

# Penerapan Algoritma Cosine Similarity Pada Sistem Pengajuan Judul Tugas Akhir Berbasis Web

Daniel<sup>1</sup>, Wilda Susanti<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Institut Bisnis dan Teknologi Pelita Indonesia  
e-mail: <sup>1</sup>daniel@student.pelitaindonesia.ac.id, <sup>2</sup>wilda@lecturer.pelitaindonesia.ac.id

## Abstrak

Karya Ilmiah merupakan syarat setiap mahasiswa untuk menyelesaikan pendidikannya di suatu Perguruan Tinggi. Tetapi, suatu karya ilmiah yang bukan atau meniru ide asli dari pemikiran peneliti lainnya dianggap sebagai *plagiarism*, yaitu tindakan yang sengaja dilakukan oleh individu untuk meniru atau mengakui sebagai karya sendiri. Tujuan penelitian ini merancang suatu system yang dapat mendeteksi kemiripan judul Tugas Akhir. Metode yang digunakan menggunakan algoritma *Term Frequency – Inverse Document Frequency* (TF-IDF) untuk pembobotan kata dan *Cosine Similarity* untuk melakukan perhitungan kemiripan antar judul tugas akhir. Hasil pengujian menunjukkan bahwa sistem pengecekan judul tugas akhir menggunakan algoritma *Term Frequency – Inverse Document Frequency* (TF-IDF) dan *Cosine Similarity* dapat mendeteksi persentase tingkat kemiripan judul tugas akhir. Aplikasi yang dibangun berbasis web yang berguna memudahkan prosedur pengajuan dan cek plagiarisme.

**Kata kunci:** *Judul Tugas Akhir, TF-IDF, Cosine Similarity, Website*

## Abstract

*Scientific work is a requirement for every student to complete his education at a university. However, a scientific work that does not or imitate the original ideas of the thoughts of other researchers is considered plagiarism, namely actions that are deliberately taken by individuals to imitate or acknowledge their own works. The purpose of this research is to design a system that can detect the similarity of the title of the Final Project. The method used is the Term Frequency - Inverse Document Frequency (TF-IDF) algorithm for weighting words and Cosine Similarity to calculate the similarity between the titles of the final project. The test results show that the checking system for the title of the final project using the Term Frequency - Inverse Document Frequency (TF-IDF) algorithm and Cosine Similarity can detect the percentage level of the similarity level of the final project title. A web-based application that is useful in facilitating the procedure for submitting and checking for plagiarism..*

**Keywords:** *Title of Thesis, TF- IDF, Cosine Similarity, Website*

---

## 1. Pendahuluan

Skripsi adalah salah satu syarat kelulusan yang ditetapkan di perguruan tinggi, menjadi satu syarat yang harus dipenuhi oleh calon sarjana untuk bisa mendapatkan gelar sarjana [1].

Pengajuan judul RUP merupakan tahap awal dalam proses pembuatan skripsi di Fakultas Ilmu Komputer Institut Bisnis dan Teknologi Pelita Indonesia (FILKOM IBTPI). Pengajuan judul masih dilakukan secara manual. Dimana mahasiswa diharuskan untuk datang ke BAAK untuk mengajukan judul RUP dan menunggu kurang lebih 3 hari untuk mendapatkan hasil dari pihak Koordinator maupun Penelaah. Jika judul tugas akhir yang diajukan ditolak oleh pihak Koordinator maupun Penelaah maka mahasiswa harus mengajukan ulang kembali. Pengajuan tersebut kurang efisien dan memakan waktu yang lama. Pengajuan judul yang masih menggunakan

form kertas dan antara satu pihak dengan pihak lain yang terlibat dalam proses ini harus bertatap muka secara langsung, sehingga membutuhkan waktu untuk bisa menyelesaikan tahap ini.

Selain proses pengajuan judul yang masih manual, mahasiswa juga kesulitan dalam menentukan judul yang akan diajukan, dikarenakan judul yang diajukan mahasiswa tersebut sudah pernah dibuat atau diteliti. Hal ini terjadi karena tidak adanya sistem yang menyediakan informasi mengenai judul-judul yang telah pernah dibuat dan tidak ada system untuk mengecek seberapa besar kemiripan terhadap judul-judul yang telah ada, sehingga terjadinya tindakan plagiarisme yang dilakukan oleh mahasiswa.

Tindakan plagiarisme merupakan sebuah bentuk tindakan kejahatan. Di Indonesia kasus-kasus plagiarisme di dunia akademik bukan lagi isu yang baru. Menurut [2] bahwa tindakan plagiarisme (menjiplak) karya tulis ilmiah orang lain yang kemudian diakui sebagai karya sendiri sebenarnya sudah lama berlangsung di berbagai institusi perguruan tinggi di negeri ini. Akibat dari Kegiatan plagiarisme menjadikan mahasiswa memetakan pola berpikir kreatif dan kritis sehingga mereka akan cenderung lebih mencari kemudahan yang akan mengakibatkan kebodohan terhadap generasi penerus bangsa.[3]

Untuk mendeteksi plagiarisme beberapa peneliti seperti [4] dengan judul Penerapan Algoritma Winnowing pada system Rekomendasi Penentuan Dosen Pembimbing Skripsi, [5] Rancang Bangun Aplikasi Halaman Web Berdasarkan Content Menggunakan Metode Cosine Similarity dan [6] Deteksi Kemiripan Dokumen Teks Menggunakan Algoritma Manber, pada penelitian ini program yang di buat hanya membandingkan satu teks dengan teks lainnya tanpa adanya memproses data teks yang ada di dalam database.

Sistem informasi berbasis web adalah seperangkat komponen yang saling berhubungan yang berfungsi untuk mengumpulkan, memproses, menyimpan dan mentransferkan informasi dalam bentuk teks, gambar, suara, dan informasi yang dipresentasikan dalam bentuk hypertext serta dapat diakses oleh perangkat lunak untuk mendukung pembuatan kegiatan dalam organisasi dalam mencapai tujuan [7]. Menurut Suwanto Raharjo S.Si, M.Kom, Web merupakan salah satu layanan internet yang paling banyak digunakan dibanding dengan layanan lain seperti ftp, gopher, news atau bahkan email[8]. Web menjadi sangat terkenal dan perkembangannya sangatlah pesat, tetapi umumnya web seperti ini masih bersifat statis dan tidak dapat mengelola data. Untuk dapat mengelola data dalam bentuk database, maka perlu dikembangkan sistem informasi berbasis web[9]. Penggunaan komputer dalam bidang pendidikan antara lain untuk media pembelajaran berbantuan komputer termasuk e-learning, alat bantu pengolahan data akademik, dan media penyampaian informasi [10].

Atas dasar permasalahan diatas maka dirancang sebuah sistem yang dapat mendeteksi kemiripan untuk menghindari penolakan karena adanya judul yang sudah pernah di buat atau diteliti dan membangun sistem pengajuan judul RUP berbasis web sehingga dapat memudahkan mahasiswa dalam mengajukan judul tugas akhir, tanpa mengharuskan mahasiswa untuk datang ke kampus dan mahasiswa dapat melihat sampai mana proses pengajuan judul tugas akhir melalui sistem yang akan dibangun.

## 2. Metode Penelitian

Dalam melakukan penelitian ini, metode yang digunakan adalah metode SDLC, memiliki beberapa tahapan :

- (1) **Analysis**, Pada tahap ini, peneliti melakukan analisis pada judul tugas akhir dan proses pengajuan judul tugas akhir mahasiswa di FILKOM IBTPI untuk mengidentifikasi masalah dalam membangun sistem pengajuan judul tugas akhir berbasis web menggunakan algoritma cosine similarity.
- (2) **Design**, Pada tahap ini, peneliti mengubah kebutuhan – kebutuhan yang peneliti dapat dari tahap analysis ke dalam bentuk “blueprint” atau rancangan sistem sebelum memulai tahap coding dimulai. Tahap ini diperlukan agar user dapat dengan mudah mengerti dan memakai program yang akan dibuat.
- (3) **Coding**, Pada tahap ini, peneliti membuat program atau aplikasi berdasarkan analysis yang telah dilakukan dan perancangan design yang sudah dibuat sebelumnya. Sistem informasi pengajuan

judul tugas akhir dan algoritma cosine similarity dibangun dengan menggunakan bahasa pemrograman PHP.

(4) **Testing**, Pada tahap ini, peneliti melakukan testing atau uji coba terhadap system yang sudah di buat, untuk mencari kesalahan atau error yang terlewatkan oleh peneliti.

(5) **Maintenance**, Merupakan tahapan terakhir dari siklus pengembangan dan dilakukan setelah sistem dipergunakan, kegiatan yang dilakukan berupa melakukan perbaikan apabila terdeteksi kesalahan sistem saat telah di gunakan, dan melakukan update terhadap sistem agar dapat mengikuti perkembangan yang terjadi.

Metode pengklasifikasian yang digunakan pada sistem ini adalah dengan cara membandingkan kesamaan atau similaritas antara judul dokumen dengan menggunakan Algoritma Cosine Similarity dengan beberapa tahapan :

(1) Dokumen A adalah dokumen yang akan diukur tingkat kemiripannya. Sedangkan dokumen pembanding merupakan dokumen yang digunakan untuk mengukur kemiripan dari dokumen uji. Sedangkan Dokumen B digunakan digunakan sebagai bahan untuk membandingkan kemiripan antar dokumen.

(2) Preprocessing, Tahapan yang dilakukan setelah menentukan dokumen uji dan dokumen pembanding selanjutnya harus dilakukan preprocessing terlebih dahulu. Berikut merupakan tahapan preprocessing:

#### A. Case Folding

Case folding merupakan proses pertama dari rangkaian preprocessing dokumen. Dalam proses ini mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf a sampai dengan z yang diterima.

#### B. Stopword removal

Tahap stopword removal adalah tahap menghilangkan katakata yang tidak penting dari teks. Stopword adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-words.

#### C. Stemming

Tahap stemming adalah tahap mencari root kata dari tiap kata hasil filtering. Pada tahap ini dilakukan proses pengembalian berbagai bentukan kata ke dalam suatu representasi yang sama. Pencarian kata dasar dilakukan dengan menghilangkan semua imbuhan dari kata, baik itu awalan, sisipan, maupun akhiran.

#### D. Tokenizing

Tahap Tokenizing adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya.

(3) Hitung Similaritas Consine Similarity dengan menggunakan rumus:

$$\text{Cos } a = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sum_{i=1}^n (A_i)^2 \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Keterangan :

A = Vektor A yang akan dibandingkan kemiripan

B = Vektor B yang akan dibandingkan kemiripan

A.B = Dot antara Vektor A dan Vektor B

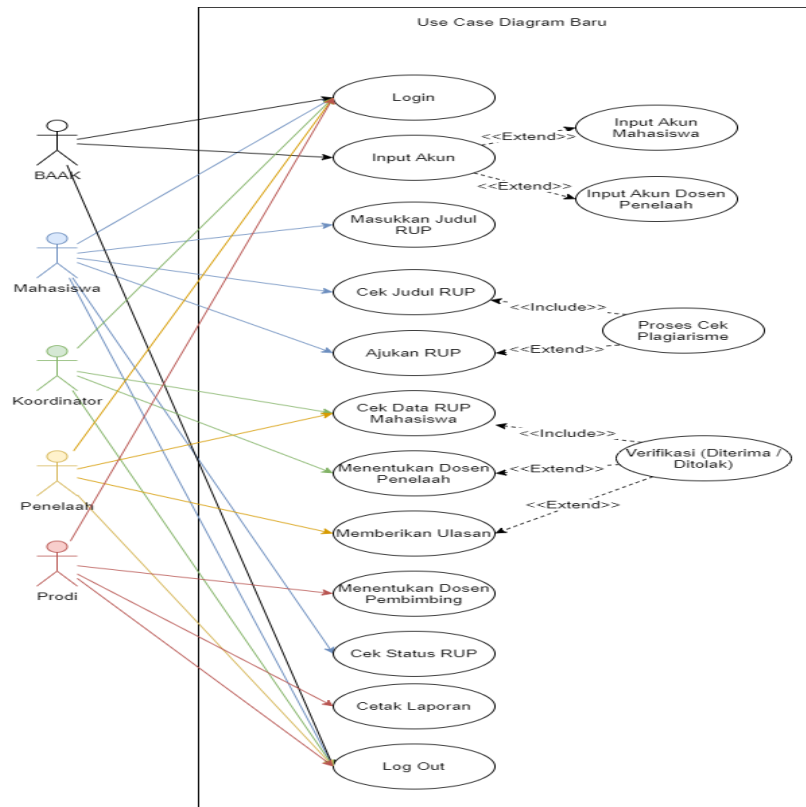
|A| = Panjang Vektor A

|B| = Panjang Vektor B

|A|B| = *Cross Product* Antara |A| dan |B|

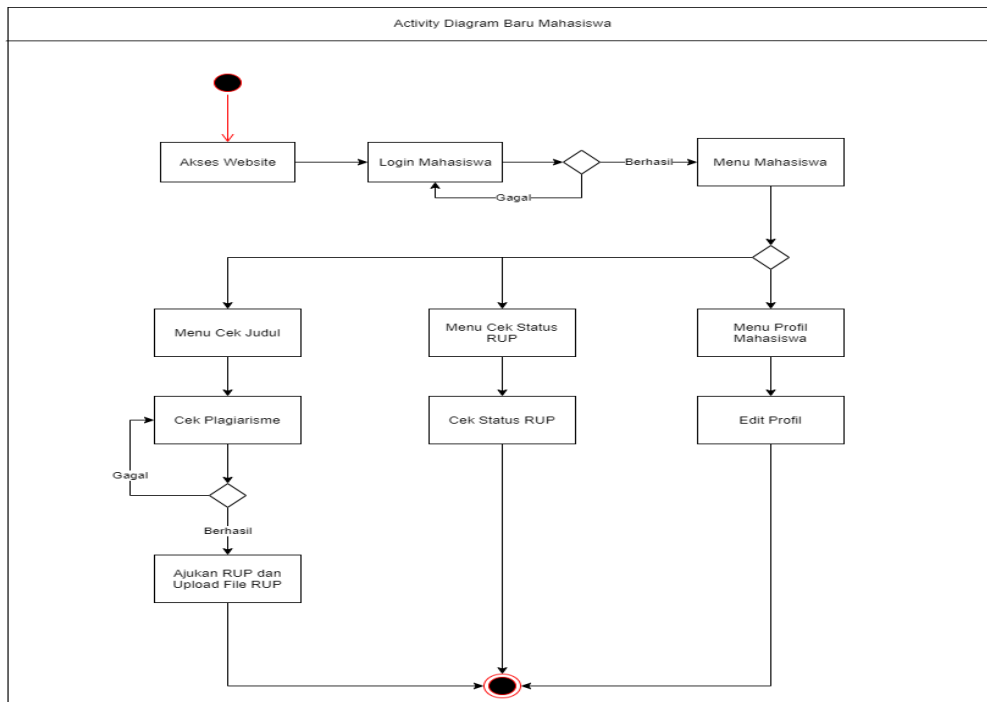
### 3. Hasil dan Pembahasan

Use case diagram menjelaskan secara umum lingkungan sistem. Gambar 2 menggambarkan rancangan use case diagram sistem baru dimana Mahasiswa, BAAK, Koordinator, Dosen Penelaah dan Prodi menjalankan aplikasi. Mahasiswa melakukan login untuk mengajukan judul. Koordinator dapat melakukan login lalu mengecek judul yang sudah diajukan oleh mahasiswa dan menentukan dosen penelaah. Dosen penelaah dapat melakukan login lalu dapat melakukan review judul mahasiswa yang telah diajukan. Setelah diterima oleh dosen penelaah maka prodi dapat menentukan dosen pembimbing untuk mahasiswa.



Gambar 1. Use Case Diagram

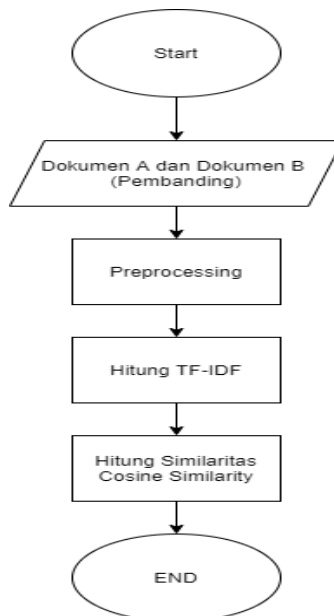
Pada Gambar 1 memperlihatkan aliran kerja yang terdapat pada sistem baru. Yang dimulai dengan start state yang menunjukkan awal aktifitas sistem. Mahasiswa mengakses website untuk melakukan login dengan id dan password masing – masing mahasiswa lalu menuju ke menu mahasiswa dimana terdapat menu cek judul, menu cek status rup dan menu profil mahasiswa.



Gambar 2. Activity Diagram

#### 4. Pembahasan

Penerapan perhitungan Algoritma *Term Frequency – Inverse Document Frequency* (TF-IDF) dan *Cosine Similarity* terletak pada fitur pengajuan judul yang dapat diakses oleh mahasiswa. Algoritma yang digunakan oleh peneliti adalah Algoritma *Term Frequency – Inverse Document Frequency* (TF-IDF) untuk menghitung bobot kata dan *Cosine Similarity* untuk menghitung tingkat kesamaan judul. Berikut alur perhitungan :



Gambar 3. Alur Perhitungan Algoritma

Berikut adalah contoh kasus mencari tingkat persentasi kemiripan judul.

### 1. Data Judul

Judul yang digunakan sebagai judul query adalah “Perancangan Sistem Informasi Akademik Didukung Pemberdayaan Kartu Pelajar Dan Guru Sebagai Media Akses Informasi (SMA Nurul Falah Pekanbaru)”. Judul pembanding dapat dilihat pada tabel 1.

Tabel 1. Tabel Judul Pembanding

ID	Judul
D1	Rancang Bangun Aplikasi Pembelajaran Bahasa Inggris Berbasis Mobile (Android) Untuk Mata Kuliah English For Office
D2	Sistem Pakar Mendiagnosa Kerusakan Sepeda Motor Injeksi Menggunakan Metode Forward Chaining Berbasis WEB
D3	Sistem Pakar Mendiagnosa Penyakit Ikan Hias Air Tawar Menggunakan Metode Forward Chaining Dan Theorema Bayes Berbasis WEB
D4	Rancang Bangun E-Commerce Mobil Tangki Dan Truk Menggunakan Api (Application Programming Interface) Dan SMS Gateway
D5	Rancang Bangun Sistem Informasi Presensi Dan Monitoring Menggunakan Radio Frequency Identification (RFID) Pada Vihara PUSDIKLAT Bumi Suci Maitreya Pekanbaru
D6	Rancang Bangun Aplikasi Sistem Pakar Untuk Pemilihan Spesifikasi Laptop Menggunakan Metode Forward Chaining Dan Certainty Factor Berbasis Mobile Android
D7	Sistem Penunjang Keputusan Rekomendasi Tenaga Kerja Menggunakan Metode Multi Factor Evaluation Process
D8	Perancangan Sistem Mesin Pencari Text Dengan Algoritma Semantik Berbasis Mobile Pada Perpustakaan Perguruan Tinggi Pelita Indonesia
D9	Rancang Bangun Sistem Penjualan Kelapa Sawit Berbasis Client Server
D10	Rancang Bangun Sistem Penjadwalan Shift Kerja Karyawan Menggunakan Metode Steepest Ascent Hill Climbing Berbasis WEB

### 2. Tahap Preprocessing

#### 1. Case Folding

Case folding merupakan proses pertama dari rangkaian preprocessing dokumen. Dalam proses ini mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf a sampai dengan z yang diterima.

#### 2. Tokenizing

Tahap Tokenizing adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya.

#### 3. Filtering

Tahap Filtering adalah pengambilan kata-kata yang penting dari hasil tokenizing. Proses yang dilakukan adalah dengan membuang kata stopword, seperti “dan”, “sebagai”, “tetapi” dan sebagainya.

#### 4. Stemming

Tahap stemming adalah tahap mencari root kata dari tiap kata hasil filtering. Pada tahap ini dilakukan proses pengembalian berbagai bentukan kata ke dalam suatu representasi yang sama. Pencarian kata dasar dilakukan dengan menghilangkan semua imbuhan dari kata, baik itu awalan, sisipan, maupun akhiran.

Hasil dari tahap preprocessing judul query dan judul pembanding dapat dilihat pada tabel 2.

Tabel 2. Tabel Hasil Preprocessing

ID.	Judul
Query	rancang sistem informasi akademik dukung daya kartu ajar guru media akses sma nurul falah pekanbaru
D1	rancang bangun aplikasi ajar bahasa inggris basis mobile android mata kuliah english for office
D2	sistem pakar mendiagnosa rusa sepeda motor injeksi guna metode forward chaining basis web
D3	sistem pakar mendiagnosa sakit ikan hias air tawar guna metode forward chaining theorema bayes basis web
D4	rancang bangun e commerce mobil tangki truk guna api application programming interface sms gateway
D5	rancang bangun sistem informasi presensi monitoring guna radio frequency identification rfid vihara pusdiklat bumi suci maitreya pekanbaru
D6	rancang bangun aplikasi sistem pakar pilih spesifikasi laptop guna metode forward chaining certainty factor basis mobile android
D7	sistem tunjang putus rekomendasi tenaga kerja guna metode multi factor evaluation process
D8	rancang sistem mesin cari text algoritma semantik basis mobile pustaka guru tinggi pelita indonesia
D9	rancang bangun sistem jual kelapa sawit basis client server
D10	rancang bangun sistem jadwal shift kerja karyawan guna metode steepest ascent hill climbing basis web

### 3. Menghitung Bobot

Menghitung bobot kriteria judul diperlukan untuk perhitungan algoritma Cosine Similarity, untuk mencari bobot dapat dilakukan dengan menggunakan algoritma TF-IDF (Term Frequency Inversed Document Frequency).

Pada tabel 3, adalah perhitungan untuk mencari *Inverse Document Frequency* (IDF) dengan rumus :

$$\log\left(\frac{n}{df}\right) + 1$$

Keterangan :

n = Jumlah Dokumen

t = Jumlah frekuensi kata terpilih

Tabel 3. Perhitungan TF dan IDF

No	Term	TF										IDF	Log(n/df)+1		
		Q	D1	D2	D3	D4	D5	D6	D7	D8	D9			D10	DF
1	air	0	0	0	1	0	0	0	0	0	0	0	0	1	2.04
2	ajar	1	1	0	0	0	0	0	0	0	0	0	0	2	1.74
3	akademik	1	0	0	0	0	0	0	0	0	0	0	0	1	2.04
4	akses	1	0	0	0	0	0	0	0	0	0	0	0	1	2.04
5	algoritma	0	0	0	0	0	0	0	0	0	1	0	0	1	2.04
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
95	text	0	0	0	0	0	0	0	0	0	1	0	0	1	2.04
96	theorema	0	0	0	1	0	0	0	0	0	0	0	0	1	2.04
97	tinggi	0	0	0	0	0	0	0	0	0	1	0	0	1	2.04
98	truk	0	0	0	0	1	0	0	0	0	0	0	0	1	2.04

99	tunjang	0	0	0	0	0	0	0	1	0	0	0	1	2.04
100	vihara	0	0	0	0	0	1	0	0	0	0	0	1	2.04
101	web	0	0	1	1	0	0	0	0	0	0	1	3	1.56

Pada tabel 4, adalah perhitungan untuk mencari *Term Frequency – Inverse Document Frequency* (TF-IDF) dengan rumus:

$$W_{dt} = tf_{dt} * IDF_t$$

Keterangan :

d = Dokumen ke-d

t = Kata ke-t dari kata kunci

w = bobot dokumen ke-d terhadap kata ke-t

Pada tahap ini dikalikan nilai TF dengan IDF pada tiap term berdasarkan hasil yang kita dapatkan.

Contoh pada term “ajar”, pada query dan dokumen pembanding d1 memiliki nilai 1, tiap nilai tersebut di kalikan dengan nilai IDF yang telah kita dapatkan di awal tadi, sehingga hasilnya menjadi  $1 \times 1.74 = 1.74$  pada query dan  $1 \times 1.74 = 1.74$  pada dokumen pembanding d1. Sedangkan yang lainnya mendapatkan nilai 0 karena tidak memiliki nilai TF.

Tabel 4. Perhitungan TF-IDF

No.	Term	Q	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
1	air	0	0	0	2.04	0	0	0	0	0	0	0
2	ajar	1.74	1.74	0	0	0	0	0	0	0	0	0
3	akademik	2.04	0	0	0	0	0	0	0	0	0	0
4	akses	2.04	0	0	0	0	0	0	0	0	0	0
5	algoritma	0	0	0	0	0	0	0	0	2.04	0	0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
95	text	0	0	0	0	0	0	0	0	2.04	0	0
96	theorema	0	0	0	2	0	0	0	0	0	0	0
97	tinggi	0	0	0	0	0	0	0	0	2.04	0	0
98	truk	0	0	0	0	2.04	0	0	0	0	0	0
99	tunjang	0	0	0	0	0	0	0	2.04	0	0	0
100	vihara	0	0	0	0	0	2.04	0	0	0	0	0
101	web	0	0.00	1.56	1.56	0	0	0	0	0	0	1.56

#### 4. Menghitung *Similarity*

Pada Tabel 5, adalah melakukan perkalian skalar dengan rumus :

$$Q * Di$$

Keterangan :

Q = Judul Query

Di = Judul ke-d



Contoh pada term “ajar” yang memiliki nilai query 1.74, di dokumen pembanding d1 memiliki nilai 1.74 dan sisa dokumen pembanding d2 sampai dengan d10 memiliki nilai 0. Perkalian scalar ini mengkalikan nilai TF yang telah di dapatkan dengan nilai IDF tiap term, misalnya  $q \times d1, q \times d2, q \times d3$

Tabel 5. Perkalian Scalar

No.	Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
1	air	0	0	0	0	0	0	0	0	0	0
2	ajar	3.03	0	0	0	0	0	0	0	0	0
3	akademik	0	0	0	0	0	0	0	0	0	0
4	akses	0	0	0	0	0	0	0	0	0	0
5	algoritma	0	0	0	0	0	0	0	0	0	0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
95	text	0	0	0	0	0	0	0	0	0	0
96	theorema	0	0	0	0	0	0	0	0	0	0
97	tinggi	0	0	0	0	0	0	0	0	0	0
98	truk	0	0	0	0	0	0	0	0	0	0
99	tunjang	0	0	0	0	0	0	0	0	0	0
100	vihara	0	0	0	0	0	0	0	0	0	0
101	web	0	0	0	0	0	0	0	0	0	0
Jumlah		4.32	1.18	1.18	1.30	8.54	2.48	1.18	5.51	2.48	2.48

Pada Tabel 6, adalah melakukan perkalian vector tiap judul query dan judul pembanding yang telah kita dapatkan pada tabel 4 dengan rumus :

$$(TF - IDF)^2$$

Tabel 6. Perkalian Vektor  $(TF-IDF)^2$

No.	Term	Q	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
1	air	0	0	0	4.17	0	0	0	0	0	0	0
2	ajar	3.03	3.03	0.00	0.00	0	0	0	0	0	0	0
3	akademik	4.17	0	0	0	0	0	0	0	0	0	0
4	akses	4.17	0	0	0	0	0	0	0	0	0	0
5	algoritma	0	0	0	0	0	0	0	0	4.17	0	0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
95	text	0	0	0	0	0	0	0	0	4.17	0	0
96	theorema	0	0	0	4.17	0	0	0	0	0	0	0
97	tinggi	0	0	0	0	0	0	0	0	4.17	0	0
98	truk	0	0	0	0	4.17	0	0	0	0	0	0
99	tunjang	0	0	0	0	0	0	0	4.17	0	0	0
100	vihara	0	0	0	0	0	4.17	0	0	0	0	0
101	web	0	0	2.45	2.45	0	0.0	0	0	0	0	2.45
Jumlah		52.10	45.03	35.33	47.83	50.16	57.40	44.28	39.64	46.89	26.34	43.38
$\sqrt{\text{Jumlah}}$		7.22	6.71	5.94	6.92	7.08	7.58	6.65	6.30	6.85	5.13	6.59

Setelah didapat hasil perkalian vektor TF-IDF, selanjutnya akan di dihitung total jumlah masing – masing dokumen query dan dokumen pembanding d1 sampai dengan d10. Lalu mengakarkan nilai total jumlah tersebut.

Selanjutnya memasukkan nilai yang telah didapatkan untuk menghitung similarity judul dengan menggunakan algoritma Cosine Similarity, dengan rumus :

$$\text{Cos } a = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Contoh perhitungan dokumen pembanding d1, dimana total jumat perkalian scalar pada dokumen pembanding d1 adalah 4.32.

$$\frac{4.32}{7.22 * 6.71} = 0.089288403$$

Hasil perhitungan dapat dilihat pada tabel 7.

Tabel 7. Hasil Persentase

d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
0.0892884	0.0275	0.0236754	0.0253	0.1560	0.0515	0.0260	0.1114	0.0668	0.0521
03	47	46	46	79	84	06	09	82	14
9%	3%	2%	3%	16%	5%	3%	11%	7%	5%

Maka dapat diambil analisis hasil pengujiannya yang dapat dilihat dari tabel 8 di bawah ini.

Tabel 8. Analisis Hasil Pengajuan

ID.	Judul	Cosine Similarity	%
D5	rancang bangun sistem informasi presensi monitoring guna radio frequency identification rfid vihara pusdiklat bumi suci maitreya pekanbaru	0.156078607	16%
D8	rancang sistem mesin cari text algoritma semantik basis mobile pustaka guru tinggi pelita indonesia	0.111409213	11%
D1	rancang bangun aplikasi ajar bahasa inggris basis mobile android mata kuliah english for office	0.089288403	9%
D9	rancang bangun sistem jual kelapa sawit basis client server	0.066881702	7%
D10	rancang bangun sistem jadwal shift kerja karyawan guna metode steepest ascent hill climbing basis web	0.052114132	5%
D6	rancang bangun aplikasi sistem pakar pilih spesifikasi laptop guna metode forward chaining certainty factor basis mobile android	0.05158377	5%
D2	sistem pakar mendiagnosa rusa sepeda motor injeksi guna metode forward chaining basis web	0.027547469	3%
D7	sistem tunjang putus rekomendasi tenaga kerja guna metode multi factor evaluation process	0.026006273	3%
D4	rancang bangun e commerce mobil tangki truk guna api application programming interface sms gateway	0.02534613	3%

---

D3	sistem pakar mendiagnosa sakit ikan hias air tawar guna metode forward chaining theorema bayes basis web	0.023675446	2%
----	--	-------------	----

---

Hasil yang didapatkan adalah bahwa judul pembanding d5 yaitu Rancang Bangun Sistem Informasi Presensi Dan Monitoring Menggunakan Radio Frequency Identification (RFID) Pada Vihara PUSDIKLAT Bumi Suci Maitreya Pekanbaru memiliki tingkat kesamaan judul sebesar 16%.

## 5. Kesimpulan

Dengan adanya sistem informasi pengajuan judul tugas akhir ini yang dapat melakukan pengecekan tingkat kesamaan pada judul dengan menggunakan algoritma *Term Frequency – Inverse Document Frequency* (TF-IDF) sebagai perhitungan bobot kata dan *Cosine Similarity* sebagai perhitungan mencari tingkat kesamaan judul. Hasil plagiat di dapat berupa persentase kesamaan judul yang sudah pernah di ajukan oleh mahasiswa sebelumnya . Aplikasi berbasis web ini memudahkan mahasiswa dan Perguruan Tinggi dalam prosedur pengajuan judul dengan menggunakan website yang sudah disediakan dimana saja dengan menggunakan internet, dosen atau staff akademik perguruan tinggi dapat memeriksa judul dan data proposal yang di ajukan mahasiswa dengan mudah bahkan dapat melihat persentasi kemiripan judul yang diajukan mahasiswa tersebut dengan database yang ada di perguruan tinggi sehingga meminimalisir terjadinya plagiarisme. Mahasiswa juga dapat menentukan judul dengan mudah dikarenakan sudah ada sistem yang menyimpan data judul-judul dari perguruan tinggi dan juga mempermudah mahasiwa dalam pengajuan judul dengan mengakses website yang sudah disediakan.

## Daftar Pustaka

- [1] P. Neuzil, “Tuntunan Penulisan Tugas Akhir Berdasarkan Prinsip Dasar Penelitian Ilmiah,” *Nucleic Acids Res.*, vol. 34, no. 11, pp. e77–e77, 2006.
- [2] M. Zalnur, “Plagiarisme Di Kalangan Mahasiswa Dalam Membuat Tugas-Tugas Perkuliahan Pada Fakultas Tarbiyah Iain Imam Bonjol Padang,” *Al-Ta lim*, vol. 19, no. 1, p. 55, 2012, doi: 10.15548/jt.v19i1.6.
- [3] R. F. Arista and R. H. Listyani, “Rasionalitas Plagiarisme Di Kalangan Mahasiswa,” *e J. UNESA*, pp. 1–5, 2010.
- [4] W. E. Y. R. Jarwati, Antonius Cahya Prihandoko, “Penerapan Algoritma Winnowing pada Sistem Rekomendasi Penentuan Dosen Pembimbing Skripsi (Studi Kasus: Prodi Sistem Informasi Universitas Jember Jember),” no. 1, pp. 11–20, 2016.
- [5] V. Prima and A. Rhosady, “Rancang bangun aplikasi pengklasifikasian halaman web berdasarkan,” 2013.
- [6] P. Nilai, P. Lingkungan, A. I. R. Pada, A. M. Sensor, and O. Aquarium, “Fakultas Ilmu Komputer Dan Teknologi Informasi,” vol. 1, no. 2, pp. 117–125, 2017.
- [7] G. Febrian, “Sistem Info,” *International*, vol. 12, no. 145, pp. 345–453, 2014.
- [8] Yuhefizar, “Pengertian Website,” *Pengertian Web*, vol. 16, no. 2, pp. 6–20, 2014.
- [9] D. Jones, “World wide web,” *Embroidery*, vol. 63, no. MAY-JUNE, pp. 24–27, 2012.
- [10] J. Jauhari, “Pengembangan Model Pembelajaran Berbasis E-Learning Sebagai Inovasi Pembelajaran Yang Tidak Mengenal Ruang dan Waktu di Kota Palembang,” 2011.