KLASIFIKASI SENTIMEN PADA *LEVEL* ASPEK TERHADAP ULASAN PRODUK BERBAHASA INGGRIS MENGGUNAKAN *BAYESIAN NETWORK* *(CASE STUDY : DATA ULASAN PRODUK AMAZON)*

ASPECT-LEVEL SENTIMENT ANALYSIS ON ENGLISH PRODUCT REVIEWS USING BAYESIAN NETWORK

(CASE STUDY : AMAZON PRODUCT REVIEW DATA)

Andri Dhika Saputra[1], Adiwijaya[2], M. Syahrul Mubarok[3]

[1,2,3]Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom
[1]andridhikasaputra@gmail.com, [2]adiwijaya@telkomuniversity.ac.id,
[3]msyahrulmubarok@telkomuniversity.ac.id

Abstrak

Dari tahun ke tahun, transaksi *online* atau *e-commerce* semakin meningkat. Dengan peningkatan tersebut, *e-commerce* dapat memberikan peluang besar bagi produsen untuk memasarkan produk dan memudahkan orang-orang untuk berbagi aktivitas yang mereka lakukan, termasuk memberikan ulasan suatu produk. Ulasan tersebut digunakan oleh calon konsumen untuk mengetahui kelebihan atau kekurangan dari suatu produk dan dapat membantu calon konsumen dalam menentukan keputusan dalam pembelian produk. Dengan meningkatnya jumlah ulasan suatu produk, calon konsumen kesulitan untuk memahami semua ulasan suatu produk dan akhirnya tidak dapat menarik kesimpulan yang tepat dari ulasan tersebut. Oleh karena itu, pada tugas akhir dibangun sistem yang mampu melakukan klasifikasi sentimen terhadap fitur dan peringkasan hasil klasifikasi sentimen terhadap fitur. Klasifikasi sentimen dan peringkasan suatu ulasan produk dilakukan pada level aspek untuk mengetahui opini konsumen suka atau tidak terhadap fitur suatu produk. Klasifikasi aspek dan sentimen menggunakan pendekatan *supervised learning*, dimana *learning* menggunakan data yang berlabel. *Bayesian Network* merupakan salah satu metode yang digunakan pada *probabilistic classifiers*. *Bayesian network* digunakan untuk menentukan aspek yang terdapat pada ulasan beserta sentimen positif atau negatif dengan memanfaatkan hubungan antar kata-kata dan variabel pada ulasan. Penerapan *Bayesian network* untuk klasifikasi aspek menghasilkan performansi *f1-score* sebesar 88,73 % dan klasifikasi aspek dan sentimen menghasilkan performansi *f1-score* sebesar 86,0408%.

Kata kunci: klasifikasi sentimen, *level* aspek, ulasan produk, *Bayesian network*.

Abstract

Nowdays the online transactions or e-commerce is increasing. This increase of e-commerce can provide great opportunities for producers to market their products and make it easier for people to share their activities, including sharing reviews. reviews on products used by consumers to know the advantages or disadventages of a product and can help consumers to make a decision in purchasing the product. The number of reviews of product makes the consumer struggling to understand the review, so that ultimately the consumer is not able to conclude from the review. So that in this Final Project made a system which can classification of sentiment on feature product. Sentiment Classification and summaries of product review are based on the aspect level, because to know consumer opinion towards an aspect product whether they likes or not. Classification aspect and sentimen using supervised learning approach which used labelled data. Bayesian network method is a method used on probabilistic classifiers. Bayesian network method used to determine the aspects in the review with positive or negative sentiment by utilizing the relationship between words and variable in the review. The implementation of Bayesian network method on aspect classification can deliver performance f1- score about 88.73% and the classification aspect and sentiment can deliver performance f1-score about 86,0408%.

Keywords: sentiment analysis, aspect level, product reviews, Bayesian network.

1 Pendahuluan

Dari tahun ke tahun, transaksi *online* atau *e-commerce* semakin meningkat. Pada tahun 2016, *e-commerce* meningkat 6% dari tahun sebelumnya[1]. Dengan peningkatan tersebut, *e-commerce* dapat memberikan peluang besar bagi produsen untuk memasarkan produk dan memudahkan orang-orang untuk berbagi aktivitas yang mereka lakukan, termasuk memberikan ulasan suatu produk[2]. Ulasan tersebut digunakan oleh calon konsumen untuk mengetahui kelebihan atau kekurangan dari suatu produk dan dapat membantu calon konsumen dalam menentukan keputusan dalam pembelian produk. Hal tersebut dibuktikan oleh survei yang dilakukan *BrightLocal* yang menyatakan bahwa 88% konsumen percaya terhadap ulasan secara *online*[3]. Jumlah ulasan suatu produk semakin meningkat karena pada e-*commerce* memberikan kebebasan kepada konsumen untuk memberikan ulasan suatu produk[4].

Dengan meningkatnya jumlah ulasan suatu produk, calon konsumen kesulitan untuk memahami semua ulasan suatu produk dan akhirnya tidak dapat menarik kesimpulan yang tepat dari ulasan tersebut. Oleh karena itu, dibutuhkan sebuah sistem yang dapat memberikan ringkasan dan klasifikasi sentimen terhadap fitur produk yang diharapkan dapat membantu calon konsumen untuk memahami dan membantu dalam mengambil kesimpulan apakah ulasan tersebut mengandung opini positif atau opini negatif.

Menganalisis sentimen dan peringkasan suatu ulasan produk dilakukan pada *level* aspek untuk mengetahui opini konsumen suka atau tidak terhadap fitur suatu aspek[4]. Klasifikasi sentimen dan peringkasan terhadap ulasan suatu produk yang dianalisis yaitu ulasan produk menggunakan Bahasa Inggris. Pada penelitian tugas akhir akan dibangun suatu sistem untuk mengklasifikasikan opini dan peringkasan hasil klasifikasi opini terhadap fitur suatu produk dengan menggunakan pendekatan *supervised learning*, dimana *learning* menggunakan data yang berlabel[5]. Pada ulasan suatu produk terdapat unsur ketidakpastian, dimana opini positif dan negatif dapat muncul bersamaan dalam satu kalimat ulasan. *Bayesian network* merupakan salah satu metode yang digunakan pada *probabilistic classifiers* yang dapat menangani permasalah ketidakpastian pada ulasan. *Bayesian network* digunakan untuk menentukan aspek yang terdapat pada ulasan beserta sentimen positif atau negatif dengan memanfaatkan hubungan antar kata-kata dan variabel pada ulasan[6]. Performansi *Bayesian network* akan dibandingkan dengan performansi FBS (*Feature-Base Summarization*)[4] dalam melakukan identifikasi aspek.

2 Landasan Teori

Analisis Sentimen & Metodologi

Analisis Sentimen atau juga disebut *opinion mining* merupakan bidang studi yang menganalisis opini, sentimen, evaluasi, sikap dan emosi dari seseorang terhadap suatu entitas seperti produk, jasa, individu, topik, dan atribut dari suatu entitas[9]. Secara umum analsis sentimen dibagi menjadi tiga level, yaitu level dokumen, level kalimat, level aspek dan entitas. Hasil performa dari level aspek lebih baik dibandingkan level dokumen dan level kalimat. Alasan tersebut mengapa dalam penelitian Tugas Akhir ini memilih melakukan analisis sentimen pada level aspek dan entitas, karena suatu opini dinilai dari setiap aspek, sehingga hasil yang didapat lebih terperinci dalam menilai suatu produk.

Metodologi yang tekait dalam pengerjaan tugas akhir ini adalah metode yang digunakan pada proses *preprocessing*. Metode yang digunakan yaitu *stop word removal, lemmatization,* dan *part-of-speech (POS) Tagging*. *Stop Word Removal* merupakan proses penghapusan atau penghilangan *stop word* pada sebuah dokumen atau teks. *Stop Word* merupakan sekumpulan kata yang tidak memilik arti dan tidak mencirikan sebuah dokumen atau teks. Dalam konteks bahasa Inggris, *Stop Word* dapat berupa *prepositions* dan *conjunctions*[10]. *Lemmatization* adalah proses normalisasi teks sesuai dengan *lemma*-nya[11]. *Lemma* sendiri merupaka kata paling dasar dari sebuah kata yang memiliki arti pada kamus. Sebagai contoh pada konteks bahasa Inggris, bentuk *lemma* dari "*am, is, are, was, were*" menjadi kata dasar yaitu "*be*". *Part-of-Speech Tagging* merupakan proses identifikasi terhadap kata-kata pada sebuah kalimat berdasarkan kata benda, kata kerja, kata sifat dsb[9]. Proses *PoS Tagging* dilakukan menggunakan library java yaitu *Stanford POS-Tagger* yang dikembangkan oleh *Standord*.

Bayesian Network

Bayesian Network adalah model yang merepresentasikan random variabel dan hubungan antar variabel (*dependencies*)[6]. *Bayesian network* banyak diminati oleh berbagai bidang dikarenakan kemampuannya yang dapat memetakan dependensi antar variabel[12]. *Bayesian network* terdiri dari dua bagian, yaitu *Drirected Acyclic Graph* (DAG) dan *Conditional Probability Table* (CPT). Untuk menghitung probabilitas setiap variabel dengang menggunakan *parameter learning*. Perhitungan menggunakan *Maximum A Posterior (*MAP*) parameters.* Rumus *Maximu A Posterior* dapat dilihat pada rumus (1). θ_{ijk} adalah parameter di node $X_i = k$,dan $PA_i = j$, α_{ijk} adalah nilai bias dari prior di node X_i, dan n_{ijk} adalah jumlah kemunculan dari $X_i = k$,dan $PA_i = j$ pada dokumen.

$$\theta_{ijk} = \frac{\alpha_{ijk} + n_{ijk}}{\sum_k (\alpha_{ijk} + n_{ijk})} \quad (1)$$

Dimana nilai α_{ijk} didapat dari rumus (2). α adalah *equivalent sample size* yang bernilai 0.1, r_i adalah jumlah value dari node X_i, dan q_i adalah jumlah intansiasi value dari parent X_i..

$$\alpha_{ijk} = \frac{\alpha}{ri * qi} \tag{2}$$

DAG yang digunakan pada *Bayesian network* dapat ukur tingkat efisiensinya dengan menggunakan *scoring function*. *Scoring Function* merupakan bagian dari *Bayesian network* yang berguna untuk menilai suatu *graph* dalam merepresentasikan suatu data. Jika terdapat lebih dari satu *graph*, akan dilakukan *scoring function* untuk menentukan graph yang lebih baik dalam merepresentasikan variabel dan hubungan antar variabel pada data. Ada berbagai macam cara perhitungan *scoring function*, salah satunya yaitu dengan menggunakan *Minimum Description Length (*MDL*).* MDL berusaha mengukur dua kemampuan, yaitu *model encoding* dan *data encoding.*

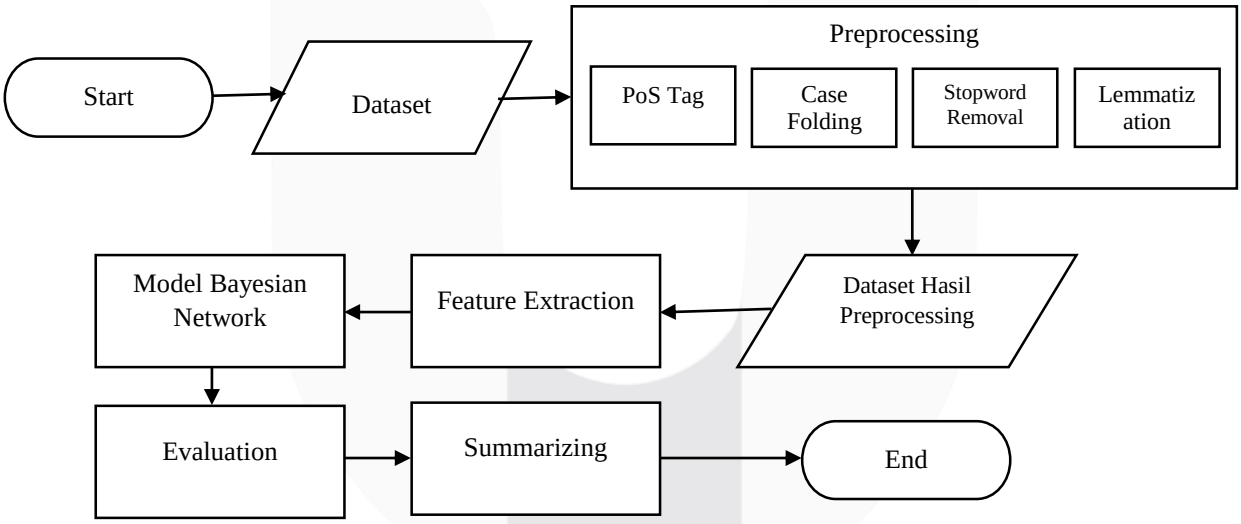
Rumus MDL dapat dilihat pada rumus (3) dimana *n* adalah jumlah *node* pada graph, q_i adalah jumlah instansiasi *value* dari node X_i, r_i adalah *value* dari node X_i, N_{ijk} adalah jumlah keunculan dari $X_i = k$,dan $PA_i = j$ pada dokumen, dan N_{ij} adalah Total kemunculan dari semua *value Xi* dimana $PA_i = j$.

$$MDL(\mathcal{N} : D) = -\sum_{i}^{n}\left\{\sum_{j}^{qi}\sum_{k}^{ri} N_{ijk}\log\frac{N_{ijk}}{N_{ij}}\right\} + \frac{\log N}{2} * (r_i - 1) * q_i \tag{3}$$

3 Skema yang Diusulkan

3.1 Gambaran Umum Sistem

Sistem yang akan dibuat pada penelitian tugas akhir ini adalah sebuah sistem yang dapat memberikan orientasi sentimen positif atau sentimen negatif pada aspek produk yang terdapat pada satu kalimat secara otomatis dan memberikan ringkasan terhadap opini-opini tersebut berdasakan fitur-fitur produk. Sistem ini dibagi menjadi lima proses utama, (1) *preprocessing* dataset, (2) *feature extraction*, (3) Model *Bayesian Network*, (4) *evaluation*, (5) *summarizing*. Gambaran umum sistem dapat dilihat pada *Gambar 1*.



Gambar 1 Gambaran Umum Sistem

3.2 Tahapan Tiap Proses

a. Preprocessing

Pada dataset yang digunakan akan dilakuakan *preprocessing* sebelum masuk ke proses *feature ekstraction*. Hasil dari tahap *preprocessing* ini adalah *Clean Data* yang siap digunakan pada tahap berikutnya. Pada tahap preprocessing ini akan dilakukan *PoSTagging*, *Case Folding*, *Stopword Removal,* dan *Lemmatization*. Berikut adalah penjelasan dari setiap tahapan yang dilakukan pada tahap *preprocessing* :

Case Folding

Pada *dataset* penggunaan huruf kapital atau huruf kecil pada kata tidak konsisten. Untuk menyelaraskan semua kata dibutuhkan *case folding* untuk mengkonversi keseluruhan kata pada *dataset* menjadi bentuk yang standar. Pada tahap ini semua kata pada *dataset* akan dikonversi menjadi huruf kecil. Contoh kalimat yang diolah pada tahap ini adalah:

Sebelum **: I have had the phone for 1 week , the signal quality has been great in the detroit area.**

Sesudah **: i have had the phone for 1 week , the signal quality has been great in the detroit area**

Stopword Removal

Stop word removal adalah proses menghilangkan kata-kata yang memiliki fungsi pada kalimat namun tidak memiliki arti dari kata yang akan diolah. Contoh kalimat yang diolah pada tahap ini adalah:

Sebelum : **i have had the phone for 1 week , the signal quality has been great in the detroit area**

Sesudah : **phone 1 week, signal quality great detroit area**

Dari contoh diatas dapat dilihat, untuk kata "*i, have, had, the, for, has, been, in*" dihilangkan karena termasuk kata stopword.

Lemmatization

Pada tahap ini semua kata akan dirubah kedalam bentuk kata dasarnya. Untuk tahap ini digunakan *tools* dalam bentuk library yaitu *StanfordNLP*. Contoh kalimat yang diolah pada tahap ini adalah:

Sebelum : **i have had the phone for 1 week , the signal quality has been great in the detroit area**

Sesudah : **i have have the phone for 1 week, the signal quality have be great in the detroit area**

Dari contoh diatas dapat dilihat untuk kata "*had, has, been*" berubah menjadi kata dasar "*have, be*".

PoS Tagging

Pada tahap ini setiap kalimat pada *dataset* akan diidentifikasi untuk mendapatkan *tag-tag* jenis kata. *POS Tagging* dilakukan dengan menggunakan *tools* berupa *library* yaitu *Stanford coreNLP*. Contoh kalimat yang diolah pada tahap ini adalah:

Sebelum : **i have had the phone for 1 week , the signal quality has been great in the detroit area**

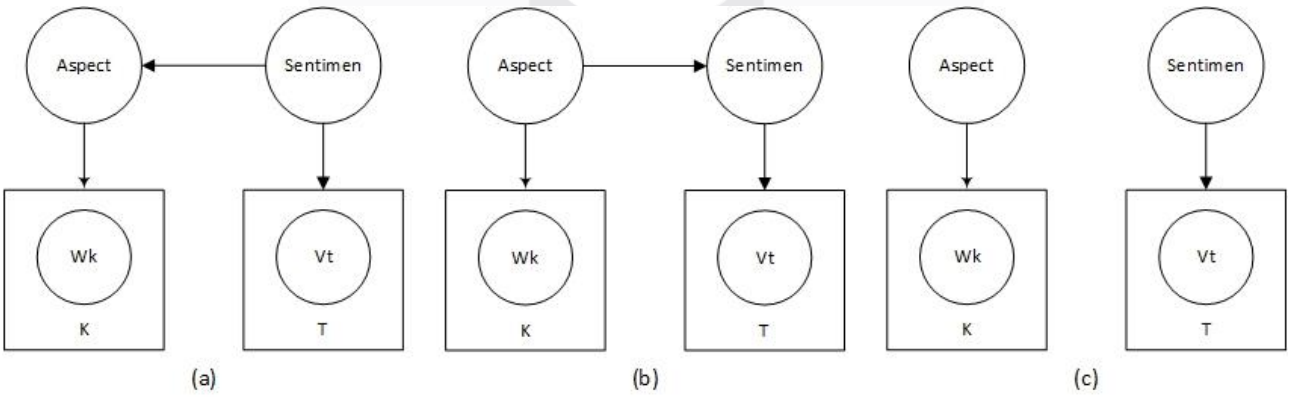
Sesudah : **i_LS have_VBP had_VBD the_DT phone_NN for_IN 1_CD week_NN the_DT signal_NN quality_NN has_VBZ been_VBN great_JJ in_IN the_DT detroit_NN area_NN**

b. Feature Extraction

Pada tahap ini, semua dataset hasil *preprocessing* akan dilakukan *Tokenization*, dimana akan dipisah menjadi perkata. Kata-kata tersebut akan dipisahkan kedalam dua *Bag of Words* yaitu *Bag of Words* W_k dan *Bag of Words* V_t. *Bag of words* V_t akan berisi kata-kata sifat atau kata-kata yang memiliki *tag "JJ, JJR,* dan *JJS"* dari hasil *PoSTagging*, sedangkan *Bag of words* B berisi kata-kata selain kata sifat atau kata-kata yang tidak memiliki *tag "JJ, JJR,* dan *JJS"* dari hasil *PoSTagging*.

c.* Model *Bayesian Network

Bag of words dari hasil *feature extraction* akan digunakan pada model *Bayesian network*. Pada penelitian ini akan digunakan tiga struktur *Bayesian Network* yang berbeda yaitu *Graph* 1, *Graph* 2, dan *Graph* 3 dapat dilihat pada Gambar 2.



Gambar 2 Struktur *graph* yang digunakan untuk membangun sistem, (a) *Graph* 1, (b) *Graph* 2, dan (c) *Graph* 3

dimana P(*Aspect*,*Sentiment*|d) adalah kemungkinan atau probabilitas suatu *node aspect* dan *node sentiment* terhadap dokumen d, P(*Aspect*) adalah kemungkinan atau probabilitas suatu *node aspect*, P(*Sentiment*) adalah kemungkinan atau probabilitas suatu *node sentiment*, P(*Aspect*|*Sentiment*) adalah kemungkinan atau probabilitas *node aspect* terhadap *node sentiment*, P(*Sentimen*|*Aspect*) adalah kemungkinan atau probabilias *node sentiment* terhadap *node aspect*, P(W_k|*Aspect*) adalah kemungkinan suatu kata non-sifat (*node* W_1,W_2,...,W_k) terhadap *node aspect*, P(V_t|*Sentimen*) adalah kemungkinan suatu kata sifat (node V_1,V_2,...,V_t) terhadap node *sentiment.*

Dari ketiga struktur diatas, yang membedakan tiap struktur adalah koneksi antara *node aspect* dan *node sentiment*. Dalam menentukan klasifikasi sentimen akan dipengaruhi oleh kumpulan kata yang saling independen, begitu pula dengan dengan menentukan aspek akan dipengaruhi oleh kumpulan kata yang saling independen. Label pada dataset yang digunakan yaitu *Multi-Label*, dimana satu ulasan pada dataset dapat memiliki lebih dari satu aspek yang dibicarakan. Solusi untuk permasalahan multi-label pada level aspect yaitu dengan mereperesentasikan dalam berbeda struktur *graph*, artinya setiap aspect memiliki struktur graph masing-masing. Contoh ketika pada dataset terdapat 95 *aspect*, maka jumlah struktur *graph* yang akan dibuat sebanyak 95x3.

Untuk *Graph* 1 yang ditunjukan pada Gambar 2(a) untuk klasifikasi aspek dan sentimen menggunakan rumus berikut

$$P(Aspect, Sentiment|d) \propto \mathrm{P}(Aspect|Sentiment)P(Sentiment)\prod_{t=1}^{d} P(Vt|Sentiment)\prod_{k=1}^{d} P(Wk|Aspect) \quad (4)$$

Untuk *Graph* 2 yang ditunjukan pada Gambar 2(b) untuk klasifikasi aspek dan sentimen menggunakan rumus berikut

$$P(Aspect, Sentiment|d) \propto \mathrm{P}(Aspect)P(Sentiment|Aspect)\prod_{t=1}^{d} P(Vt|Sentiment)\prod_{k=1}^{d} P(Wk|Aspect) \quad (5)$$

Untuk *Graph* 3 yang ditunjukan pada Gambar 2(c) untuk klasifikasi aspek dan sentimen menggunakan rumus berikut

$$P(Aspect, Sentiment|d) \propto \mathrm{P}(Aspect)P(Sentiment)\prod_{t=1}^{d} P(Vt|Sentiment)\prod_{k=1}^{d} P(Wk|Aspect) \quad (6)$$

dimana P(*Aspect*,*Sentiment*|d) adalah kemungkinan atau probabilitas suatu *node aspect* dan *node sentiment* terhadap dokumen d, P(*Aspect*) adalah kemungkinan atau probabilitas suatu *node aspect*, P(*Sentiment*) adalah kemungkinan atau probabilitas suatu *node sentiment*, P(*Aspect*|*Sentiment*) adalah kemungkinan atau probabilitas *node aspect* terhadap *node sentiment*, P(*Sentiment*|*Aspect*) adalah kemungkinan atau probabilias *node sentiment* terhadap *node aspect*, P(W_k|*Aspect*) adalah kemungkinan suatu kata non-sifat (*node* W_1,W_2,...,W_k) terhadap *node aspect*, P(V_t|*Sentimen*) adalah kemungkinan suatu kata sifat (*node* V_1,V_2,...,V_t) terhadap node *sentiment.*

d. Evaluasi

Untuk mengetahui performansi dari sistem yang dibangun akan dilakukan evaluasi untuk setiap tahapan. Ada dua evaluasi pada sistem yaitu, (1) evaluasi pada aspek produk yang dilakukan dengan cara mengetahui selisih antara standar yang telah ditetapkan pada corpus dengan aspek produk yang dapat terambil oleh sistem dengan menggunakan *Bayesian network*, (2)evaluasi yang dilakukan pada klasifikasi untuk mengetahui performansi sistem dalam pemberian orientasi sentimen pada aspek dengan menggunakan *Bayesian network* Untuk evaluasi pada penelitian tugas akhir ini dilakukan dengan menggunakan *Precision, Recall dan F-Score*. Hasil prediksi akan direpresentasikan oleh *confusion matrix*, dijelaskan pada Tabel 1.

Tabel 1 *Confusion Matrix*

		Actual Class	
		+	-
Predicted Class	+	*True Positive (TP)*	*False Positive (FP)*
	-	*False Negative (FN)*	*True Negative (TN)*

Ada 4 pengkategorian dalam *matriks confusion*, yaitu:

1. *True positive* (TP) adalah target yang memiliki kelas positif dan hasil prediksi menyatakan kelas positif.
2. *False positive* (FP) adalah target yang memiliki kelas negatif tetapi hasil prediksi menyatakan kelas positif.
3. *True negative* (TN) adalah target yang memiliki kelas negatif dan hasil prediksi menyatakan kelas negatif.
4. *False negative* (FN) adalah target yang memiliki kelas positif tetapi hasil prediksi menyatakan kelas negatif.

Dari confusion Matrix ini lah *Precision, Recall,* dan *F-Score* dapat diukur dengan menggunakan rumus sebagai berikut:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (9)$$

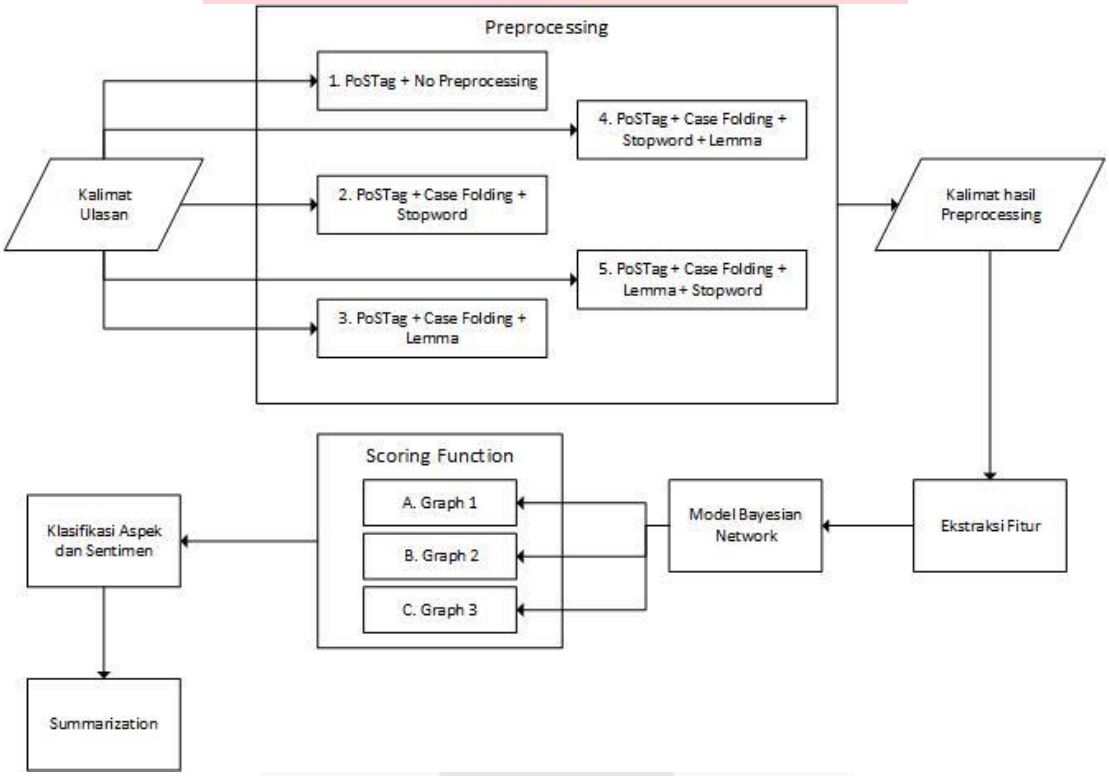
4 Analisis Hasil Pengujian

Secara garis besar hasil sistem yang dibagun adalah meringkas opini mengenai fitur suatu produk melalui kumpulan data ulasan dengan menjabarkan fitur produk dan orientasi sentimennya. Proses pengujian dilakukan pada setiap tahapan untuk mengetahui nilai keberhasilan pada ringkasan opini yang didapat. Tujuan pengujian tersebut, yaitu :

1. Menganalisis hasil metode *preprocessing* yang dapat mendukung proses klasifikasi aspek menggunakan metode *Bayesian network.*
2. Menentukan struktur *graph* yang baik dalam mendukung klasifikasi aspek dan sentimen menggunakan *scoring function.*
3. Menganalisis hasil pembangkitan ringkasan (*summarization*) dengan menampilkan nama produk dan lima aspek teratas beserta banyaknya opini positif maupun opini negatif dari hasil klasifikasi aspek dan sentimen.

4.1 Skenario Pengujian

Gambaran umum senario pengujian dapat dilihat pada Gambar 3.



Gambar 3 Gambaran umum skenario pengujian

Skenario pengujian yang akan dilakukan pada setiap tahapan program pada penelitian ini, yaitu (1) Pemilihan Metode *preprocessing*, pengujian ini dilakukan untuk mendapatkan hasil proses preprocessing yang dapat mendukung proses klasifikasi aspek dan sentimen, (2) Pemilihan Struktur *Graph*, pengujian ini dilakukan untuk menentukan graph yang terbaik dalam memodelkan variabel dan hubungan antar variabel pada data dan dalam melakkukan klasifikasi aspek dan sentimen.

4.2 Analisis Hasil Pengujian

4.2.1 Analisis Skema *Preprocessing*

Proses pengujian dilakukan dengan 5 kombinasi *preprocessing* yaitu, (1)*PoS Tag + No Preprocessing, (2) PoS Tag + Lemmatization, (3) PoS Tag + Stopword Removal,(4) PoS Tag + Stopword Removal + Lemmatization, dan (5) PoS Tag + Lemmatization + Stopword Removal.* Perbandingan nilai rata-rata *f1-score* dari tiap *graph* dapat dilihat pada Tabel 2, dan 4.

Tabel 2 Perbandingan nilai rata-rata *f1-score* pada *graph* 1

No	Dataset	F1-Score Graph 1				
		1	2	3	4	5
1	Apex dvd player	87,33%	84,17%	86,33%	79,73%	78,78%
2	Canon G3	88,03%	76,80%	88,04%	75,73%	75,81%
3	Zen Mp3 Player	83,89%	73,20%	82,43%	71,58%	71,42%
4	Nikon Coolpix 6610	88,52%	84,49%	88,63%	84,34%	84,71%
5	Nokia 6610	88,12%	74,80%	87,04%	73,96%	73,22%
Rata-rata		81,17%	78,69%	86,49%	77,06%	76,78%

Tabel 3 Perbandingan nilai rata-rata *f1-score* pada *graph* 2

No	Dataset	F1-Score Graph 2				
		1	2	3	4	5
1	Apex dvd player	90,91%	90,77%	89,04%	87,34%	86,28%
2	Canon G3	89,42%	81,49%	90,21%	80,78%	81%
3	Zen Mp3 Player	86,69%	82,47%	86,47%	80,66%	80,48%
4	Nikon Coolpix 6610	90,56%	90,40%	89,75%	90,01%	90,47%
5	Nokia 6610	90,57%	83,26%	89,74%	81,89%	90,47%
Rata-rata		89,63%	85,67%	89,04%	84,13%	85,74%

Tabel 4 Perbandingan nilai rata-rata *f1-score* pada *graph* 3

No	Dataset	F1-Score Graph 3				
		1	2	3	4	5
1	Apex dvd player	90,80%	90,75%	89,10%	87,44%	86,28%
2	Canon G3	89,54%	81,48%	90,20%	90,78%	90,99%
3	Zen Mp3 Player	86,85%	82,47%	86,70%	80,60%	80,48%
4	Nikon Coolpix 6610	89,57%	90,40%	90,55%	90%	90,46%
5	Nokia 6610	90,16%	83,26%	89,74%	81,89%	81,50%
Rata-rata		89,38%	85,67%	89,25%	86,14%	85,94%

Berdasarkan Tabel 2, 3, dan 4 setelah dilakukan pengujian terhadap 5 kombinasi preprocessing didapatkan bahwa nilai rata-rata f1-score terbaik adalah 89,38% dari hasil kombinasi nomer 1 yaitu *PoS Tag + No Preprocessing.* Hal tersebut terjadi karena saat tidak menggunakan stopword removal pada *preprocessing* tidak akan menghilangkan kata-kata apapun dan dapat digunakan sebagai ciri dari suatu aspek. Hasil rata-rata *f1-score* pada kombinasi yang tidak menggunakan *stopword removal* lebih baik dibanding menggunakan *stopword removal*, karena pada penggunaan stopword removal akan menghilangkan kata-kata yang ada pada kata stopword dan mungkin kata-kata tersebut adalah ciri dari aspek tertentu. Pada *Bayesian Network* dalam menentukan suatu aspek akan dipengaruhi oleh kata-kata yang terkait terhadap aspek tersebut. Semakin banyak kata-kata yang terkait terhadap aspek, maka semakin baik pula *bayesian network* dalam melakukan klasifikasi aspek.

Pada dataset Apex DVD Player, ada beberapa kalimat yang sistem tidak dapat melakukan prediksi aspek secara tepat, pada kalimat "*the dvd player is fine*" memiliki aspek *player* pada corpus tetapi sistem memprediksi pada kalimat tersebut memiliki aspek *dvd player* dan *player*. Hal tersebut terjadi karena pada kedua aspek tersebut terdapat kata *dvd* dan *player* yang memiliki jumlah kemunculan yang banyak pada dataset dan nilai probabilitas dari kata tersebut lebih besar dibanding kata-kata yang lainnya. Akhirnya sistem tidak dapat memprediksi secara tepat ketika ada kata *dvd* dan *player* muncul bersamaan pada kalimat.

Pada dataset Canon G3, ada beberapa kalimat yang sistem tidak dapat melakukan prediksi aspek secara tepat, pada kalimat "*4x zoom is nice*" memiliki aspek *zoom* pada corpus tetapi sistem memprediksi pada kalimat tersebut memiliki aspek *zoom, zoom optical,* dan *digital zoom.* Hal tersebut terjadi karena pada kata *zoom* memiliki nilai probabilitas yang tinggi pada aspek yang diprediksi. Kalimat yang mengulas aspek *zoom, zoom optical,* dan *digital zoom* hanya muncul satu kali dan kata *zoom* muncul di kalimat tersebut.

Pada dataset Zen Mp3 Player, ada beberapa kalimat yang sistem tidak bisa melakukan prediksi aspek secara tepat, pada kalimat "*great price*" memiliki aspek *price* pada corpus tetapi sistem memprediksi pada kalimat tersebut memiliki aspek *size, sound, screen, battery life, price, navigation, battery, control, player, use, look, sound quality,*

playback quality, storage, firmware, feature, value, product, capacity, dan *quality*. Hal tersebut terjadi karena pada kalimat tersebut hanya ada 2 kata yaitu *great* dan *price.* Kata *price* memiliki nilai probabilitas yang tinggi pada aspek *price* dan pada aspek lainnya memiliki nilai probabilitas yang rendah, tetapi pada kata *great* memiliki nilai probabilitas yang tinggi pada aspek yang diprediksi.

Pada dataset Nikon Coolpix, ada beberapa kalimat yang sistem tidak bisa melakukan prediksi aspek secara tepat, pada kalimat "*the quality is super*"memiliki aspek *quality* pada corpus tetapi sistem memprediksi pada kalimat tersebut memiliki aspek *picture quality, movie, customer service, optical zoom, quality, design, construction,* dan *optic.* Kalimat tersebut memiliki 4 kata, pada kata *the* memiliki probabilitas yang tinggi pada aspek yang diprediksi, kata *quality* memiliki probabilitas yang tinggi pada aspek *picture quality* dan *quality.* ketika jumlah kata pada kalimat sedikit dan terdapat kata *the* akan memprediksi aspek yang tidak tepat karena kata *the* memiliki probabilitas yang tinggi, tetapi ketika jumlah kata pada pada kalimat banyak maka kata *the* tidak berpengaruh banyak pada semua aspek. Pada kalimat " *the manual is easy to understand and it is mostly idiot proof*" terdapat kata *the* dan sistem berhasil memprediksi aspek yang dimiliki kalimat tersebut adalah aspek *manual.*

Pada dataset Nokia 6610, ada beberapa kalimat yang sistem tidak bisa melakukan prediksi aspek secara tepat, pada kalimat "*great battery life, perfect size*" memiliki aspek *battery life* dan *size* pada corpus, tetapi sistem memprediksi pada kalimat tersebut memiliki aspek *battery life, size, weight, battery.* Hal tersebut terjadi karena pada kalimat tersebut terdapat kata *battery, life* dan *size.* Kata *battery* memiliki probabilitas yang tinggi pada aspek *battery life* dan *battery,* kata *life* memiliki nilai probabilitas yang tinggi pada aspek *battery life,* pada kata *size* memiliki nilai probabilitas yang tinggi pada aspek *size* dan *weight.*

Performansi *Bayesian network* yang menggunakan metode *PoS Tagging + No Preprocessing* akan dibandingkan dengan metode FBS (*Feature-Based Summarization*)[4]. Hasil perbandingan dapat dilihat pada

Tabel 5 Perbandingan nilai *precision* dan *recall* antara *Bayesian network* dan *FBS*

Dataset	Bayesian network		FBS			
			Frequent feature		Infrequent feature identification	
	Precision	Recall	Precision	Recall	Precision	Recall
Digital Camera 1	90,54%	87,80%	55,2%	67,1%	74,7%	82,2%
Digital Camera 2	91,01%	87,28%	59,4%	59,4%	71%	79,2%
Celluler Phone	86,76%	93,05%	56,3%	73,1%	71,8%	76,1%
Mp3 Player	87,41%	84,51%	57,3%	65,2%	69,2%	81,8%
Dvd Player	90.25%	89,42%	53,1%	75,4%	74,3%	79,7%
Rata- rata	88,93%	88,41%	56,26%	68,04%	72,20%	79,80%

Berdasarkan Tabel 5, dapat dilihat nilai precision dan recall dari *Bayesian network* lebih baik dibanding dengan *FBS*. Pada FBS dalam melakukan identifikasi aspek menggunakan *association mining* untuk menemukan semua *frequent itemset*. Jika pada *frequent feature* masih ada *feature* yang tidak teridentifikasi maka akan menggunakan *infrequent feature.* Pada kalimat "*The pictures are absolutely amazing."* dan "*The software that comes with it is amazing.",* pada kata opini *amazing* menggambarkan dua feature yang berbeda yaitu *software* dan *pictures*. *Frequent feature* tidak dapat mengidentifikasi *feature* tersebut dan menggunakan *infrequent feature* agar dapat mengidentifikasi *feature* tersebut. Pada *Bayesian network* dapat mengidentifikasi *feature* dari kalimat diatas, karena pada *Bayesian network* menggunakan probabilitas dari setiap kata-kata yang ada pada kalimat terhadap *feature* yang dibicarakan.

4.2.2 Analisis Skema *Scoring Function* dan Klasifikasi aspek beserta sentimennya

Proses *Scoring Function* untuk menilai seberapa baik struktur *graph* dalam merepresentasikan data. Pada skema ini menilai dari struktur graph yang telah ditentukan yaitu *Graph* 1, *Graph* 2, dan Graph 3 yang dapat dilihat pada Gambar 2(a), (b), dan (c). Penentuan struktur terbaik akan ditentukan berdasarkan nilai *Scoring Function* yang terkecil dan nilai rata-rata *f1-score* yang terbesar dari hasil klasifikasi aspek dan sentimen. Hasil nilai *scoring function* dari ketiga *graph* yang digunakan dapat dilihat pada Tabel 6.

Tabel 6 Hasil nilai *scoring function*

No	Dataset	Nilai Scoring Function		
		1	2	3
1	Apex dvd player	7002,239	7167,274	7406,151
2	Canon G3	6236,528	6318,846	6545,772
3	Zen Mp3 Player	13580,97	13930,34	14458,75
4	Nikon Coolpix 6610	4248,715	4290,665	4462,611

5	Nokia 6610	6178,464	6286,428	6549,625
Rata-rata		7449,383	7598,71	7884,581

Berdasarkan 6, secara konsisten nilai scoring function yang tekecil adalah pada graph 1 dengan rata-rata sebesar 7449,383. Hal tersebut membuktikan pada graph 1 dapat memodelkan hubungan antar variabel pada data dengan menggunakan bit yang sedikit dibanding dengan *graph* 2 dan 3. Yang membedakan dari ketiga graph tersebut adalah nilai probabilitas pada *node aspect* dan *node sentiment.*

Setelah dilakukan scoring function akan dilakukan perhitungan nilai rata-rata *f1-score* dari hasil klasifikasi aspek dan sentimen terhadap ketiga struktur yang diujikan. Tabel 7adalah tabel yang menampilkan hasil perbandingan nilai *f1-score* terhadap ketiga graph yang diujikan.

Tabel 7 Perbandingan nilai *f1- score* dari hasil klasifikasi aspek dan sentimen

No	Dataset	F1-Score		
		1	2	3
1	Apex dvd player	80,20%	82,92%	82,82%
2	Canon G3	87,55%	89,96%	89,06%
3	Zen Mp3 Player	78,62%	81,18%	81,47%
4	Nikon Coolpix 6610	87,13%	87,58%	88,19%
5	Nokia 6610	86,59%	88,54%	88,64%
Rata-rata		84,02%	86,0408%	86,0401%

Berdasarkan Tabel 7 didapatkan bahwa nilai rata-rata *f1-score* yang terbaik adalah pada *graph* 2 sebesar 86,0408%. Pada *graph* 2 berhasil memprediksi aspek dan sentimen dengan benar sesuai dengan aspek dan sentimen yang ada pada corpus dibandingkan denga *graph* 1 dan 3. Hal tersebut terjadi karena perbedaan nilai probabilitas pada node Sentiment. Pada *graph* 2, nilai probabilitas pada *node sentiment* akan dipengaruhi oleh *node aspect.*

Berdasarkan hasil nilai *scoring function* dan nilai rata-rata *f1-score* dapat disimpulkan bahwa graph yang terbaik dalam memodelkan hubungan antar variabel pada data dan melakukan klasifikasi aspek dan sentimen adalah *graph 2*. Dapat lihat antara *graph* 1 dan *graph* 2, walaupun nilai scoring function pada *graph* 2 lebih besar dibandingkan *graph* 1 tetapi dalam melakukan klasifikasi *graph* 2 lebih baik dibandingkan graph 1 dengan selisih 2,02%. Kemudian jika dilihat antara *graph* 2 dan 3, *graph* 2 unggul dalam nilai *scoring function* dan nilai rata-rata *f1-score*.

5 Kesimpulan

Sistem yang dibangun dapat memodelkan hubungan antara variabel pada data yaitu dengan bentuk struktur *graph* 2. terdapat hubungan antara variabel aspek dan variabel sentimen, dimana variabel sentimen akan dipengaruhi dipengaruhi oleh variabel aspek. Sistem yang dibangun juga dapat melakukan identifikasi kelas aspek dan klasifikasi sentimen terhadap aspek.

Penggunaan metode preprocessing yang berbeda untuk identifikasi kelas aspek dapat memberikan hasil yang berbeda pada nilai *f1-score* berkisar 0,46% - 5,90%. Metode *preprocessing* yang dapat mendukung identifikasi kelas aspek adalah *PoS Tagging + No Preprocessing* dengan nilai *f1-score* sebesar 88,73%.

Bentuk sturktur graph dapat memberikan pengaruh pada hasil klasifikasi sentimen terhadap aspek. Perubahan bentuk struktur dapat mempengaruhi nilai probabilitas dari tiap variabel. Pada ketiga graph yang digunakan terdapat selisih nilai *f1-score* sebesar 0,0006% - 2,01%. Hasil nilai rata-rata f1-score yang terbaik adalah pada struktur *graph 2* sebesar 86,0408%.

Daftar Pustaka

[1] "Worldwide Retail Ecommerce Sales Will Reach $1.915 Trillion This Year - eMarketer." [Online]. Available: http://www.emarketer.com/Article/Worldwide-Retail-Ecommerce-Sales-Will-Reach-1915-Trillion-This-Year/1014369. [Accessed: 10-Oct-2016].

[2] A. R. Naradhipa and A. Purwarianti, "Sentiment classification for Indonesian message in social media," in *2012 International Conference on Cloud Computing and Social Networking (ICCCSN)*, 2012, pp. 1–5.

[3] "88% Of Consumers Trust Online Reviews As Much As Personal Recommendations," *Search Engine Land*, 07-Jul-2014. [Online]. Available: http://searchengineland.com/88-consumers-trust-online-reviews-much-personal-recommendations-195803. [Accessed: 10-Oct-2016].

[4] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2004, pp. 168–177.

[5] D. K. Gupta and A. Ekbal, "IITP: Supervised Machine Learning for Aspect based Sentiment Analysis," *Proc. 8th Int. Workshop Semantic Eval. SemEval 2014 Deep. Kumar Gupta*, 2014.

[6] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014.

[7] O. Maimon and L. Rokach, Eds., *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer US, 2010.

[8] S. M. Weiss, N. Indurkhya, and T. Zhang, *Fundamentals of Predictive Text Mining*. London: Springer London, 2010.

[9] L. Zhang and B. Liu, "Sentiment Analysis and Opinion Mining," in *Encyclopedia of Machine Learning and Data Mining*, C. Sammut and G. I. Webb, Eds. Springer US, 2016, pp. 1–10.

[10] C. Silva and B. Ribeiro, "The importance of stop word removal on recall values in text categorization," in *Proceedings of the International Joint Conference on Neural Networks, 2003*, 2003, vol. 3, pp. 1661–1666 vol.3.

[11] S. Nirenburg and S. Nirenburg, *Language Engineering for Lesser-Studied Languages*. Amsterdam, The Netherlands, The Netherlands: IOS Press, 2009.

[12] A. H. R. Z. Arifin, M. S. Mubarok, and Adiwijaya, "Learning Struktur Bayesian Networks menggunakan Novel Modified Binary Differential Evolution pada Klasifikasi Data," *Indones. Symp. Comput. IndoSC 2016*, p. 2016.

[13] B. Malone, "Scoring Functions for Learning Bayesian Networks and Parameter Estimation with Complete Data," 2014.

[14] D. Jurafsky and C. Manning, "Text Classification and Naïve Bayes." stanford.edu, 11-Jan-2012.

[15] Adiwijaya, *Aplikasi Matriks dan Ruang Vektor*. Yogyakarta: Graha Ilmu, 2014.

[16] Adiwijaya, *Matematika Diskrit dan Aplikasinya*. Bandung: Alfabeta, 2016.

[17] M. S. Mubarok, Adiwijaya, and M. D. Aldhi, "Aspect-based sentiment analysis to review products using Naïve Bayes," *AIP Conf. Proc. 1867 020060 2017*, p. 2017.

[18] R. A. Aziz, M. S. Mubarok, and Adiwijaya, "Klasifikasi Topik pada Lirik Lagu dengan Metode Multinomial Naive Bayes," *Indones. Symp. Comput. IndoSC 2016*, 2016.