



Evaluasi Topik Tersembunyi Berdasarkan *Aspect Extraction* menggunakan Pengembangan *Latent Dirichlet Allocation*

Dinda Adimanggala¹, Fitra A. Bachtiar², Eko Setiawan³

^{1,2,3}Jurusan Magister Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Brawijaya

¹dindadinho@student.ub.ac.id, ²fitra.bachtiar@ub.ac.id, ³ekosetiawan@ub.ac.id

Abstract

Recently, *Sentiment Analysis* is used for expression detection of products or services. *Sentiment Analysis* is one category type with a level of aspect focused on extracting product aspects. One of the common methods used for aspect extraction is *Latent Dirichlet Allocation (LDA)* using random topic identification, but this method has not been able to find an acceptable topic with some aspects having been found. Undeterminable topics are referred to as the hidden topics. This study purpose is to evaluate and compare the suitability of identifying hidden topics between human and computer evaluation. The study is also focused on aspect extraction using a variety of LDA innovations. The data used in this study used case studies on e-Commerce. Data were processed using feature selection and grouped using LDA development. Then the data results are processed using *Latent Topic Identification* based on subjective and objective evaluations. The identification of hidden topic results was evaluated using several semantic and lexicon tests. The evaluation results indicate the comparison of two hidden topic identification assessment values is quite relevant with the average difference in value reaching 6%. As a result, computer calculations assist humans in determining topics if each topic has a low coherence value.

Keywords: *sentiment analysis, aspect, topic, extraction, LDA, evaluation*

Abstrak

Pada beberapa tahun terakhir, *Sentiment Analysis* sering dijumpai sebagai pendeteksi ekspresi pengguna terhadap suatu barang atau jasa. *Sentiment Analysis* mempunyai beberapa klasifikasi dan tingkatan, salah satu jenis *Sentiment Analysis* tersebut adalah tingkat aspek yang berfokus pada ekstraksi aspek suatu produk. Penelitian ini menggunakan metode yang populer saat ini, yaitu, *Latent Dirichlet Allocation (LDA)* yang menggunakan identifikasi topik secara acak. Akan tetapi metode ini belum dapat menentukan topik yang sesuai dengan aspek-aspek yang diperoleh. Topik yang belum bisa ditentukan tersebut juga bisa disebut dengan topik tersembunyi. Fokus penelitian ini adalah evaluasi dan perbandingan kesesuaian dari identifikasi topik yang tersembunyi antara penilaian manusia dan perhitungan komputer. Penelitian ini didasarkan pada ekstraksi aspek menggunakan beberapa pengembangan LDA. Data yang digunakan pada penelitian ini menggunakan studi kasus pada e-Commerce dengan jumlah ulasan sebanyak 28.222. Pengolahan data menggunakan seleksi fitur dengan nilai α sebesar 0,99 dan dikelompokkan menggunakan pengembangan LDA yang ditentukan dengan nilai *hyperparameter* α sebesar 0,9 dan β sebesar 0,05. Kemudian hasil data tersebut diolah menggunakan *Latent Topic Identification* berdasarkan penilaian subjektif dan objektif. Hasil identifikasi topik tersembunyi dievaluasi menggunakan beberapa pengujian semantik dan leksikon. Hasil evaluasi mengindikasikan bahwa nilai perbandingan kedua penilaian identifikasi topik tersembunyi cukup relevan dengan rata-rata selisih nilai mencapai 6%. Sehingga perhitungan komputer dapat membantu manusia untuk menentukan topik jika setiap aspek memiliki nilai koherensi yang rendah.

Kata kunci: *sentiment analysis, aspek, topik, ekstraksi, LDA, evaluasi.*

1. Pendahuluan

Sentiment Analysis (SA) merupakan sebuah teknologi yang berfungsi untuk mendeteksi opini atau ekspresi dari seorang *user* terhadap entitasnya atau sebuah aspek. SA dapat menambahkan kemampuan dalam hal sistem rekomendasi dan hubungan terhadap pelanggan [1].

Pelanggan dapat mencari fitur dari sebuah barang yang baik berdasarkan sentimen positif atau negatif dari opini pelanggan lain. Pada dasarnya, SA merupakan cabang dari penelitian dari *affective computing* dan salah satu bidang *Natural Language Processing (NLP)* [2]. SA terdiri dari *Sentiment* dan *Analysis* yang memiliki jenis dan tingkatan pada tiap-tiap kata. Pada umumnya

sentiment yang digunakan terdiri dari positif, negatif, dan netral. Sementara itu, *analysis* terdiri dari tingkat dokumen, kalimat, dan aspek [3,4]. SA dengan tingkat dokumen dan kalimat menganalisa polaritas dokumen secara keseluruhan, sedangkan SA dengan tingkat aspek berfokus pada ekstraksi aspek produk yang disebutkan pada ulasan terhadap setiap aspek [4].

Ekstraksi aspek atau *Aspect Extraction* merupakan bagian penting dari SA pada tingkat aspek yang memiliki kinerja yang lebih baik [5]. Selain mengacu pada pengumpulan kata dalam setiap subtopik pada topik [6], tujuan dari ekstraksi aspek adalah mendeteksi aspek tertentu dari produk atau layanan yang mendapatkan respon positif atau negatif dari pengguna [7]. Secara umum, ekstraksi aspek mengandung frasa dan kata benda [8,9]. Pada penelitian sebelumnya, terdapat beberapa jenis pendekatan yang dapat diaplikasikan pada ekstraksi aspek, yaitu *supervised learning* [5,10], *unsupervised learning* [11–13], dan *rule-based learning* [9,14]. *Supervised learning* membutuhkan pelabelan data, sedangkan pada *unsupervised learning* tidak membutuhkannya.

Latent Dirichlet Allocation (LDA) merupakan metode yang populer pada *aspect extraction*, khususnya pada pendekatan *unsupervised learning* [6]. Metode *topic modelling* ini diperkenalkan oleh Blei [15,16]. Metode ini dapat mengidentifikasi topik secara acak pada suatu dokumen. Saat ini, banyak peneliti melakukan pengembangan LDA seperti, LDA- *Global Vector for Word Representation* (GloVe) yang menambahkan word embedding pada saat proses seleksi fitur atau *Biterm Topic Modelling* (BTM) yang menambahkan jumlah *term* pada proses LDA [12]. Pada saat menggunakan pendekatan *topic modelling*, peneliti berfokus pada sejauh mana topik yang dipelajari sesuai dengan penilaian manusia dan mampu menentukan topik tersebut. Faktanya, topik tidak memberikan jaminan penafsiran yang baik sehingga pengukuran koheren diusulkan untuk membedakan antara topik tersebut. *Topic Coherence* merupakan pengukuran yang dapat membantu membedakan antara kata yang bisa ditafsirkan secara semantik dan statistik [17]. Pada pendekatan ini terdapat beberapa evaluasi yang dapat diajukan, yaitu ekstrinsik dan intrinsik yang dapat dideskripsikan pada metode *Point-wise Mutual Information* (PMI) dan Umass. Namun, perhitungan evaluasi tersebut hanya dapat mengukur hubungan antar kata dan belum mampu untuk menentukan topik yang tersembunyi atau *latent topic* pada kumpulan aspek kata yang ditentukan. Jelodar melakukan survei bahwa LDA efektif mengkalkulasi rekomendasi topik ulasan teratas, namun langkah ini masih belum bisa menentukan topik tersembunyi sesungguhnya [16].

Proses penentuan topik yang tersembunyi diperlukan sebuah pendekatan, yaitu pendekatan dengan penilaian

manusia [16,18]. Namun, penilaian manusia hanya efektif jika *topic coherence* memiliki nilai yang tinggi, penilai akan mengalami kesulitan karena hubungan antar aspek kata atau *topic coherence* yang rendah, sehingga diperlukan metode lain, yaitu, identifikasi *latent topic* atau *Latent Topic Identification* melalui perhitungan komputer. Perhitungan dilakukan beberapa tahapan, yaitu, *text preprocessing*, *Rank & Score*, dan *Topic Selection* [19].

Penelitian ini bertujuan untuk mengevaluasi identifikasi *latent topic* yang dilakukan melalui penilaian manusia dan perhitungan komputer, sehingga perhitungan komputer dapat membantu manusia dalam menentukan topik tersembunyi jika setiap aspek mempunyai nilai koherensi yang rendah. Evaluasi tersebut dilakukan dengan membandingkan kedua pendekatan, penilaian manusia dan perhitungan komputer dengan beberapa pengujian, pengujian tersebut terdiri dari pengujian *Cosine Similarity* dengan *Word Embedding*, pengujian *Word2vec*, dan pengujian leksikon. Beberapa pengujian dilakukan berdasarkan ekstraksi aspek dengan sebuah dataset dengan jumlah dokumen tertentu dan metode pengembangan LDA. Hasil perbandingan menentukan kesesuaian topik pada penilaian manusia dan perhitungan komputer secara semantik.



Gambar 1. Diagram Metode Penelitian

2. Metode Penelitian

Pada penelitian ini, terdapat dua metode yang digunakan, yaitu Ekstraksi Aspek dan *Latent Topic Identification*. Ekstraksi aspek digunakan untuk memecah dan membagi kata dari review dokumen menjadi beberapa aspek dalam suatu kelompok atau *cluster*. Sedangkan *Latent Topic Identification* digunakan untuk menentukan nama topik yang sesuai dan relevan dengan aspek-aspek dari suatu kelompok atau *cluster*. Gambar 1 merupakan diagram atau abstrak yang mendeskripsikan tahapan penelitian. Sebelum melakukan ekstraksi aspek, langkah awal yang dilakukan adalah menentukan dataset yang berisi ulasan dengan topik tertentu, kemudian dilakukan pengujian ekstraksi aspek melalui proses *topic coherence*. Langkah keempat yang dilakukan pada penelitian adalah *Latent Topic Identification* yang diikuti dengan proses pengujian *word vector* dan leksikon.

2.1. Data dan Pengaturan Parameter

Data yang diambil merupakan ulasan untuk SA yang mengabaikan pelabelan data positif atau negatif. Data yang digunakan dalam penelitian ini diperoleh dari web Kaggle yang memiliki tema *e-commerce*. Dataset *e-commerce* diambil pada web kaggle.com dengan link

/nicapotato/womens-ecommerce-clothing-reviews dan mempunyai jumlah 28.222 ulasan, Pada tahapan *latent topic identification*, kamus yang diperlukan untuk mendapatkan kata benda ini diperoleh dari situs www.desiquintans.com/nounlist. Kamus ini digunakan untuk menentukan nama topik atau kategori. Untuk memvalidasinya, kata tersebut harus merepresentasikan *pos tag* “NN” yang berarti *Noun singular* dan “NNS” yang berarti *Noun plural* yang dipasang pada tahapan *text preprocessing* menggunakan proses *POS tagging*.

Selanjutnya pada pengaturan parameter seleksi fitur, nilai α adalah 0,99 yang mempunyai nilai maksimum dari *threshold*. Sedangkan dalam proses LDA, nilai *hyperparameter* α yang ditentukan adalah 0,9, dan nilai *hyperparameter* β yang ditentukan adalah 0,05. Sementara itu, proses iterasi LDA yang ditentukan adalah 10. Pada tahapan evaluasi ekstraksi aspek, parameter *topic coherence* γ yang ditentukan adalah 0,99. Parameter yang ditentukan berperan sebagai indikator dari hasil perhitungan yang optimal.

2.2. Aspect Extraction

Pada ekstraksi aspek, terdapat beberapa metode ekstraksi yang dilakukan, yaitu, LDA, LDA- *Term Frequency- Inverse Document Frequency* (TFIDF), LDA-GloVe, dan BTM. LDA merupakan versi Bayesian dari pLSA dan memiliki tipe model statistik generatif yang memungkinkan setiap dokumen terdiri dari campuran topik kecil yang tidak terlihat dan setiap kehadiran kata berkontribusi pada salah satu topik dokumen [15]. LDA menggunakan *dirichlet priors* untuk distribusi dokumen-topik dan kata-topik, *dirichlet* dianggap sebagai distribusi dokumen-topik terhadap distribusi kata-topik. Metode ini menggunakan model *Collapsed Gibbs Sampling*. LDA-TFIDF merupakan gabungan dari LDA dan TFIDF berperan dalam proses seleksi fitur [20]. LDA-GloVe merupakan gabungan dari LDA dan GloVe yang bertugas sebagai *word vector* atau mengubah kata menjadi angka, GloVe juga mempunyai peran untuk menghitung probabilitas kata yang muncul dengan kata lain berdasarkan *co-occurrence matrix* [21]. BTM merupakan metode *topic modelling* yang dirancang untuk teks pendek, seperti review, menggunakan dua kata atau *bi-term* [12].

Pada metode penelitian ekstraksi aspek, secara umum terdapat beberapa tahapan untuk dilakukan, yaitu, proses pengumpulan data yang diperoleh dari dataset, proses *text preprocessing* yang terdiri dari tokenisasi, penghapusan *stopword* dan tanda baca, dan penyaringan kata benda, proses *word embedding*, proses seleksi fitur, proses LDA yang menggunakan *hyperparameter* α dan β yang telah ditentukan, dan proses evaluasi. Pada proses seleksi fitur, terdapat beberapa langkah, yaitu, proses mencari persamaan kata yang menggunakan *cosine similarity*, proses mencari titik *cluster* kata, dan menyeleksi kata dengan *threshold*. *Threshold* dapat

dikalkulasi menggunakan *Cosine Similarity Threshold* (CST) [22] dan diformulasikan seperti rumus 1.

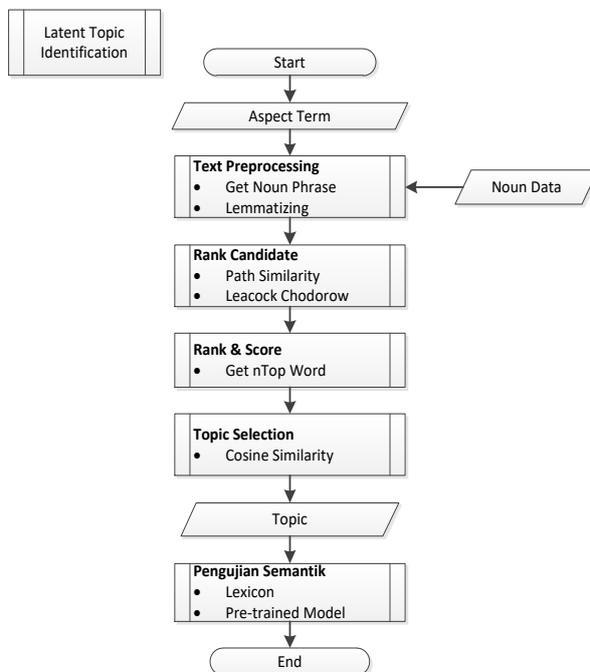
$$CST = avg(CS) + \alpha * std(CS) \quad (1)$$

$$p(z_i = k | w_i, d_i, \alpha, \beta, \cdot) \propto \frac{n_{kw_i} + \beta}{\sum_{w=1}^W n_{kw} + W\beta} \cdot \frac{n_{d_i k} + \alpha}{\sum_{k'=1}^K n_{d_i k'} + K\alpha} \quad (2)$$

Variabel *CS* adalah *Cosine Similarity*, *avg* adalah rata-rata dari nilai *cosine similarity*, *std* merupakan standar deviasi, dan α merupakan nilai parameter yang telah ditentukan, batas parameter α antara 0,01 dan 0,99

Sedangkan pada proses LDA dapat diformulasikan dengan rumus 2. Variabel n_{kw_i} merepresentasikan jumlah kata w yang muncul pada topik t , sedangkan $\sum_{w=1}^W n_{kw}$ merepresentasikan jumlah kata pada topik t . $n_{d_i k}$ merupakan representasi dari jumlah dokumen d pada topik t , sedangkan $\sum_{k'=1}^K n_{d_i k'}$ merepresentasikan total seluruh dokumen t dalam semua topik. W adalah panjang dari kamus kata. K merupakan jumlah topik. α dan β merupakan *Dirichlet prior hyperparameter*. Sementara itu, pada proses evaluasi terdapat beberapa langkah, pengujian ekstrinsik yang direpresentasikan sebagai PMI yang erat kaitannya dengan probabilitas bersyarat dan frekuensi kata, dan pengujian intrinsik yang direpresentasikan sebagai Umass yang bertugas untuk membandingkan kata sebelumnya dengan kata berikutnya. Secara umum, proses evaluasi ini digunakan untuk mengukur nilai semantik antar aspek kata dalam suatu kelompok topik atau biasa disebut *topic coherence*.

2.3. Latent Topic Identification



Gambar 2. Diagram Alir Identifikasi *Latent Topic*

Setelah melakukan metode ekstraksi aspek dan evaluasi *topic coherence*, langkah selanjutnya yang dilakukan adalah proses *Latent Topic Identification*. Langkah ini bertujuan untuk menentukan kata topik atau kategori dari kumpulan kata berdasarkan proses sebelumnya. Pada langkah ini terdapat dua cara, yaitu cara subjektif dan objektif. Cara subjektif berdasarkan penilaian manusia, cara ini dilakukan berdasarkan asumsi manusia terhadap kumpulan aspek yang telah diperoleh. Cara objektif berdasarkan perhitungan dan penilaian komputer. Gambar 2 merupakan diagram alir tahapan identifikasi *latent topic* berdasarkan perhitungan komputer.

Langkah pertama yang dilakukan adalah *text preprocessing*, terdapat dua langkah pada tahapan ini, yaitu, *Get Noun Phrase* dan *Lemmatizing*. Secara umum, langkah ini bertujuan untuk menyaring kata yang mempunyai jenis kata benda. Langkah selanjutnya adalah *Rank Candidate*, pada langkah ini terdapat dua metode yang dilakukan, yaitu, *Path Similarity* dan *Leacock Chodorow*, secara umum kedua metode tersebut berguna untuk mencari kata hipernim. Setelah *Rank Candidate* dilakukan, langkah selanjutnya adalah *Rank & Score*, langkah ini berguna untuk mendapatkan beberapa kata dengan nilai tertinggi dari nilai rata-rata proses sebelumnya. Terakhir adalah langkah *Topic Selection*, langkah ini digunakan untuk mencari dan mendapatkan kata terdekat di antara kandidat kata yang diperoleh dari proses sebelumnya. Langkah ini didapatkan dari mencari nilai *cosine similarity* tertinggi.

2.4. Skenario Pengujian

Setelah mendapatkan topik, langkah selanjutnya adalah melakukan dan menentukan pengujian. Pengujian dilakukan untuk mengetahui kinerja dari metode identifikasi *latent topic* yang diusulkan. Pada pengujian ini terdapat 6 jenis pengujian yang dilakukan, yaitu, pengujian *cosine similarity* GloVe, *cosine similarity FastText*, *Word2Vec*, *Wordnet*, *Lin Similarity*, dan *Path Similarity*. Keenam pengujian ini dilakukan dengan menggunakan dua langkah, yaitu, penilaian manusia dan perhitungan komputer. Pengujian ini juga akan dilakukan pada dataset *e-Commerce* yang dipilih berdasarkan ekstraksi aspek yang dilakukan pada metode pengembangan LDA sebelumnya.

Pengujian *cosine similarity* GloVe dan *FastText* merupakan pengujian semantik yang menggunakan metode *cosine similarity* dengan mengkonversikan kata menjadi *word vector* atau angka dari *pre-trained model* GloVe dan *FastText*, pengujian ini menggunakan *word co-occurrence matrix* dari topik dengan aspek sebagai media perhitungan. Berikut algoritme pengujian *cosine similarity* yang memiliki konsep sama pada pengujian lainnya yang direpresentasikan pada *pseudocode* pengujian *cosine similarity*. Pertama, mengubah kata topik dan aspek menjadi *vector*, selanjutnya menghitung

kedua kata topik dan aspek ke dalam rumus *cosine similarity*. Kemudian cari nilai rata-rata dari *cosine similarity*. Sedangkan nilai input pada *pseudocode* ini adalah topik dan daftar aspek-aspek setiap topik. Selain itu, nilai output *pseudocode* ini adalah nilai rata-rata pengujian.

Pengujian Cosine Similarity

```
Input: topic, list aspect
Output: average test
Initialization k, i
Get model_vec, cos_similarity, mean
for k, word_aspect in list aspect do
    vector_word[k] = model_vec[word_aspect]
end for
vector_topic = model_vec[topic]
for i, cosine_word in vector_word do
    word[i] = cos_similarity(vector_topic,
cosine_word)
    average_test = mean(word[i])
end for
```

Pengujian *word2vec* merupakan pengujian semantik yang menggunakan GloVe sebagai *word vector* dan mengkonversikannya menjadi *word2vec* yang diperoleh dari Google, pengujian ini hampir mirip dengan pengujian *cosine similarity*. Selanjutnya, pengujian *wordnet* merupakan pengujian yang bertujuan untuk mencari dan menemukan persamaan kata berdasarkan leksikon. Langkah pertama yang dilakukan pada algoritme *wordnet* adalah mengkonversikan kata topik dan aspek menjadi *synset*, kemudian menghitung kedua kata ke dalam rumus *wordnet*, langkah terakhir adalah cari nilai rata-rata dari *wordnet* secara keseluruhan. Pengujian ini memiliki beberapa rumus *wordnet* yang berbeda-beda, salah satunya adalah *wup similarity*. *Path similarity* dilakukan untuk menghitung jalur terpendek antara hipernim dan hiponim dari kedua kata topik dan aspek. *Path similarity* mempunyai algoritme yang mirip dengan *wordnet*, yaitu mengkonversikan kata aspek dan topik ke dalam *synset*, kemudian menghitung nilai *synset* dari aspek dan topik ke dalam *path similarity*, dan menghitung nilai rata-rata *path similarity* secara keseluruhan. *Lin similarity* dilakukan untuk menghitung kemiripan kata topik dan aspek berdasarkan *Information Center (IC)* dari *wordnet*. Langkah pertama yang dilakukan pada algoritme *lin similarity* adalah mengubah kata aspek dan topik ke dalam *synset*, selanjutnya menghitung nilai *synset* dari aspek dan topik ke dalam *lin similarity*, langkah terakhir yang dilakukan adalah menghitung nilai rata-rata dari *lin similarity*.

Pada proses penentuan kesesuaian topik, dilakukan perhitungan selisih dari kedua nilai antara perhitungan komputer dan manusia. Perhitungan tersebut diformulasikan pada rumus 3. Pada rumus tersebut variabel V_1 merepresentasikan jumlah nilai terendah, sedangkan variabel V_2 mewakili jumlah nilai tertinggi.

$$((V_2 - V_1) / V_2) \times 100 \quad (3)$$

3. Hasil dan Pembahasan

Bab penelitian ini, dijelaskan tentang implementasi data dari proses ekstraksi aspek, identifikasi *latent topic*, dan enam pengujian dari dua cara yang dilakukan.

3.1. Implementasi Data

Pada eksperimen proses *text preprocessing* ekstraksi aspek mengambil 25000 ulasan pada setiap dataset. Pada dataset *e-commerce*, waktu komputasi proses tersebut mencapai 48,9 detik. Proses selanjutnya adalah seleksi fitur dan pengelompokan menggunakan LDA. Pada metode LDA, LDA-TFIDF, dan LDA-GloVe memiliki perhitungan seleksi fitur ekstraksi aspek yang berbeda, sehingga hasil implementasi pada setiap metode juga berbeda. Tabel 1 merupakan salah satu contoh hasil implementasi ekstraksi aspek menggunakan metode LDA-GloVe pada dataset *e-commerce* berjumlah 500 dokumen.

Tabel 1. Tabel Hasil Ekstraksi Aspek Dataset E-commerce

		Topik		
1	2	3	4	5
length	material	size	quality	color
waist	person	dress	way	sleeve
medium	issue	fit	time	cut
hip	weight	shirt	love	shoulder
shape	compliment	skirt	summer	pattern
bust	review	style	day	lace
pocket	thing	store	work	body
room	winter	jacket	online	chest

Tabel 2. Tabel Hasil Identifikasi *Latent Topic* Dataset E-commerce

Jumlah Dokumen	Hasil
500	structure, someone, shop, quality, cord
1000	chest, person, day, denim, shop
1500	cloth, shoulder, way, shop, day

Pada Tabel 1 terdapat 5 topik dan setiap kolom terdiri dari 8 aspek kata pada setiap topiknya. Pada topik 1 terdapat beberapa contoh aspek kata, seperti, “length”, “waist”, “medium”, dan “hip”, kata-kata tersebut dapat merepresentasikan hubungan pada topik kata *length*. Pada topik 2 terdapat beberapa contoh aspek kata, seperti, “person”, “issue”, “review”, dan “weight”, beberapa kata tersebut dapat mendeskripsikan hubungan pada topik kata *problem*. Pada topik 3 terdapat beberapa contoh aspek kata seperti, “dress”, “fit”, “shirt”, “size”, dan “skirt”, kata-kata tersebut dapat mendeskripsikan keterkaitan pada topik kata *dress*. Pada topik 4 terdapat contoh kata-kata seperti, “day”, “summer”, “time”, dan “work”, beberapa aspek kata tersebut dapat mendeskripsikan hubungan pada topik kata *time*. Selanjutnya pada topik 5 terdapat beberapa contoh kata seperti, “chest”, “body”, “shoulder”, dan “sleeve”, beberapa kata tersebut dapat mendeskripsikan keterkaitan pada topik kata *body*. Beberapa topik

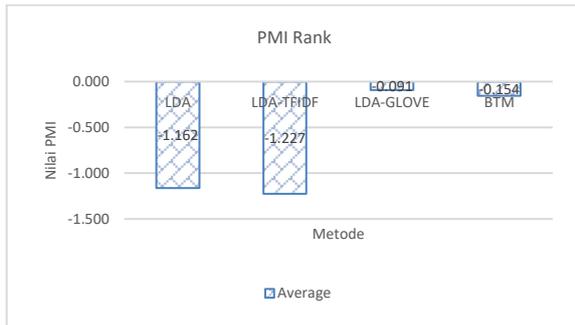
tersebut diperoleh dari identifikasi *latent topic* berdasarkan penilaian manusia.

Sedangkan Tabel 2 merupakan salah satu contoh hasil implementasi dari identifikasi *latent topic* berdasarkan perhitungan komputer dengan ekstraksi aspek metode LDA-GloVe menggunakan 500, 1000, dan 1500 dokumen dataset *e-commerce*. Pada dokumen 500 terdapat 5 topik, topik 1 mempunyai beberapa aspek kata yang dapat mendeskripsikan hubungan pada topik kata *structure*. Topik 2 memiliki beberapa aspek kata yang mendeskripsikan hubungan pada topik kata *someone*. Topik 3 mempunyai beberapa aspek kata yang dapat mendeskripsikan hubungan pada topik kata *shop*. Topik 4 mempunyai beberapa aspek kata yang dapat mendeskripsikan hubungan pada topik kata *quality*. Sedangkan topik 5 mempunyai beberapa aspek kata yang dapat mendeskripsikan hubungan pada topik kata *cord*.

Selanjutnya, pada dokumen 1000 juga mempunyai 5 topik yang mempunyai keterkaitan pada beberapa aspek kata. Topik 1 mendeskripsikan hubungan pada topik kata *chest*. Topik 2 mendeskripsikan hubungan pada topik kata *person*. Topik 3 mendeskripsikan hubungan pada topik kata *day*. Topik 4 juga dapat mendeskripsikan hubungan pada topik kata *denim*. Sedangkan topik 5 mendeskripsikan hubungan pada topik kata *shop*. Terakhir, pada dokumen 1500 juga memiliki 5 topik yang mempunyai keterkaitan pada beberapa aspek kata. Topik 1 dapat mendeskripsikan hubungan pada topik kata *cloth*. Topik 2 mendeskripsikan hubungan pada topik kata *shoulder*. Topik 3 mendeskripsikan hubungan pada topik kata *way*. Topik 4 mendeskripsikan hubungan pada topik kata *shop*. Sedangkan topik 5 juga dapat mendeskripsikan hubungan pada topik kata *day*.

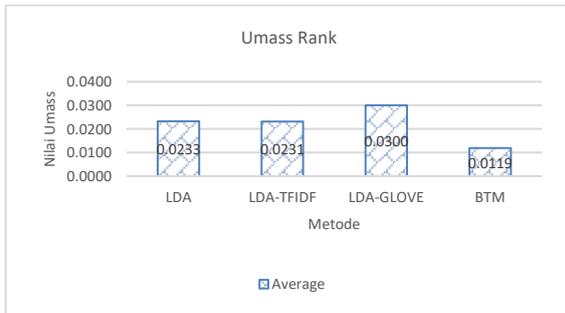
3.2. Evaluasi Data

Selanjutnya, tahapan evaluasi dijelaskan secara singkat tentang *topic coherence* ekstraksi aspek yang diperoleh dari metode PMI yang merepresentasikan pengujian ekstrinsik dan Umass yang merepresentasikan pengujian intrinsik, sedangkan tahapan ini juga dijelaskan lebih detail tentang keenam pengujian dari identifikasi *latent topic* yang dilakukan berdasarkan penilaian manusia dan perhitungan komputer, pengujian tersebut adalah pengujian *cosine similarity* GloVe, pengujian *cosine similarity FastText*, pengujian *word2vec*, pengujian *wordnet*, pengujian *lin similarity*, dan pengujian *path similarity*. Beberapa pengujian tersebut dikalkulasikan menggunakan metode LDA dasar, LDA-TFIDF, LDA-GloVe, dan BTM. Jumlah dokumen pengujian ini menggunakan 500, 1000, dan 1500 dokumen pada dataset *e-commerce*. Jumlah topik juga ditentukan hanya 5 dan 7 topik saja.



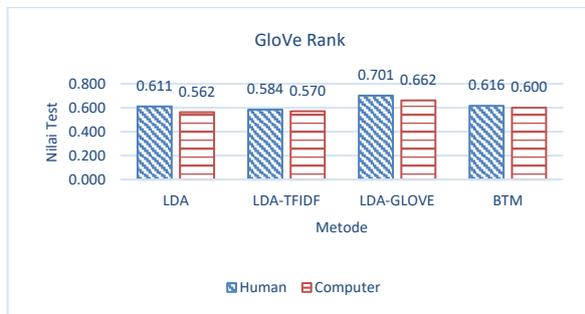
Gambar 3. Hasil Pengujian PMI Topic Coherence

Gambar 3 merupakan grafik hasil pengujian PMI *topic coherence* seluruh dataset. Pengujian ini menggunakan grafik batang dengan nilai PMI dan metode sebagai axisnya. Hasil menunjukkan bahwa metode LDA-GloVe memiliki nilai PMI tertinggi sebesar -0,091. Sedangkan metode LDA-TFIDF mempunyai nilai PMI paling rendah sebesar -1,227. Hasil tersebut merupakan nilai rata-rata dari *topic coherence* ekstraksi aspek berjumlah 5 dan 7 topik.



Gambar 4. Hasil Pengujian Umass Topic Coherence

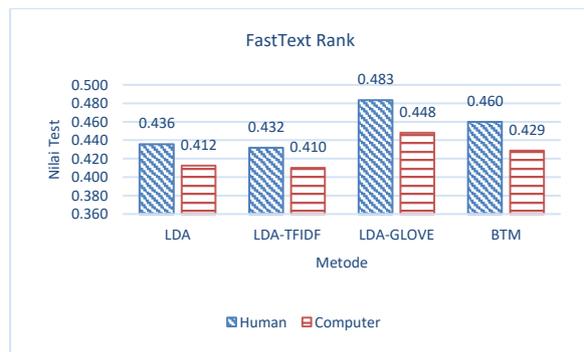
Gambar 4 merupakan grafik hasil pengujian Umass *topic coherence* seluruh dataset. pengujian ini juga menggunakan grafik batang dengan axis nilai Umass dan metode. Hasil pengujian menunjukkan bahwa metode LDA-GloVe mempunyai nilai tertinggi sebesar 0,03. Sedangkan metode BTM memiliki nilai terendah sebesar 0,0119. Hasil tersebut juga merupakan nilai rata-rata *topic coherence* ekstraksi aspek berjumlah 5 dan 7 topik.



Gambar 5. Hasil Pengujian Cosine Similarity GloVe

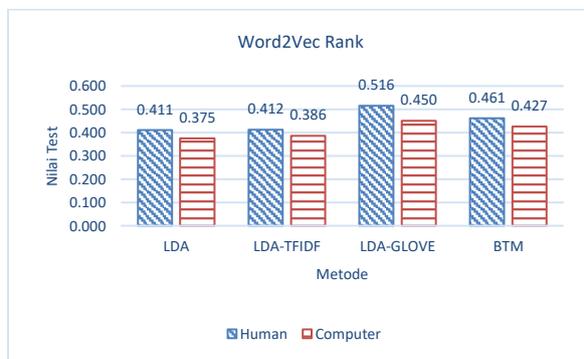
Selanjutnya pada identifikasi *latent topic*, Gambar 5 merupakan grafik hasil pengujian *cosine similarity*

dengan *word embedding* GloVe pada seluruh dataset. pengujian ini menggunakan grafik batang dengan nilai *test* dan metode sebagai axisnya. Hasil menunjukkan bahwa secara umum metode LDA-GloVe memiliki nilai pengujian tertinggi, sedangkan metode LDA-TFIDF mempunyai nilai pengujian paling rendah. Hasil juga menunjukkan bahwa nilai perbandingan penilaian manusia dan perhitungan komputer dari metode LDA adalah 0,611 dan 0,562 atau selisih 8,09%. Sedangkan nilai perbandingan metode LDA-TFIDF adalah 0,584 dan 0,57 atau selisih 2,41%. Metode LDA-GloVe mempunyai nilai perbandingan 0,701 dan 0,662 atau selisih 5,51%. Sementara metode BTM mempunyai nilai perbandingan dengan selisih 2,5%.



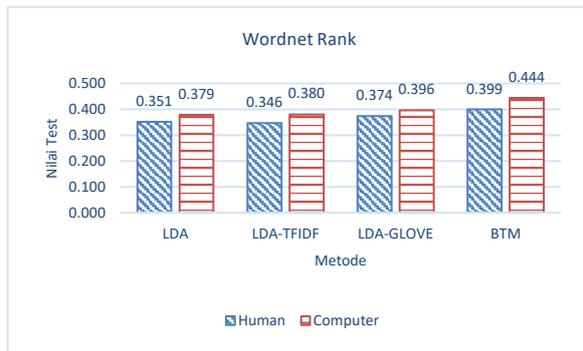
Gambar 6. Hasil Pengujian Cosine Similarity FastText

Gambar 6 merupakan grafik hasil pengujian *cosine similarity* dengan *word embedding* FastText dari Facebook pada seluruh dataset, pengujian ini menggunakan grafik batang dengan axis nilai *test* dan metode. Hasil menunjukkan bahwa secara umum metode LDA-GloVe mempunyai nilai pengujian paling tinggi, sedangkan metode LDA-TFIDF memiliki metode pengujian terendah. Hasil juga menunjukkan bahwa nilai perbandingan antara penilaian manusia dan perhitungan komputer metode LDA adalah 0,436 dan 0,412 atau selisih 5,35%. Sedangkan nilai perbandingan metode LDA-TFIDF adalah 0,432 dan 0,41 atau selisih 5,01%. Metode LDA-GloVe mempunyai nilai perbandingan 0,483 dan 0,448 atau selisih 7,31%. Sementara metode BTM mempunyai nilai perbandingan 0,46 dan 0,429 atau selisih 6,81%.



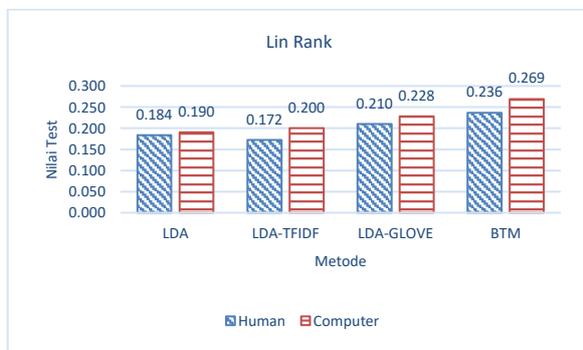
Gambar 7. Hasil Pengujian Word2vec

Gambar 7 merupakan grafik hasil pengujian *word2vec* yang diperoleh dari Google, pengujian ini menggunakan grafik batang dengan nilai *test* dan metode sebagai axisnya. Secara umum, hasil menunjukkan bahwa metode LDA-GloVe mempunyai nilai pengujian tertinggi, sedangkan metode LDA dasar juga mempunyai nilai pengujian terendah. Hasil juga menunjukkan bahwa nilai perbandingan antara penilaian manusia dan komputer pada metode LDA adalah 0,411 dan 0,375 atau selisih 8,8%. Sedangkan metode LDA-TFIDF mempunyai nilai perbandingan 0,412 dan 0,386 atau selisih 6,29%. Nilai perbandingan LDA-GloVe adalah 0,516 dan 0,45 atau selisih 12,73%. Sementara metode BTM mempunyai nilai perbandingan 0,461 dan 0,427 atau selisih 7,52%.



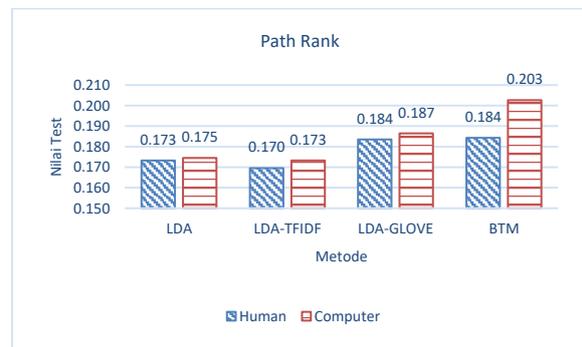
Gambar 8. Hasil Pengujian Wordnet

Gambar 8 merupakan grafik hasil pengujian *wordnet* sebagai informasi relasi semantik antar *synset* pada seluruh dataset, pengujian ini juga menggunakan grafik batang dengan axis nilai *test* dan metode. Secara umum, hasil menunjukkan bahwa metode BTM mempunyai nilai pengujian tertinggi, sedangkan metode LDA-TFIDF mempunyai nilai pengujian terendah. Hasil tersebut juga menunjukkan bahwa nilai perbandingan antara penilaian manusia dan komputer pada metode LDA adalah 0,351 dan 0,379 atau selisih 7,35%. Sedangkan nilai perbandingan metode LDA-TFIDF adalah 0,346 dan 0,38 atau selisih 9%. Metode LDA-GloVe mempunyai nilai perbandingan 0,374 dan 0,396 atau selisih 5,5%. Metode BTM memiliki nilai perbandingan 0,399 dan 0,44 atau selisih 10,12%.



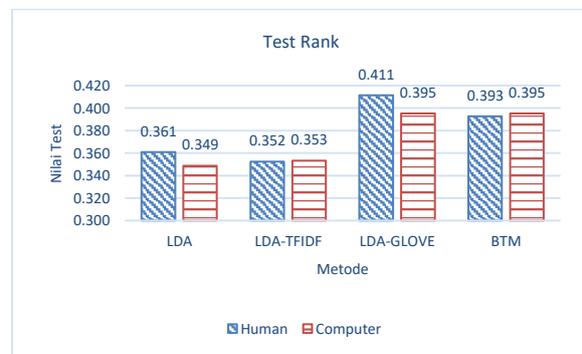
Gambar 9. Hasil Pengujian Lin Similarity

Selanjutnya, Gambar 9 merupakan grafik hasil pengujian *lin similarity* sebagai relasi semantik berdasarkan node secara spesifik berdasarkan IC, pengujian ini juga menggunakan grafik batang dengan nilai *test* dan metode sebagai axisnya. Secara umum, hasil menunjukkan bahwa metode BTM memiliki nilai pengujian tertinggi, sedangkan metode LDA-TFIDF memiliki nilai pengujian terendah. Hasil juga menunjukkan bahwa nilai perbandingan antara penilaian manusia dan komputer pada metode LDA adalah 0,184 dan 0,175 atau selisih 3,41%. Sedangkan metode LDA-TFIDF mempunyai nilai perbandingan 0,172 dan 0,2 atau selisih 14,1%. Nilai perbandingan metode LDA-GloVe adalah 0,21 dan 0,228 atau selisih 8,03%. Sementara metode BTM mempunyai nilai perbandingan 0,236 dan 0,269 atau selisih 12,3%.



Gambar 10. Hasil Pengujian Path Similarity

Gambar 10 merupakan grafik hasil pengujian *path similarity* sebagai relasi semantik hubungan terpendek antara hiponim dan hipernim pada seluruh dataset, pengujian ini juga menggunakan grafik batang dengan axis metode dan nilai *test*. Secara umum, hasil menunjukkan bahwa BTM mempunyai nilai pengujian tertinggi, sedangkan LDA-TFIDF memiliki nilai pengujian terendah. Hasil tersebut juga menunjukkan bahwa nilai perbandingan antara penilaian manusia dan komputer pada metode LDA adalah 0,173 dan 0,175 atau selisih 0,69%. Sedangkan nilai perbandingan metode LDA-TFIDF adalah 0,17 dan 0,173 atau selisih 2,17%. LDA-GloVe memiliki nilai perbandingan 0,184 dan 0,187 atau selisih 1,55%. Metode BTM memiliki nilai perbandingan 0,184 dan 0,203 atau selisih 9,05%.



Gambar 11. Hasil Pengujian Secara Keseluruhan

Secara keseluruhan, nilai rata-rata selisih semua pengujian adalah 6,7%. Sedangkan nilai rata-rata perbandingan pada seluruh pengujian antara penilaian manusia dan komputer dapat dideskripsikan pada Gambar 11. Gambar tersebut merupakan grafik seluruh pengujian pada seluruh dataset yang menggunakan grafik batang dengan axis metode dan nilai *test*. Secara umum, hasil menunjukkan bahwa metode LDA-GloVe mempunyai nilai pengujian tertinggi, sedangkan metode LDA-TFIDF memiliki nilai pengujian terendah. Hasil tersebut juga menunjukkan bahwa nilai perbandingan antara penilaian manusia dan komputer pada metode LDA adalah 0,361 dan 0,349 atau selisih 3,38%. Sedangkan nilai perbandingan metode LDA-TFIDF adalah 0,352 dan 0,353 atau selisih 0,22%. Metode LDA-GloVe memiliki nilai perbandingan 0,411 dan 0,395 atau selisih 3,92%. Metode BTM mempunyai nilai perbandingan 0,393 dan 0,395 atau selisih 0,63%.

Dari penjelasan beberapa pengujian dapat membuktikan bahwa nilai perbandingan penilaian manusia memiliki relevansi terhadap perhitungan komputer dengan selisih nilai antara 0,6% dan 14%, sedangkan nilai rata-rata selisih kedua perbandingan tersebut hanya sekitar 6%. Walaupun demikian, penilaian manusia masih memiliki nilai yang lebih tinggi daripada perhitungan komputer.

4. Kesimpulan

Berdasarkan penelitian dan evaluasi identifikasi latent topic yang dihasilkan dari ekstraksi aspek menunjukkan bahwa dari beberapa pengujian nilai perbandingan identifikasi melalui perhitungan komputer masih relevan dalam membantu manusia untuk menentukan topik jika setiap aspek memiliki nilai koherensi yang rendah, dan memiliki kesesuaian topik yang tinggi terhadap identifikasi melalui penilaian manusia, walaupun hasil penilaian manusia masih lebih unggul daripada perhitungan komputer. Hasil tersebut juga membuktikan bahwa metode LDA-GloVe mempunyai beberapa nilai *topic coherence*, pengujian *word vector*, dan pengujian leksikon yang cenderung stabil daripada metode lainnya dengan jarak selisih antara 0,8 dan 8,4. Pada penelitian selanjutnya, peneliti diharapkan mengembangkan jumlah dokumen dan mencoba dataset ulasan lain. Peneliti juga dapat mengembangkan proses seleksi fitur dalam ekstraksi aspek dan identifikasi *latent topic* untuk kesesuaian topik yang lebih maksimal.

Daftar Rujukan

- [1] S. Poria, E. Cambria, and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network," *Knowledge-Based Syst.*, vol. 108, pp. 42–49, Sep. 2016, doi: 10.1016/j.knosys.2016.06.009.
- [2] M. Tubishat, N. Idris, and M. Abushariah, "Explicit aspects extraction in sentiment analysis using optimal rules combination," *Futur. Gener. Comput. Syst.*, vol. 114, pp. 448–480, Jan. 2021, doi: 10.1016/j.future.2020.08.019.

- [3] B. Liu, *Sentiment analysis: mining opinions, sentiments, and emotions*. New York: Cambridge University Press, 2015.
- [4] T. A. Rana and Y.-N. Cheah, "Aspect extraction in sentiment analysis: comparative analysis and survey," *Artif. Intell. Rev.*, vol. 46, no. 4, pp. 459–483, Dec. 2016, doi: 10.1007/s10462-016-9472-z.
- [5] A. S. Manek, P. D. Shenoy, M. C. Mohan, and V. K. R., "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier," *World Wide Web*, vol. 20, no. 2, pp. 135–154, Mar. 2017, doi: 10.1007/s11280-015-0381-x.
- [6] M. Shams and A. Baraani-dastjerdi, "Enriched LDA (ELDA): combination of latent Dirichlet allocation with word co-occurrence analysis for aspect extraction," *Expert Syst. Appl.*, vol. 80, pp. 136–146, 2017, doi: 10.1016/j.eswa.2017.02.038.
- [7] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, Eds., *A Practical Guide to Sentiment Analysis*, vol. 5. Cham: Springer International Publishing, 2017.
- [8] L. Chen, J. Martineau, D. Cheng, and A. Sheth, "Clustering for Simultaneous Extraction of Aspects and Features from Reviews," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 789–799, doi: 10.18653/v1/N16-1093.
- [9] C. Wu, F. Wu, S. Wu, Z. Yuan, and Y. Huang, "A hybrid unsupervised method for aspect term and opinion target extraction," *Knowledge-Based Syst.*, vol. 148, pp. 66–73, 2018, doi: 10.1016/j.knosys.2018.01.019.
- [10] Y. Rubtsova and S. Koshelnikov, "Aspect Extraction from Reviews Using Conditional Random Fields," in *Knowledge Engineering and Semantic Web*, 2015, pp. 158–167, doi: 10.1007/978-3-319-24543-0.
- [11] Y. Yang, C. Chen, and F. S. Bao, "Aspect Extraction from Product Reviews Using Category Hierarchy Information," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 2017, vol. 2, pp. 675–680.
- [12] X. Yan, J. Guo, Y. Lan, and X. Cheng, "A Biterm Topic Model for Short Texts," in *International World Wide Web Conference Committee*, 2013, pp. 1445–1455, doi: 10.1145/2488388.2488514.
- [13] D.-H. Pham and A.-C. Le, "Exploiting multiple word embeddings and one-hot character vectors for aspect-based sentiment analysis," *Int. J. Approx. Reason.*, vol. 103, pp. 1–10, Dec. 2018, doi: 10.1016/j.ijar.2018.08.003.
- [14] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "An Unsupervised Neural Attention Model for Aspect Extraction," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2017, pp. 388–397.
- [15] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
- [16] H. Jelodar *et al.*, "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimed. Tools Appl.*, vol. 78, no. 11, pp. 15169–15211, Jun. 2019, doi: 10.1007/s11042-018-6894-4.
- [17] K. Stevens, P. Kegelmeyer, D. Andrzejewski, and D. Buttler, "Exploring Topic Coherence over many models and many topics," in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 2012, no. July, pp. 952–961.
- [18] N. Aletras and M. Stevenson, "Evaluating Topic Coherence Using Distributional Semantics," in *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Long Papers*, 2013, pp. 13–22.

- [19] D. Nolasco and J. Oliveira, "Mining social influence in science and vice-versa: A topic correlation approach," *Int. J. Inf. Manage.*, vol. 51, p. 102017, Apr. 2020, doi: 10.1016/j.ijinfomgt.2019.10.002.
- [20] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, Jul. 2018, doi: 10.1016/j.neucom.2017.11.077.
- [21] A. M. Priyatno, M. M. Muttaqi, F. Syuhada, and A. Z. Arifin, "Deteksi Bot Spammer Twitter Berbasis Time Interval Entropy dan Global Vectors for Word Representations Tweet's Hashtag," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 5, no. 1, p. 37, Jan. 2019, doi: 10.26594/register.v5i1.1382.
- [22] A. Panchenko *et al.*, "A Graph-Based Approach to Skill Extraction from Text," in *Proceedings of TextGraphs-8 Graph-based Methods for Natural Language Processing*, 2013, pp. 79–87.