

PERACANGAN SISTEM KLASIFIKASI SURAT ELEKTRONIK (*E-MAIL*) MENGUNAKAN METODE *COSINE SIMILARITY*

Sadly Syamsuddin, Ahyuna dan Kalfin Alloto'dang

STMIK Dipaenagara, Indonesia

Email: sadlyg2@gmail.com, ahyuna@dipanegara.ac.id dan celvindp02@gmail.com

INFO ARTIKEL	ABSTRAK
Diterima 24 Agustus 2020 Diterima dalam bentuk revisi Diterima dalam bentuk revisi	Saat ini banyak kantor pemerintah dan swasta yang melakukan pengolahan data surat menyurat disetiap harinya, dimana ada ratusan surat yang masuk untuk keperluan harian di kantor tersebut. Namun kami melihat ada beberapa kekurangan yang terdapat di dalamnya antara lain (1) Banyak kantor dalam pencatatan data surat masuk dan surat keluar masih harus dicatat dan dibukukan dalam pembuatan surat masuk dan surat keluar sehingga jika nomor surat ingin dicocokkan nantinya agak sulit untuk menemukannya, (2) Proses sortir surat kebagian-bagian yang ada di perusahaan masih memakan waktu, karena harus dibuka terlebih dahulu dan di periksa satu-persatu tujuan tembusan dari surat tersebut hal tersebut juga mengakibatkan proses kerja dari persuratan disana masih terbilang lambat. Maka berdasarkan latar belakang permasalahan tersebut diatas kami melakukan perancangan sistem klasifikasi surat elektronik (e-mail) menggunakan metode cosine similarity, dimana metode cosine similarity yang digunakan berfokus pada teks berbahasa Indonesia dan berkonsepkan text mining yang dapat mengukur kesamaan teks berdasarkan kemunculan kata-kata dalam teks tersebut. Hasil penelitian ini menunjukkan sistem telah bekerja dengan baik dan mampu melakukan pemisahan antara surat yang masuk, dimana surat tersebut secara otomatis akan di pisahkan berdasarkan analisis isi pesan yang ada menggunakan tahapan penentuan stopwords, keyword tiap surat, dan klasifikasi.
Kata kunci: Surat; <i>cosine similarity</i> dan text mining	

Pendahuluan

Dalam suatu organisasi, surat menurut prosedur pengurusannya dibagi menjadi dua yaitu surat masuk dan surat keluar. Surat masuk adalah semua jenis surat yang diterima dari instansi lain maupun perorangan, baik yang diterima melalui pos, melalui kurir dengan mempergunakan buku pengiriman, atau melalui media surat elektronik (email), sedangkan surat keluar adalah surat yang sudah lengkap (bertanggal, bernomor,

berstempel, dan telah ditanda tangani oleh pejabat yang berwenang) yang dibuat oleh suatu instansi, kantor atau lembaga untuk ditujukan atau dikirim kepada instansi, kantor atau lembaga lain melalui cara yang sama (Wursanto, 2006).

Namun kami melihat ada beberapa kekurangan yang terdapat di dalamnya antara lain (1) Banyak kantor dalam pencatatan data surat masuk dan surat keluar masih harus dicatat dan dibukukan dalam pembuatan surat masuk dan surat keluar sehingga jika nomor surat ingin dicocokkan nantinya agak sulit untuk menemukannya, (2) Proses sortir surat kebagian-bagian yang ada di perusahaan masih memakan waktu, karena harus dibukaterlebih dahulu dan di periksa satu-persatu tujuan tembusan dari surat tersebut hal tersebut juga mengakibatkan proses kerja dari persuratan disana masih terbilang lambat.

Berdasarkan latar belakang permasalahan tersebut maka dilakukanlah penelitian dengan berfokus pada proses perancangan dan pembuatan aplikasi persuratan yang dapat memecahkan permasalahan yang ada, dimana aplikasi persuratan ini dapat melakukan klasifikasi surat elektronik (e-mail) menggunakan metode *Cosine Similarit*. Cosine similarity berfokus pada teks berbahasa Indonesia dan berkonsepkan text mining yang dapat mengukur kesamaan teks berdasarkan kemunculan kata-kata dalam teks tersebut (Imbar et al., 2014).

Metode Penelitian

A. Surat Elektronik (E-Mail)

Di dalam pengelolaan surat, terlebih dahulu surat dibedakan menjadi dua macam yaitu surat masuk dan surat keluar. Menurut (Saminah, 2015) “Surat masuk adalah semua surat yang diterima oleh organisasi kantor.” Sedangkan menurut (Mardiana, 2017) “Surat keluar adalah surat-surat yang dikirimkan sebagai jawaban atau tanggapan atas isi surat masuk yang diterima dari organisasi, kantor lain, atau perorangan, agar terjalin rangkaian hubungan timbal balik yang serasi yang berakibat menguntungkan kedua belah pihak.” Adapun prosedur pengelolaan surat masuk dan surat keluar adalah sebagai berikut:

1. Prosedur pengelolaan surat masuk

Pengelolaan surat masuk adalah serangkaian proses pengelolaan surat-surat yang masuk atau diterima oleh suatu organisasi dari organisasi yang lainnya. Prosedur pengelolaan surat masuk adalah sebagai berikut: (Rustamin & Dewi, 2016)

- a) Penerimaan;
- b) Penyortiran;
- c) Pencatatan;
- d) Tindak lanjut;
- e) Penyimpanan.

2. Prosedur pengelolaan surat keluar

Pengelolaan surat keluar adalah serangkaian proses pengelolaan surat-surat yang dikirim oleh organisasi ke organisasi lain. Prosedur pengelolaan surat keluar adalah sebagai berikut: (Mardiana, 2016)

- a) Pembuatan konsep surat;
- b) Meminta persetujuan pimpinan;
- c) Pengetikan;
- d) Penandatanganan;
- e) Pencatatan;
- f) Penyimpanan;
- g) Pengiriman.

Pengelolaan surat masuk maupun surat keluar tidak selalu memiliki prosedur pengelolaan yang sama pada setiap organisasi. Prosedur pengelolaan surat disesuaikan dengan tingkat aktivitas surat menyurat di organisasi yang bersangkutan. Pengelolaan surat elektronik menurut Peraturan Kepala Arsip Nasional Republik Indonesia Nomor 15 Tahun 2012 antara lain surat elektronik perlu diklasifikasikan sesuai dengan klasifikasi surat yang digunakan pada organisasi atau instansi yang bersangkutan agar surat dapat ditemukan dengan mudah dan cepat. Kemudian surat elektronik dicatat ke dalam sistem pengelolaan surat. Selanjutnya, surat elektronik di simpan ulang sebagai cadangan (*backup*). *Backup* surat elektronik ke dalam sistem informasi surat elektronik serta disimpan ke dalam direktori atau folder atau dapat pula disimpan dalam kaset. Langkah terakhir yakni mencetak dan memberkaskan surat elektronik, lampiran serta data-data lainnya ke dalam sistem pengelolaan berbasis kertas.

Berdasarkan pendapat-pendapat di atas dapat diketahui bahwa kegiatan surat-menyurat dalam suatu organisasi diselesaikan dengan beberapa langkah agar nilai guna surat tetap terjaga sehingga apabila suatu saat surat yang bersangkutan diperlukan dapat ditemukan secara cepat, tepat, dan dalam kondisi baik. Prosedur pengelolaan surat dalam setiap organisasi tidaklah sama. Setiap organisasi memiliki prosedur pengelolaan yang telah disesuaikan dengan keadaan yang ada dalam organisasi tersebut.

B. Text Mining

Text mining (disebut juga dengan *text data mining*), adalah suatu proses untuk mengambil informasi dari teks yang ada. *Text mining* mencari pola-pola yang ada di teks teks dalam bahasa natural yang tidak terstruktur seperti buku, email, artikel, halaman web, dll. Kegiatan yang biasa dilakukan oleh *text mining* adalah *text categorization*, *text clustering*, *conception/entity extraction*, dan lain-lain. Ada 3 proses yang biasanya ada dalam sebuah kegiatan *text mining* (Syahroni, 2017).

1. Characterization of Data

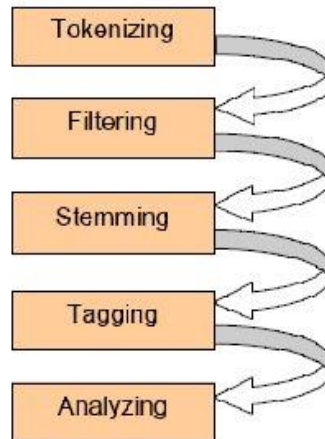
Teks yang ada distrukturkan dengan proses seperti parsing, dan di masukkan ke dalam sebuah database;

2. Data Mining

Dari data yang ada lalu dilakukan sebuah pencarian dengan algoritma tertentu untuk mendapatkan pola dari data tersebut;

3. Data *visualization*

Hasil pencarian yang ada akan diinterpretasi dan dikeluarkan dalam bentuk Output yang dapat dimengerti dengan mudah;



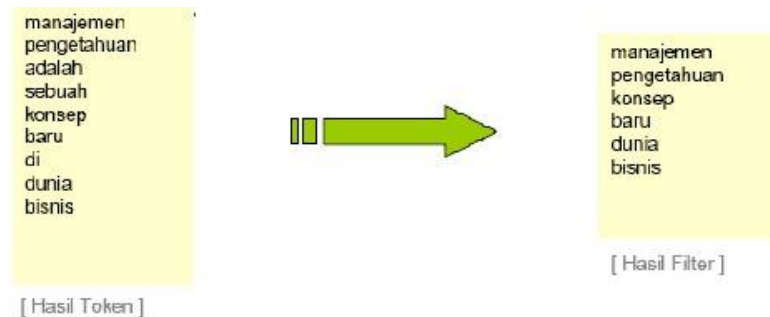
Gambar 1. Tahapan yang Dilakukan Secara Umum

(Sumber : (Han et al., 2011))



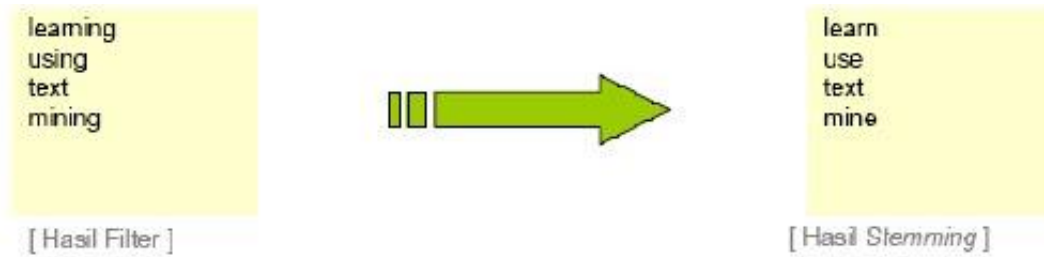
Gambar 2. Contoh Tahapan Tokenizing

(Sumber : (Han et al., 2011))



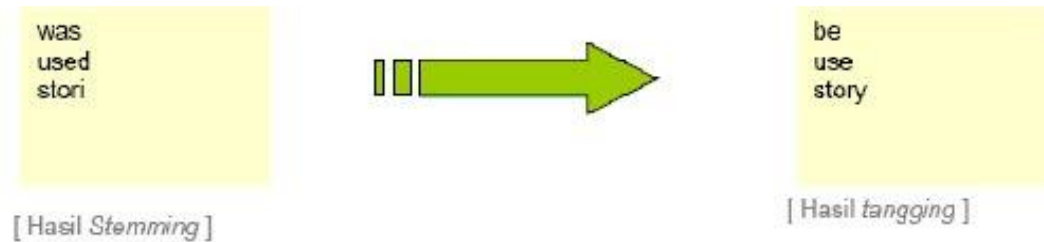
Gambar 3. Contoh Tahapan Filtering

(Sumber : (Han et al., 2011))



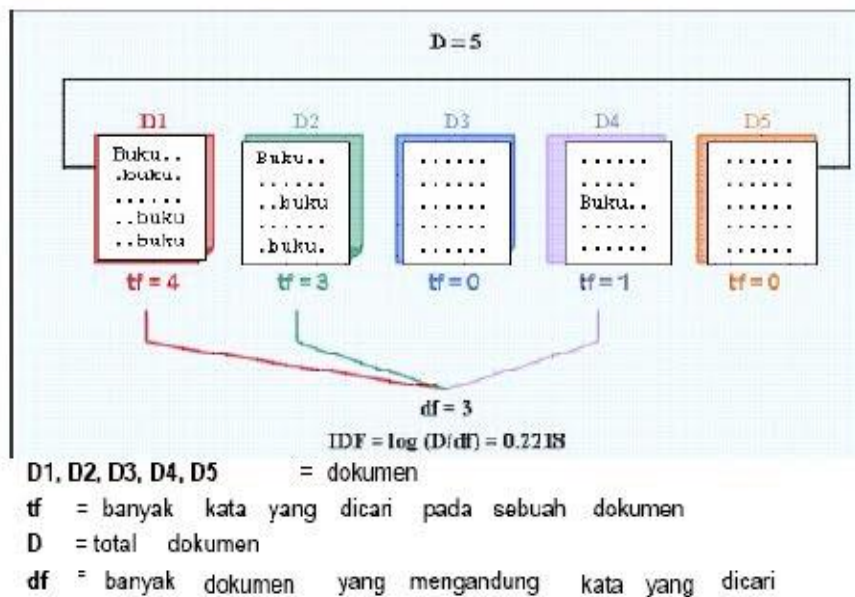
Gambar 4. Contoh Tahapan Stemming

(Sumber : (Han et al., 2011))



Gambar 5. Contoh Tahapan Tagging

(Sumber : (Han et al., 2011))

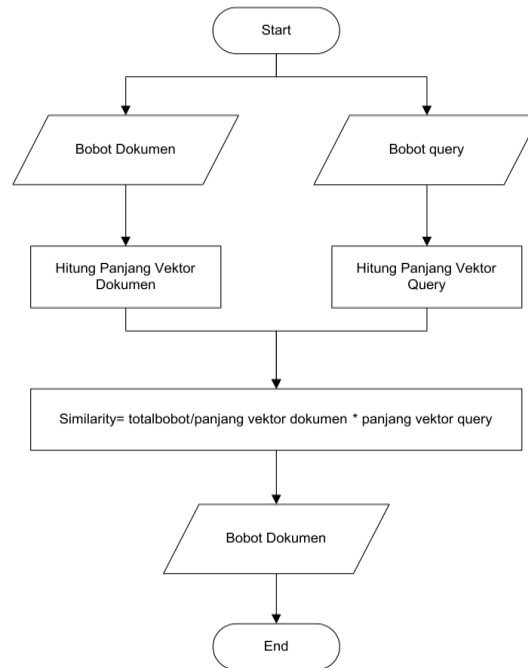


Gambar 6. Contoh Tahapan Analizing

(Sumber : (Han et al., 2011))

C. Cosine Similarity

Dalam proses cosine similarity yang menjadi masukan adalah bobot dari term setiap data, bobot term tersebut di gunakan dala proses perhitungan jarak kemiripan dengan kata klaster, kemudian dari setiap nilai akan menentukan centroid setiap klaster. Berikut merupakan proses perhitungan cosine similarity: (Sumber : (Han et al., 2011))



Gambar 7. Flowchart Cosine Similarity

(Sumber : (Han et al., 2011))

Cosine similarity adalah ukuran kesamaan yang lebih umum digunakan dalam information retrieval dan merupakan ukuran sudut antara vektor dokumen (titik (ax, bx)) dan Db (titik (ay, by)). Tiap vektor tersebut merepresentasikan setiap kata dalam setiap dokumen (teks) yang dibandingkan dan membentuk sebuah segitiga, sehingga dapat diterapkan hukum kosinus untuk menyatakan bahwa : (Sumber : (Han et al., 2011))

$$\cos(C) = \frac{a^2 + b^2 - c^2}{2ab} \tag{1}$$

dimana

$$a^2 = a_x^2 + a_y^2, b^2 = b_x^2 + b_y^2 \tag{2}$$

dan

$$c^2 = (b_x - a_x)^2 + (a_y - b_y)^2 \tag{3}$$

Gantikan nilai-nilai tersebut untuk a, b, dan c, sehingga didapatkan:

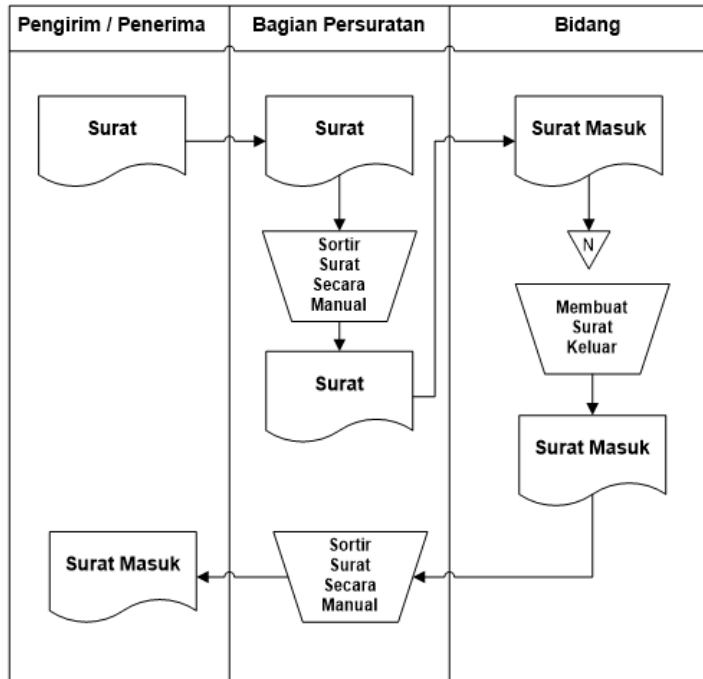
$$\cos C = \frac{a_x b_x + a_y b_y}{\sqrt{a_x^2 + a_y^2} \times \sqrt{b_x^2 + b_y^2}} \tag{4}$$

Ketika dua dokumen identik, sudutnya adalah nol derajat (0°) dan kesamaannya adalah satu (1); dan ketika dua dokumen tidak identik sama sekali, sudutnya adalah 90 derajat (90°) dan kesamaannya adalah nol (0).

Hasil dan Pembahasan

A. Sistem Berjalan

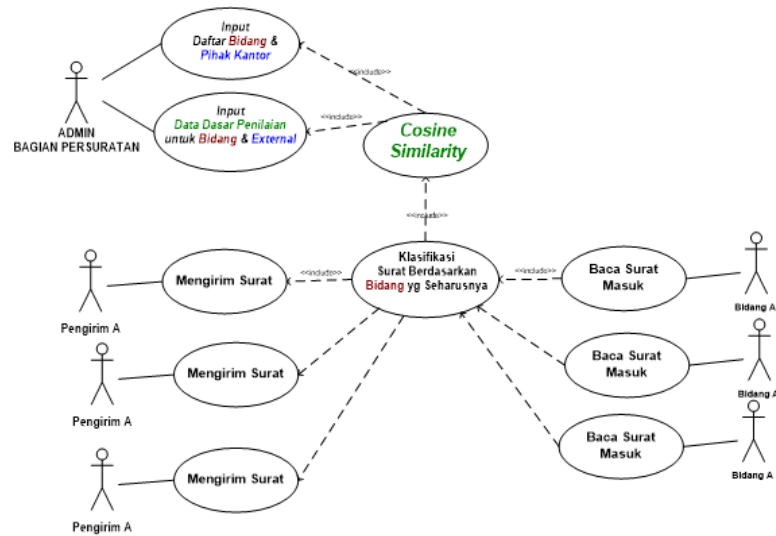
Setelah kami melakukan penelitian dan terhadap beberapa kantor, perusahaan, dan instansi denagar. Saat ini sistem yang sedang berjalan di beberaoa kantor adalah kurang lebih seperti yang kami gambarkan pada Gambar 8. berikut:



Gambar 8. Bagan Alir Dokumen yang Berjalan

B. Desain Use Case Diagram

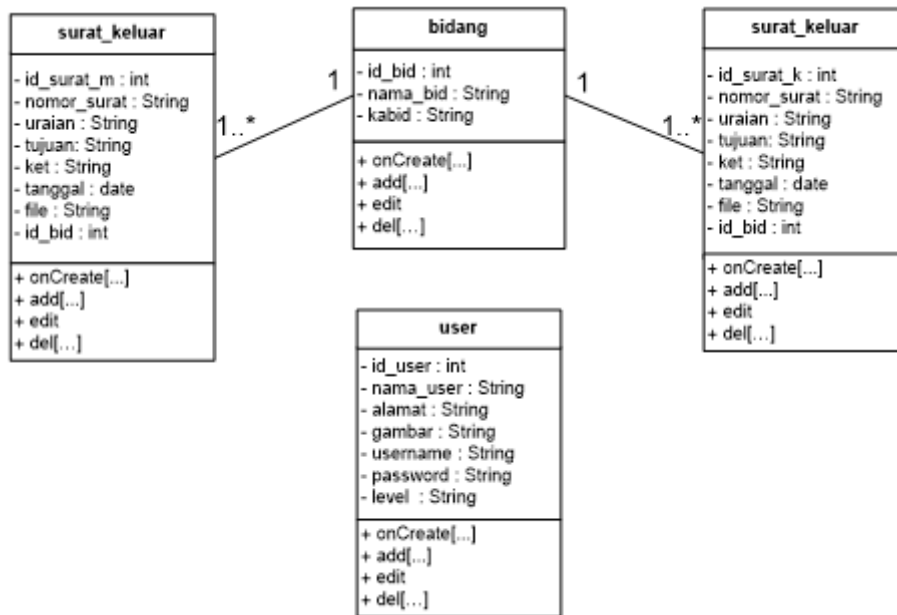
Use case diagram menggambarkan fungsionalitas yang diharapkan dari sebuah sistem. Yang ditekankan adalah “apa” yang diperbuat sistem, dan bukan “bagaimana”. Sebuah use case merepresentasikan sebuah interaksi antara aktor dengan sistem. Use case merupakan sebuah pekerjaan tertentu. Bentuk use case untuk sistem ini kami paparkan pada Gambar 8. berikut:



Gambar 9. Use Case Diagram

C. Desain Class Diagram

Class adalah sebuah spesifikasi yang jika diinstansiasi akan menghasilkan sebuah objek dan merupakan inti dari pengembangan dan desain berorientasi objek. Class menggambarkan keadaan (atribut/properti) suatu sistem, sekaligus menawarkan layanan untuk memanipulasi keadaan tersebut (metoda/fungsi). Gambar 9. berikut akan menjelaskan bagaimana bentuk class diagram pada system ini.

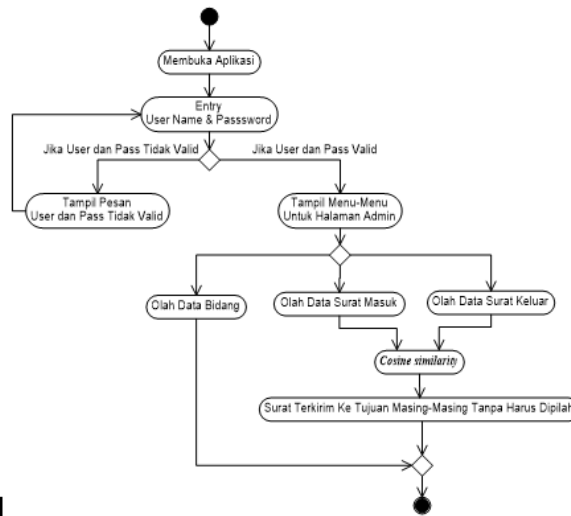


Gambar 10. Class Diagram Aplikasi

D. Activity Diagram

Activity diagram menggambarkan berbagai alir aktivitas dalam aplikasi yang sedang dirancang, bagaimana masing-masing alir berawal, decision yang mungkin

terjadi, dan bagaimana mereka berakhir. Activity diagram untuk sistem yang dirancang terdapat pada Gambar 10. berikut.

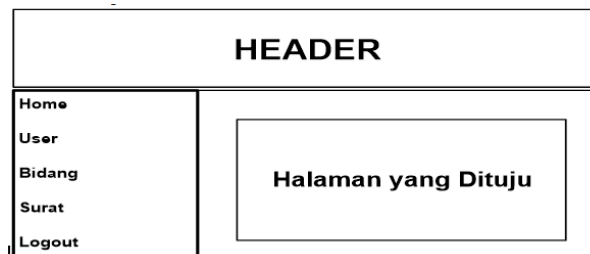


Gambar 11. Activity Diagram Sistem

E. Desain Input Output

1. Rancangan *Output* Halaman Utama Aplikasi

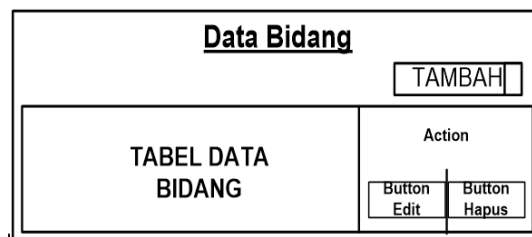
Rancangan ini dibuat untuk menggambarkan bentuk halaman utama aplikasi dimana terdapat beberapa menu utama yang dapat digunakan untuk mengakses fungsi fungsi utama dari aplikasi.



Gambar 12. Rancangan Output Halaman Utama Aplikasi

2. Rancangan *Output* Pengolahan Data Bidang

Rancangan ini dibuat untuk menggambarkan bentuk halaman pengolahan data bidang dimana admin dapat menambah, mengedit, menghapus data bidang melalui halaman ini.



Gambar 13. Rancangan Output Pengolaha Data Bidang

3. Rancangan *Output* Pengolahan Data Surat Masuk

Rancangan ini dibuat untuk menggambarkan bentuk halaman pengolahan data surat masuk.

Data Surat Masuk	
	TAMBAH
Tabel Data Surat Masuk	Action
	Button Acc Button Hapus

Gambar 14. Rancangan Output Pengolahan Data Surat Masuk

4. Rancangan *Output* Pengolahan Data Surat Keluar

Rancangan ini dibuat untuk menggambarkan bentuk halaman pengolahan data Surat Keluar.

Data Surat Keluar	
	TAMBAH
Tabel Data Surat Keluar	Action
	Button Acc Button Hapus

Gambar 15. Rancangan Output Pengolahan Data Surat Keluar

5. Rancangan *Input* Data Bidang

Rancangan ini dibuat untuk menggambarkan bentuk halaman Input data bidang.

Input Data Bidang	
Id Bidang	<input type="text"/>
Nama Bidang	<input type="text"/>
Nama Kabid	<input type="text"/>
	SAVE CANCEL

Gambar 16. Rancangan Input Data Bidang

6. Rancangan *Input* Data Surat Masuk

Rancangan ini dibuat untuk menggambarkan bentuk halaman Input data surat masuk.

Data Surat Masuk	
No. Surat	<input type="text"/>
Uraian	<input type="text"/>
Keterangan	<input type="text"/>
Tanggal Surat	<input type="text"/>
File	<input type="text"/> <input type="checkbox"/>
	SAVE CANCEL

Gambar 17. Rancangan Input Data Surat Masuk

7. Rancangan *Input* Surat Keluar

Rancangan ini dibuat untuk menggambarkan bentuk halaman Input data surat keluar.

The image shows a web form titled "Data Surat Keluar". It contains the following fields and controls:

- No. Surat**: A single-line text input field.
- Uraian**: A multi-line text area.
- Keterangan**: A multi-line text area.
- Tanggal Surat**: A date input field.
- File**: A file upload input field with a small square icon to its right.
- At the bottom, there are two buttons: **SAVE** and **CANCEL**.

Gambar 18. Rancangan Input Data Surat Keluar

F. Proses Implementasi *Cosine similarity*

Berikut ini adalah implementasi dari metode *Cosine similarity* :

a. Penentuan *Stopwords*

Stop words adalah kata umum (*common words*) yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. *Stop words* umumnya dimanfaatkan dalam task *information retrieval*, termasuk oleh Google. Contoh *stop words* untuk bahasa Inggris diantaranya "*of*", "*the*". Sedangkan untuk bahasa Indonesia diantaranya "*yang*", "*di*", "*ke*".

Dalam tahapan *text mining* yang kami bangun, ada tahapan yang bernama *filtering*, tahapan ini bertugas untuk menghilangkan atau membuat kata-kata yang termasuk data *stopword* dari teks yang diproses. Dalam sistem yang dibangun kami memasukkan semua data *stopword* ke dalam tabel *stopword*. Inputan data dan view data *stopword*.

b. Pembangkitan *Keyword* Tiap Surat

Setelah tahapan penentuan *stopword*, kami melakukan klasifikasi dengan pembangkitkan *keyword* dengan nilai TF yang disimpan untuk jadi pengali kemunculan dari kata-kata yang di analisis dalam satu dokument. Proses pembangkitan *keyword* dilakukan oleh sistem secara otomatis. Dimana *textarea keyword* diisi dengan judul, abstrak, dan penggalan *keyword* yang dipastikan termasuk dalam jenis penelitian yang ingin dibangkitkan *keywordnya*.

c. Klasifikasi

Setelah tahapan penentuan *keyword* untuk masing-masing jenis surat, tahap selanjutnya analisis untuk masing-masing surat masuk/keluar, berapa nilai total TF yang dimiliki untuk masing-masing jenis surat.

1. Pada surat yang masuk untuk salah satu bidang, sistem akan mengambil data uraian dan keterangan dalam satu variabel.

2. Kemudian nilai / isi kalimat variabel tersebut dipecah berdasarkan spasi dan spesial karakter yang ada ke dalam *array*. Tahap ini disebut dengan nama *tokenizing*.
3. Langkah selanjutnya adalah *filtering* dimana semua kata-kata yang termasuk ke dalam *stopwords* di hilangkan dari variabel array tersebut.

Langkah terakhir adalah menghitung nilai TF dari kata yang ada dikalikan dengan nilai TF dari Keyword masing-masing surat masuk/keluar. Nilai total TF yang terbesar akan dianggap sebagai surat masuk yang sah mewakili surat yang akan di tujukan ke salah satu bidang.

Kesimpulan

Sistem Klasifikasi Surat Elektronik (e-mail) menggunakan metode cosine similarity ini dibangun menggunakan alat perancangan UML yang terdiri dari use case, class diagram dan activity diagram, kemudian penulis merancang desain input ouput, lalu merancang tahapan implementasi metode cosine similarity yang digunakan

Hasil penelitian menunjukkan sistem telah bekerja dengan baik dan mampu melakukan pemisahan antara surat yang masuk, dimana surat tersebut secara otomatis akan di pisahkan berdasarkan analisis isi pesan yang ada menggunakan tahapan penentuan stopwords, keyword tiap surat, dan klasifikasi.

BIBLIOGRAFI

- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Imbar, R. V., Adelia, A., Ayub, M., & Rehatta, A. (2014). Implementasi Cosine Similarity dan Algoritma Smith-Waterman untuk Mendeteksi Kemiripan Teks. *Jurnal Informatika*, 10(1), 31–42.
- Mardiana, G. (2016). *Sistem Informasi Pendataan Rumah Tangga Miskin Pada Badan Pusat Statistik (Studi Kasus: Kabupaten Garut)*. Universitas Komputer Indonesia.
- Mardiana, W. (2017). *Laporan Praktik Kerja Lapangan Pada Sub Direktorat Pencegahan Kanker Alat Reproduksi Dan Penanggulangan Infertilitas Di Badan Kependudukan Dan Keluarga Berencana Nasional Pusat Jakarta*.
- Rustamin, Z., & Dewi, A. P. (2016). Sistem Pengarsipan Surat Masuk Dan Surat Keluar Pada Kantor Sekretariat DPRD Provinsi Sulawesi Tenggara Menggunakan Borland Delphi 7. *Simtek: Jurnal Sistem Informasi Dan Teknik Komputer*, 1(2), 165–172.
- Saminah, S. (2015). *Ulama dan guru ngaji sebagai prioritas utama penerima zakat fitrah (studi kasus di Desa Gaji Kecamatan Guntur Kabupaten Demak)*. UIN Walisongo.
- Syahroni, M. (2017). *Implementasi Data Mining untuk Pencarian Pola Asosiasi Penggunaan Internet Universitas Muhammadiyah Jember Menggunakan Algoritma Apriori*. Universitas Muhammadiyah Jember.
- Wursanto, I. (2006). *Kompetensi sekretaris profesional*. Yogyakarta: Andi.