

ANALISA DAN PERANCANGAN SISTEM PENDETEKSI PLAGIARISME SKRIPSI PADA STMIK INSAN PEMBANGUNAN MENGGUNAKAN METODE COSINE SIMILARITY

¹ Hery Herlambang, ² Jaka Suwita, ³ Beby Tiara

¹ Mahasiswa STMIK Insan Pembangunan

^{2,3} Dosen STMIK Insan Pembangunan

Program Studi Sistem Informasi, STMIK Insan Pembangunan,
Jln. Raya Serang Km.10 Bitung-Tangerang

heryherlambang.hh@gmail.com, jaka@ipem.ac.id, bebytiara27@gmail.com

ABSTRAK

Plagiarisme adalah penjiplakan atau pengambilan karya, karangan, pendapat, dan sebagainya dari orang lain serta menjadikannya seolah karya miliknya. Pelaksanaan pemeriksa plagirisme skripsi di STMIK Insan Pembangunan masih dilakukan secara manual. Petugas melakukan pengecekan skripsi satu persatu. Jumlah skripsi yang terus bertambah dan jumlah halaman skripsi yang banyak membuat proses pengecekannya nguras waktu dan tenaga.

Penelitian ini akan menggunakan metode cosinus similarity untuk melakukan perhitungan pesentase kemiripan antar dokumen. Sistem ini di buat menggunakan framework Laravel, Mysql. Metode pemodelannya menggunakan UML(Unified Modeling Language) dan pengembangan sistem menggunakan metode waterfall. Sistem ini memiliki fitur seperti : cek plagiarisme skripsi, kelola data skripsi, data mahasiswa, data tema skripsi, data jurusan, data stopword dan laporan data (skrisipsi, hasil pengecekan dan mahasiswa).

Kata Kunci: *Cosinus Similarity, Plagiarisme, skripsi, Perancangan Sistem, Pendeteksi.*

PENDAHULUAN

Latar Belakang Masalah

Perkembangan teknologi informasi saat ini semakin pesat, dibuktikan dengan peranan teknologi yang telah memasuki seluruh aspek kehidupan manusia dengan maksud membantu pekerjaan manusia. Kecepatan serta ketepatan dalam memproses data merupakan salah satu kelebihan yang dimiliki teknologi untuk menghasilkan informasi. Di dukung dengan adanya internet sebagai jaringan terbesar didunia memudahkan manusia untuk mendistribusikan ilmu dan pengetahuan. Pertukaran informasi yang begitu mudah membuat manusia mendapatkan informasi yang tepat dan relevan. Akan tetapi kemudahan tersebut malah berdampak negatif jika manusia tidak bijak dalam menyikapi perkembangan teknologi. Salah satu dampak negatif perkembangan teknologi adalah munculnya plagiarisme atau biasa disebut dengan plagiat.

Plagiarisme adalah perbuatan secara sengaja atau tidak sengaja dalam memperoleh atau mencoba memperoleh atau mencoba memperoleh kredit atau nilai untuk suatu karya ilmiah, dengan mengutip sebagian atau seluruh karya dan/atau karya ilmiah pihak lain yang diakui sebagai karya ilmiahnya, tanpa menyatakan sumber secara tepat dan memandai. Orang yang melakukan tindakan plagiat disebut plagiator. (PERMENDIKNAS No. 17 tahun 2010). Perilaku ini sangat merugikan, selain melanggar hak cipta plagiat juga dapat menurunkan daya kreatifitas seseorang. Tindakan plagiat juga terjadi dilingkungan perkuliahan, dimana skripsi yang

merupakan syarat dalam memperoleh kelulusan yang harusnya dibuat dengan sebenar-sebenarnya dan juga berdasarkan fakta yang ada. Kini skripsi bahkan menjadi karya yang rentan terhadap plagiat, akibat kurangnya pengecekan yang intensif dan waktu yang cukup lama dalam melakukan pengecekan skripsi secara manual. Seperti yang kita tahu bahwa skripsi merupakan karya ilmiah yang memiliki banyak jumlah halaman. Akibatnya jika kita ingin melakukan pengecekan kemiripan skripsi yang satu dengan yang lainnya secara manual, akan memakan waktu yang cukup lama. Ditambah lagi jumlah skripsi yang terus bertambah juga akan menambah pekerjaan jika harus dilakukan pengecekan secara manual.

STMIK Insan Pembangunan adalah salah satu perguruan tinggi yang ada di Kabupaten Tangerang, Provinsi Banten. Setiap tahun nya perguruan tinggi tersebut meluluskan sekitar 200-300 wisudawan-wisudawati (Data Wisuda 2018-2019). Ini menunjukan bahwa setiap tahunnya skripsi dikampus tersebut akan terus bertambah. Selama ini skripsi akan di periksa satu persatu oleh petugas untuk melakukan pengujian terhadap keaslian karya tersebut. Tentunya itu sangat tidak efektif jika dilakukan pengecekan secara manual. Menimbang berapa lama waktu yang dibutuhkan untuk membaca suatu skripsi dan berapa banyak petugas yang harus melakukan pengecekan.

Berdasarkan latar belakang tersebut, maka penulis tertarik untuk melakukan penelitian mengenai “Sistem Pendeteksi Plagiarisme Skripsi Menggunakan Metode

Cosine Similarity Pada STMIK Insan Pembangunan”.

Tujuan

Tujuan dari penelitian ini adalah adalah :

- a. Untuk mengetahui sistem yang digunakan oleh STMIK Insan Pembangunan dalam melakukan pengecekan plagiarisme pada skripsi yang saat ini berjalan.
- b. Untuk membangun sistem pendeteksi plagiarisme pada skripsi dan mengimplementasikan metode *cosine similarity*.

LANDASAN TEORI

Pengertian Sistem

Menurut Ahmad dan Munawir (2018:3-4), “Sistem adalah suatu susunan yang teratur dari kegiatan-kegiatan yang saling berkaitan dan susunan prosedur yang saling berhubungan, sinergi dari semua unsur-unsur dan elemen-elemen yang ada didalamnya, yang menunjang dan mempermudah kegiatan-kegiatan utama tercapai dari suatu organisasi atau pun kesatuan kerja”.

Berdasarkan pengertian di atas dapat disimpulkan bahwa sistem adalah kumpulan elemen atau komponen yang saling berhubungan dan saling terkait satu sama lain untuk mencapai tujuan tertentu.

Pengertian Analisa Sistem

Menurut Sukanto dan Shalahuddin (2016:18) menyatakan “analisa sistem adalah kegiatan untuk melihat sistem yang sudah berjalan, melihat bagian mana yang bagus dan tidak bagus, kemudian mendokumentasikan

kebutuhan yang akan dipenuhi dalam sistem yang baru”.

Dari definisi diatas, maka dapat disimpulkan analisa sistem adalah proses mengamati sistem informasi untuk mengidentifikasi dan mengevaluasi segala kekurangan yang terjadi agar sistem sesuai dengan yang diharapkan.

Pengertian Perancangan Sistem

Menurut Sukanto dan Shalahuddin (2016:23) menyatakan “Perancangan sistem merupakan upaya untuk mengonstruksi sebuah sistem yang memberikan kepuasan (mungkin informal) akan spesifikasi kebutuhan fungsional, memenuhi target, memenuhi kebutuhan secara implicit atau eksplisit dari segi performansi maupun penggunaan sumber daya, kepuasan pada proses desain dari segi biaya, waktu dan perangkat”.

Dari definisi diatas dapat disimpulkan bahwa Perancangan Sistem adalah perencanaan pembangunan sistem untuk memenuhi kebutuhan user dan memberikan gambaran yang jelas sistem apa yang akan dibangun.

Pengertian Pendeteksi

Menurut Kamus Besar Bahasa Indonesia (KBBI) *online*. “ Deteksi adalah usaha menemukan dan menentukan keberadaan, anggapan, atau kenyataan.”

Deteksi adalah suatu proses untuk memeriksa atau melakukan pemeriksaan terhadap sesuatu dengan menggunakan cara dan teknik tertentu. Deteksi dapat digunakan untuk berbagai masalah, misalnya dalam sistem pendeteksi plagiat, dimana sistem mengidentifikasi masalah-masalah yang

berhubungan dengan kasus tersebut. Tujuan dari deteksi adalah memecahkan suatu masalah dengan berbagai cara tergantung metode yang diterapkan sehingga menghasilkan sebuah solusi.

Pengertian Plagiarisme.

Plagiarisme adalah perbuatan secara sengaja atau tidak sengaja dalam memperoleh atau mencoba memperoleh atau mencoba memperoleh kredit atau nilai untuk suatu karya ilmiah, dengan mengutip sebagian atau seluruh karya dan/atau karya ilmiah pihak lain yang diakui sebagai karya ilmiahnya, tanpa menyatakan sumber secara tepat dan memandai. (PERMENDIKNAS No. 17 tahun 2010).

Menurut Pratama, dkk (2019).

“Plagiarisme merupakan tindakan mengambil gagasan, mengambil hasil riset, mengakuisisi hasil riset, dan meringkas suatu tulisan tanpa menyebutkan sumbernya.”

Cosine Similarity

Menurut Ariantini, dkk (2016). “*Cosine Similarity* adalah ukuran kesamaan antara dua buah vektor dalam sebuah ruang dimensi yang didapat dari nilai cosinus sudut dari perkalian dua buah vektor yang dibandingkan karena cosinus dari 0° adalah 1 dan kurang dari 1 untuk nilai sudut yang lain, maka nilai similarity dari dua buah vektor dikatakan mirip ketika nilai dari *cosine similarity* adalah 1.”

Menurut Hadi, dkk (2017). “Metode *cosine similarity* merupakan metode yang digunakan untuk menghitung similarity (tingkat kesamaan) antar dua buah objek. Secara umum

perhitungan metode ini didasarkan pada *vector space similarity measure*.”

Menurut Apriyanto dan Aribowo (2018). “Metode Cosine Similarity merupakan metode yang dapat melakukan pendeteksian secara cepat dan tepat, *cosine similarity* juga merupakan salah satu metode yang dapat digunakan untuk mengukur tingkat kemiripan dokumen. Sebelum melakukan perhitungan kemiripan, terhadap beberapa tahap processing text mining yang dilakukan, yaitu tokenizing, stoplist dan stemming. Term hasil proses tersebut direpresentasikan dalam model ruang vector dengan pembobotan TF-IDF, setelah itu, dilakukan perhitungan jarak *cosinus* antar vector menggunakan *cosine similarity*.”

Dalam Jurnal Apriyanto dan Aribowo (2018) Karhendana menjelaskan “secara umum, fungsi *similarity* adalah fungsi yang menerima dua buah objek dan mengembalikan nilai kemiripan (*similarity*) antara kedua objek tersebut berupa bilangan riil. Umumnya, nilai yang dihasilkan oleh fungsi *similarity* berkisar pada interval $[0...1]$. Untuk memetakan hasil fungsi tersebut pada interval $[0...1]$ dapat dilakukan normalisasi. Semakin besar hasil fungsi *similarity*, maka kedua objek yang dievaluasi dianggap semakin mirip. Sebaliknya, semakin kecil hasil similarity, maka kedua objek tersebut dianggap semakin berbeda. Pada fungsi yang menghasilkan nilai jangkauan $[0...1]$, nilai 1 melambangkan kedua objek persis sama, sedangkan nilai 0 melambangkan kedua objek sama sekali berbeda.”

Dari beberapa penjelasan diatas dapat disimpulkan bahwa cosine similarity adalah

metode yang digunakan untuk menghitung jarak dari 2 buah vektor dimana nilainya adalah berkisar 0 sampai 1. Jika nilainya mendekati 1 maka dokumen tersebut dikatakan mirip jika hasilnya mendekati 0 maka dokumen tersebut dikatakan tidak mirip. Berikut adalah rumus *cosine similarity*.

$$\text{Similarity} = \text{Cos}(\theta) = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Keterangan :

A = Vektor A

B = Vektor B

A_i = Bobot *term* i dalam blok A_i

B_i = Bobot *term* i dalam blok B_i

i = index elemen dalam vektor

n = jumlah elemen vektor

Metode Waterfall.

Menurut Sukamto dan Shalahuddin (2015:28). “Model SDLC air terjun (*waterfall*) sering juga disebut model skuensial linier (*sequential linier*) atau alur hidup klasik (*classic life cycle*). Model air terjun (*waterfall*) menyediakan pendekatan air hidup perangkat lunak secara sekuensial atau terurut dimulai dari analisis, desain, pengkodean, pengujian dan tahap pendukung (*support*).”

Unified Modeling Language Waterfall (UML).

Pengertian *Unified Modeling Language* (UML) menurut Sukamto dan Shalahuddin (2016:137) “merupakan bahasa visual untuk pemodelan dan komunikasi mengenai sebuah sistem dengan menggunakan diagram dan teks-teks pendukung”.

Diagram yang akan digunakan adalah sebagai berikut :

- a. *Usecase Diagram*
- b. *Activity Diagram*
- c. *Squence Diagram*
- d. *Class Diagram*

METODE PENELITIAN

Langkah-langkah penelitian

Langkah-langkah yang akan dilakukan dalam penelitian sistem pendeteksi plagiarisme pada skripsi di STMIK Insan Pembangunan yaitu :

- a. Mencari dan membaca literatur mengenai proses perhitungan nilai *Cosinus Similarity* pada dokumen.
- b. Melakukan observasi langsung ke lokasi untuk mengetahui sistem yang sedang berjalan.
- c. Mempelajari proses pengelolaan data skripsi guna mengetahui kelebihan dan kekurangan sistem yang sedang berjalan.
- d. Melakukan pengumpulan data, seperti : data skripsi, data mahasiswa, dan data petugas.
- e. Mencari solusi agar masalah yang terjadi di STMIK Insan Pembangunan dapat teratasi.
- f. Membuat perancangan sistem pendeteksi plagiarisme pada skripsi di STMIK Insan Pembangunan.

Data dan Sumber Data

Langkah-langkah yang akan dilakukan dalam penelitian sistem pendeteksi plagiarisme skripsi di STMIK Insan Pembangunan yaitu :

Dalam penelitian ini membutuhkan dua jenis data berikut:

- a. Data Primer

Data primer meliputi data yang berasal langsung dari objek penelitian. Seperti: informasi mengenai sistem yang berjalan, data orang terlibat dalam pengecekan skripsi, data mengenai kendala yang terjadi selama ini dan data pendukung lainnya yang diperlukan. Data primer didapatkan melalui Observasi langsung ke STMIK Insan Pembangunan dan Wawancara kepada petugas yang terlibat dalam proses pengecekan skripsi.

b. Data Skunder

Data skunder meliputi data yang didapatkan tidak secara langsung dari objek penelitian, bisa berasal dari buku atau pun sumber referensi lainnya. Seperti : cara melakukan perhitungan kemiripan dokumen skripsi, cara membangun sebuah sistem dan data sekunder lainnya yang dibutuhkan. Data sekunder diperoleh dari buku, jurnal, buku dan internet.

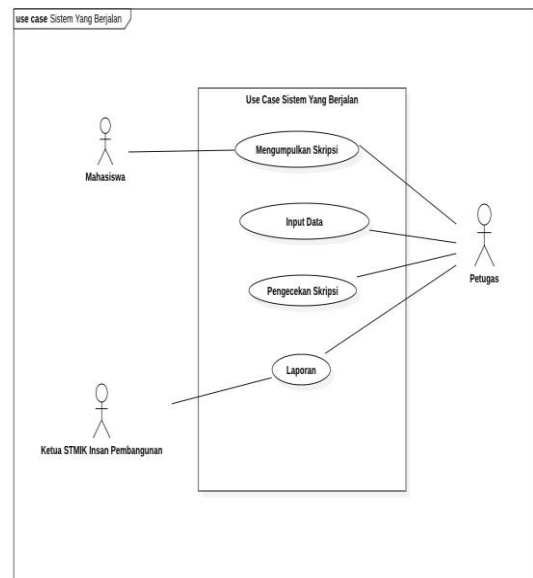
Tempat dan Waktu Penelitian

Penelitian ini dilakukan pada STMIK Insan Pembangunan dan waktu penelitiannya dilakukan selama 4 (empat) bulan, yaitu: Mulai bulan Juni 2020 sampai dengan September 2020.

PEMBAHASAN

Sistem Yang Berjalan

Adapun sistem yang saat ini berjalan di STMIK Insan Pembangunan adalah sebagai berikut.



GAMBAR 1 Usecase Sistem Yang Berjalan

Deskripsi Usecase:

- Mahasiswa mengumpulkan skripsi kepada petugas.
- Petugas melakukan input data ke excel.
- Proses pengecekan yang dilakukan oleh petugas ketika data skripsi telah di input.
- Proses pelaporan dari petugas kepada ketua STMIK mengenai data skripsi dan skripsi yang terindikasi sebagai plagiarisme.

Rancangan Sistem Usulan

- Perhitungan *Cosine Similarity*

Langkah-langkah perhitungan *cosine similarity*

- Menentukan Q dan D.

Q adalah dokumen yang akan di uji, sedangkan D adalah dokumen lain yang akan menjadi pembanding.

- Case Folding

Case Folding adalah proses mengubah semua kata menjadi huruf kecil.

Sebagai contoh kalimat "INFORMASI merupakan data yang telah dikelola menjadi bentuk yang berguna untuk mengambil

KEPUTUSAN.” akan diubah menjadi “informasi merupakan data yang telah dikelola menjadi bentuk yang berguna untuk mengambil keputusan.”

3. *Filtering*

Filtering adalah proses menyaring kalimat dari dokumen agar tidak memiliki spesial karakter seperti: seperti tanda titik, tanda koma, tanda seru, dan sebagainya.

4. *Tokenizing*

Tokenizing adalah proses memecah kalimat kedalam kata-kata. Berikut contoh hasil kalimat yang sudah di lakukan tokenizing.

Tabel 1 Hasil *Tokenizing*

No	Kata
1	informasi
2	merupakan
3	data
4	yang
5	telah
6	dikelolah
7	menjadi
8	bentuk
9	yang
10	berguna
11	untuk
12	mengambil
13	keputusan

5. *Remove Stopword*

Stopword adalah kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh *stopword* : “agar”, ”amat”, ”agak”, ”adalah”, ”atau”, “bagaimanapun” dan sebagainya.

Pada proses ini akan menghapus semua kata yang termasuk kedalam *stopword*.

6. *Stemming*

Stemming adalah proses menghilangkan imbuhan (awalan dan akhiran) sehingga semua kata menjadi kata dasar. Sebagai contoh: “pengumpulan” menjadi “kumpul”, “belajar” menjadi “ajar”, “memakan” menjadi “makan” dan lain sebagainya.

7. *Hitung Nilai Cosine Similarity*

Setelah melakukan beberapa tahap tersebut baru kemudian dilakukan perhitungan nilai *cosine similarity*.

Misalkan ada dua kalimat yang ingin dibandingkan yang masing-masing nilainya sebagai berikut.

Q = “informasi merupakan data yang telah dikelola menjadi bentuk yang berguna untuk mengambil keputusan.”

D1 = “informasi adalah data yang telah dikelola dan diproses untuk memberikan arti dan memperbaiki proses pengambilan keputusan.”

Setelah melakukan tahap 1-6. Kemudian kita dapat membuat tabel sebagai berikut.

Tabel Perhitungan										
No.	Term	tf		df	idf	tf * idf		Q*D	(TF/IDF(Q,D)) ²	
		Q	D1		Log(n/df)+1	Q	D1	Q*D1	Q ²	D1 ²
1	ambil	1	1	2	1	1	1	1	1	1
2	arti	0	1	1	1.30103	0	1.30103	0	0	1.692679
3	baik	0	1	1	1.30103	0	1.30103	0	0	1.692679
4	bentuk	1	0	1	1.30103	1.30103	0	0	1.692679	0
5	beri	0	1	1	1.30103	0	1.30103	0	0	1.692679
6	data	1	1	2	1	1	1	1	1	1
7	guna	1	0	1	1.30103	1.30103	0	0	1.692679	0
8	informasi	1	1	2	1	1	1	1	1	1
9	jadi	1	0	1	1.30103	1.30103	0	0	1.692679	0
10	kelola	1	1	2	1	1	1	1	1	1
11	proses	0	2	1	1.30103	0	2.60206	0	0	6.770716
12	putus	1	1	2	1	1	1	1	1	1
13	rupa	1	0	1	1.30103	1.30103	0	0	1.692679	0
SUM								5	11.770716	16.848753
SQRT(SUM(TF/IDF(Q,D)) ²)									3.430848	4.104723

GAMBAR 2 Perhitungan Cosine Similarity

IDF = Log(n/df)+1

- TF : Jumlah kemunculan suatu kata dalam 1 dokumen.
- DF : Jumlah dokumen yang mengandung kata tersebut.
- n : Jumlah dokumen.

TF/IDF = tf * idf

- TF : Jumlah kemunculan suatu kata dalam 1 dokumen.
- IDF : Inverse Document Frequency

$$\text{COS}(A,B) = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

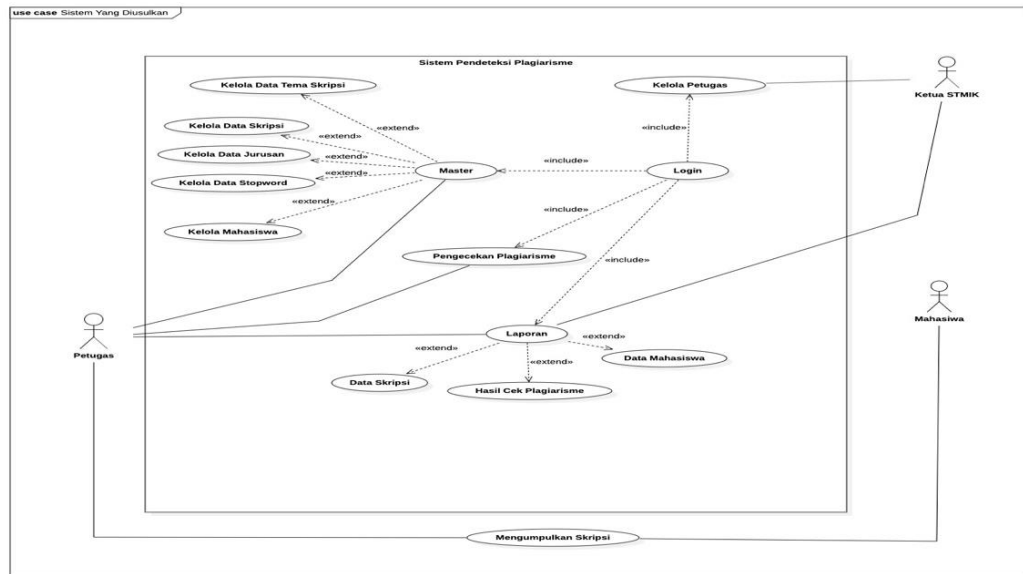
$$\text{COS}(Q,D1) = \frac{5}{3.430848 \times 4.104723}$$

$$\text{COS}(Q,D1) = 0.35504601$$

$$\text{COS}(Q,D1) = 35.5 \%$$

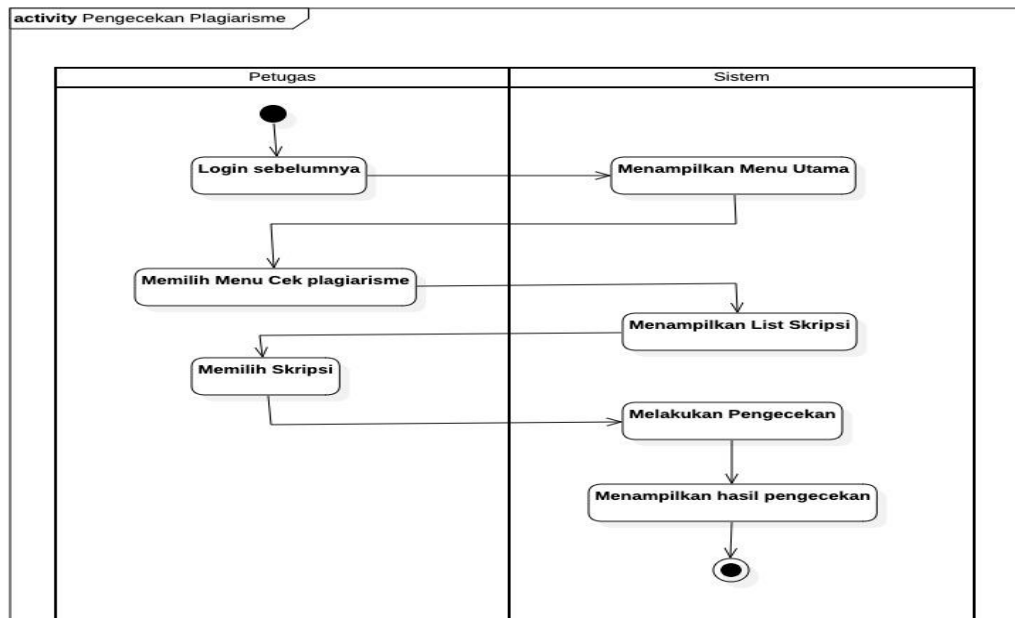
Desain Sistem Yang Diusulkan

1. Activity Diagram Cek Plagiarisme



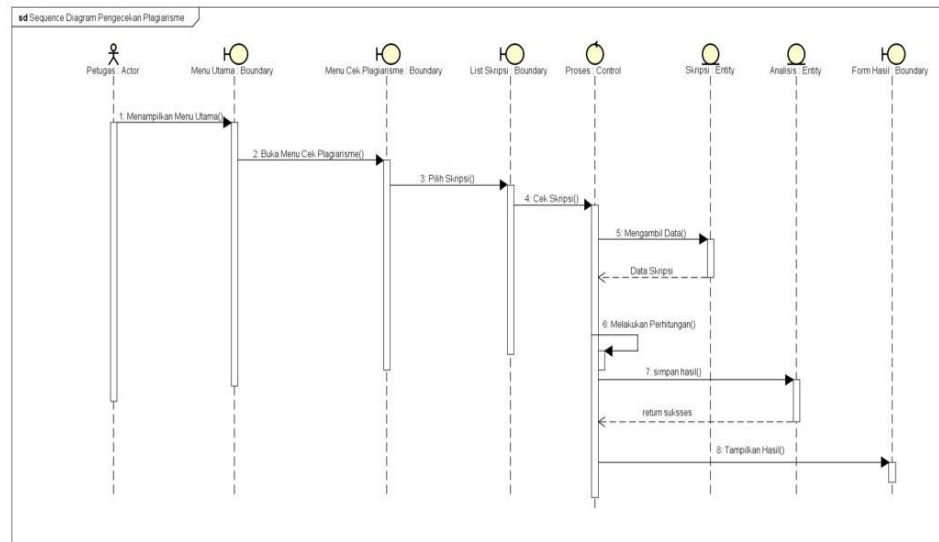
GAMBAR 3 Usecase Sistem Yang Diusulkan

2. Activity Diagram Cek Plagiarisme



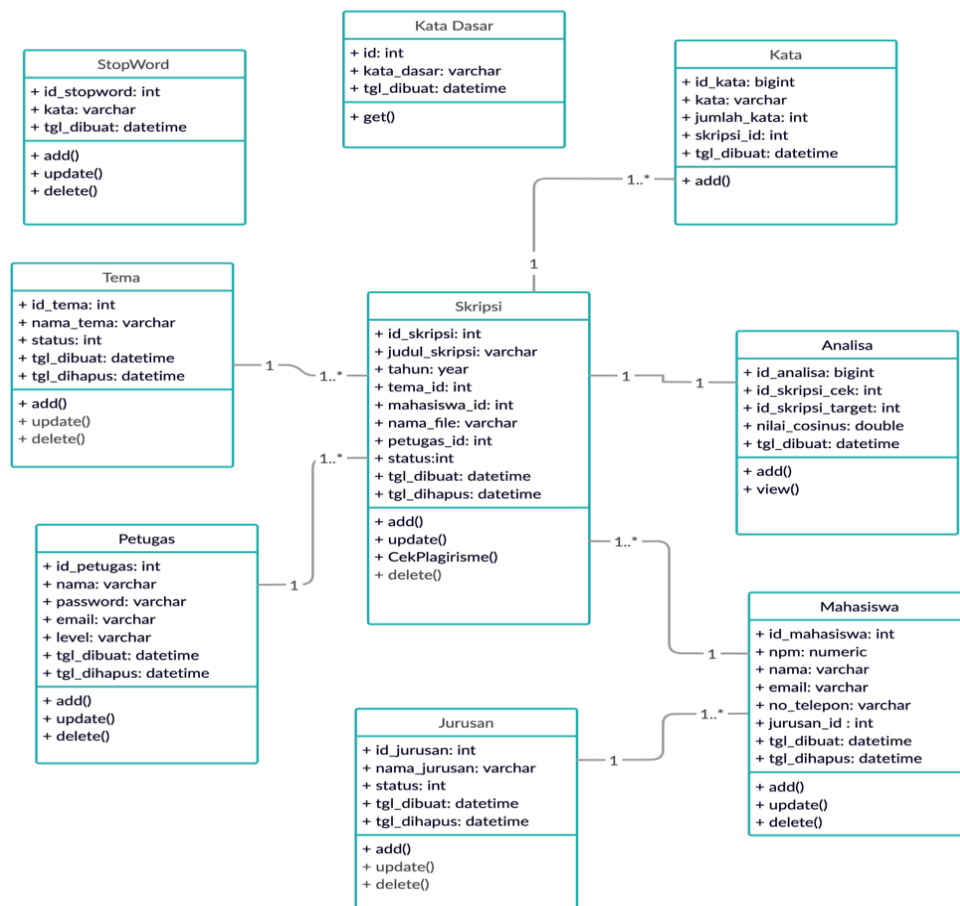
GAMBAR 4 Activity Diagram Cek Plagiarisme

3. Sequence Diagram Cek Plagirisme



GAMBAR 5 Activity Sequence Cek Plagirisme

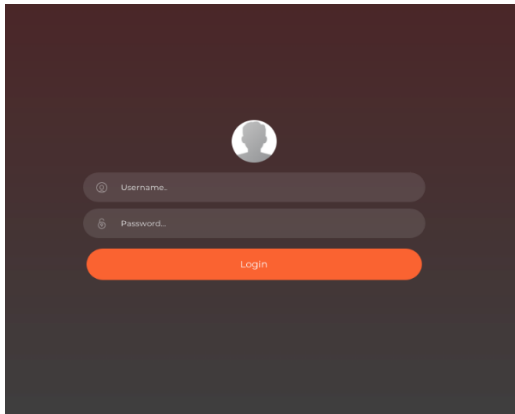
4. Class Diagram



GAMBAR 6 Class Diagram

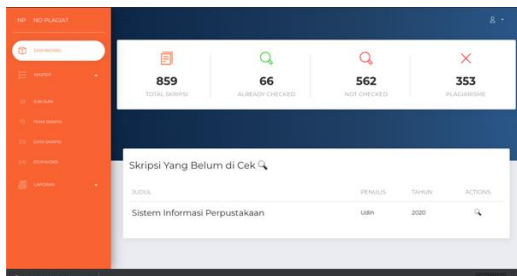
5. Desain Tampilan

Form login



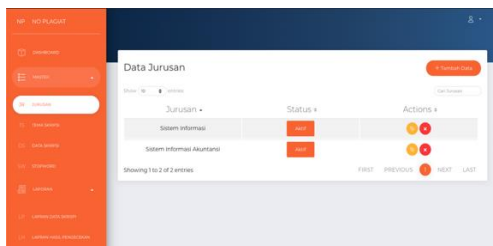
GAMBAR 7 Form Login

Halaman Utama:



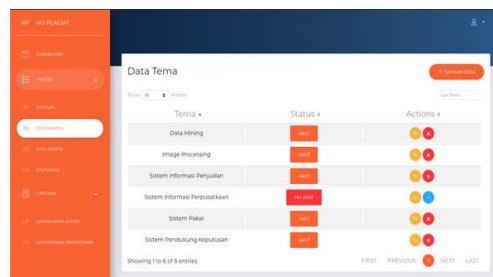
GAMBAR 8 Halaman Utama

Master Jurusan



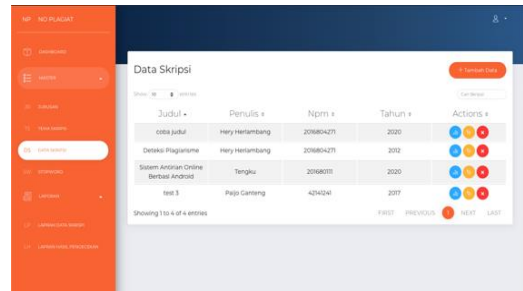
GAMBAR 9 Master Jurusan

Master Tema Skripsi



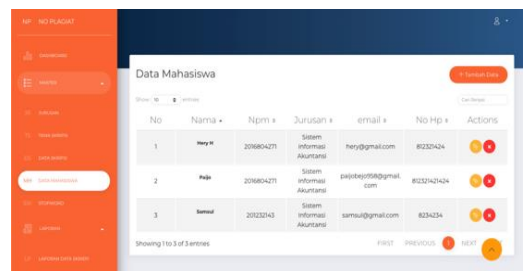
GAMBAR 10 Master Tema Skripsi

Master Skripsi



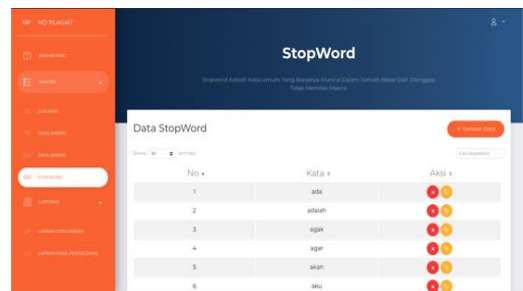
GAMBAR 11 Master Skripsi

Master Mahasiswa



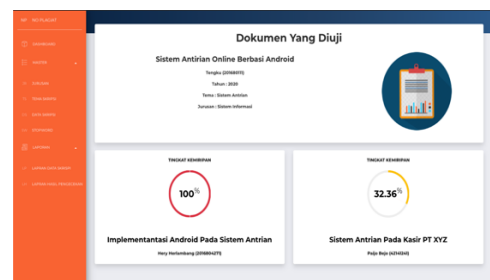
GAMBAR 12 Master Mahasiswa

Master Stopword



GAMBAR 13 Master Mahasiswa

Hasil Pengecekan



GAMBAR 14 Hasil Pengecekan

6. Laporan Data Skripsi

No	Judul	Mahasiswa	Npm	Tahun	Tanggal
1	ANALISA DAN PERANCANGAN SISTEM INFORMASI APLIKASI E-HENRI PADA RUANG MAKAN BERBASIS ANDROID	Hery H	201804271	2020	2020-08-18 00:38:27
2	ANALISA DAN PERANCANGAN SISTEM PERPUSTAKAAN BERBASIS ANDROID	Paji	201804271	2020	2020-08-20 01:08:45
3	Sistem Aplikasi Online Berbasis Android	Samud	20202143	2020	2020-08-20 01:10:29

GAMBAR 15 Laporan Data Skripsi

7. Laporan Data Hasil Pengecekan

No	Judul doc	Judul target	Nilai Cosine	Tanggal
1	Pety YE ANALISA DAN PERANCANGAN SISTEM INFORMASI APLIKASI E-HENRI PADA RUANG MAKAN BERBASIS ANDROID	(Pety) ANALISA DAN PERANCANGAN SISTEM PERPUSTAKAAN BERBASIS ANDROID	66 %	2020-08-12 22:52:52
2	(Pety) ANALISA DAN PERANCANGAN SISTEM PERPUSTAKAAN BERBASIS ANDROID	Pety YE ANALISA DAN PERANCANGAN SISTEM INFORMASI APLIKASI E-HENRI PADA RUANG MAKAN BERBASIS ANDROID	66 %	2020-08-17 02:22:45
3	(Pety) ANALISA DAN PERANCANGAN SISTEM PERPUSTAKAAN BERBASIS ANDROID	(Samud) Sistem Aplikasi Online Berbasis Android	31 %	2020-08-18 21:29:37
4	(Samud) Sistem Aplikasi Online Berbasis Android	(Pety) ANALISA DAN PERANCANGAN SISTEM PERPUSTAKAAN BERBASIS ANDROID	31 %	2020-08-18 20:40:06

GAMBAR 16 Laporan Hasil Pengecekan

PENUTUP

Kesimpulan

Setelah melakukan penelitian dengan judul “Analisa Dan Perancangan Sistem Pendeteksi Plagiarisme Skripsi Pada STMIK Insan Pembangunan Dengan Metode *Cosinus Similarity*. Adapun kesimpulan nya adalah sebagai berikut.

- a. Sistem pengecekan yang berjalan pada STMIK Insan Pembangunan masih secara manual dan hanya melakukan pengecekan pada skripsi yang judulnya mirip. Petugas memeriksa dokumen satu-persatu untuk mengecek dokumen terdapat kesamaan

atau tidak. Hal tersebut sangat memakan waktu dan tenaga.

- b. Sistem yang di rancang pada pada STMIK Insan Pembangunan berbasis website dan menggunakan *framework* Laravel. Sistem menghitung nilai kemiripan dokumen skripsi dengan metode *cosinus similary*. Metode pemodelannya menggunakan UML(*Unified Modeling Language*) dan pengembangan sistem menggunakan metode *waterfall*. Sistem ini memiliki fitur seperti : cek plagiarisme skripsi, kelola data skripsi, data mahasiswa, data tema skripsi, data jurusan, data *stopword* dan laporan data (skripsi, hasil pengecekan dan mahasiswa).

- c. Kekurangan sistem ini hanya dapat memberikan persentasi kemiripan dokumen tidak dengan detail memberikan informasi pada bagian yang memiliki kemiripan.

Saran

Penulis menyadari bahwa sistem ini masih jauh dari kata sempurna, maka dari itu penulis berharap peneliti selanjutnya dapat menutupi kurang-kekurangan tersebut dan di jadikan pertimbangan dalam membangun sistem *text mining* yang lebih baik.

Selanjutnya saran yang dapat penulis berikan, yaitu :

- a. Menggunakan bahasa pemrograman yang lebih *powerfull* seperti *python*.
- b. Melakukan pemenggalan dokumen skripsi per bab. karena skripsi terdiri dari banyak halaman maka perlu dilakukan pemenggalan setiap babnya. Dengan begitu

diharapkan akan membuat proses lebih ringan dan lebih spesifik pada target yang ingin dibandingkan.

- c. Dapat memberikan hasil pengecekan secara detail.

DAFTAR PUSTAKA

- Ahmad, L., & Munawir. (2018). *Sistem Informasi Manajemen : Buku Referensi*. Banda Aceh: KITA.
- Apriyanto, I. A., & Aribowo, E. (2016). Rancang Bangun Aplikasi Pengecekan Kemiripan Judul Skripsi Dengan Metode Cosine Similarity (Studi Kasus : Program Studi Teknik Informatika UAD), *Jurnal Sarjana Teknik Informatika Universitas Ahmad Dahlan*, 6.
- Ariantini, D. A., Lumenta, A. S., & Jacobus, A. (2016). Pengukuran Kemiripan Dokumen Teks Bahasa Indonesia Menggunakan Metode Cosine Similarity. *E-Journal Tekni Informatika Universitas Sam Ratulangi Manado*, 9.
- Hadi, C., & Ma'arif, M. R. (2017). Implementasi Cosine Similarity Dalam Aplikasi Pencarian Ayat Al-Qur'an Berbasis Android. *Jurnal Teknologi Informasi dan Komunikasi STMIK Jenderal A. Yani Yogyakarta*, 6.
- Pratama, R. P., Faisal, M., & Hanani, A. (2019). Deteksi Plagiarisme pada Artikel Jurnal Menggunakan Metode Cosine Similarity. *SMARTICS Journal*, 5.
- Sukanto R. A., & Shalahuddin, M. (2016). *Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek*. Bandung: Informatika.
- PERMENDIKNAS No. 17 tahun 2010 tentang Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi.
- Kamus Besar Bahasa Indonesia Online. Tersedia di: <https://kbbi.web.id/deteksi>. Diakses 5 Juni 2020