Analisis Pengaruh Algoritma *Stopwords* Dan *Stemmer* Terhadap Model *Information Retrieval*

Nurilmi Amalia Marda[1], Tiwi Nurhidayatullah[2], Ir. Rismayani, S. Kom., MT[3], Usman, SE., M.Kom[4].
[1,2] Jurusan Teknik Informatika Universitas Dipa Makassar
Jln. Perintis Kemerdekaan KM. 9 Makassar
e-mail: [1]nurilmiilmi71@gmai.com, [2] tiwinur24@gmai.com, [3]rismayani@dipanegara.ac.id
[4]usman@dipanegara.ac.id

Abstrak

Information retrieval ialah mendapatkan kembali suatu document dari document-document. Salah satu masalah yang dihadapi oleh pengguna dalam mencari suatu informasi berita ialah terlalu banyak nya hasil yang tampil dari suatu website yang ada namun belum relevan atau tidak sesuai yang di inginkan dari database/kumpulan dokumen sehingga menyulitkan pengguna untuk memilih berita yang relevan. Solusi dari masalah tersebut maka dilakukan analisis pengaruh algortima stopwords dan stemmer terhadap model information retrieval. Metode yang ingin digunakan penyusun ini dengan 4 skenario pengujian dengan melihat presisi yang besar. Hasil dari penelitian di dapatkan bahwa model 1 yang menerapkan stopwords dan stemmer merupakan model yang paling baik dengan nilai average precision 0.89604715. Model terbaik 2 yang menerapakan stopwords tapi tanpa Stemmer dengan nilai average precision 0.889867042. Model terbaik 3 yang menerapakan stemmer tapi tanpa stopwords dengan nilai average precision 0.879008822. Model terbaik 4 tidak menerapakan stopwords dan tanpa stemmer dengan nilai average precision 0.876263991.

Kata kunci*— Information retrieval, stopwords, stemmer.*

Abstract

Information retrieval is getting back a document from documents. One of the problems faced by users in searching for news information is that too many results are displayed from an existing website but are not yet relevant or do not match what is desired from the database/set of documents so that it attacks the user to select relevant news. The solution to this problem is to analyze the influence of the stopwords and stemmer algorithms on the information retrieval model. The method that this author wants to use with 4 test scenarios by looking at great precision. The results of the study found that model 1 which applied stopwords and stemmers was the best model with an average precision value of 0.89604715. The 2nd best model applies stopwords but without Stemmer with an average precision value of 0.889867042. The 3rd best model applies the stemmer but without stopwords with an average precision value of 0.879008822. The best model 4 does not apply stopwords and without a stemmer with an average precision value of 0.876263991.

Keywords— *Information retrieval, stopwords, temmer.*

1. PENDAHULUAN

Di era *global* saat ini, pencarian informasi artikel berita sangatlah mudah didapatkan karena sudah tersedia berbagai macam *searching* atau Sistem Temu Kembali Informasi (STKI). Berita ialah informasi yang mengungkapkan kejadian yang sedang terjadi, misalnya berita pada

umumnya dibuat seorang wartawan atau jurnalis[1]. Artikel berita memiliki banyak jenis yang memungkinkan untuk diterapkan sebagi objek penelitian analisis *information retrieval* pada artikel berita dengan menggunakan *Latent Semantic Precision* (LSI). *Information retrieval* (IR) ialah merupakan sebuah pemodelan yang digunakan untuk melaksanakan pencarian dokumen yang relevan pada kumpulan dokumen yang besar[2].

Namun algoritma LSI membutuhkan pengoptimalan dalam melakukan proses analisa *teks* yang mendorong penelitian ini dilakukan. Dalam *preprocessing* dibutuhkan teknik *stopwords* untuk menghapuskan term yang bukan butuhkan atau term yang bukan memiliki interpretasi yang dibuthkan misalkan kata ganti orang, kata sambung, kata pertentangan dan lain-lain[3]. Adapun juga dibutuhkan teknik stemmer untuk merubah *term* yang mempunyai imbuhan menjadi *term* dasar agar model LSI tidak memiliki perspektif berbeda diantara *term* yang mempunyai imbuhan dan tidak mempunyai imbuhan[4]. Teknik yang digunakan untuk *stopwords* adalah Natural *Language Tool Kit* (NLTK) sedangkan *stemmer* menggunakan Teknik sastrawi yang merupakan hasil pengebangan algoritma nazief-adriani[5].

Namun belum tentu juga jika penerapan teknik *stopwords* dan *stemmer* memiliki pengaruh yang cukup berarti maka untuk menghasilkan model LSI terbaik dalam sistem temu kembali atau *information retrieval*, maka akan dibuat 4 skenario pengujian dengan melihat *Precision* yang besar. Keempat skenario yang dimaksud adalah penerapan model LSI dengan *stopwords* & *stemmer*, model LSI dengan *stopwords* & tanpa *stemmer*, model LSI tanpa *stopword* & tanpa *stemmer* serta model LSI tanpa *stopwords* & tanpa *stemmer*. Sehingga menghasilkan model terbaik untuk dapat disajikan dalam penelitian ini.

Information text retrieval ialah cara yang dipakai untuk menaruh data dengan model mengolahnya (menghapuskan *stopword*) dan menaruh tiap *term* beserta informasi dari *term* tersebut baik letak *term*, total bobot, dan lain-lain. Pada kasus ini penulis menginginkan menganalisis berita dengan algoritma *stopwords* dan *stemmer* tertentu dengan suatu kata kunci, misalkan kita mencari sesuatu *keyword* yang relevan[2]. Sistem pencarian informasi pada zaman teknologi ini memunculkan revolusi dan inovasi dalam ilmu pengtahuan, salah salah satunya adalah teknologi *information retrieval* (Sistem Temu Kembali Informasi)[6]. *Information Retrieval* umumnya dapat di kembangkan dengan *Latent simantic indenxing* (LSI), Tujuan dari *Latent Semantic Indexing* (LSI) ialah memperoleh sebuah pemodelan yang efektif untuk meperepsentasikan keterkaitan dengan kunci kunci dan *document* yang dicari. Dari sekumpulan *keyword*, yang sebelumnya tidak lengkap dan tidak sesuai, menjadi sekumpulan objek yang berkaitan[7]. LSI dapat menemukan hubungan *semantic* antara kata dan keterkaitan makna dengan kata[7].

Berdasarkan latar belakang di atas maka penulis akan menganalisis pengaruh algoritma *stopwords* dan *stemmer* terhadap model *information retrieval* menggunakan LSI pada artikel berita. Maka dapat memberikan kemudahan bagi pengguna dalam mendapatkan informasi atau bahkan menganalisa artikel berita dan untuk mencari kterkaitan makna antar kata pada sebuah kalimat yang diharapkan agar hasil pencarian menggunakan LSI dapat lebih relevan.

2. METODE PENELITIAN

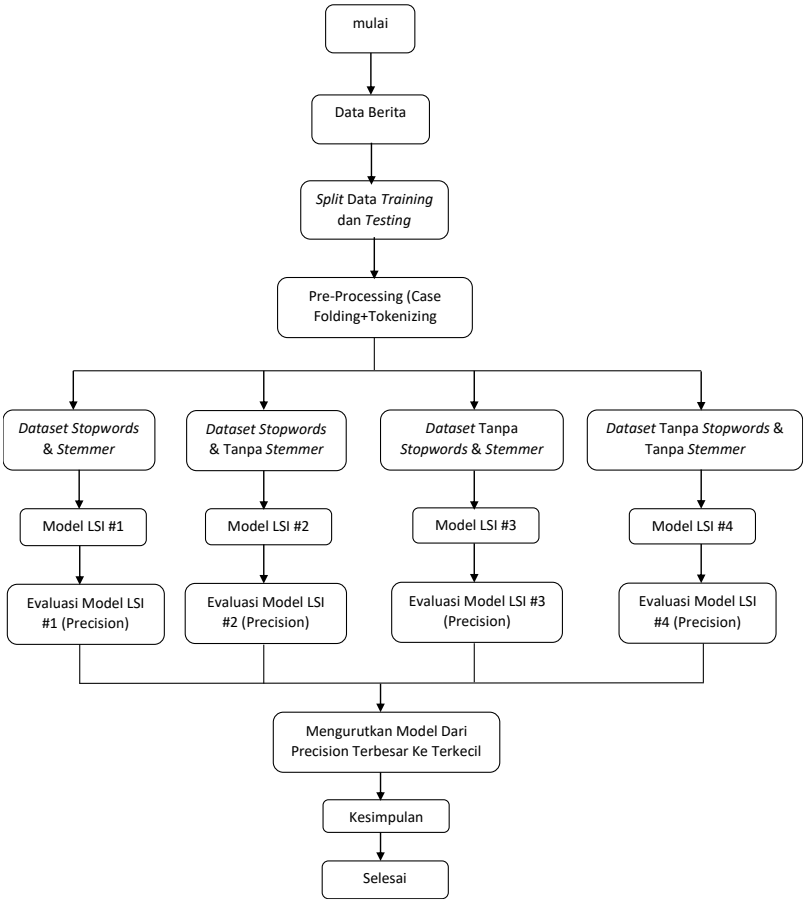
2.1 Jenis Penelitian

Jenis penelitian yang dilaksanakan oleh penulis dalam penelitian ini, ialah *Library research*, ialah penelitian yang dilaksanakan dengan Teknik membaca buku dan referensi-referensi lainnya untuk mendapatkan ilmu pengetahuan dan landasan teori yang berkaitan dengan *problem* yang dijelaskan oleh penulis[8].

2. 2 Pengumpulan Data

Pengumpulan data dilaksanakan dengan mempelajari buku dan jurnal yang membantu pada penelitian ini, tercantum di dalamnya kepustakaan tentang penulisan dan berhubungan dengan hal-hal yang membantu implementasi temu kembali pada analisis[9]. Metadata kumpulan document judul berita yang dipakai. Data tercatat tidak berurutan, bermula hasil penelusuran information, dihasilkan beberapa *document* judul artikel berita, pada langkah berikutnya penelitian ini memetik bebrapa dari *document* judul artikel berita tersebut menjadi data pada penelitian ini. Cara yang dipakai pada penelitian ini memakai cara *Latent Semantic Indexing*.

2. 3 Perancangan Solusi



Gambar 1 *Flowcart* alur penelitian

Pada gambar *flowchart* alur penelitian, data judul berita di lakukan *split* data *training* dan *testing* dan di lanjutkan dengan *preprocesing* yang dimana terdapat *case folding, tokenzing, stopword,* dan *stemmer*. Pada *split* data *training* 70% dan untuk data *testing* 30% dari *dataset*. *Case folding* adalah untuk menghapus angka mengubah ke huruf kecil dan mengapus tanda baca[10]. *Tokenezing* adalah pemisahan tiap kata. *Stopword* adalah menghapus kata yang tidak memiliki interpretasi[10]. *Stemmer* adalah mengubah kata menjadi kata dasar[10]. Metode yang akan digunakan penyusun ini dengan 4 skenario pengujian dengan melihat *precision* yang besar. Keempat skenario yaitu penerapan model LSI dengan *stopwords* dan *stemmer*, model LSI dengan *stopwords* dan tanpa *stemmer*, model LSI tanpa *stopwords* dan tanpa *stemmer* serta

model LSI tanpa *stopwords* dan tanpa *stemmer*. Dilakukan pencarian nilai evalusi model LSI terhadap *Precision*. Melakukan pengurutan model dari *Precision* terbesar ke terkecil. Dan di lanjutkan dengan mendapakan kesimpulan menganalisis pengaruh algoritma *stopwords* dan *stemmer* terhadap model *information retrieval* menggunakan *Latent semantice indenxing* (LSI) dan nilai besar akurasi pada analisis pengaruh algoritma *stopwords* dan *stemmer* terhadap model *information retrieval* menggunakan *Latent semantice indenxing* (LSI).

2. 4 Split Data *Training* dan *Testing*

Teknik *Split* data Training dan testing yang digunakan pada percobaan ini adalah sebesar 70 persen dari dataset untuk data traning sedangkan untuk data testing yang digunakan adalah sebesar 30 persen dari dataset.

2. 5 *Pre - procesing*

Pre-processing test ialah gerakan menghapuskan simbol kategori yang dimiliki *document*, misalnya koma, tanda petik dan lain-lain juga menukar seluruh huruf kapital jadi huruf kecil. Dalam langka *pre-processing test* ini dikerjakan *tokenization. tokenization* ialah tahapan pengerjaan *token* yang ditemukan dalam susunan *text*, kemudian *document* akan dibagi-bagi menjadi kata[11].

2. 5.1 Case Folding

Proses *case folding* yaitu merubah semua karakter huruf pada sebuah kalimat menjadi huruf kecil dan menghilangkan karakter yang dianggap tidak valid seperti angka, tanda baca, dan *Uniform Resources Locator* (URL) ataupun nama[11].

2. 5.2 Tokenzing

Proses *tokenization* berguna untuk memecah setiap kalimat dari seluruh dokumen pengetahuan ke dalam kata-kata *(term)* dengan menggunakan pembatas tab dan karakter spasi. Hal yang perlu dilakukan juga adalah menjadikan kata menjadi huruf kecil menghilangkan karakter tanda baca seperti tanda titik (.), koma (,), petik (“’), kurung (()), tanda tanya (?), tanda seru (!) dan tanda baca lainnya. Data hasil dari proses *tokenization*[11].

2. 5.3 Stopwrds Stopremoval

Proses diatas merupakan proses penghapusan kata yang tidak memiliki interpretasi yang sangat tidak dibutuhkan dalam proses traning model sehingga perlu dihapus agar proses traning berjalan lancar[11].

2. 5.4 Stemmer

Stemming ialah langkah memulangkan *term* yang didapatkan dari dapatkan saringan ke term dasarnya, menghapuskan imbuhan pertama *(prefix)* dan imbuhan belakang *(sufix)* selanjutnya diperoleh kata dasar[11].

2. 5.5 Document to Bag-Of-Words (Doc2bow)

Doc2bow ialah melakukan perhitungan jumlah setiap kata unik yang muncul di dokumen kemudian mengkonversi nya menjadi berformat *array* dan mengembalikan nilai nya menjadi sebuah vektor[12].

2. 5.6 Latent Semantic Indexing (LSI)

LSI ialah memperoleh sebuah pemodelan yang efektif untuk meperepsentasik keterkaitan dengan kunci kunci dan *document* yang dicari. Dari sekumpulan *keyword*, yang sebelumnya tidak lengkap dan tidak sesuai, menjadi sekumpulan objek yang berkaitan[7].

2. 5.7 Precision

Merupakan bobot yang membuktikan seberapa jelas predisksi yang dilaksanakan[13].

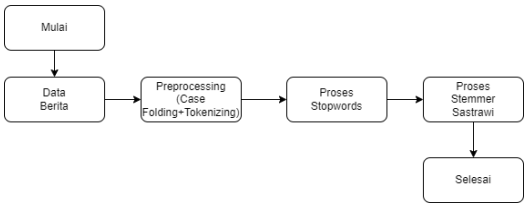
$$Precision = \frac{TP}{TP+FP}$$

Dimana:
TP = True Positive (TP), decision yang dapat cara mengklasifikasi data menjadi *(TRUE)* dan tanggapan aktualnya ialah ya *(TRUE)*.
FP = False Positive (FP), decision yang dapat cara mengklasifikasi data menjadi ya *(TRUE)* dan tanggapan aktualnya ialah tidak *(FALSE)*.

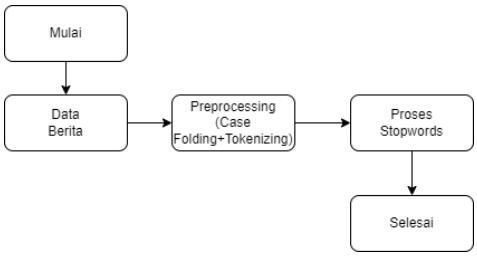
2. 6 Pembagian Processing Data 4 Skenario

Teknik *split* data traning dan testing yang digunakan pada percobaan ini adalah sebesar 70 persen dari dataset untuk data traning sedangkan untuk data testing yang digunakan adalah sebesar 30 persen dari dataset.

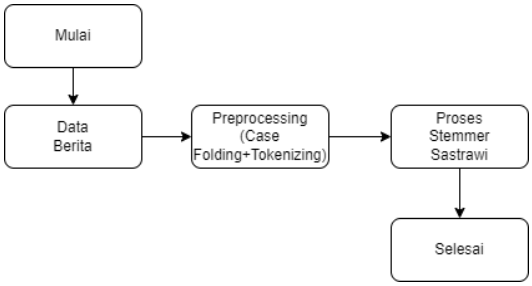
Pada *Preprocessing* yang dilakukan pada data skenario pertama dilakukan proses *stopwords* dan *stemmer* sebelum dilanjutkan pada proses pembuatan model, seperti yang tampak pada gambar dibawah ini.



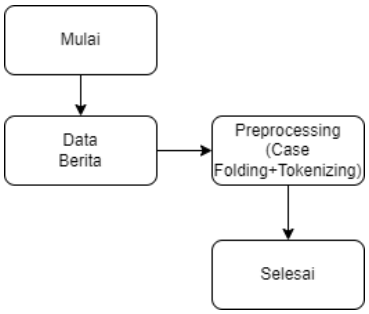
Pada *Preprocessing* yang dilakukan pada data skenario kedua akan diterapkan stopwords namun tanpa stemmer sebelum dilanjutkan pada proses pembuatan model, seperti yang tampak pada gambar dibawah ini.



Pada *Preprocessing* yang dilakukan pada data skenario ketiga akan diterapkan proses *stemmer* tapi tanpa proses *stopwords* sebelum dilanjutkan pada proses pembuatan model, seperti yang tampak pada gambar dibawah ini.



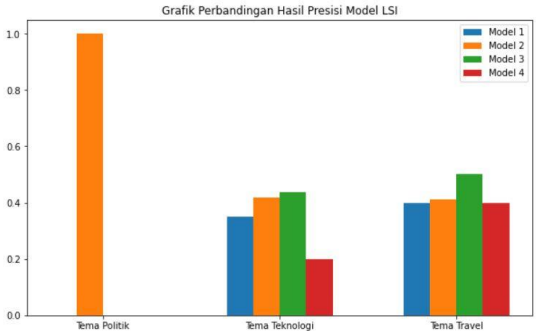
Pada *Preprocessing* yang dilakukan pada data skenario keempat tidak dilakukan sama sekali proses *stopwords* dan *stemmer* sebelum dilanjutkan pada proses pembuatan model, seperti yang tampak pada gambar dibawah ini.



3. HASIL DAN PEMBAHASAN

3.1 Model Terbaik Berdasarkan Precision

Berikut ini hasil dari penelitian pencarian model terbaik berdasarkan *Precision* tertinggi.



Gambar 2 Grafik Perbandingan Hasil Presisi Model LSI

Setelah memprediksi dan di masukan kedalam *doc2bow* dengan cara membagi data trening dan data testing 70% dan 30% di dapatkan hasil pada tema politik yaitu model 2 yang menerapkan *stopwords* tanpa *stemmer* dengan nilai *precision* 1.0.

Pada tema teknologi di dapatkan model 1 yang menerepakan *stopwords* dan *stemmer* dengan nilai *precision* pada rentang 0.2 – 0.4, model 2 menerapkan *stopwords* tanpa *stemmer*

didapatkan dengan nilai *precision* pada rentang 0.4-0.6, model 3 menerapkan *stemmer* tanpa *stopwords* didapatkan dengan nilai *precision* pada rentang 0.4 - 0.6 dan pada model 3 menerapkan *stemmer* tanpa *stopwords* didapatkan dengan nilai *precision* 0.2.

Pada tema travel didapatkan model 1 menerapkan *stopwords* dan *stemmer* dengan nilai precision 0.4, model 2 menerapkan stopwords tanpa stemmer dengan nilai *precision* pada rentang 0.4-0.6, model 3 menerapkan *stemmer* tanpa *stopwords* dengan nilai *precision* pada rentang 0.4-0.6 dan model 4 menerapkan tanpa *stopwords* dan tanpa *stemmer* dengan nilai *precision* 0.4.Berikut adalah ditail *precision* yang didapatkan dalam bentuk tabel.

Tabel 1 Hasil 1,2,3 dan 4 Berdasarkan LSI

Query	**Model 1**	**Model 2**	**Model 3**	**Model 4**
pinjaman online	0.6195307	0.63314974	0.6195307	0.63314974
teknologi komunikasi dalam interaksi	0.73456013	0.85386956	0.8572235	0.85386956
menjajal terapi oksigen hiperbarik	0.91313	0.8656118	0.91313	0.8656118
kini museum broadway di amerika serikat	0.9131439	0.96957856	0.9203252	0.96957856
pemilu di malaysia	0.9778593	0.79574084	0.79043746	0.79574084
warisan beracun politik apartheid	0.9051819	0.93978316	0.9051819	0.93978316
menyaksikan sejarah pengobatan tradisional korea	0.83036774	0.8265017	0.83036774	0.8265017
buka semua pintu	0.8165723	0.80578554	0.80711365	0.80578554
simak cara beli tiket	0.9267984	0.9907904	0.9796669	0.9907904
beri diskon tiket pesawat	0.9756766	0.9889176	0.9825974	0.9889176
melihat merunya acara dubai *design week*	0.9880978	0.9887657	0.9880978	0.9887657
kembangkan pusat kajian halal	0.9975788	0.9990315	0.9975788	0.9990315
ajak masyarakat	0.9790709	0.9454106	0.954818	0.9454106

peduli pada lingkungan				
orang-orang berbahaya	0.8994504	0.89826924	0.8994504	0.89826924
syarat yang diperlukan agar elektromagnet	0.85631293	0.89701277	0.9315858	0.89701277
dalam ilmu fisika	0.89462775	0.869716	0.8511073	0.869716
destinasi wisata terbaik	0.9409898	0.9710774	0.9409898	0.9710774
jembatan kereta api terpanjang	0.7801091	0.78699684	0.7801091	0.78699684
Average precision	**0.89604715**	**0.879008822**	**0.876263991**	**0.889867042**

Berdasarkan hasil diatas di dapatkan bahwa model 1 yang menerapkan *stopwords* dan *stemmer* merupakan model yang paling baik dengan nilai *average precision* 0.89604715. Model terbaik 2 yang menerapakan *stopwords* tapi tanpa *Stemmer* dengan nilai *average precision* 0.889867042. Model terbaik 3 yang menerapakan *stemmer* tapi tanpa *stopwords* dengan nilai *average precision* 0.879008822. Model terbaik 4 tidak menerapakan *stopwords* dan tanpa *stemmer* dengan nilai *average precision* 0.876263991.

4. KESIMPULAN

Setelah penulis melakukan penelitian analisis pengaruh algoritma *Stopwords* dan *Stemmer* terhadap model *information retrieval* pada artikel berita menggunakan VSM maka dapat dihaslikan kesimpulan sebagai berikut:

1. Hasil analisis berupa analisis pengaruh algoritma *Stopwords* dan *Stemmer* terhadap model *information retrieval* pada artikel berita menggunakan LSI di dapatkan bahwa model 1 yang menerapkan *stopwords* dan *stemmer* merupakan model yang paling baik dengan nilai *average precision* 0.89604715. Model terbaik 2 yang menerapakan *stopwords* tapi tanpa *Stemmer* dengan nilai *average precision* 0.889867042. Model terbaik 3 yang menerapakan stemmer tapi tanpa *stopwords* dengan nilai *average precision* 0.879008822. Model terbaik 4 tidak menerapakan *stopwords* dan tanpa *stemmer* dengan nilai *average precision* 0.876263991.
2. Hasil penerapan testing pada ketepatan mencari *(indexing)* document dapat dilakukan pencarian *(indexing) document* dengan menggunakan model LSI pada sistem *information retrival.*
3. Setelah membagi data trening dan data testing 70% dan 30% di dapatkan hasil pada tema politik yaitu model 2 yang menerapkan *stopwords* tanpa *stemmer* dengan nilai *precision* 1.0. Pada tema teknologi di dapatkan model 1 yang menerepakan *stopwords* dan *stemmer* dengan nilai *precision* pada rentang 0.2 – 0.4, model 2 menerapkan *stopwords* tanpa *stemmer* didapatkan dengan nilai *precision* pada rentang 0.4-0.6, model 3 menerapkan *stemmer* tanpa *stopwords* didapatkan dengan nilai *precision* pada rentang 0.4 - 0.6 dan pada model 3 menerapkan *stemmer* tanpa *stopwords* didapatkan dengan nilai *precision* 0.2. Pada tema travel didapatkan model 1 menerapkan *stopwords* dan *stemmer* dengan nilai *precision* 0.4,

model 2 menerapkan *stopwords* tanpa *stemmer* dengan nilai *precision* pada rentang 0.4-0.6, model 3 menerapkan *stemmer* tanpa *stopwords* dengan nilai *precision* pada rentang 0.4-0.6 dan model 4 menerapkan tanpa *stopwords* dan tanpa *stemmer* dengan nilai *precision* 0.4.

4. SARAN

Adapun saran yang ingin disampaikan penulis yaitu selain LSI, Penelitian ini juga masih jauh dari kata sempurna oleh karena itu penulis menyarankan agar penelitian selanjutnya dapat dikembangkan pada sistem berbasis *web* ataupun *android.*

UCAPAN TERIMA KASIH

Pada kesempatan ini penulis hendak mengucapkan terimakasih kepada keluarga, teman – teman, dan dosen pembimbing yang sudah mendukung, membantu dan memberi banyak masukan selama proses penyusunan skripsi ini berlangsung khususnya kepada:

1. Dr. Y. Jhony Wijaya. Soetikno, SE, MM. selaku Ketua Universitas Dipa Makassar.
2. Ir. H. Irsal, MT. selaku Ketua Jurusan Teknik Informatika program studi strata satu (S1) Universitas Dipa Makassar.
3. Ir. Rismayani, S.Kom., MT. selaku Pembimbing I, yang telah membimbing penulis dalam penyelesaian skripsi ini.
4. Usman, SE., M.Kom. selaku Pembimbing II, yang telah banyak memberikan bimbingan dan arahan penulis dalam menyelesaikan skirpsi ini.
5. Dosen Universitas Dipa Makassar yang telah mendidik dan mengajarkan berbagai disiplin ilmu kepada penulis.
6. Kedua orang tua tercinta yang memberikan *suport* dan dukungan yang tidak bisa kami nilai dalam bentuk apapun. Semoga Allah selalu melimpahkan kesehatan dan kebahagian, Aamin.
7. Untuk seluruh teman-teman tanpa terkecuali yang tidak dapat disebutkan namanya, terimakasih atas *suport* sistem dan bantuan yang telah kalian berikan kepada penulis.

DAFTAR PUSTAKA

[1] K. Nisa, “ANALISIS KESALAHAN BERBAHASA PADA BERITA DALAM MEDIA SURAT KABAR SINAR INDONESIA BARU,” *J. Bindo Sastra*, vol. 2, no. 2, Art. no. 2, Oct. 2018, doi: 10.32502/jbs.v2i2.1261.

[2] “Information Retrieval Sistem Kearsipan Pencarian Dokumen Di Dinas Pemberdayaan Perempuan Dan Perlindungan Anak Kota Semarang Menggunakan Metode Vector Space Model | JURNAL MAHAJANA INFORMASI,” Jun. 2022, Accessed: Mar. 07, 2023. [Online]. Available: http://e-journal.sari-mutiara.ac.id/index.php/7/article/view/2538

[3] A. F. Hidayatullah, “Pengaruh Stopword Terhadap Performa Klasifikasi Tweet Berbahasa Indonesia,” *JISKA J. Inform. Sunan Kalijaga*, vol. 1, no. 1, Art. no. 1, May 2016, doi: 10.14421/jiska.2016.11-01.

[4] I. M. A. Agastya, “PENGARUH STEMMER BAHASA INDONESIA TERHADAP PEFORMA ANALISIS SENTIMEN TERJEMAHAN ULASAN FILM,” *J. Tekno Kompak*, vol. 12, no. 1, Art. no. 1, Feb. 2018, doi: 10.33365/jtk.v12i1.70.

[5] “Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir | Jurnal Teknik Informatika dan Sistem Informasi.” https://journal.maranatha.edu/index.php/jutisi/article/view/2892 (accessed Mar. 07, 2023).

[6] S. Y. Baskoro, A. Ridok, and M. T. Furqon, “PENCARIAN PASAL PADA KITAB UNDANG-UNDANG HUKUM PIDANA (KUHP) BERDASARKAN KASUS MENGGUNAKAN METODE COSINE SIMILARITY DAN LATENT SEMANTIC

INDEXING (LSI)," *J. Environ. Eng. Sustain. Technol.*, vol. 2, no. 2, Art. no. 2, Dec. 2015, doi: 10.21776/ub.jeest.2015.002.02.4.

[7] G. Susanto and H. L. Purwanto, "Information Retrieval Menggunakan Latent Semantic Indexing Pada Ebook," *SMATIKA J. STIKI Inform. J.*, vol. 8, no. 02, Art. no. 02, Oct. 2018, doi: 10.32664/smatika.v8i02.204.

[8] R. F. Pringgar and B. Sujatmiko, "PENELITIAN KEPUSTAKAAN (LIBRARY RESEARCH) MODUL PEMBELAJARAN BERBASIS AUGMENTED REALITY PADA PEMBELAJARAN SISWA," *IT-Edu J. Inf. Technol. Educ.*, vol. 5, no. 01, pp. 317–329, 2020.

[9] I. Fahrezi, M. Taufiq, Akhwani, and Nafiah, "Meta-analisis Pengaruh Model Pembelajaran Project Based Learning Terhadap Hasil Belajar Siswa Pada Mata Pelajaran IPA Sekolah Dasar," *J. Ilm. Pendidik. Profesi Guru*, vol. 3, no. 3, Art. no. 3, 2020.

[10] "INFORMATION RETRIEVAL TUGAS AKHIR DAN PERHITUNGAN KEMIRIPAN DOKUMEN MENGACU PADA ABSTRAK MENGGUNAKAN VECTOR SPACE MODEL | Mas'udia | Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer." https://jurnal.umk.ac.id/index.php/simet/article/view/1016 (accessed Mar. 07, 2023).

[11] H. Bunyamin and C. Negara, "Aplikasi Information Retrieval (IR) CATA Dengan Metode Generalized Vector Space Model," vol. 4, Oct. 2011.

[12] W. T. H. Putri and R. Hendrowati, "PENGGALIAN TEKS DENGAN MODEL BAG OF WORDS TERHADAP DATA TWITTER," *J. Muara Sains Teknol. Kedokt. Dan Ilmu Kesehat.*, vol. 2, no. 1, Art. no. 1, Sep. 2018, doi: 10.24912/jmstkik.v2i1.1560.

[13] S. Hendrian, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan," *Fakt. Exacta*, vol. 11, no. 3, Art. no. 3, Oct. 2018, doi: 10.30998/faktorexacta.v11i3.2777.