision*, *recall*, *f1-score* dan *accuracy* seperti yang disajikan di Tabel 3. Nilai akurasi paling kecil adalah 52,70% pada kondisi data pelatihan berjumlah 10% dan jumlah data uji sebesar 90%. Dan nilai akurasi tertinggi diperoleh dari data pelatihan berjumlah 90% dan jumlah data uji sebesar 10% dengan nilai akurasi adalah 75,47%. Nilai akurasi yang dihasilkan dari eksperimen tersebut menunjukkan bahwa performa model multinomial naïve bayes yang cukup baik untuk mengenali kelas-kelas emosi majemuk pada kalimat-kalimat majemuk Bahasa Indonesia.

Kalimat-kalimat Bahasa Indonesia yang digunakan sebagai dataset dalam eksperimen ini berjumlah 1143 kalimat. Dan jumlah kalimat Bahasa Indonesia setelah oversampling adalah 2187 kalimat. Beberapa contoh kalimat majemuk yang digunakan sebagai dataset disajikan di Tabel 4. Kalimat majemuk Bahasa Indonesia yang digunakan sebagai dataset telah dipilih oleh bidang linguistik dan bidang psikologi yang memiliki kata-kata sifat yang merepresentasikan kondisi emosi tertentu.

Di Tabel 3 dijelaskan bahwa nilai *recall*, *precision*, *f1-score* dan *accuracy* memiliki hasil perhitungan bervariasi berdasarkan prosentase jumlah data pelatihan dan data ujinya. Dalam 9 kali eksperimen yang dilakukan, Jumlah data pelatihan dan data uji diberikan bervariasi mulai dari 10% sampai dengan 90%. Dan demikian juga dengan data uji. Nilai akurasi paling kecil yang diperoleh adalah 52,70% untuk jumlah data pelatihan 10% dan jumlah data uji sebesar 90%. Sedangkan nilai akurasi yang paling tinggi adalah 75,47% untuk jumlah data pelatihan sebesar 90% dan jumlah data uji sebesar 10%. Kondisi ini menunjukkan bahwa performa model MNB untuk dataset yang digunakan dalam penelitian adalah baik yaitu dengan akurasi sebesar 75,47%. Hal ini berarti bahwa model MNB mampu melakukan klasifikasi terhadap kalimat-kalimat berbahasa Indonesia kedalam kelas-kelas emosi.

Setelah nilai probabilitas untuk setiap kelas-kelas emosi dari masing-masing kalimat Bahasa Indonesia di Tabel 4 yang digunakan sebagai data uji berhasil dihitung seperti yang ditunjukkan di Tabel 5, maka langkah selanjutnya adalah melakukan pemetaan terhadap kelas-kelas emosi yang dominan dengan menggunakan persamaan batas dominan. Langkah penting yang harus dilakukan agar penentuan kelas-kelas dominan dapat dilakukan adalah mengkonversi nilai-nilai dari Tabel 5 kedalam nilai prosentase dengan menggunakan persamaan 3. Hasil konversi ke nilai prosentase ditunjukkan di Tabel 6.

Tabel 3. Nilai *precision*, *recall*, *F1-score* dan *accuracy* dari tiap eksperimen

Data Latih	Data Uji	Precision	Recall	F1-Score	Accuracy
10%	90%	52,70%	52,70%	52,70%	52,70%
20%	80%	61,51%	58,40%	58,23%	58,40%
30%	70%	62,51%	61,29%	60,86%	61,29%
40%	60%	63,04%	61,91%	61,57%	61,76%
50%	50%	66,56%	65,79%	65,59%	64,24%
60%	60%	69,52%	67,29%	66,57%	67,29%
70%	70%	70,94%	69,28%	69,50%	69,28%
80%	20%	73,98%	73,58%	73,62%	73,58%
90%	10%	77,44%	75,47%	75,43%	75,47%

Tabel 4. Sampel beberapa kalimat majemuk Bahasa Indonesia sebagai dataset

No	Kalimat Majemuk Bahasa Indonesia
#1	kurang ajar siapa yang berani melakukan ini, apa jangan-jangan itu hantu ya
#2	aku lega dia tidak mengetahuinya, namun aku khawatir dengan resiko yang akan datang
#3	tempat itu memang horor, tetapi aku harus mengambil tasku yang tertinggal
#4	aku mendapat kado dari ayah, namun betapa terkejutnya aku ternyata ini adalah sepatu yang selama ini aku inginkan
#5	aku sedih berat badanku naik, aku kesal dibully terus sama teman-temanku
#6	aku ingin menghajar para gerombolan itu, tapi nyaliku ciut aku hanya sendirian
#7	akhirnya kegiatan jelajah malam kali ini selesai, namun ketika aku hendak ke toilet aku dikejutkan ular yang tiba-tiba di depanku akupun lari terbirit birit
#8	Aku tidak menyangka dia akhirnya pulang, aku memeluknya sambil menangis

No	Kalimat Majemuk Bahasa Indonesia
#9	kasihan sejak usia delapan tahun andi telah ditinggalkan ayahnya tetapi sekarang dia sudah sukses dan bahagia
#10	malam ini malam jumat kliwon cintya tidak berani tidur sendirian
#11	Presiden Jokowi hari ini melantik menteri baru bernama Handoko Budianto

Tabel 5. Nilai probabilitas tiap kelas emosi

No.	Senang	Sedih	Marah	Takut	Jijik	Terkejut
#1	0,0119	0,0358	0,5442	0,2256	0,0122	0,1702
#2	0,1441	0,0152	0,0076	0,8176	0,0077	0,0077
#3	0,0499	0,1002	0,0503	0,6978	0,0508	0,0508
#4	0,5051	0,1273	0,0107	0,0052	0,0054	0,3461
#5	0,0170	0,5833	0,3477	0,0170	0,0175	0,0175
#6	0,0283	0,0283	0,4315	0,4540	0,0289	0,0289
#7	0,0009	0,0008	0,0035	0,1789	0,0045	0,8112
#8	0,0284	0,5119	0,0286	0,0283	0,0289	0,3739
#9	0,3229	0,5678	0,0273	0,0268	0,0275	0,0275
#10	0,0007	0,0006	0,0013	0,9958	0,0007	0,0007
#11	0,1647	0,1660	0,1673	0,1646	0,1687	0,1687

Tabel 6. Hasil konversi nilai probabilitas ke nilai Persentase

No.	Senang	Sedih	Marah	Takut	Jijik	Terkejut
#1	1,19%	3,58%	54,43%	22,56%	1,22%	17,02%
#2	14,41%	1,52%	0,76%	81,76%	0,77%	0,77%
#3	4,99%	10,03%	5,04%	69,79%	5,08%	5,08%
#4	50,51%	12,74%	1,08%	0,53%	0,54%	34,61%
#5	1,70%	58,33%	34,77%	1,70%	1,75%	1,75%
#6	2,83%	2,83%	43,15%	45,41%	2,89%	2,89%
#7	0,09%	0,09%	0,36%	17,90%	0,45%	81,12%
#8	2,84%	51,19%	2,86%	2,84%	2,89%	37,39%
#9	32,29%	56,78%	2,73%	2,69%	2,75%	2,75%
#10	0,07%	0,07%	0,14%	99,58%	0,07%	0,07%
#11	16,47%	16,60%	16,73%	16,47%	16,87%	16,87%

Setelah mendapat nilai prosentase dari masing-masing kelas emosi, maka selanjutnya menentukan nilai batas dominan dengan menggunakan persamaan batas dominan seperti yang ditulis dipersamaan 4. Kelas-kelas emosi yang memiliki nilai prosentase diatas nilai batas dominan maka kelas emosi tersebut dianggap sebagai kelas-kelas emosi yang representatif terhadap kondisi emosi dari suatu kalimat Bahasa Indonesia. Apabila terdapat lebih dari satu kelas emosi yang dominan maka berarti kalimat majemuk Bahasa Indonesia tersebut memiliki kelas-kelas emosi majemuk (kelas emosi gabungan) seperti yang ditunjukkan di Tabel 10.

Persamaan 4 yang digunakan untuk menghitung nilai batas dominan harus ditambahkan dengan nilai β yang optimal. Nilai β yang optimal diperoleh dengan cara melakukan pengujian terhadap salah satu kalimat majemuk Bahasa Indonesia seperti yang ditunjukkan di Tabel 7. Nilai

yang diperoleh dalam bentuk nilai probabilitas dan nilai persentase. Selanjutnya diambil salah satu kelas yang memiliki nilai terkecil. Nilai ini digunakan sebagai nilai untuk parameter $min(p)$ pada Persamaan 4.

Langkah selanjutnya adalah melakukan eksperimen terhadap nilai β dari 0 sampai dengan 10%. Penentuan nilai β nantinya digunakan untuk menentukan nilai d . Penentuan kelas emosi dominan dengan syarat bahwa nilai kelas emosi \geq nilai batas dominan dan hasil eksperimennya ditunjukkan di Tabel 8.

Tabel 7. Kalimat majemuk untuk pengujian persamaan batas dominan

Kalimat	Kelas	Probabilitas	Persentase
kurang ajar siapa yang berani melakukan ini, apa jangan-jangan itu hantu ya	Senang	0,011875	1,19%
	Sedih	0,035824	3,58%
	Marah	0,544285	54,43%
	Takut	0,225631	22,56%
	Jijik	0,012172	1,22%
Terkejut	0,170213	17,02%	

Tabel 8. Hasil eksperimen penentuan nilai β dan dominan

$Min(p)$	β	Dominan	Emosi Dominan
1,19%	0%	1,19%	Sedih, Marah, Takut, Jijik, Terkejut
1,19%	1%	2,19%	Sedih, Marah, Takut, Terkejut
1,19%	2%	3,19%	Sedih, Marah, Takut, Terkejut
1,19%	3%	4,19%	Marah, Takut, Terkejut
1,19%	4%	5,19%	Marah, Takut, Terkejut
1,19%	5%	6,19%	Marah, Takut, Terkejut
1,19%	6%	7,19%	Marah, Takut, Terkejut
1,19%	7%	8,19%	Marah, Takut, Terkejut
1,19%	8%	9,19%	Marah, Takut, Terkejut
1,19%	9%	10,19%	Marah, Takut, Terkejut
1,19%	10%	11,19%	Marah, Takut, Terkejut

Berdasarkan Tabel 8 ditunjukkan nilai $min(p)$ yang diperoleh adalah 1,19%. Nilai ini ditambahkan ke nilai β sehingga menghasilkan nilai d . Nilai diperoleh ini merupakan nilai batas dominan. Di Tabel 8 ditunjukkan hasil dari beberapa eksperimen. Eksperimen dengan nilai $\beta = 0\%$ dapat menghasilkan lima kelas emosi yang dominan. Sedangkan, dengan nilai β antara 1 sampai dengan 2% dapat menghasilkan empat kelas emosi dominan. Pada nilai β sama dengan 3% dapat menghasilkan tiga kelas emosi dominan. Kondisi ini tidak berubah jumlah kelas emosi dominannya sampai dengan nilai β sama dengan 10%. Oleh karena itu, nilai β yang digunakan adalah 3 sampai dengan 10% sehingga pada penelitian ini nilai β yang digunakan adalah 3%.

Identifikasi kelas-kelas emosi majemuk yang dilakukan dalam penelitian ini menggunakan model MNB dan persamaan batas dominan untuk memilih kelas-kelas emosi yang representatif berdasarkan nilai ambang batas tertentu. Kinerja identifikasi kelas-kelas emosi dengan menggunakan model MNB dan persamaan batas dominan menunjukkan hasil yang lebih baik

dibandingkan dengan identifikasi kelas-kelas emosi hanya dengan model MNB saja tanpa menyertakan persamaan batas dominan. Yang dimaksud dengan kinerja yang lebih baik adalah bahwa kelas-kelas emosi yang berhasil diidentifikasi lebih representatif terhadap kondisi emosi yang terkandung dari kalimat majemuk yang diklasifikasi seperti ditunjukkan di Tabel 9.

Tabel 9. Perbandingan Kinerja Identifikasi menggunakan Model MNB dengan Model MNB + Persamaan Batas Dominan

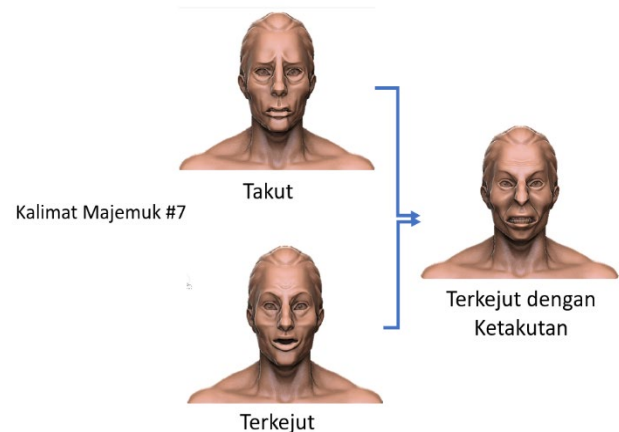
Kalimat	Pemetaan Emosi	
	MNB	MNB + PBD
kurang ajar siapa yang berani melakukan ini, apa jangan-jangan itu hantu ya	Marah (54,43%)	Marah (54,43%), Takut (22,56%), Terkejut (17,04%)

Hasil identifikasi terhadap seluruh kalimat majemuk yang tertera di Tabel 4 dengan proses klasifikasi menggunakan model MNB dan penentuan kelas yang dominan dengan persamaan batas dominan dapat dilihat di Tabel 10.

Tabel 10. Nilai batas dominan dan kelas-kelas dominan

No.	Batas Dominan		Kelas-kelas Emosi Dominan
	Nilai Probabilitas	Nilai Prosentase	
#1	0,041934	4,19%	Marah (54,42%) Takut (22,56%) Terkejut (17,02%)
#2	0,037688	3,76%	Senang (14,41%) Takut (81,76%)
#3	0,079944	7,99%	Sedih (10,03%) Takut (69,79%)
#4	0,035331	3,53%	Senang (50,51%) Sedih (12,74%) Terkejut (34,61%)
#5	0,047056	4,70%	Sedih (58,33%) Marah (34,77%)
#6	0,058295	5,83%	Marah (43,15%) Takut (45,41%)
#7	0,030841	3,09%	Takut (17,90%) Terkejut (81,12%)
#8	0,058368	5,84%	Sedih (51,19%) Terkejut (37,39%)
#9	0,056827	5,69%	Senang (32,29%) Sedih (56,78%)
#10	0,030664	3,07%	Takut (99,58%)
#11	0,194420	19,47%	Netral

Salah satu contoh implementasi kelas-kelas emosi majemuk kedalam animasi karakter wajah adalah ekspresi majemuk seperti ditampilkan di Gambar 2.



Gambar 2. Visualisasi ekspresi majemuk yang merupakan kombinasi dari emosi takut dan terkejut

4. Kesimpulan

Berdasarkan hasil dari beberapa eksperimen yang telah dijelaskan diatas, maka dapat diambil kesimpulan bahwa Penggabungan model MNB dan Persamaan Batas Dominan mampu melakukan proses klasifikasi dan penentuan kelas-kelas emosi dominan terhadap kalimat-kalimat majemuk Bahasa Indonesia dengan hasil akurasi yang cukup bagus yaitu 75.47%. Kelas-kelas emosi dominan yang dihasilkan dari penelitian ini selanjutnya dapat berkontribusi untuk mengotomatisasi animasi ekspresi karakter wajah yang majemuk. Berdasarkan kalimat majemuk Bahasa Indonesia. Kelas-kelas emosi yang dominan dapat divisualisasikan dalam bentuk animasi 3D wajah. Dengan otomatisasi ekspresi wajah ini nantinya dapat mengefisienkan dan mempermudah pengembangan animasi karakter wajah sehingga dapat mendukung pengembangan film animasi wajah dan produksi game.

Ucapan terima kasih

Kami mengucapkan banyak terima kasih kepada Kementerian Riset, Teknologi dan Pendidikan Tinggi melalui Direktorat Riset dan Pengabdian Kepada Masyarakat yang telah mendukung pendanaan penelitian ini melalui program hibah bersaing dengan skim Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT). Ucapan terima kasih juga kami sampaikan kepada seluruh pimpinan Universitas Dian Nuswantoro yang telah memberikan dorongan, dukungan, spirit dan fasilitas sarana prasarana sehingga penelitian ini dapat diselesaikan.

Daftar pustaka

- Aripin, (2012), Classification of Emotions in Indonesian Text using K-NN Method, *Int. J. Inf. Electron. Eng.*, vol. 2, no. 6, pp. 899–903, 2012, doi: 10.7763/ijee.2012.v2.237.
- Badan Pusat Statistik, (2019), Jumlah kunjungan wisman ke Indonesia Oktober 2019 mencapai 1,35 juta kunjungan, Perkembangan Pariwisata Oktober 2019. [Online]. Available: <https://www.bps.go.id/pressrelease/2019/12/02/1618/jumlah-kunjungan-wisman-ke-indonesia-oktober-2019-mencapai-1-35-juta-kunjungan.html>. [Accessed: 01-Mar-2020].
- E. Martiana, R. Y. Hakkun, N. R. M, and M. Firodh, (2011), Machine learning dalam Program Chatting untuk Merespon Emosi Teks Berbahasa Indonesia Menggunakan Text Mining dan Naïve Bayes, *Electron. Eng. Polytech. Inst. Surabaya (EEPIS), Indones.*, vol. 6, no. Ies, pp. 978–979.

- J. E. Prawitasllri, (2016), Mengenal Emosi Melalui Komunikasi Nonverbal, *Bul. Psikol.*, vol. 3, no. 1, pp. 27–43, doi: 10.22146/bpsi.13384.
- L. Sofiyana, Z. Abidin, and H. Nurhayati, (2012), Klasifikasi Emosi Untuk Teks Berbahasa Indonesia Dengan Menggunakan K-Nearest Neighbor, vol. 1, no. January. pp. 194–299.
- M. Arief, M. I. Wardhana, S. Sumpeno, and M. Hariadi, (2010), *Emotion Expression of Three Dimensional Face Model Using Naive Bayes and Fuzzy Logic*, *Int. J. Comput. Sci. Netw. Secur.*, vol. 10, no. 5, pp. 208–214.
- M. C. Noviardini, A. B. Osmond, and C. Setianingsih, (2018), Klasifikasi Emosi Pada Lirik Lagu Menggunakan Metode Naive Bayes Classifier, *e-Proceeding Eng.*, vol. 5, no. 3, pp. 6187–6194.
- P. Ekman and D. Cordaro, (2011), *What is Meant by Calling Emotions Basic*, *Emot. Rev.*, vol. 3, no. 4, pp. 364–370, doi: 10.1177/1754073911410740.
- R. Sitorus and H. S. Dachlan, (2017), Deteksi Emosi Dari Teks Menggunakan Vector Space Model, *J. EECCIS*, vol. 11, no. 1, pp. 41–47.
- S. Widodo, (2020), Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan dan Kebudayaan, Bahasa Indonesia Menuju Bahasa Internasional. [Online]. Available: <http://badanbahasa.kemdikbud.go.id/lamanbahasa/content/bahasa-indonesia-menuju-bahasa-internasional>. [Accessed: 01-Mar-2020].
- S. S. T. W. Sasangka, (2014), Seri Penyuluhan Bahasa Indonesia : Kalimat
- Shivam Bachhety, Shubham Dhingra, Rachna Jain and Nikita Jain, (2018), Improved Multinomial Naïve Bayes Approach for Sentiment Analysis on Social Media, *International Journal of Information Systems & Management Science*, Vol. 1, No. 1.