

Klasifikasi Emosi Pada Twitter Menggunakan Bayesian Network

Mohamad Syahrul Mubarak, Muhammad Surya Asriadi and Adiwijaya

Fakultas Informatika, Telkom University, Jalan Telekomunikasi No. 1, Bandung, ID

E-mail: msyahrulmubarak@telkomuniversity.ac.id,

suryaa@student.telkomuniversity.ac.id, adiwijaya@telkomuniversity.ac.id

Ringkasan. Bahasa digunakan tidak hanya untuk mengutarakan fakta, akan tetapi juga emosi. Emosi tersebut juga dapat terlihat mulai dari perilaku hingga tulisan yang ditulis olehnya. Analisis emosi di dalam teks sendiri dapat dilakukan pada berbagai media, salah satunya adalah Twitter. Pendeteksian emosi bisa memiliki berbagai macam pengaplikasian salah satunya adalah sebagai bahan pertimbangan keputusan politik dalam suatu pemerintahan. Pada tugas akhir ini, penulis meneliti klasifikasi emosi pada twitter menggunakan Bayesian Network. Metode tersebut digunakan karena kemampuannya dalam memodelkan ketidakpastian dan relasi antar fitur. Hasil penelitian menunjukkan bahwa, metode yang digunakan untuk melatih jaringan Bayesian Network belum cukup efektif untuk menghasilkan model terbaik, dengan nilai F1-Score tertinggi adalah 53.71%. Model alternatif berbasis Bayesian Network juga dipelajari. Hasil percobaan menunjukkan hasil yang lebih baik dibandingkan Multinomial Naive Bayes dengan kompleksitas inferensi yang tidak jauh berbeda. Nilai F1-Score untuk model Multinomial Naive Bayes adalah 51.49%, sedangkan model alternatif berbasis Bayesian Network adalah 52.14%.

1. Pendahuluan

Bahasa digunakan tidak hanya untuk mengutarakan fakta, akan tetapi juga emosi. Emosi tersebut juga dapat terlihat mulai dari perilaku hingga tulisan yang ditulis olehnya. Analisis emosi di dalam teks sendiri dapat dilakukan pada berbagai media, salah satunya adalah Twitter. Twitter merupakan salah satu jaringan sosial dengan sistem microblogging yang populer dikalangan masyarakat saat ini. Twitter dapat digunakan untuk berbagai hal, mulai dari mengobrol, berbagi informasi berita hingga sebagai salah satu media curhat. Setiap informasi yang dibagikan di twitter disebut juga dengan istilah tweet. Pengguna twitter sendiri terbagi menjadi dua jenis, yaitu pengguna yang berbagi informasi dirinya sendiri dan pengguna yang membagikan informasi umum. Pada kedua kasus tersebut, tweet dapat memberikan informasi tentang emosi dari pengguna itu sendiri [4] [15]. Oleh karena itu, twitter dianggap memiliki sumber data yang sangat mendukung untuk dijadikan bahan penelitian terkait klasifikasi emosi.

Pendeteksian emosi memiliki berbagai macam pengaplikasian. Emosi dapat digunakan untuk mengukur indeks kebahagiaan suatu masyarakat, yang dimana dapat menjadi pertimbangan dalam mengambil keputusan politik pemerintah dalam meningkatkan kesejahteraan suatu negara. Di Indonesia, indeks kebahagiaan diukur melalui Survei Pengukuran Tingkat Kebahagiaan (SPTK) yang diselenggarakan Badan Pusat Statistik (BPS) Indonesia [2]. Contoh

lainnya adalah pendeteksian emosi yang dapat digunakan untuk memeriksa konten dari pesan sebelum dikirim agar pesan yang dikirim benar-benar dapat disampaikan tanpa adanya kesalahpahaman. Selain itu pendeteksian emosi pada teks juga dapat digunakan untuk meningkatkan pelayanan customer yang dapat diterapkan kedalam sistem Customer Relation Management (CRM) [6] dan juga dapat digunakan untuk memperkirakan popularitas dari suatu produk atau merek [14].

Klasifikasi emosi merupakan salah satu bentuk dari klasifikasi sentiment. Beberapa penelitian klasifikasi sentimen menggunakan Bayesian network adalah penelitian yang dilakukan oleh Yun Wan dkk [21] dan Jonathan Ortigosa-Hernandez dkk [16]. Naive Bayes merupakan salah satu metode yang umum digunakan dalam klasifikasi teks. Namun metode ini memiliki kelemahan dimana setiap kata diasumsikan independen terhadap kata lainnya [10]. Pada tugas akhir ini, penulis meneliti klasifikasi emosi pada twitter menggunakan Bayesian Network yang dapat memodelkan relasi antar kata. Penelitian ini diharapkan dapat menjadi dasar dari pengaplikasian ke berbagai bidang terkait emotion analysis.

2. Penelitian terkait

Analisis sentimen adalah analisis yang bertujuan untuk mengekstraksi informasi subjektif dari suatu sumber. Secara umum analisis dapat menentukan sikap dari pembicara atau penulis terhadap suatu topik ataupun dokumen. Salah satu contoh permasalahan analisis sentimen adalah menemukan polaritas sentimen dari suatu teks pada sebuah dokumen atau kalimat untuk menentukan apakah teks tersebut beraspek positif, negatif ataupun netral. Aspek yang dapat ditentukan juga tidak hanya sebatas polaritas, melainkan juga bisa berupa sesuatu yang bersifat non-fakta misalnya status emosional [17]. Beberapa penelitian terkait klasifikasi emosi diantaranya adalah penelitian oleh R C Balabantaray [3] dan K. Roberts [20].

R C Balabantaray [3] dkk meneliti klasifikasi Emosi pada dataset Twitter yang diperoleh dari web dan dianotasi secara manual oleh lima orang juri. Dataset dikelompokkan kedalam 7 kelas, yaitu enam kelas emosi Ekman (joy, sadness, anger, disgust, surprise dan fear) ditambahkan dengan satu kelas netral. Klasifikasi dilakukan menggunakan SVM dengan fitur yang didapatkan dari kombinasi fitur Unigrams, Bigrams, Personal-pronouns, adjectives, Word-net Affect emotion lexicon, Word-net Affect emotion lexicon with left/right context, Word-net Affect emotion POS, POS, POS-Bigrams, Dependency-Parsing Feature dan Emoticons. Hasil eksperimen menggunakan Leave-one-out cross-validation memberikan akurasi terbesar 73.24%. Akurasi tersebut didapatkan dengan menggunakan mengkombinasikan semua fitur-fitur yang tersedia.

K. Roberts [20] dkk meneliti Klasifikasi Emosi pada twitter kedalam 7 kelas yaitu anger, disgust, fear, joy, love, sadness, dan surprise. Dataset yang digunakan didapatkan dengan memilih 14 topik khusus yang dipercaya sering mengandung emosi. Dataset kemudian dianotasi oleh tiga orang juri. Tujuh Binary SVM digunakan untuk membangun sistem klasifikasi, masing masing Binary SVM berperan untuk mengklasifikasikan satu kelas. Fitur yang digunakan adalah Unigrams, Bigrams, Trigrams, Contains !, Contains ?, WordNet synsets, WordNet hypernyms, Topic scores dan Significant words. Untuk setiap emosi, kombinasi fitur terbaik dipilih menggunakan greedy additive feature selection. Eksperimen menggunakan 10-fold cross-validation performa 0.668 untuk micro-averaged F1-Score.

Yun Wan dkk [21] meneliti klasifikasi sentimen pada twitter untuk studi kasus pelayanan maskapai penerbangan. Klasifikasi dilakukan menggunakan metode Ensemble yang diaplikasikan menggunakan teknik majority vote dari bermacam teknik klasifikasi. Teknik klasifikasi yang digunakan dalam ensemble adalah Naive Bayes, SVM, Bayesian Network, C4.5 Decision Tree dan Random Forest. Hasil eksperimen menunjukkan bahwa metode Ensemble dapat meningkatkan akurasi dari sistem klasifikasi. Selain itu, secara umum Bayesian Network juga memberikan performa yang menyaingi dan bahkan melebihi metode lainnya.

Jonathan Ortigosa-Hernandez dkk [16] meneliti klasifikasi sentimen menggunakan Multidimensional Bayesian Network Classifier untuk menangani permasalahan klasifikasi sentimen multi kelas. Klasifikasi dilakukan secara semi-supervised menggunakan dataset ASOMO SA yang memiliki 14 atribut dengan 150 dokumen yang telah diberi label dan 2.392 dokumen yang belum diberi label. Uji coba dilakukan dengan membandingkan fitur yang digunakan yaitu n-grams, PoS dan ASOMO feature serta pendekatan uni-dimensional dan multi-dimensional. Ujicoba menunjukkan bahwa fitur ASOMO lebih baik dibandingkan dengan n-grams dan PoS. Pendekatan Multi-dimensional juga memberikan hasil lebih baik saat menggunakan fitur ASOMO.

3. Desain Sistem

Dalam penelitian ini, penulis meneliti dua model Bayesian Network yaitu Full Bayesian Network dan Bayesian Network dengan indikator mood. Tahapan-tahapan yang dilakukan untuk melatih model Bayesian Network adalah sebagai berikut:

- (i) Data
Dataset dipersiapkan dengan menggunakan skema K-fold Cross-Validation. Jumlah K-fold yang digunakan adalah 5, sehingga 20% dari keseluruhan data menjadi data uji untuk model yang dilatih terhadap 80% data yang tersisa. Data terlebih dulu diacak urutan kemunculannya sebelum dipisahkan menggunakan skema diatas.
- (ii) Pre-processing
Data di pre-processing sesuai dengan tahapan yang terdapat pada bagian sebelumnya.
- (iii) Binary Term Frequencies
Setelah data teks di pre-processing, data tersebut kemudian diubah menjadi fitur Binary Term Frequencies.
- (iv) Feature Selection
Tahapan ini dilakukan untuk menyeleksi fitur-fitur yang dapat dijadikan simpul. Seleksi fitur dilakukan menggunakan algoritma chi-squared.
- (v) Structure Learning
Struktur dari Bayesian Network dilatih menggunakan data latih. Algoritma struktur learning yang digunakan adalah Simulated Annealing dengan fungsi optimasi PCLL.
- (vi) Parameter Learning
Setelah struktur yang mendekati optimal ditemukan. Parameter pada setiap simpul dilatih terhadap data latih menggunakan MAP-estimation.
- (vii) Classification
Data diklasifikasikan menggunakan model Bayesian Network yang telah dilatih pada tahapan sebelumnya. Selanjutnya performa dari model dievaluasi menggunakan F1-Score.

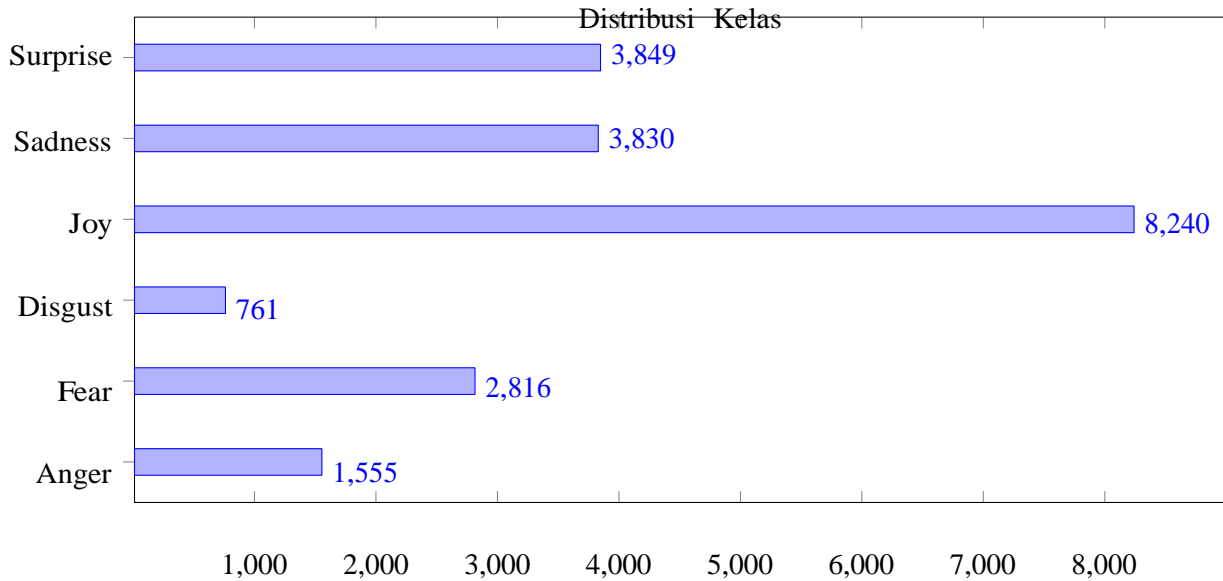
4. Dataset

Dataset yang digunakan pada penelitian ini adalah Dataset #Emotional Tweets yang dipublikasikan oleh Saif Mohammad dkk. Dataset tersebut didapatkan dengan cara memanfaatkan fitur hashtag pada twitter [13]. Dataset tersebut berjumlah 21051 data dan terdiri dari 6 kelas (Gambar 1).

4.1. Pre-processing

Dataset yang diperoleh pada umumnya tidak bisa digunakan begitu saja dan membutuhkan tahapan pre-processing sebelum dapat digunakan. Dengan menggunakan teks berikut sebagai contoh:

@samtravels I love you!!! :) I've learned to explore my city in style & with enthusiasm because of you! #travel



Gambar 1: Terdapat 6 kelas dari 21051 total data yang merupakan emosi dasar Ekman, yaitu: anger, fear, disgust, joy, sadness dan surprise.

Tabel 1: Sampel dataset #Emotional Tweets untuk masing-masing kelas emosi.

Tweet	Emosi
Soooo dooowwnn!! Move on, get some sleep... Me deserve better.#forgetit #yawning	anger
Making art and viewing art are different at their core!	fear
Can parents teach kids how to cover their mouths when theycough? Ewah.	disgust
the moment when you get another follower and you cheer.	joy
I said no to the same boy like 5 times last night!	sadness
Tomorrow I'm going to get @HerBraceFaceAss nd Jaela frmschool early nd takin my grls to lunch. LOL	surprise

Tahapan yang menjadi bagian processing adalah sebagai berikut:

(i) HTML Escape

Data yang didapatkan dari internet biasanya mengandung code html seperti < dan > dan & amp; yang merupakan karakter yang telah di escape agar dapat ditampilkan dalam format HTML. Proses ini bertujuan untuk mengkonversi kode HTML tersebut menjadi bentuk karakter asli.

Hasil:

@samtravels I love you!!! :) I've learned to explore my city in style & with enthusiasm because of you! #travel

(ii) Apostrophe Lookup

Pada tahap ini, setiap kata yang memiliki imbuhan kepemilikan dipisahkan menjadi kata-kata berbeda. contoh kata I've menjadi I have.

Hasil:

@samtravels I love you!!! :) I have learned to explore my city in style & with enthusiasm because of you! #travels

(iii) Username Lookup

Pada twitter, seorang pengguna dapat menandai pengguna lain pada tweet miliknya, dengan menggunakan simbol @ diikuti dengan username pengguna yang ingin ditandai. Kata tersebut disebut juga dengan istilah mention. Setiap mention tersebut memiliki arti sama namun dapat memiliki username yang berbeda-beda. Tahap ini mengelompokan mention tersebut kedalam suatu tag khusus yaitu `{{username}}`.

Hasil:

`{{username}}` I love you!!! :) I have learned to explore my city in style & with enthusiasm because of you! #travels

(iv) URL Lookup

Tidak jarang twitter digunakan sebagai tempat berbagi informasi ataupun tautan ke situs lain. Tahap ini mengubah tautan tersebut menjadi tag tertentu yaitu `{{url}}`.

Hasil:

`{{username}}` I love you!!! :) I have learned to explore my city in style & with enthusiasm because of you! #travels

(v) Money Lookup

Pada tahapan ini, kata-kata yang diawali karakter mata uang (contoh: \$), dianggap sebagai uang dan diubah menjadi tag `{{money}}`.

Hasil:

`{{username}}` I love you!!! :) I have learned to explore my city in style & with enthusiasm because of you! #travels

(vi) Remove Number

Tahap ini dilakukan untuk mengkonversikan angka yang terdapat pada teks menjadi tag tertentu yaitu `{{number}}`.

Hasil:

`{{username}}` I love you!!! :) I have learned to explore my city in style & with enthusiasm because of you! #travels

(vii) Remove Repeats

Pada tahapan ini, karakter yang berulang pada teks dihapus sehingga tidak terjadi perulangan yang tidak seharusnya.

Hasil:

`{{username}}` I love you! :) I have learned to explore my city in style & with enthusiasm because of you! #travels

(viii) Stemming

Tahapan ini bertujuan untuk mengubah suatu kata menjadi kata dasar, proses tersebut dilakukan dengan cara menghapus sebagian huruf dari sebuah kata dengan harapan dapat memberikan hasil yang bagus. Algoritma yang digunakan pada proses ini adalah Porter stemmer [18].

Hasil:

`{{username}}` I love you! :) I have learn to explor my citi in style & with enthusiasm becaus of you! #travel

(ix) Remove Punctuations

Pada tahapan ini, tanda-tanda baca dihilangkan dari teks.

Hasil:

`{{username}}` I love you I have learn to explor my citi in style with enthusiasm becaus of you #travel

(x) Hashtag Removal

Hashtag adalah fitur khusus pada twitter untuk menandai suatu topik tertentu. Hashtag dimulai dengan karakter ”#” diikuti kata kunci dari topik yang dimaksud. Untuk mereduksi fitur yang ditimbulkan akibat adanya hashtag tanpa membuang makna yang

terkandung dalam hashtag tersebut, hashtag yang memiliki total kemunculan dibawah jumlah minimum dikonversikan menjadi tag khusus yaitu {{hashtag}}, sedangkan hashtag yang sering muncul dibiarkan tanpa diubah.

Hasil:

{{username}} I love you I have learn to explor my citi in style with enthusiasm becaus of you {{hashtag}}

(xi) Stop Words Removal

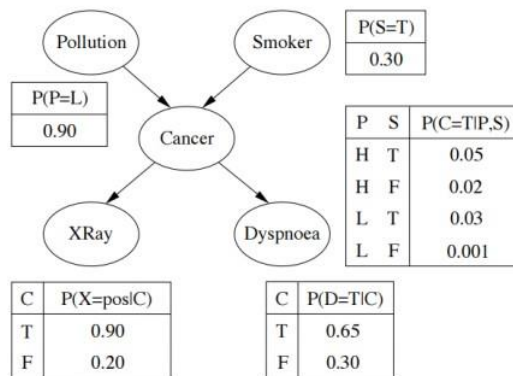
Pada tahapan ini, kata yang tidak memiliki arti signifikan dihilangkan.

Hasil:

{{username}} love have my in style enthusiasm {{hashtag}}

5. Bayesian Network

Bayesian Network adalah bentuk model graphical yang merepresentasikan struktur dari suatu model probablistik. Dimana terdapat set variabel yang dapat bergantung pada variabel lainnya. Bayesian Network sangat cocok untuk merepresentasikan domain yang mempunyai struktur sebab-musabab. Bayesian Network direpresentasikan sebagai directed acyclic graph (DAG), dimana setiap simpul pada graf merupakan sebuah variabel acak yang menandakan sebuah attribut, fitur atau hipotesis tentang sesuatu kemungkinan (gambar 2).



Gambar 2: Struktur Bayesian Network untuk permasalahan kanker paru-paru [11].

5.1. Parameter Learning

Parameter learning bertujuan menentukan probability distribution dari setiap simpul yang terdapat pada suatu Bayesian Network. Parameter learning dapat dilakukan baik dengan menggunakan data yang lengkap maupun tidak. Parmeter learning dengan data yang lengkap dapat dilakukan dengan beberapa cara diantaranya Maximum Likelihood Estimate (ML-estimation) dan Maximum A Posteriori Estimation (MAP-estimation) [19]. ML-estimation dirumuskan sebagai berikut:

$$\theta_{ijk} = \frac{N_{ijk}}{N_{ij}} \tag{1}$$

MAP-estimation dapat dilihat sebagai bentuk umum dari ML-estimation, dimana MAP-estimation menggabungkan prior distributions terhadap jumlah yang ingin diestimasi. Rumus dari MAP-estimation dengan equivalent sample size adalah:

$$\theta_{ijk} = \frac{\frac{\alpha}{r_i \cdot q_i} + N_{ijk}}{\sum_k \frac{\alpha}{r_i \cdot q_i} + N_{ijk}} \quad (2)$$

Penjelasan dari masing-masing simbol terdapat dibagian selanjutnya.

5.2. Structure Learning

Structure learning dilakukan untuk memodelkan data menjadi Bayesian Network. Algoritma structure learning pada umumnya bisa dibagi menjadi dua kategori, yaitu yang berbasis analisis dependency dan berbasis strategi search dan scoring [19]. Beberapa jenis algoritma search yang dapat digunakan diantaranya hill climbing, tabu search dan simulated annealing.

Simulated Annealing adalah sebuah algoritma metaheuristic yang terinspirasi dari proses annealing pada metallurgy (pendinginan bertahap pada proses penempaan besi). Teknik ini sering digunakan ketika ruang solusi berbentuk diskrit. Contoh permasalahan yang dapat diselesaikan dengan algoritma ini adalah Traveling Salesman Problem dan structure learning pada Bayesian Network.

Algorithm 1 Algoritma simulated annealing untuk structure learning pada Bayesian Network [5]

```

1: function SimulatedAnnealing( $\tau_0, T_{ni}, ni$ )
2:   for  $i = 1, \dots, n$  do
3:      $\pi_i \leftarrow \emptyset$ 
4:      $\pi_{best,i} \leftarrow \emptyset$ 
5:   end for
6:    $T \leftarrow \tau_0 \cdot N$ 
7:    $\alpha \leftarrow (T_{ni}/T_0)^{1/ni}$ 
8:    $k \leftarrow 0$ 
9:   repeat
10:    repeat
11:      Pilih dua indeks  $i, j$  secara acak
12:    until  $v_j \in \pi_j$  tidak mengandung siklus
13:    if  $v_j \in \pi_i$  then
14:      if  $\exp(m(v_i, \pi_i \setminus v_j) \Leftrightarrow m(v_i, \pi_i)) > \text{random}[0..1]$  then
15:         $\pi_i \leftarrow \pi_i \setminus v_j$ 
16:      end if
17:    else
18:      if  $\exp(m(v_i, \pi_i v_j) \Leftrightarrow m(v_i, \pi_i)) > \text{random}[0..1]$  then
19:         $\pi_i \leftarrow \pi_i v_j$ 
20:      end if
21:    end if
22:    if  $\prod_{i=1}^n m(v_i, \pi_{best,i}) < \prod_{i=1}^n m(v_i, \pi_i)$  then
23:       $\forall_{i \in \{1, \dots, n\}} \pi_{best,i} \leftarrow \pi_i$ 
24:    end if
25:  until  $k = ni$ 
26:  return  $\pi_{best,1}, \dots, \pi_{best,n}$ 
27:   $T \leftarrow T * \alpha$ 
28:   $k \leftarrow k + 1$ 
29: end function

```

Dimana π_i adalah simpul orang tua untuk simpul v_i , $\pi_{best,i}$ adalah solusi simpul orang tua

terbaik. T adalah suhu dan τ_0 adalah faktor awal yang menginisialisasi suhu awal T_0 berdasarkan jumlah keseluruhan simpul N . α adalah percepatan perubahan suhu yang dihitung berdasarkan suhu akhir T_{ni} , suhu awal T dan jumlah iterasi maksimal ni . m adalah fungsi optimasi yang digunakan untuk membandingkan kualitas solusi.

Untuk menentukan struktur Bayesian Network terbaik, setiap struktur hasil dari proses pencarian dihitung skornya menggunakan fungsi optimasi (fitness). Salah satu fungsi skor yang tersedia adalah factorized Conditional Log Likelihood (FCLL). FCLL didefinisikan sebagai berikut [8]:

$$\begin{aligned}
 FCLL(G|D) = & (\alpha + \beta) \mathcal{L}(B|D) - \beta \lambda \sum_{i=1}^n \sum_{j=1}^{q_i^*} \sum_{k=1}^{r_i} \sum_{c=0}^1 N_{ijk} (\log(\frac{N_{ijck}}{N_{ij*k}}) - \log(\frac{N_{ijc}}{N_{ij*}})) \\
 & - \beta \lambda \sum_{c=0}^1 N_c \log(\frac{N_c}{N}) - \beta N \rho
 \end{aligned} \tag{3}$$

Dimana n adalah jumlah simpul, r_i adalah jumlah kemungkinan nilai dari sebuah variabel acak, q_i adalah jumlah kemungkinan konfigurasi dari simpul-simpul orang tua, q_i^* adalah q_i

tanpa simpul kelas sebagai pada set simpul orang tua dan C adalah jumlah kelas.

α , β dan Λ adalah konstanta yang diasumsikan agar dapat meminimalisasi mean square error (MSE) fungsi $f(U_t, V_t) = \alpha \log(U_t) + \beta \log(V_t) + \lambda$ terhadap $f = \log(\frac{U_t}{U_t+V_t})$. Dimana fungsi f adalah aproksimasi dari fungsi f .

$$\alpha = \frac{\pi^2 + 6}{24} \tag{4}$$

$$\beta = \frac{\pi^2 - 18}{24} \tag{5}$$

$$\lambda = \frac{\pi^2}{12 \ln 2} - (2 + \frac{(\pi^2 - 6) \log p}{12}) \tag{6}$$

\mathcal{L} adalah fungsi skoring log-likelihood yang didefinisikan sebagai berikut [7]:

$$\mathcal{L}(G|D) = \sum_{i=1}^n \sum_{j=1}^k \sum_{k=1}^l N_{ijk} \log(\frac{N_{ijk}}{N_{ij}}) \tag{7}$$

Parameter-parameter observasi frekuensi didefinisikan sebagai berikut:

$$N_{ij} = \sum_{k=1}^l N_{ijk} \tag{8}$$

$$N_{ij*k} = \sum_{c=1}^1 N_{ijck} \tag{9}$$

$$N_{ijc} = \sum_{k=1}^l N_{ijck} \tag{10}$$

$$N_{ij*} = \sum_{k=1}^l \sum_{c=1}^1 N_{ijck} \tag{11}$$

$$N_c = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^m N_{ijck} \quad (12)$$

Dimana N adalah jumlah dataset, X_i adalah variable yang diobservasi, $P A_i$ adalah simpul orang tua dari X_i dan C_i adalah label dari X_i . N_{ijk} adalah jumlah kemunculan ketika $X_i = k$ dan $P A_i = j$ pada dataset. N_{ijck} adalah jumlah kemunculan ketika $X_i = k$, $P A_i = j$ dan $C_i = c$ pada dataset.

6. Model Bayesian Network

Bagian ini membahas lebih rinci mengenai detail dari masing-masing model yang diteliti.

6.1. Full Bayesian Network

Pada model ini, setiap kata dijadikan random variable yang kemudian menjadi simpul-simpul pada Bayesian Network. Untuk mengurangi jumlah kata yang dijadikan simpul, kata-kata terlebih dahulu diseleksi menggunakan teknik seleksi fitur chi-squared.

Algoritma metaheuristic simulated annealing digunakan untuk melakukan pencarian pada ruang solusi dari struktur Bayesian Network yang mungkin. Nilai fitness dihitung menggunakan fungsi optimasi PCLL. Pencarian struktur dipermudah dengan menginisialisasi struktur sebagai Naive Bayes sebelum memulai algoritma Simulated Annealing. Untuk menghindari overfitting yang dapat disebabkan oleh fungsi PCLL, jumlah simpul orang tua yang dibatasi.

6.2. Bayesian Network dengan indikator mood

Model ini bekerja dengan cara menghubungkan simpul indikator ke simpul kelas pada model Multinomial Naive Bayes. Simpul tersebut berpengaruh terhadap probabilitas prior kelas sehingga dapat memberikan hasil inferensi yang lebih baik. Simpul indikator yang ditambahkan adalah dua simpul indikator yang menandakan bahwa suatu tweet memiliki mood yang positif dan mood yang negatif. Indikator mood didapatkan dengan cara mengekstraksi emoticon yang terdapat pada setiap tweet. Emoticon yang menunjukkan ekspresi senang dianggap sebagai mood positif dan mood yang tidak menunjukkan ekspresi senang dianggap sebagai mood negatif.

Dikarenakan simpul yang hanya berjumlah maksimum tiga, model ini tidak memerlukan algoritma pencarian struktur untuk mencari struktur terbaik. Struktur yang terbaik didapatkan dengan cara menguji coba semua struktur yang mungkin.

7. Hasil Pengujian dan Analisis

Bagian ini memaparkan hasil pengujian yang berupa performansi dari setiap model Bayesian Network.

7.1. Full Bayesian Network

Analisis Performansi Pada Data Latih dan Data Uji

Tabel 2: Performansi Naive Bayes (NB) dan Bayesian Network (BN) dengan maksimum dua dan tiga simpul pada data latih untuk setiap significant level.

Model	Significant Level	Precision	Recall	F1-Score
NB	0.001	62.76%	61.45%	60.28%
NB	0.01	65.09%	63.88%	62.96%
NB	0.1	72.27%	70.92%	70.37%
BN 2 Simpul	0.001	63.02%	61.69%	60.5%
BN 2 Simpul	0.01	65.5%	64.28%	63.35%
BN 2 Simpul	0.1	72.67%	71.29%	70.73%
BN 3 Simpul	0.001	63.2%	61.85%	60.66%
BN 3 Simpul	0.01	65.69%	64.45%	63.52%
BN 3 Simpul	0.1	72.89%	71.48%	70.92%

Tabel 2 menunjukkan performansi dari Naive Bayes dan Bayesian Network untuk setiap significant level pada data latih. Terlihat bahwa Bayesian Network memiliki performansi yang lebih baik dibandingkan Naive Bayes pada untuk setiap significant level. Hal ini menunjukkan pengaruh nilai conditional likelihood yang dihitung menggunakan fungsi skor P_{CLL} sebagai fungsi optimasi. Optimasi dari nilai P_{CLL} tersebut dapat meningkatkan kemampuan model Bayesian Network untuk klasifikasi data.

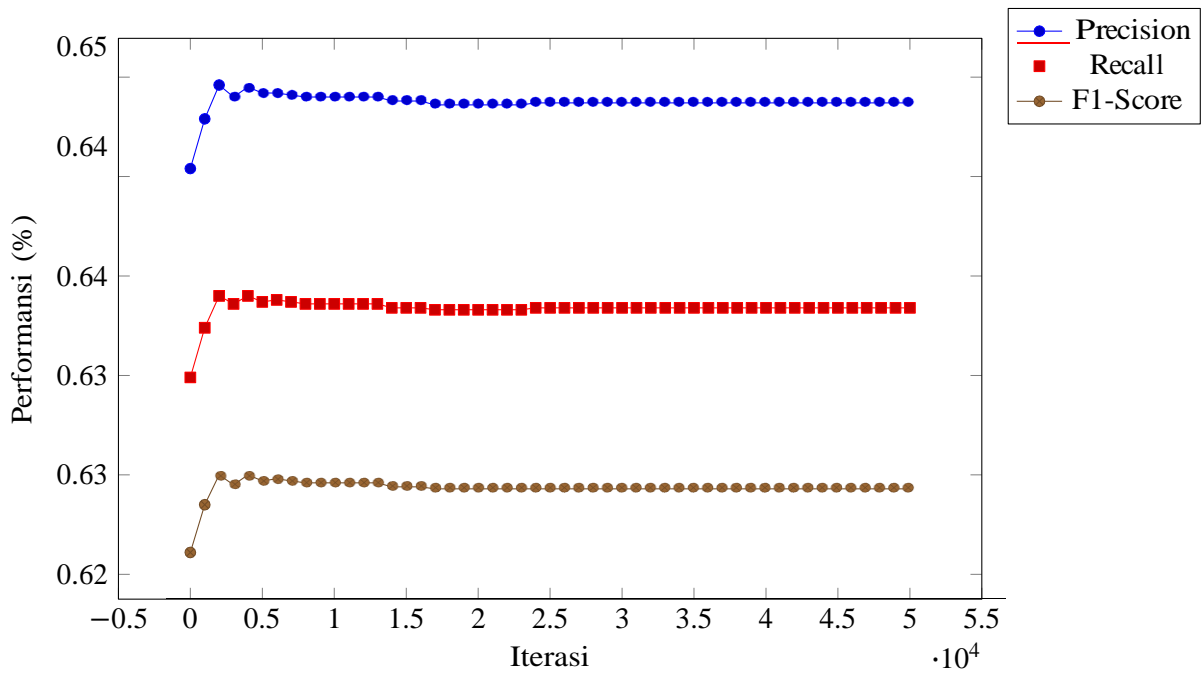
Tabel 3: Performansi Naive Bayes (NB) dan Bayesian Network (BN) dengan maksimum dua dan tiga simpul pada data uji untuk setiap significant level.

Model	Significant Level	Precision	Recall	F1-Score
NB	0.001	54.74%	54.56%	52.75%
NB	0.01	55.58%	55.46%	53.89%
NB	0.1	55.46%	55.74%	54.07%
BN 2 Simpul	0.001	54.60%	54.50%	52.68%
BN 2 Simpul	0.01	55.32%	55.30%	53.71%
BN 2 Simpul	0.1	54.37%	54.89%	53.24%
BN 3 Simpul	0.001	54.48%	54.42%	52.60%
BN 3 Simpul	0.01	55.00%	55.09%	53.49%
BN 3 Simpul	0.1	54.11%	54.65%	53.01%

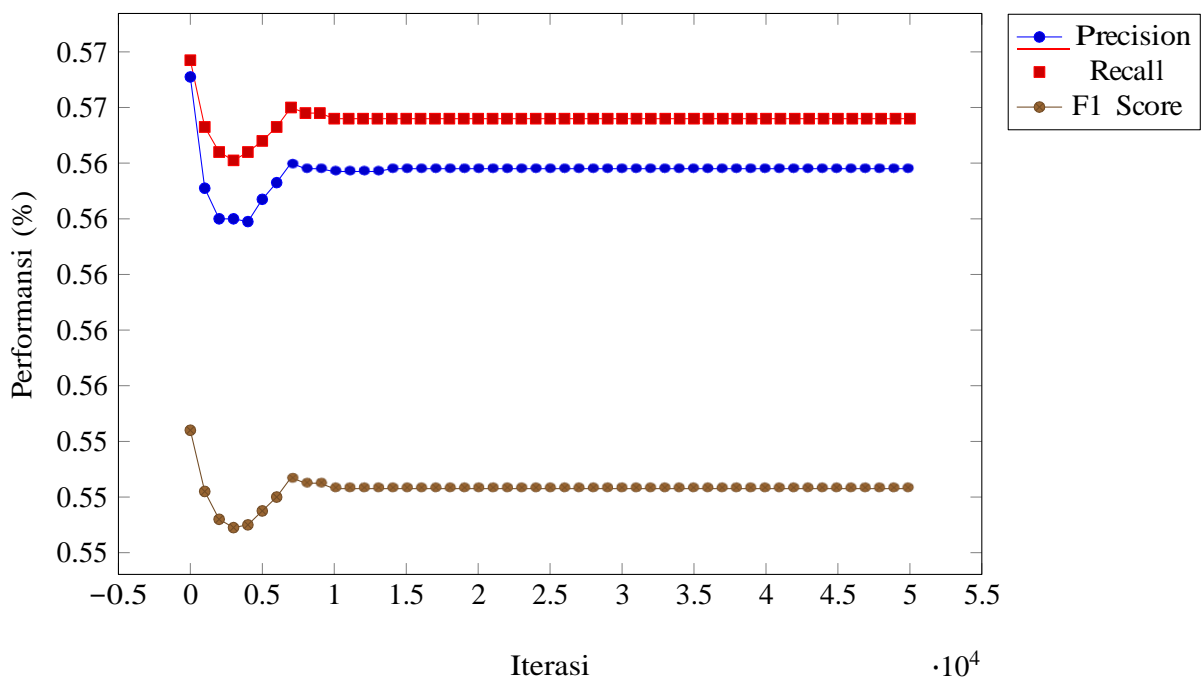
Tabel 2 menunjukkan performansi dari Naive Bayes dan Bayesian Network untuk setiap significant level pada data uji. Terlihat bahwa Naive Bayes memiliki performansi lebih baik untuk setiap significant level. Bayesian Network dengan maksimum simpul dua, lebih baik dibandingkan dengan maksimum tiga simpul. Model Bayesian Network yang dilatih dengan struktur awal Naive Bayes justru memiliki performansi lebih rendah dari pada Naive Bayes. Hal ini menunjukkan bahwa, peningkatan nilai conditional likelihood tidak berpengaruh secara langsung terhadap peningkatan performansi suatu model.

Analisis Performansi Pada Setiap Iterasi

Untuk meneliti penyebab lebih buruknya performansi Bayesian Network dibandingkan dengan Naive Bayes. Penulis membandingkan performansi Bayesian Network terbaik terhadap data latih dan data uji pada setiap iterasinya.



Gambar 3: Performansi Bayesian Network terhadap data latih dengan significant level 0.01 pada setiap iterasi.



Gambar 4: Performansi Bayesian Network terhadap data uji dengan significant level 0.01 pada setiap iterasi.

Berdasarkan Gambar 3 dan 4, terlihat tren perubahan performansi terhadap data latih dan data uji disetiap iterasi. performansi terhadap data latih meningkat tajam diawal iterasi namun sedikit menurun sebelum akhirnya stabil hingga iterasi maksimum. Sedangkan untuk performansi terhadap data uji memiliki penurunan signifikan diawal iterasi dan meningkat

tajam setelahnya, namun performansi sedikit menurun sebelum akhirnya stabil hingga iterasi maksimum. Hal ini menunjukkan bahwa, meskipun dapat meningkatkan performansi dari model, skor $\mathcal{P}CLL$ tidak berpengaruh langsung terhadap performansi dari model. Hal

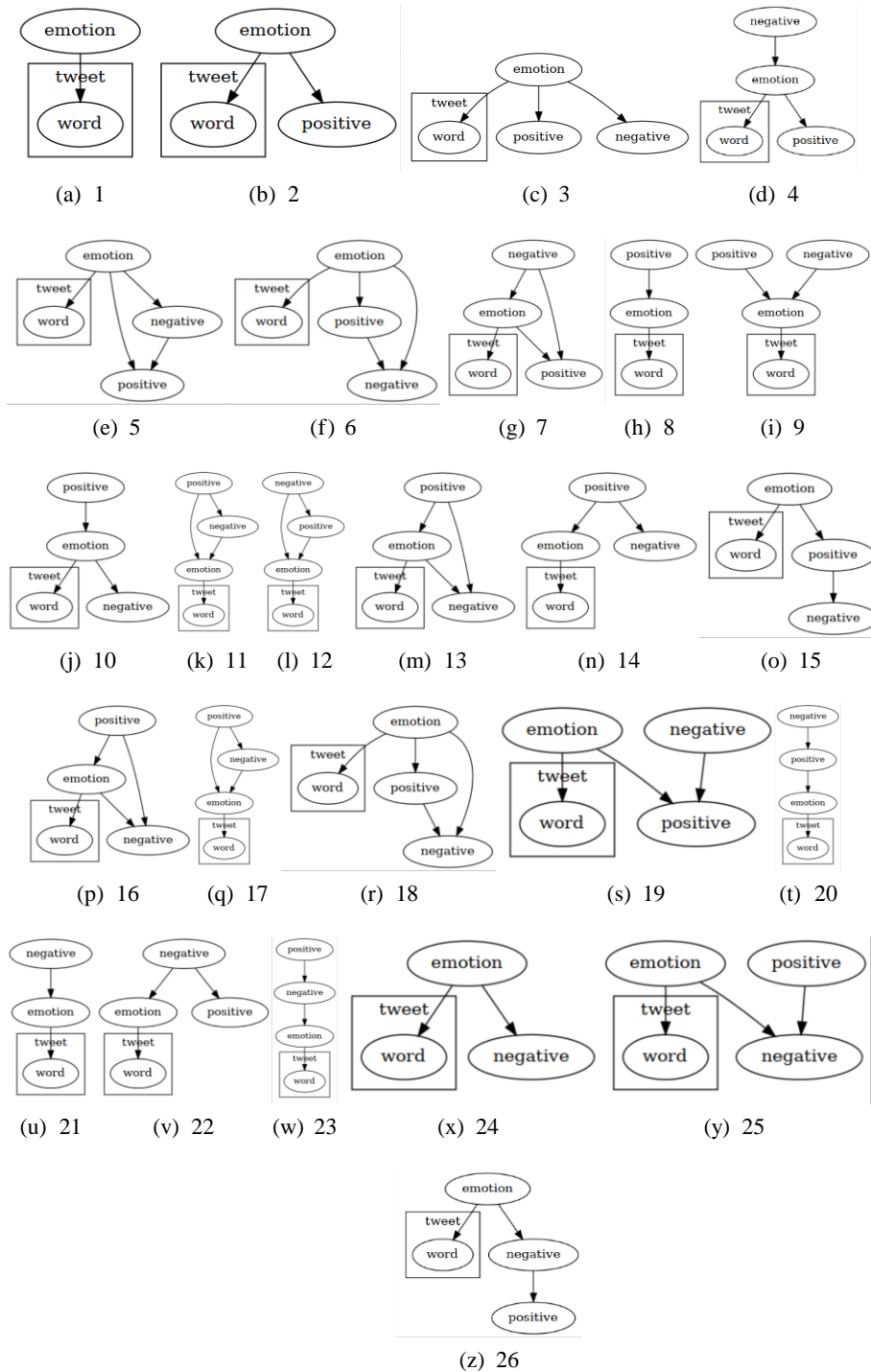
tersebut disebabkan oleh karakteristik fungsi $\mathcal{P}CLL$ yang hanya mengoptimasi nilai conditional likelihood suatu model. Peningkatan nilai conditional likelihood tidak hanya dapat meningkatkan performansi dari model, akan tetapi juga dapat menurunkan performansi dari model. Penurunan performansi terjadi jika nilai conditional likelihood meningkat dengan cara meningkatkan kemampuan mengenali suatu kelas lebih baik dengan mengorbankan sedikit kemampuannya untuk mengenali kelas lainnya.

Analisis Kompleksitas Inferensi

Dengan pemodelan data tweet menggunakan Binary Term Frequency, maka dapat diketahui bahwa suatu variabel kata hanya dapat memiliki nilai ada (true) atau tidak ada (false). Hal tersebut menyebabkan keseluruhan variabel kecuali variabel kelas diketahui nilainya. Dengan diketahui semua nilai variabel, maka marginalisasi pada Bayesian Network tidak perlu dilakukan. Dengan begitu kompleksitas Full Bayesian Network dan Naive Bayes sama, yaitu $O(CN)$. Dimana C adalah banyaknya kelas yang memungkinkan dan N adalah banyaknya simpul pada Bayesian Network.

7.2. Bayesian Network dengan Indikator Mood

Bagian ini menganalisis performansi dari model Bayesian Network dengan indikator mood. Gambar 5 menunjukkan struktur-struktur yang diteliti. Tabel 4 menunjukkan performansi dari setiap struktur yang diteliti.



Gambar 5: Struktur Bayesian Network dengan Indikator Mood. Simpul word yang digambar menggunakan plate notation merupakan simpul Multinomial Naive Bayes, emotion merupakan kelas dari tweet yang diprediksi, Positive dan negative merupakan simpul indikator mood positif dan negatif. Untuk mempermudah ilustrasi, hanya simpul yang terhubung dengan simpul emotion yang diperlihatkan

Tabel 4: Performansi setiap struktur bayesian network pada gambar 5.

Struktur	Precision	Recall	F1-Score
1	54.97%	54.98%	51.49%
2	55.08%	55.0%	51.64%
3	55.56%	55.44%	52.13%
4	55.56%	55.44%	52.13%
5	55.57%	55.44%	52.14%
6	55.57%	55.44%	52.14%
7	55.57%	55.44%	52.14%
8	55.08%	55.0%	51.64%
9	55.57%	55.44%	52.14%
10	55.56%	55.44%	52.13%
11	55.57%	55.44%	52.14%
12	55.57%	55.44%	52.14%
13	55.57%	55.44%	52.14%
14	55.08%	55.0%	51.64%
15	55.08%	55.0%	51.64%
16	55.57%	55.44%	52.14%
17	55.57%	55.44%	52.14%
18	55.57%	55.44%	52.14%
19	55.08%	55.0%	51.64%
20	55.08%	55.0%	51.64%
21	55.49%	55.3%	51.97%
22	55.49%	55.3%	51.97%
23	55.49%	55.3%	51.97%
24	55.49%	55.3%	51.97%
25	55.5%	55.32%	51.99%
26	55.49	55.3%	51.97%

Berdasarkan Tabel 4 terlihat bahwa stuktur 5, 6, 7, 9, 11, 12, 13, 16, 17 dan 18 memiliki performansi terbaik dibandingkan dengan struktur-struktur lainnya. Setiap struktur dengan indikator juga memberikan peningkatan performansi dibandingkan dengan struktur multinomial naive bayes tanpa indikator (struktur 1). Hal ini menunjukkan bahwa, simpul yang ditambahkan pada model multinomial naive bayes dapat meningkatkan performansi model sehingga melebihi performansi semula. Peningkatan tersebut terjadi karena modifikasi yang dilakukan pada prior kelas dengan cara memberikan informasi tambahan. Informasi tambahan tersebut kemudian berpengaruh terhadap nilai probabilitas prediksi yang dapat membantu meningkatkan kemampuan model untuk mengenali suatu kelas tertentu.

Analisis Kompleksitas Inferensi

Multinomial Naive Bayes memiliki kompleksitas inferensi yang mengikuti dengan besarnya data yang di inferensi. Kompleksitas inferensi Multinomial Naive Bayes adalah $O(CW)$, dimana C adalah banyaknya kelas yang memungkinkan dan W adalah banyaknya kata yang menjadi data inferensi. Pada model Bayesian Network dengan Indikator Mood, hanya kalkulasi prior pada Multinomial Naive Bayes yang berubah sehingga tidak memberikan perubahan signifikan pada kompleksitas inferensi. Kompleksitas dari model tersebut adalah $O((2 + C)W)$, dimana 2 berasal dari banyaknya simpul yang mempengaruhi simpul kelas, yaitu simpul positive dan negative.

8. Kesimpulan dan Saran

Berdasarkan hasil pengujian yang telah dilakukan, dapat dilihat bahwa model Full Bayesian Network yang dilatih dengan Simulated Annealing dengan fungsi optimasi PCLL memiliki

performa lebih buruk dibandingkan Naive Bayes pada data uji. Sedangkan pada data latih, Full Bayesian Network justru lebih baik dari pada Naive Bayes. Hal ini disebabkan karena skor PCLL tidak berpengaruh secara langsung terhadap performansi dari model Full Bayesian Network. Untuk penelitian selanjutnya, penulis menyarankan untuk mencoba fungsi skor yang berpengaruh langsung terhadap performansi seperti classification rate. Adapun algoritma pencarian struktur lainnya yang lebih baik dibandingkan Simulated Annealing juga patut diteliti.

Untuk model Bayesian Network dengan indikator mood. Meskipun tidak memiliki selisih yang besar, terlihat bahwa model yang dihasilkan dengan menambahkan simpul indikator mood dapat meningkatkan performa dari multinomial naive bayes. Untuk penelitian selanjutnya, penulis ingin menggunakan indikator lain sebagai simpul untuk meningkatkan performansi Multinomial Naive Bayes. Salah satu indikator yang mungkin adalah hasil kmeans clustering yang dapat dilakukan terhadap teks.

References

- [1] Grimace. <http://www.grimace-project.net/>. Accessed: 2016-11-10.
- [2] Indeks kebahagiaan indonesia tahun 2014. Badan Pusat Statistik Indonesia, feb 2015.
- [3] Balabantaray, R. C., Mohammad, M., and Sharma, N. Multi-class twitter emotion classification: A new approach. *International Journal of Applied Information Systems* 4, 1 (September 2012), 48–53. Published by Foundation of Computer Science, New York, USA.
- [4] Bollen, J., Mao, H., and Pepe, A. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena, 2011.
- [5] Bouckaert, R. Bayesian Belief Networks: from Construction to Inference. PhD thesis, Utrecht, Netherlands, 1995.
- [6] Bougie, R., Pieters, R., and Zeelenberg, M. Angry customers don't come back, they get back: The experience and behavioral implications of anger and dissatisfaction in services. *Journal of the Academy of Marketing Science* 31, 4 (2003), 377–393.
- [7] Carvalho, A. M. Scoring functions for learning bayesian networks, 2009.
- [8] Carvalho, A. M., Roos, T., Oliveira, A. L., and Myllymäki, P. Discriminative learning of bayesian networks via factorized conditional log-likelihood. *J. Mach. Learn. Res.* 12 (July 2011), 2181–2210.
- [9] Ekman, P. Facial expression and emotion. *American Psychologist* 48, 4 (1993), 384–392.
- [10] Gamallo, P., and Garcia, M. Citius: A naive-bayes strategy for sentiment analysis on english tweets. pp. 171–175.
- [11] Korb, K. B., and Nicholson, A. E. Bayesian Artificial Intelligence, Second Edition, 2nd ed. CRC Press, Inc., Boca Raton, FL, USA, 2010.
- [12] Manning, C. D., Raghavan, P., and Schütze, H. Introduction to Information Retrieval. Cambridge University Press, New York, NY, USA, 2008.
- [13] Mohammad, S. M. #emotional tweets. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (Stroudsburg, PA, USA, 2012), SemEval '12, Association for Computational Linguistics*, pp. 246–255.
- [14] Mohammad, S. M., and Yang, T. W. Tracking sentiment in mail: How genders differ on emotional axes. In *Proceedings of the 2Nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (Stroudsburg, PA, USA, 2011), WASSA '11, Association for Computational Linguistics*, pp. 70–79.
- [15] Naaman, M., Boase, J., and Lai, C.-H. Is it really about me?: Message content in social awareness streams. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (New York, NY, USA, 2010), CSCW '10, ACM*, pp. 189–192.
- [16] Ortigosa-Hernandez, J., Rodriguez, J. D., Alzate, L., Lucania, M., Inza, I., and Lozano, J. A. Approaching sentiment analysis by using semi-supervised learning of multi-dimensional classifiers. *Neurocomputing* 92 (2012), 98 – 115. Data Mining Applications and Case Study.
- [17] Pang, B., and Lee, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2, 12 (2008), 1–135.
- [18] Porter, M. F. Readings in information retrieval. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1997, ch. An Algorithm for Suffix Stripping, pp. 313–316.

- [19] Rish, I. A tutorial on inference and learning in bayesian networks. IBM T.J.Watson Research Center, 2000.
- [20] Roberts, K., Roach, M. A., Johnson, J., Guthrie, J., and Harabagiu, S. M. Empatweet: Annotating and detecting emotions on twitter. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12) (Istanbul, Turkey, may 2012), N. C. C. Chair), K. Choukri, T. Declerck, M. U. Doan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, Eds., European Language Resources Association (ELRA).
- [21] Wan, Y., and Gao, Q. An ensemble sentiment classification system of twitter data for airline services analysis. In 2015 IEEE International Conference on Data Mining Workshop (ICDMW) (Nov 2015), pp. 1318–1325.