

# Topic Modeling Pada Abstrak Skripsi Menggunakan Metode Latent Semantic Analysis

<sup>1</sup> Rifqi Hakim,<sup>2</sup>Khalid, <sup>3</sup>Dwi Rolliawati

Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Ampel<sup>1,2,3</sup>

Jl. Ahmad Yani No.117, Jemur Wonosari, Kec. Wonocolo, Kota Surabaya, Jawa Timur <sup>1,2,3</sup>

rifqihakim5889@gmail.com<sup>1</sup>, khalid@uinsby.ac.id<sup>2</sup>, dwi\_roll@uinsby.ac.id<sup>3</sup>

*Abstract - Thesis is the final research for undergraduate students. With the increasing number of thesis documents, information will be formed from the collection of documents. This research was conducted to determine the topic modeling and topic trend analysis from the collection of thesis abstracts of the UINSA English Literature Study Program from 2014 to 2019. From 720 thesis abstract datasets, topic modeling was carried out using the Latent Semantic Analysis method which included preprocessing, term weighting, and Singular Value Decomposition calculations. . Topic Modeling resulted in 20 linguistic topics and 17 literature topics. Then on the topic trend analysis, 7 topic trends were obtained for each type of research. The research is dominated by speech act linguistic research which is included in the field of sociolinguistics. Based on the results of the analysis of the type of research compared to the real data of the type of research at the UINSA English Literature Study Program, the results of the analysis of linguistic research have 80% precision and 90% recall, while the number of literature studies has 74% precision and 57% recall, the accuracy of this type of research analysis has an average of 79%.*

**Keywords:** Topic Classification, Latent Semantic Analysis.

**Abstrak –** Skripsi merupakan penelitian akhir bagi mahasiswa strata-1. Dengan semakin bertambahnya dokumen skripsi, maka akan terbentuk informasi dari kumpulan dokumen tersebut. Penelitian ini dilakukan untuk menentukan pemodelan topik dan analisis tren topik dari kumpulan abstrak skripsi Program Studi Sastra Inggris UINSA tahun 2014 sampai 2019. Dari 720 dataset abstrak skripsi dilakukan pemodelan topik dengan metode Latent Semantic Analysis yang meliputi preprocessing, pembobotan term, dan perhitungan Singular Value Decomposition. Pemodelan Topik menghasilkan 20 topik linguistik dan 17 topik literatur. Kemudian pada analisis tren topik, diperoleh 7 tren topik untuk setiap jenis penelitian. Penelitian didominasi oleh penelitian linguistik tindak tutur yang termasuk dalam bidang sosiolinguistik. Berdasarkan hasil analisis jenis penelitian dibandingkan dengan data real jenis penelitian Program Studi Sastra Inggris UINSA, menghasilkan hasil analisis penelitian linguistik memiliki presisi 80% dan recall 90%, sedangkan jumlah penelitian literatur memiliki presisi 74% dan recall 57%, tingkat akurasi analisis jenis penelitian memiliki rata-rata 79%.

**Kata kunci:** Klasifikasi Topik, Latent Semantic Analysis.

## I. PENDAHULUAN

Dalam dunia pendidikan tinggi, skripsi atau tugas akhir mahasiswa merupakan salah satu syarat yang digunakan oleh perguruan tinggi di Indonesia untuk memperoleh gelar Sarjana Strata-1. Dalam skripsi terdapat abstrak yang berisi penjelasan singkat terhadap keseluruhan isi penelitian, dimana sebuah kalimat dalam skripsi akan diubah menjadi kalimat yang lebih ringkas [1]. Dengan melihat dokumen skripsi yang selalu bertambah setiap tahunnya, maka seharusnya terdapat informasi dari kumpulan dokumen skripsi tersebut. Namun pada umumnya jumlah informasi yang dapat disarikan dari dokumen skripsi selama ini tidak terdapat kajian lebih lanjut, walaupun jumlah dokumen skripsi semakin bertambah [2]. Sehingga, dengan adanya lulusan Sarjana Strata-1 setiap tahun dan dokumen skripsi yang terus bertambah, perlu untuk menerapkan metode text mining dalam penggalian dan pengelolaan informasi pada kumpulan dokumen skripsi tersebut [3].

Pada penelitian sebelumnya dengan judul “Topic Modelling Skripsi Menggunakan Metode Latent Dirichlet Allocation” [4] dihasilkan sejumlah topik penelitian dari dokumen skripsi tahun 2014 sampai 2019 pada Program Studi Sastra Inggris UIN Sunan Ampel Surabaya. Hal ini yang menjadi landasan dalam penelitian ini menggunakan salah satu teknik text mining yakni selain menemukan sejumlah topik penelitian juga untuk mencari informasi jumlah tren topik dalam penelitian skripsi setiap tahun.

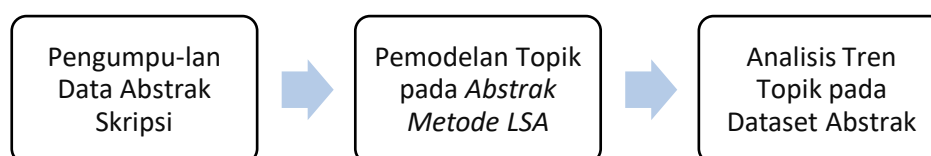
Usaha untuk pemodelan topik pada dokumen disebut ekstraksi topik. Salah satu metode untuk mengekstrak topik adalah Latent Semantic Analysis (LSA) [5]. Pada penelitian sebelumnya telah dibahas pemodelan topik dengan metode

Latent Semantic Analysis (LSA), seperti pada penelitian dengan judul “Algoritma Latent Semantic Analysis (LSA) Pada Peringkat Dokumen Otomatis untuk Proses Clustering Dokumen” memiliki tingkat akurasi clustering yang baik pada suatu dokumen dengan persentase summary document 40% menghasilkan nilai F-measure rata-rata sebesar 71.04% [6]. Kemudian pada penelitian “Klasifikasi Topik Multi Label pada Hadis Shahih Bukhari Menggunakan K-Nearest Neighbor dan Latent Semantic Analysis” mendapatkan hasil akurasi dengan kombinasi LSA-KNN sebesar 90.28% dengan waktu komputasi sebesar 19 menit 21 detik (Hidayati et al., 2020). Sedangkan pada studi kasus Program Studi Sastra Inggris UIN Sunan Ampel Surabaya, pemodelan topik dilakukan dengan metode Latent Dirichlet Allocation yang dibahas dalam penelitian “Topic Modelling Skripsi Menggunakan Metode Latent Dirichlet Allocation” dan berhasil menghasilkan sejumlah topik penelitian dari dokumen skripsi tahun 2014 sampai 2019 [4].

Dengan melihat penelitian dari studi kasus yang sama yakni pada penelitian [4], maka penelitian ini dimaksudkan untuk mengetahui informasi tren dari pemodelan topik yang terbentuk menggunakan metode LSA pada penelitian skripsi Program Studi Sastra Inggris UINSA setiap tahun antara tahun 2014 sampai 2019.

## II. METODE

*Flowchart* atau diagram alir merupakan penggambaran alur penyelesaian suatu masalah [7]. *Flowchart* pada penelitian ini secara runtut dapat dilihat pada Gambar 1, berikut:



Gambar 1. Flowchart Penelitian

## III. TAHAP PENGUMPULAN DATA

Tahap pengumpulan data, berisi proses pengumpulan data abstrak skripsi yang menjadi objek penelitian dari website <http://digilib.uinsby.ac.id>. Proses pengambilan data menggunakan salah satu tool dalam google chrome yaitu *web scrapper*. Dari proses pengumpulan data dengan teknik *web scrapping* didapatkan kumpulan teks abstrak sebagai data penelitian. Prinsip dari *web scrapper* ini adalah membentuk suatu jaring yang memuat urutan halaman dalam suatu website, yang mana jaring tersebut dapat digunakan untuk mengambil *element text* pada suatu website. Urutan dari proses pengumpulan data dengan *web scrapper* adalah sebagai berikut.

- a. Mengakses Website Target
- b. Pembuatan *Sitemap Web Scrapper*
- c. Pembuatan *Selector Web Scrapper*
- d. Eksekusi *Selector/ Scrape Data*
- e. Penyimpanan Data

## IV. TAHAP PEMODELAN TOPIK METODE LSA

Pada tahapan pemodelan topik dengan metode *Latent Semantic Analysis (LSA)* bertujuan untuk melakukan pemodelan topik. *Topic modeling* merupakan teknik untuk menyimpulkan topik yang tersembunyi dalam dokumen. *Topic modeling* mewakili setiap dokumen sebagai gabungan dari beberapa topik dan setiap topik merupakan kombinasi gabungan dari beberapa kata, maka *topic modeling* merupakan alat dalam *text mining* untuk mengklasifikasikan dokumen berdasarkan topik nya [8]. Sedangkan *Latent Semantic Analysis (LSA)* adalah suatu teori atau metode statistik aljabar yang melakukan ekstraksi struktur semantik yang tersembunyi dari kalimat berupa himpunan *term* dari dokumen [9] Tahap ini secara berurut meliputi tahap *preprocessing data, feature selection*, dan pembentukan topik dengan perhitungan *singular value decomposition*.

### A. PREPROCESSING DATA

Pada tahap ini dilakukan beberapa pengolahan pada data yang bertujuan untuk menormalkan dataset, diantaranya dengan menghilangkan karakter spesial dan mengubah semua huruf menjadi huruf kecil (*case folding*), memotong tiap kata dalam teks (*tokenizing*), mengubah kata imbuhan menjadi kata dasar (*stemming*), dan menghilangkan kata yang tidak menggambarkan isi tulisan (*filtering*). [10].

B. *FEATURE SELECTION*

Tahap Dalam proses *feature selection*, bertujuan untuk menemukan *term*/kata kunci yang menjadi ciri dari suatu dokumen yang membedakan dokumen tersebut dengan dokumen yang lain dalam satu korpus. Tahap ini memiliki peran yang penting dan signifikan dalam akurasi *text analytic*. Salah satu pendekatan yang paling umum digunakan dalam *feature selection* adalah *Term Frequency/Inverse Document Frequency (TF/IDF)* [11]. Persamaan dari TF-IDF (lihat Persamaan 1).

$$W_{i,j} = t_{f_{i,j}} \times idf = t_{f_{i,j}} \times \log\left(\frac{D}{df_i}\right) \quad (1)$$

Keterangan:

$W_{i,j}$  = Bobot kata  $t_j$  terhadap dokumen  $d_i$

$t_{f_{i,j}}$  = Jumlah kemunculan kata  $t_j$  dalam dokumen  $d_i$

$D$  = Jumlah dokumen

$df_i$  = Jumlah kemunculan kata dalam dokumen

C. *PERHITUNGAN SINGULAR VALUE DECOMPOSITION*

*Singular Value Decomposition (SVD)* sendiri adalah salah satu teknik aljabar linear yang digunakan untuk menguraikan (dekomposisi) suatu matriks menjadi tiga buah matriks baru, yaitu matriks orthogonal  $U$ , matriks diagonal  $S$ , dan Transpose matriks orthogonal  $V$ . Pembentukan topik dengan metode LSA ini didasarkan pada teknik matematika yang diberi nama *Singular Value Decomposition (SVD)* [12]. Persamaan dari SVD (lihat Persamaan 2).

$$A_{m \times n} = U_{m \times n} \cdot S_{n \times n} \cdot V_{n \times n}^T \quad (2)$$

Keterangan:

$A_{m \times n}$  = matriks  $A$  dengan nilai  $m > n$

$U_{m \times n}$  = matriks ortogonal berukuran  $m \times n$

$S_{n \times n}$  = matriks diagonal berukuran  $n \times n$ , dengan elemen matriks positif atau nol

$V_{n \times n}^T$  = matriks ortogonal berukuran  $n \times n$ , yang merupakan transpose matriks  $V$

D. *TAHAP ANALISIS TREN TOPIK*

Tahap ini berisi proses pemilahan topik yang terbentuk sesuai dengan jenis penelitian yang ada pada Program Studi Sastra Inggris UINSA. Serta proses perhitungan jumlah dokumen yang diklasifikasikan sesuai topik yang terbentuk. Jumlah dari klasifikasi topik tersebut kemudian diambil topik yang memiliki tren paling tinggi dan divisualisasikan hasilnya. Visualisasi dibuat dalam bentuk grafik yang dapat menunjukkan informasi statistik topik yang diangkat dalam penelitian skripsi mahasiswa Program Studi Sastra Inggris UIN Sunan Ampel Surabaya.

V. *HASIL DAN PEMBAHASAN*

A. *HASIL PENGUMPULAN DATA*

Proses dari pengumpulan data berlangsung sekitar 45 menit, dan Dari proses pengambilan data berhasil dikumpulkan berjumlah 720 dataset abstrak skripsi Program Studi Sastra Inggris UINSA. Dengan rincian data setiap tahunnya dapat dilihat pada Tabel 1 berikut.

Tabel 1. Jumlah Dataset Abstrak Per Tahun

Tahun	Jumlah Abstrak
2014	71
2015	165
2016	152
2017	120
2018	95
2019	117

Hasil dari pengumpulan data secara keseluruhan berjumlah 720 dataset abstrak. Dengan jumlah dataset tiap tahun seperti pada Tabel 1.

B. *HASIL PEMODELAN TOPIK*

1. *Preprocessing Data*

Hasil dari *preprocessing data* yang terdiri dari menghilangkan karakter spesial dan mengubah semua huruf menjadi huruf kecil (*case folding*), memotong tiap kata dalam teks (*tokenizing*), mengubah kata imbuhan menjadi kata dasar (*stemming*), dan menghilangkan kata yang tidak menggambarkan isi tulisan (*filtering*).

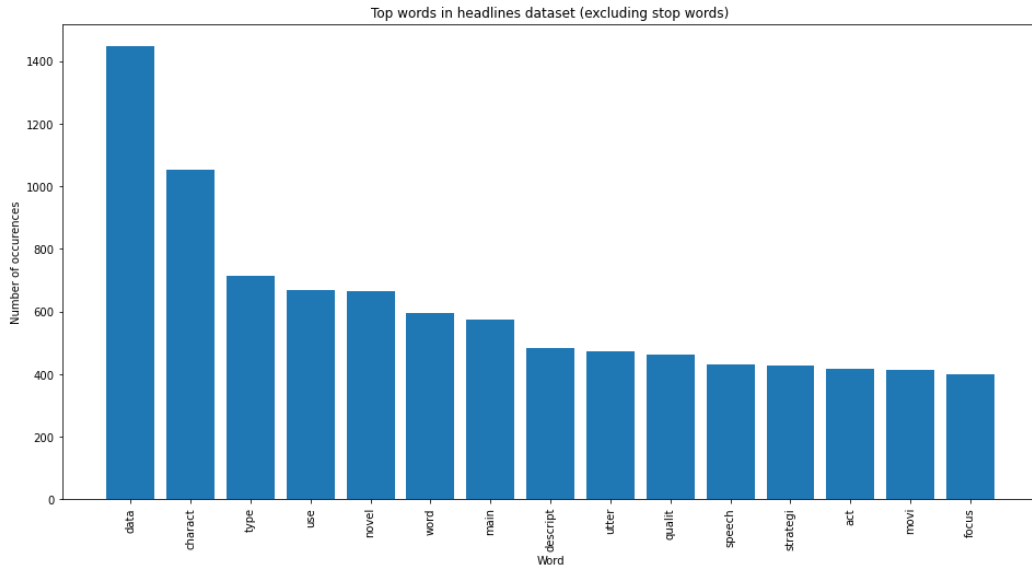
Tabel 2. Output Preprocessing Data

Sebelum Preprocessing	Setelah Preprocessing
In this study, the author discusses the "swear words" used by characters in a film by Ben Stiller\'s Tropic Thunder. This study consisted of two discussion of the problem, namely: a variety of swear words, and functions as well as the reasons for the use of swear words. The author uses descriptive research method because the author presents the results of the analysis in the form of descriptive text. The steps of this research is to categorize words that contain swear words, analyze the functions and reasons use of swear words. The results of this analysis, the authors found many swear words used by a character belonging to a "sex-term". There are four functions used by the characters in the use of swear words that expletive expression, abusive expression, humorous and swear words as an auxiliary function. Based on analysis of the most frequently used functions are functions abusive. And the reasons used in the use of swear words is due to psychological factors and social factors. Psychology is the reason that often appear based on the analysis results.	author discuss swear word use charact film ben stiller tropic thunder consist two discuss problem name varieti swear word function well reason use swear word author use descript author present result form descript text step categor word contain swear word function reason use swear word result author mani swear word use charact belong sexterm four function use charact use swear word explet express abus express humor swear word auxiliari base frequent use function function abus reason use use swear word due psycholog factor social factor psycholog reason often appear base result

Dari hasil *preprocessing data* Tabel 2 menghasilkan *dataset* yang tidak mengandung karakter seperti simbol dan tanda baca, *stopwords* (kata umum dalam jumlah besar yang tidak memiliki makna), serta kata imbuhan.

2. Feature Selection

Hasil dari feature selection adalah setiap kata dalam dataset diberi bobot sesuai dengan Persamaan 1. Hasil kata yang telah diberi bobot pada 15 kata tertinggi sebagai berikut. Gambar 1.



Gambar 2. Reranking 15 Kata Tertinggi

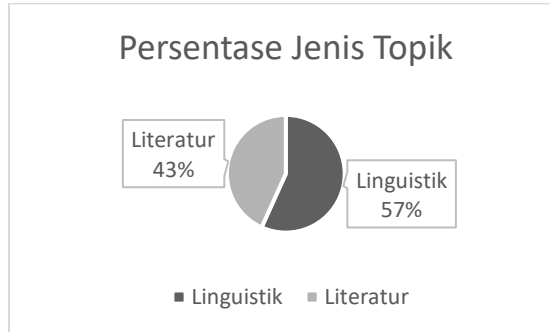
Sumbu x (horizontal) menunjukkan daftar 15 kata tertinggi yang divisualisasikan. Dan sumbu y (vertikal) menunjukkan jumlah kemunculan kata dalam dokumen.

3. Perhitungan *Singular Value Decomposition*

Hasil dari *singular value decomposition* adalah setiap kata dalam dataset dihitung sesuai dengan Persamaan 2. Hasil dari perhitungan ini membentuk 3 pasangan kata yang disusun dalam tiap topik. Hasil 3 pasangan kata yang merepresentasikan topik pada dataset yaitu; *conjunct function type, persuas utter speech, word swear javanes, featur catalog letter, code switch mix, slogan metaphor type, illocutionari act hamm, great wisdom*

*imply, interrupt talk particip, children sensit phonology, conjunct max sandelin, trump structur speech, pond beast technique, strategi refus polit, cope swango delus, violat data maxim, grammat cohes type, flout percentag maxim, strategi polit posit, brick data discuss, data polit movi, restaur food fast, carri susan reveng, grenouil psychosi alter, gossip charact male, masculin henchard character, georg milton ranch, ambit marchmil ella, guilt creasi charact, islam sharia reject, differ movi effort, choic main charact, roper pine girlfriend, challeng august new, russel pickett daisi, ashputtel cinder concept, siddhartha existenti process.*

Daftar topik yang terbentuk berjumlah 37 topik dengan presentase perbandingan jenis topik berdasarkan jenis penelitian di Program Studi Sastra Inggris UINSA sebagai berikut. Gambar 2.



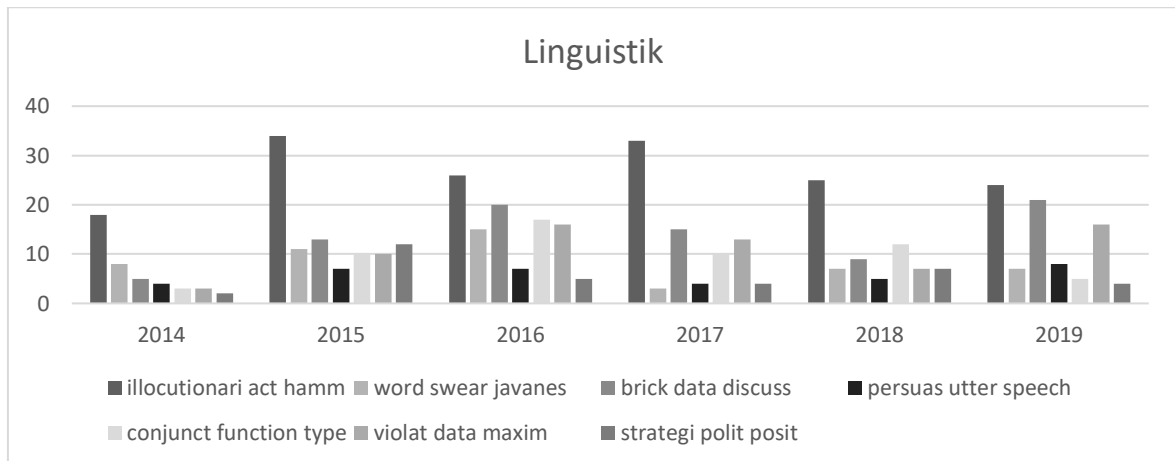
Gambar 3. Presentase Jenis Topik

Dari hasil penentuan jenis topik berdasarkan jenis penelitian di Program Studi Sastra Inggris UINSA terdiri dari 16 topik literatur dan 21 topik linguistik dengan presentase seperti pada Gambar 3. Presentase Jenis Topik.

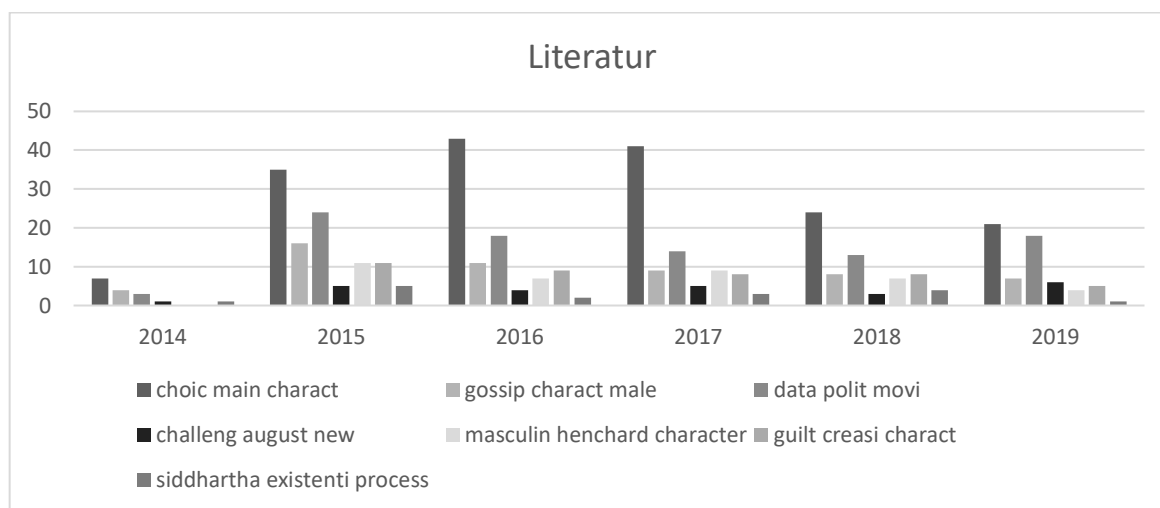
4. Analisis Tren Topik

Topik yang telah terbentuk digunakan untuk membaca setiap kata topik dalam setiap *dataset* abstrak skripsi. Sehingga diperoleh setiap abstrak tergolong pada topik penelitian tertentu.

Dari pembacaan *dataset* abstrak dengan kata kunci topik yang terbentuk diketahui topik terbanyak dimiliki oleh topik *illocutionari act ham* (linguistik). Sehingga jika divisualisasikan dalam 7 tren topik tertinggi sesuai dengan jenis penelitian linguistik sebagai berikut. Gambar 4.



Gambar 4. Top 7 Topik Linguistik Tertinggi Setiap Tahun



Gambar 5. Top 7 Topik Literatur Tertinggi Setiap Tahun

Dari hasil tren diketahui untuk jenis literatur (Gambar 5), topik *choice main character* sebagai topik tertinggi dengan rata-rata 28 penelitian setiap tahun. Sedangkan pada hasil tren jenis linguistik (Gambar 4), topik *illocutionary act hamm* sebagai topik tertinggi dengan rata-rata 27 penelitian setiap tahun.

Dengan melihat jenis penelitian dari topik yang terkandung dalam seluruh dataset maka diperoleh perbandingan jenis penelitian setiap tahun pada Program Studi Sastra Inggris UINSA sebagai berikut. Tabel 3.

Tabel 3. Jumlah Jenis Penelitian Setiap Tahun

Tahun	Dataset	Linguistik	Literatur
2014	71	55	16
2015	165	114	51
2016	152	105	47
2017	120	77	43
2018	95	66	29
2019	117	88	29

Dari hasil perbandingan jenis penelitian setiap tahun, diketahui bahwa jenis penelitian linguistik mendominasi jumlah penelitian setiap tahun.

#### 5. Pengujian Hasil

Pengujian hasil dilakukan untuk mengukur tingkat akurasi dari jumlah jenis penelitian setiap tahun (Tabel 3). Data Tabel 3 dibandingkan dengan data *real* jenis penelitian setiap tahun Program Studi Sastra Inggris UINSA. Perbandingan dari dua data tersebut sebagai berikut. Tabel 4.

Tabel 4. Perbandingan Hasil analisis dan Data Real Jumlah Jenis Penelitian

Tahun	Dataset	Hasil Analisis		Data Real	
		Linguistik	Literatur	Linguistik	Literatur
2014	71	54	16	55	15
2015	165	99	66	114	51
2016	152	96	56	105	47
2017	120	72	48	77	43
2018	95	57	38	66	29
2019	117	82	35	88	29

Dari perbandingan hasil analisis dengan data real jenis penelitian setiap tahun, diperoleh tingkat akurasi sebagai berikut. Tabel 7.

Tabel 5. Persentase Tingkat Presisi, Recall, dan Akurasi Jenis Penelitian

Tahun	Linguistik		Literatur		Accuracy (%)
	Presisi (%)	Recall (%)	Presisi (%)	Recall (%)	
2014	0,782	0,796	0,267	0,250	0,671
2015	0,772	0,752	0,431	0,458	0,667
2016	0,848	0,967	0,936	0,733	0,875
2017	0,779	0,968	0,953	0,707	0,842
2018	0,848	1,000	1,000	0,744	0,895
2019	0,795	0,946	0,862	0,581	0,812

Perbandingan hasil analisis dengan *data real* diperoleh tingkat Presisi dan *Recall* untuk Linguistik sebesar 80 dan 90 persen, sedangkan untuk Literatur sebesar 74 dan 57 persen. Serta tingkat akurasi keseluruhan sebesar 79 persen.

### C. PEMBAHASAN

Dari hasil pemodelan topik didapatkan 37 topik (Tabel3), yang mana jumlah pemodelan topik lebih banyak daripada penelitian [4]. Hal ini dikarenakan jumlah dataset yang digunakan dalam penelitian ini lebih banyak yakni 720 dataset, sedangkan pada penelitian sebelumnya menggunakan 584 dataset.

Pada hasil topik tertinggi yang didapat yakni *illocutionary act hamm* dengan makna semantik aksi ilokusi yang termasuk pada bidang sosiolinguistik [13].

*Scope* sosiolinguistik diketahui merupakan ilmu yang berhubungan dengan sosiokultural dalam [14] dan [15]. Dan *scope* sosiolinguistik yang menjadi tren dalam penelitian ini sesuai dengan hasil [16] yang menyebutkan dari 42 *Social Science Citation Index (SSCI)-indexed journals of applied linguistics*, sebagian besar topik yang sering dibahas tetap populer selama 12 tahun, beberapa (terutama masalah sosiokultural/ fungsional/ identitas). Sehingga penelitian yang ada pada Program Studi Sastra Inggris UINSA dalam 5 tahun terakhir memiliki tren yang mirip secara umum.

## VI. KESIMPULAN

Dari penelitian ini disimpulkan bahwa dalam kurun waktu 5 tahun dihasilkan total 37 topik skripsi Sastra Inggris, yang terbagi menjadi 2 jenis yaitu 21 topik berjenis linguistik dengan rata-rata presisi 80 persen, dan *recall* 90 persen, dan 16 topik berjenis literatur dengan rata-rata presisi 74 persen, dan *recall* 57 persen. Serta untuk tingkat akurasi kedua jenis topik yang menunjukkan kedekatan hasil prediksi dan *data real* memiliki hasil yang cukup baik yakni 79 persen. Dalam proses analisis tren topik menghasilkan 7 topik tertinggi untuk setiap jenis penelitian. Topik *illocutionary act hamm* (Sosiolinguistik) sebagai topik dengan rata-rata paling tinggi sebesar 27 penelitian setiap tahun pada jenis linguistik dan topik *choice main character* (literatur) sebagai topik dengan rata-rata paling tinggi sebesar 28 penelitian tiap tahun pada jenis literatur.

## REFERENSI

- [1] A. Setiawan, J. Darmanta, Tinaliah, and Yoannita, "Peringkasan dokumen berita Bahasa Indonesia menggunakan metode Cross Latent Semantic Analysis," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 3, no. 2, p. 94, 2017, doi: 10.26594/register.v3i2.1161.
- [2] K. R. Prilianti and H. Wijaya, "Aplikasi Text Mining untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering," vol. 2, no. 1, pp. 1–6, 2014.
- [3] C. S. Hudaya, H. Fakhurroja, and A. Alamsyah, "Analisis Persepsi Konsumen Terhadap Brand Go-Jek Pada Media Sosial Twitter Menggunakan Metode Sentiment Analysis Dan Topic Modelling," *J. Mitra Manaj.*, vol. 2, no. 4, pp. 273–285, 2018, [Online]. Available: <http://ejournalmitramanajemen.com/index.php/jmm/article/view/125/69>.
- [4] A. I. Alfanzar, "Topic modelling skripsi menggunakan metode latent dirichlet allocation," Universitas Islam Negeri Sunan Ampel, 2019.
- [5] I. N. Farida, M. Kom, D. Swanjaya, and M. Kom, "ARTIKEL IMPLEMENTASI METODE LATENT DIRICHLET ALLOCATION UNTUK MENENTUKAN TOPIK TEKS BERITA Oleh : DIMAS ARYANTO SAPUTRO Dibimbing oleh : UNIVERSITAS NUSANTARA PGRI KEDIRI TAHUN 2019 SURATPERNYATAAN ARTIKEL SKRIPSI TAHUN 2019," 2019.
- [6] A. Luthfiarta, J. Zeniarja, and A. Salam, "Algoritma Latent Semantic Analysis (LSA) Pada Peringkasan Dokumen Otomatis Untuk Proses Clustering Dokumen," *Semin. Nas. Teknol. Inf. Komun. Terap. 2013 (SEMANTIK 2013)*,

vol. 2013, no. November, pp. 13–18, 2013.

- [7] Wikipedia, “Diagram Alir,” 2019. [https://id.wikipedia.org/wiki/Diagram\\_alir](https://id.wikipedia.org/wiki/Diagram_alir) (accessed Nov. 27, 2019).
- [8] M. A. Putra *et al.*, “ANALISIS BRAND AWARENES S PRODUK OTOMOTIF PADA MASA RILIS MENGGUNAKAN DYNAMIC NETWORK ANALYSIS ( STUDI KASUS PADA ALL NEW NISSAN LIVINA ) BRAND AWARENESS ANALYSIS OF AUTOMOTIVE PRODUCTS DURING THE RELEASE USING DYNAMIC NETWORK ANALYSIS ( CASE STUDY ON ALL,” vol. 6, no. 2, pp. 1926–1933, 2019.
- [9] A. A. Jadhira, M. A. Bijaksana, and B. A. Wahyudi, “Deteksi Kemiripan Bagian-bagian Terjemah Al-Qur’an dengan Menggunakan Metode Latent Semantic Analysis,” *eProceedings Eng.*, vol. 5, no. 3, pp. 7649–7657, 2018, [Online]. Available: <https://libraryproceeding.telkomuniversity.ac.id/index.php/engineering/article/view/7286>.
- [10] A. Rahman, “Online News Classification Using Multinomial Naive Bayes,” *Itsmart*, vol. 6, no. 1, pp. 32–38, 2017, doi: 10.1177/1096348015584441.
- [11] A. Priyanto and M. R. Ma’arif, “Implementasi Web Scrapping dan Text Mining untuk Akuisisi dan Kategorisasi Informasi dari Internet (Studi Kasus: Tutorial Hidroponik),” *Indones. J. Inf. Syst.*, vol. 1, no. 1, pp. 25–33, 2018, doi: 10.24002/ijis.v1i1.1664.
- [12] D. C. Hidayati, S. Al Faraby, and A. Adiwijaya, “Klasifikasi Topik Multi Label pada Hadis Shahih Bukhari Menggunakan K-Nearest Neighbor dan Latent Semantic Analysis,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 1, p. 140, 2020, doi: 10.30865/jurikom.v7i1.2013.
- [13] S. Fauziah, “Kesantunan Sebagai Kajian Sociolinguistik,” *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2018, doi: 10.1017/CBO9781107415324.004.
- [14] D. A. N. P. E. B. Ahasa, “HUBUNGAN VARIASI BAHASA DENGAN KELOMPOK SOSIAL,” pp. 12–13, 1986, [Online]. Available: <https://media.neliti.com/media/publications/78731-ID-hubungan-variasi-bahasa-dengan-kelompok.pdf>.
- [15] M. H. Yendra, S. S., “PENERAPAN SOSIOLINGUISTIK DALAM MEMAHAMI SOSIOKULTURAL MINANGKABAU UNTUK PENDIDIKAN KARAKTER; CIME’EH DAN INSYA ALLAH ORANG MINANGKABAU,” *J. IPTEKS Terap. Res. Appl. Sci. Educ.*, vol. 10, pp. 71–80, 2016, doi: <http://dx.doi.org/10.22216/jit.2016.v10i1.466>.
- [16] L. Lei and D. Liu, “Research Trends in Applied Linguistics from 2005 to 2016: A Bibliometric Analysis and Its Implications,” *Appl. Linguist.*, vol. 40(3), pp. 540–561, 2019, doi: doi:10.1093/applin/amy003.