

Implementasi Algoritma Cosine Similarity Dan Metode TF-IDF Berbasis PHP Untuk Menghasilkan Rekomendasi Seminar

Muhtar Yusuf¹, Anis Cherid²

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana^{1,2}

Jl. Raya Meruya Selatan, Kembangan, Jakarta, 11650

¹muctaryusuf@gmail.com, ²anis.cherid@mercubuana.ac.id

Abstrak

Sistem rekomendasi bertujuan membantu user untuk mengidentifikasi produk yang sesuai dengan kebutuhan, kesenangan, dan keinginan user. Sistem rekomendasi akan membimbing user untuk menemukan produk yang relevan dan berguna dari banyaknya produk yang tersedia. Tujuan dalam penelitian ini adalah bagaimana membangun sistem rekomendasi suatu seminar dengan memanfaatkan pendekatan melalui teks judul dan deskripsi seminar dengan Metode TF-IDF untuk memberikan bobot frekuensi hubungan pada kemunculan suatu kata(*term*) di dalam perhitungan penggabungan judul dan deskripsi seminar. Dan algoritma Cosine Similarity digunakan sebagai metode perbandingan dalam mengetahui seberapa besar kemiripan antara ke dua seminar. Menggunakan data dummy sebagai data latih yang digunakan dalam penelitian ini, berjumlah 100 seminar yang sudah terlaksana dengan beberapa kategori bidang ilmu, fakultas dan universitas yang berbeda. Dan untuk implementasinya menggunakan bahasa pemrograman PHP, dan dengan bantuan data resource penggunaan text mining yang sudah berbahasa PHP yang berguna untuk tahap *text preprocessing*, dan kemudian akan dilanjutkan dengan proses implementasi Metode TF-IDF dan Algoritma Cosine Similarity.

Kata Kunci: *Cosine Similarity, Sistem Rekomendasi, TF-IDF, Website Seminar.*

Abstract

Recommendation system aims to help users to identify products that suit the needs, pleasures, and desires of the user. The recommendation system will guide the user to find relevant and useful products from the many products available. The purpose of this research is how to build a seminar recommendation system by utilizing the approach through the text of the seminar title and description with the TF-IDF Method to give the weight of the frequency of the relationship to the appearance of a word (term) in the calculation of combining the title and description of the seminar. And the Cosine Similarity algorithm is used as a comparison method to find out how much the similarity is between the two seminars. Using dummy data as training data used in this study, a total of 100 seminars have been carried out with several categories of different fields of science, faculties and universities. And for its implementation using the PHP programming language, and with the help of data resource the use of text mining that is already in PHP which is useful for the text preprocessing stage, and then will be continued with the process of implementing the TF-IDF Method and Algoritma Cosine Similarity.

Keywords: *Cosine Similarity, TF-IDF, Recommendation System, Website Seminar.*

1. Pendahuluan

Perkembangan teknologi memiliki dampak yang sangat signifikan dalam kehidupan sehari-hari, contoh seperti pengembangan sistem rekomendasi yang dilakukan perusahaan online shop guna menarik minat belanja dari customer. Fungsi dari Sistem Rekomendasi adalah alat bantu berupa sistem yang dirancang untuk mengidentifikasi serta memprediksi suatu produk yang sesuai dengan kebutuhan, kesenangan, dan keinginan dari user, yang nantinya produk tersebut akan direkomendasikan pada user. Penerapan rekomendasi di dalam suatu sistem dilakukan dengan cara memprediksi suatu item atau berupa produk, seperti rekomendasi film, musik, buku, berita dan lain sebagainya yang akan membimbing user untuk menemukan produk yang relevan dan berguna dari banyaknya produk yang tersedia. Semakin pesatnya perkembangan zaman sistem rekomendasi sangat berdampak pada perilaku user, dan telah dilaporkan bisa meningkatkan penjualan produk dan bisa membangun loyalitas user dalam melakukan pembelian produk[1], sehingga semakin tepat sistem rekomendasi akan semakin nyaman bagi user tersebut.

Berbeda dengan penelitian yang sudah ada dengan melalui pendekatan *Collaborative Filtering*, adalah pendekatan yang umum digunakan dalam perancangan sistem rekomendasi, bekerja dengan cara menjumlahkan rating atau pilihan

dari suatu produk, prinsip dasar dari algoritma ini adalah menghitung tingkat kesamaan antara beberapa produk yang berbeda berdasarkan penilaian user[2].

Pada penelitian ini untuk menghasilkan suatu sistem rekomendasi seminar pendekatan yang akan dilakukan yaitu pendekatan melalui teks atau kata dari penggabungan judul dan deskripsi pada suatu seminar, dengan memanfaatkan metode *teks mining*, yang berfungsi untuk mencari sesuatu informasi yang berguna dan bermanfaat di dalam suatu teks, yang nantinya akan digunakan untuk pembuatan perancangan sistem rekomendasi yang bermanfaat bagi user.

Berdasarkan permasalahan di atas maka dibuat suatu website yang bertujuan untuk tempat berbagai informasi tentang penyelenggaraan event seminar. Dan akan dirancang sistem rekomendasi seminar yang berguna untuk user dalam menemukan event seminar yang sesuai dengan keinginannya dengan cara yang efisien tanpa harus mencari di list daftar seminar yang banyak tersedia. Untuk user yang sudah memiliki akun pada web tersebut, sistem rekomendasi seminar diberikan berdasarkan tingkat relevansi perbandingan dan kemiripan teks, dari hasil perhitungan penggabungan antara judul dan deskripsi tiga riwayat seminar terakhir yang sudah pernah diikuti sebelumnya.

Algoritma yang akan diterapkan untuk bisa membuat sistem rekomendasi pada web tersebut yaitu algoritma cosine similarity yaitu salah satu algoritma di dalam teks mining, yang berfungsi untuk mengklasifikasikan suatu dokumen atau teks. Dan Metode TF-IDF (*Term Frequency Inverse Document Frequency*) merupakan metode untuk menghitung bobot setiap kata yang muncul di dalam keseluruhan dokumen[4].

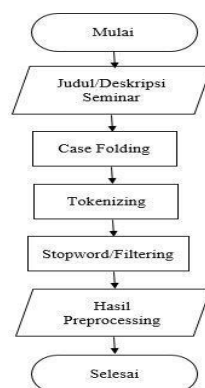
2. Landasan Teori Penelitian Terkait

Pada penelitian terkait yang menggunakan Algoritma *Cosine Similarity* dan Metode TF-IDF sebagai bahan referensi pada penelitian ini, yaitu penelitian yang membangun suatu sistem yang dapat melakukan klasifikasi dokumen skripsi secara otomatis, untuk memudahkan dalam melakukan penyimpanan sesuai dengan kategori bidang ilmu yang telah dibuat dan memudahkan dalam pencarian dokumen apa bila suatu saat dibutuhkan, hal ini dibutuhkan agar lebih efisien waktu[4].

Penelitian yang serupa juga terdapat pada penelitian sebelumnya tetapi berbeda pembahasan dalam penelitian yang dilakukan. Dengan tema yang sama yaitu melakukan pembuatan website bertemakan seminar, sebagai tempat penyedia event seminar tetapi admin hanya sebatas pengelola website saja, dan terdapat dua tipe user yang berbeda yaitu user penyelenggara seminar yang bisa mengupload event seminar dan dari sisi user yang hanya sedang mencari event seminar, yang nantinya akan diberi fitur notifikasi terdapat seminar baru dengan pilihan seleksi kategori lokasi yang diinginkan oleh usernya[5].

Text Mining

Text Mining mempunyai pengertian menggali informasi data dari dokumen yang berupa teks, banyaknya data berupa teks yang terdapat pada dokumen bertujuan mencari kata-kata yang dapat mewakili isi dari dokumen sehingga mendapatkan informasi yang berguna dari sekumpulan dokumen melalui teks dan dapat dilakukan analisa keterhubungan antar dokumen yang bermanfaat untuk tujuan tertentu. Dalam *text mining*, data teks akan diproses menjadi data *numerik* agar dapat dilakukan proses lebih lanjut. Sehingga dalam *text mining* ada istilah *preprocessing* data, yaitu proses pendahulu yang diterapkan terhadap data teks yang bertujuan untuk menghasilkan data *numerik* [4]. Tahap *preprocessing* dapat dilihat pada Gambar 1.

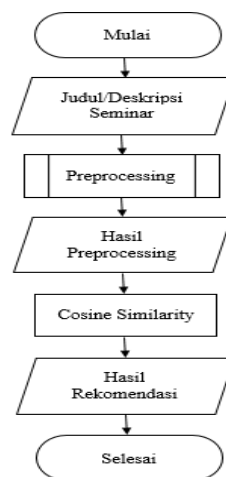


Gambar 1. Flowchart preprocessing.

Pada tahap *preprocessing* yang digunakan dalam penelitian ini antara lain:

- 1) *Case Folding*
Merupakan tahap perubahan huruf dari huruf kapital menjadi huruf kecil.
- 2) *Tokenizing*
Proses tahapan memecah dokumen menjadi kumpulan kata perkata, juga berfungsi untuk menghilangkan tanda baca, spesial karakter dan memisahkannya per spasi.
- 3) *Stopwords/Filtering*
Stopwords removal merupakan proses penghilangan kata tidak penting pada deskripsi melalui pengecekan kata-kata hasil *parsing* deskripsi apakah termasuk di dalam daftar kata tidak penting (*stoplist*) atau tidak. Jika termasuk di dalam *stoplist* maka kata-kata tersebut akan di-*remove* dari deskripsi sehingga kata-kata yang tersisa di dalam deskripsi dianggap sebagai kata-kata penting atau *keywords* [4].

Setelah melakukan tahap *preprocessing* maka data akan disimpan dalam memori sementara dan nantinya akan diproses lebih lanjut ke dalam tahap *analyzing* menggunakan TF-IDF untuk memberi bobot nilai frekuensi kemunculan suatu kata(*term*) dan pengukuran tingkat kemiripan dengan algoritma *cosine similarity* [4]. Proses dapat dilihat pada Gambar 2.



Gambar 2. Flowchart proses klasifikasi.

Metode TF-IDF(Term Frequency Inverse Document Frequency)

Menggabungkan dua konsep untuk perhitungan bobot nilai frekuensi kemunculan satu kata(*term*) di dalam suatu dokumen tertentu. Frekuensi kemunculan kata(*term*) di dalam dokumen yang berarti menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Jika Frekuensi kemunculan kata di dalam dokumen itu tinggi menunjukkan bahwa kata tersebut menjadi kata yang penting di dalam suatu dokumen. Sehingga bobot hubungan antara suatu kata dan suatu dokumen akan semakin tinggi apabila frekuensi kata tersebut tinggi kemunculannya di dalam dokumen dan jika kemunculan kata di dalam suatu dokumen itu rendah maka bobot hubungan antar kata dan dokumen akan menjadi rendah. Rumus yang akan digunakan untuk TF-IDF [4][6].

$$IDF = \log\left(\frac{D}{DF}\right)$$

$$Weight = TF \times IDF$$

Keterangan:
 D = jumlah dokumen yang ada di dalam database
 DF = banyak dokumen yang mengandung kata(*term*)
 W = hasil bobot nilai dokumen *Frequency*
 tf = banyaknya kata(*term*) pada suatu dokumen
 idf = *Inversed Document Frequency*

Cosine Similarity

Cosine Similarity adalah salah satu algoritma di dalam teks mining, yang berfungsi untuk mengklasifikasikan suatu dokumen atau teks. *Cosine similarity* mempunyai konsep normalisasi panjang *vektor* dengan membandingkan dua dokumen antara dokumen A dan B. Rumus yang umum digunakan untuk *Cosine Similarity*[4][6].

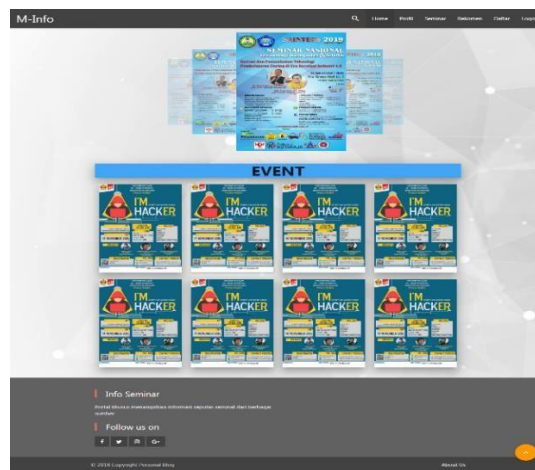
$$\text{Cos } \alpha = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Keterangan :
 A = Vektor A, yang akan dibandingkan kemiripannya
 B = Vektor B, yang akan dibandingkan kemiripannya
 A • B = dot product antara vektor A dan vektor B
 |A| = panjang vektor A
 |B| = panjang vektor B
 |A||B| = cross product antara |A| dan |B|

Penggunaan *cosine similarity* pada sistem ini adalah sebagai metode untuk membandingkan kesamaan atau similaritas perhitungan penggabungan judul dan deskripsi pada seminar yang sudah pernah diikuti oleh user, sistem rekomendasi berjalan mencari relevansi kemiripan seminar berdasarkan 3 seminar terakhir yang sudah pernah diikuti oleh user.

3. Hasil Dan Pembahasan
Implementasi Antarmuka Aplikasi

Pada tahap ini *Interface* yang dibuat yaitu menu home, pada menu home hanya bersifat tampilan dummy yang bertujuan untuk memberikan tampilan awal pada saat user mengunjungi web seminar ini.



Gambar 3. Tampilan Home.

Testing

Pada tahap ini pengujian dilakukan dengan menggunakan metode *Black Box*, di mana metode ini bertujuan untuk mengetahui dan memastikan fungsi-fungsi dari system yang dibangun dapat berjalan dengan baik dan untuk mengetahui kesalahan program ada atau tidak.

Disimulasikan pada contoh pengujian ini terdapat total 23 seminar yang berbeda-beda bidang ilmu pembahasan, fakultas dan universitas yang berbeda, termasuk seminar Q sebagai seminar yang akan di bandingkan dengan seluruh seminar yang ada di database, dan 2 seminar dengan judul dummy “seminar percobaan non aktif 1 dan 2” dimaksudkan untuk pengujian sistem untuk seminar yang sudah tidak aktif, tidak bisa muncul di dalam list daftar seluruh seminar dan di seminar yang akan direkomendasikan.

Tabel 1. Tahapan Testing.

DOC	Judul Seminar	Hasil Penggabungan Penjumlahan Judul dan deskripsi	Hasil Dalam Bentuk %
Q	Inovasi Dan Pemanfaatan Teknologi Pembelajaran Daring Di Era Revolusi Industri 4.0	1	100%
S1	Tren Aplikasi Multi-Platform Dalam Transformasi Digital Untuk Meningkatkan Perdagangan Di Era Ekonomi Digital	0,103920182	10.39 %
S2	Peran Artificial Intelligence Dalam Era Revolusi Industry 4.0	0,036381830	3.64 %
S3	I'm Hacker	0,073408527	7.34 %
S4	Arah, Model, Desain, Dan Problematika Pendidikan Guru Dalam Menghadapi Perkembangan Revolusi Industri	0,062738804	6.27 %

S5	Industry Revolution 4.0 Impact On Ohs	0,058179976	5.82 %
S6	Reformasi Tata Kelola Pajak Di Indonesia Pasca Tax Amnesty Dan Workshop	0,023034646	2.3 %
S7	Peran Akuntan Dalam Mewujudkan Jiwa Enterpreneurship Muda Indonesia	0,008039591	0.8 %
S8	Mengupas Teknologi Android Os	0,051395908	5.14 %
S9	Meningkatan Kapasitas Psikososial Masyarakat Dalam Situasi Krisis	0,016465119	1.65 %
S10	Menjadi Engineer Cerdas Digital Di Era Industri 4.0	0,036913579	3.69 %
S11	Kewirausahaan: Tinjauan Ekonomi,Manajemen, Akuntansi Dan Spiritualitas	0,085615126	8.56 %
S12	Peningkatan Mutu Pendidikan Kejuruan Mengacu Kerangka Kualifikasi Nasional Indonesia	0,032505189	3.25 %
S13	Pengelolaan sumber daya manusia dalam mengantisipasi implementasi industri 4.0	0,025102129	2.51 %
S14	Menghijaukan Sistem Ekonomi Bisnis Dan Akutansi	0,085603300	8.56 %
S15	Penerapan Dan Inovasi Energi Alternatif Dan Material Dalam Dunia Kewirausahaan	0,028854940	2.89 %
S16	Peran Perguruan Tinggi Meningkatkan Kompetensi Lokal Menghadapi Revolusi Industri 4.0	0,235157978	23.52 %
S17	Pengaruh Revolusi Industri 4.0 Terhadap Penerapan Teknologi Informasi Berbasis Sistem Pendukung Keputusan	0,493476198	49.35 %
S18	Arah Baru Pendidikan Dan Kepemimpinan Nasional Dalam Menghasilkan Sumber Daya Manusia Yang Berkompet	0,017741571	1.77 %
S19	Penerapan Dan Peluang Kerja Teknologi Blockhain	0,047992430	4.8 %
S20	Implementasi Sistem Manajemen K3 Dalam Revolusi Industri 4.0 Di Berbagai Sektor	0,081292220	8.13 %

Setelah mendapatkan 10 seminar dengan persen tertinggi, sistem rekomendasi yang nantinya akan bekerja berdasarkan 3 riwayat seminar terakhir user, dari masing-masing seminar itu akan dihitung tingkat kemiripannya dengan seminar yang ada di database dan akan mengambil masing-masing 10 seminar dari 3 seminar yang menjadi riwayat seminar yang pernah diikuti user, sehingga akan menghasilkan 30 seminar rekomendasi. Dan untuk sistem filtering nya seminar yang sudah pernah diikuti tidak akan muncul sebagai rekomendasi seminar, dan seminar yang muncul dari hasil hitung salah satu dari 3 seminar riwayat user terjadi kemiripan hasil seminar yang sama, maka hanya akan muncul salah satu nya saja.

Penerapan TF-IDF pada penelitian ini yaitu berfungsi memberikan bobot nilai pada kemunculan kata penggabungan pada judul dan deskripsi seminar, TF-IDF mengoptimalkan perhitungan jika terdapat deskripsi seminar dengan kalimat yang panjang dan banyak, tetapi jika terdapat kalimat yang pendek penggunaan TF-IDF tidak akan membuat hasil perhitungan menjadi lebih buruk, karena cara kerja TF-IDF memberi bobot nilai pada kata yang muncul dan merubahnya kedalam bentuk desimal dengan penjabaran yang lebih detail. Penjabaran perhitungan TF-IDF bisa dilihat di dalam tabel 2, dan dengan rumus di bawah.

$$IDF = \frac{1}{DF} \log(D)$$

$$Weight = TF \times IDF$$

Dan kemudian dilanjutkan dengan implementasi menghitung tingkat kemiripan antara dua dokumen menggunakan rumus cosine similarity di bawah ini dengan penjabaran yang sudah disederhanakan.

$$Sim(d, q) = \frac{\sum(weight_{wq} \times weight_{w17})}{\sqrt{\sum(weight_{wq}^2 \times weight_{w17}^2)}}$$

Pada tabel 2 menunjukkan hasil perhitungan detail mencari tingkat kemiripan antara kedua dokumen dengan menggunakan algoritma cosine similarity pada $W17^2$ adalah hasil kuadrat 2 dari $W17$ kemudian hasilnya di SUM(jumlah) dan di SQRT(akar kuadrat), setelah itu pada $WQ.WS17$ adalah hasil dari $W17$ di kali WQ setelah itu di SUM(jumlah). Dan untuk mengetahui hasil akhirnya yaitu SUM(jumlah) dari $WQ.WS17$ di bagi dengan hasil penjumlahan dari SQRT(akar kuadrat) dari $W17^2$ di kali kan dengan SQRT(akar kuadrat) dari WQ^2 .

Dan Pada tabel 2 memperlihatkan kata yang muncul di seminar Q juga banyak yang muncul di seminar S17 serta jumlah kata yang muncul di antara kedua nya sangat berimbang, dan hal tersebut mempengaruhi hasil perhitungan bagaimana seminar S17 bisa menghasilkan perhitungan yang paling tinggi seperti yang ditunjukkan pada tabel no 1, pada seminar Q terdapat banyak kata(*term*) yang sama muncul di seminar S17 yang bisa membuat hasil perhitungan antara IDF di kali TF menjadi tinggi.

Tabel 2. Perhitungan TF/IDF dan Cosine.

Kata	TF		DF	IDF	Weight		Weight ²		Weight
	Q	S17			WQ	W17	WQ ²	W17 ²	WQ x W17
teknologi	10	8	11	0,3203351509	3,203351509	2,562681207	10,26146089	6,56733496807	8,20916871
revolusi	1	1	8	0,4586378490	0,458637849	0,458637849	0,210348677	0,21034867656	0,210348677
industri	2	3	9	0,4074853266	0,814970653	1,22245598	0,664177166	1,49439862239	0,996265748
mbang	2	2	10	0,3617278360	0,723455672	0,723455672	0,523388109	0,52338810940	0,523388109
informasi	3	7	6	0,5835765856	1,750729757	4,085036099	3,065054682	16,68751993371	7,151794257
sangat	2	2	5	0,6627578317	1,325515663	1,325515663	1,756991774	1,75699177382	1,756991774
luas	1	1	5	0,6627578317	0,662757832	0,662757832	0,439247943	0,43924794346	0,439247943
bidang	5	4	8	0,4586378490	2,293189245	1,834551396	5,258716914	3,36557882494	4,206973531
khusus	1	1	4	0,7596678447	0,759667845	0,759667845	0,577095234	0,57709523426	0,577095234
bukan	1	1	4	0,7596678447	0,759667845	0,759667845	0,577095234	0,57709523426	0,577095234
pengaruh	2	3	4	0,7596678447	1,519335689	2,279003534	2,308380937	5,19385710830	3,462571406
komputer	4	2	6	0,5835765856	2,334306343	1,167153171	5,448986101	1,36224652520	2,72449305
semua	1	1	5	0,6627578317	0,662757832	0,662757832	0,439247943	0,43924794346	0,439247943
ikut	1	1	5	0,6627578317	0,662757832	0,662757832	0,439247943	0,43924794346	0,439247943
hadap	1	2	10	0,3617278360	0,361727836	0,723455672	0,130847027	0,52338810940	0,261694055
baik	1	1	4	0,7596678447	0,759667845	0,759667845	0,577095234	0,57709523426	0,577095234
cil	1	1	3	0,8846065813	0,884606581	0,884606581	0,782528804	0,78252880368	0,782528804
tengah	1	1	4	0,7596678447	0,759667845	0,759667845	0,577095234	0,57709523426	0,577095234
mau	1	1	3	0,8846065813	0,884606581	0,884606581	0,782528804	0,78252880368	0,782528804
besar	1	1	6	0,5835765856	0,583576586	0,583576586	0,340561631	0,34056163130	0,340561631
lepas	1	1	3	0,8846065813	0,884606581	0,884606581	0,782528804	0,78252880368	0,782528804
seminar	2	2	21	0,0395085413	0,079017083	0,079017083	0,006243699	0,00624369934	0,006243699

nasional	2	2	14	0,215599800 3	0,431199601	0,431199601	0,185933096	0,18593309563	0,185933096
sains	3	3	3	0,884606581 3	2,653819744	2,653819744	7,042759233	7,04275923308	7,042759233
ynote	1	1	5	0,662757831 7	0,662757832	0,662757832	0,439247943	0,43924794346	0,439247943
speaker	1	1	4	0,759667844 7	0,759667845	0,759667845	0,577095234	0,57709523426	0,577095234
dr	1	2	14	0,215599800 3	0,2155998	0,431199601	0,046483274	0,18593309563	0,092966548
dosen	1	1	7	0,516629796 0	0,516629796	0,516629796	0,266906346	0,26690634612	0,266906346
bu	1	1	4	0,759667844 7	0,759667845	0,759667845	0,577095234	0,57709523426	0,577095234
kaji	1	1	5	0,662757831 7	0,662757832	0,662757832	0,439247943	0,43924794346	0,439247943
manajemen	2	3	5	0,662757831 7	1,325515663	1,988273495	1,756991774	3,95323149110	2,635487661
akuntansi	1	1	7	0,516629796 0	0,516629796	0,516629796	0,266906346	0,26690634612	0,266906346
ekonomi	1	1	7	0,516629796 0	0,516629796	0,516629796	0,266906346	0,26690634612	0,266906346
hukum	1	1	5	0,662757831 7	0,662757832	0,662757832	0,439247943	0,43924794346	0,439247943
sistem	1	2	7	0,516629796 0	0,516629796	1,033259592	0,266906346	1,06762538447	0,533812692
bioteknologi	1	1	3	0,884606581 3	0,884606581	0,884606581	0,782528804	0,78252880368	0,782528804
ilmu	1	2	9	0,407485326 6	0,407485327	0,814970653	0,166044291	0,66417716551	0,332088583
informatika	1	1	6	0,583576585 6	0,583576586	0,583576586	0,340561631	0,34056163130	0,340561631
terap	1	1	6	0,583576585 6	0,583576586	0,583576586	0,340561631	0,34056163130	0,340561631
indones	0	1	11	0,320335150 9	0	0,320335151	0	0,10261460888	0
teknik	0	1	10	0,361727836 0	0	0,361727836	0	0,13084702735	0
muhammad	0	1	3	0,884606581 3	0	0,884606581	0	0,78252880368	0
prof	0	1	8	0,458637849 0	0	0,458637849	0	0,21034867656	0
nilai	0	1	2	1,060697840 4	0	1,06069784	0	1,12507990853	0
bisnis	0	1	4	0,759667844 7	0	0,759667845	0	0,57709523426	0
bas	0	1	1	1,361727836 0	0	1,361727836	0	1,85430269939	0
dukung	0	1	2	1,060697840 4	0	1,06069784	0	1,12507990853	0
putus	0	2	1	1,361727836 0	0	2,723455672	0	7,41721079754	0
sensasi	0	1	1	1,361727836 0	0	1,361727836	0	1,85430269939	0
latar	0	1	2	1,060697840 4	0	1,06069784	0	1,12507990853	0
belakang	0	1	2	1,060697840 4	0	1,06069784	0	1,12507990853	0
ambil	0	1	1	1,361727836 0	0	1,361727836	0	1,85430269939	0
penting	0	1	1	1,361727836 0	0	1,361727836	0	1,85430269939	0

pihak	0	1	1	1,3617278360	0	1,3617278360		1,85430269939	0
zarlis	0	1	1	1,3617278360	0	1,3617278360		1,85430269939	0
darmawan	0	1	1	1,3617278360	0	1,3617278360		1,85430269939	0
napitupulu	0	1	1	1,3617278360	0	1,3617278360		1,85430269939	0
bahasa	0	1	1	1,3617278360	0	1,3617278360		1,85430269939	0
buda	0	1	1	1,3617278360	0	1,3617278360		1,85430269939	0
pariwisata	0	1	1	1,3617278360	0	1,3617278360		1,85430269939	0
by	0	1	2	1,0606978404	0	1,0606978404	0	1,12507990853	0
stmik	0	1	1	1,3617278360	0	1,3617278360		1,85430269939	0
darma	0	1	1	1,3617278360	0	1,3617278360		1,85430269939	0
ikat	0	1	1	1,3617278360	0	1,3617278360		1,85430269939	0
republik	0	1	1	1,3617278360	0	1,3617278360		1,85430269939	0
idri	0	1	1	1,3617278360	0	1,3617278360		1,85430269939	0
Kata	Q	S17	DF	IDF	WQ	W17	101,94178293	106,35642191083	51,38355904
AKAR	Q	S17	DF	IDF	WQ	W18	10,09662235	10,31292499298	Hasil
Kata	Q	S17	DF	IDF	WQ	W19	WQ ²	WS17 ²	0,493476198

**4. Penutup
Kesimpulan**

Dari sistem rekomendasi yang sudah dibuat, sudah bisa berjalan dan sesuai dengan apa yang diprediksi, seperti contoh pada tahapan testing Q sebagai seminar pembanding yang akan dibandingkan dengan seminar yang ada di database, seminar Q yang berasal dari bidang ilmu teknik informatika yang bertema membicarakan inovasi teknologi masa depan, kemudian dibandingkan dengan tema seminar yang ada di database, sesuai dengan apa yang diprediksi dengan melakukan pendekatan melalui teks atau kata, seminar-seminar yang bertemakan sama dengan seminar Q yaitu membicarakan tentang teknologi masa depan akan mendapat tingkat kemiripan yang tinggi.

Akan tetapi jika dilihat dari sisi keterhubungan ilmu pembahasan sistem ini akan menjadi kurang relevan, di contohkan kembali pada tahapan testing terdapat seminar dari ilmu pembahasan lain tetapi sama-sama membicarakan tema yang sama dengan seminar Q dan seminar tersebut akan mendapat poin persentase yang cukup tinggi, karena tingkat kemiripan dari sistem pendekatan teks, dan seminar tersebut masuk di antara 10 seminar dengan persen tertinggi yang akan dimunculkan oleh sistem rekomendasi.

Tujuan dilakukan hasil perhitungan secara terpisah pada riwayat seminar 1, 2, dan 3 adalah untuk mencari solusi bagaimana sistem ini bisa terlihat relevan jika dilihat dari sisi bidang ilmu pembahasan. Seperti contoh pada riwayat seminar 1 pada perhitungannya akan menghasilkan 10 seminar rekomendasi, begitu juga dengan seminar 2 dan 3 akan menghasilkan 10 rekomendasi seminar, dan dari masing-masing hasil tersebut akan digabungkan menjadi 1 sehingga berjumlah 30 seminar rekomendasi, dengan sistem filtering yaitu tidak ada hasil seminar yang sama, jika seminar yang muncul dihasil perhitungan riwayat seminar 1 muncul juga dihasil perhitungan riwayat seminar 2 ataupun 3 maka yang akan dimunculkan hanya salah satunya saja, ini bertujuan agar seminar yang tidak termasuk dalam 10 seminar hasil perhitungan riwayat 1, 2 dan 3 bisa masuk kedalam 10 besar hasil perhitungan, karena seminar dengan hasil kembar dihilangkan karena hasil filtering, dan diharapkan seminar dengan bidang keilmuan yang sama bisa masuk kedalam hasil 30 seminar rekomendasi.

Saran

Penelitian Web seminar ini masih butuh banyak perbaikan di sistem rekomendasi seminar bagi user nya agar menjadi lebih relevan dari sistem yang sudah dibuat, dan web seminar ini juga bisa berfungsi sebagai sumber informasi jika diadakannya suatu event seperti seminar, jobfair dan lain-lain. Dan ada beberapa fitur yang mungkin masih bisa dikembangkan di web ini salah satunya yaitu memberikan fitur notifikasi adanya seminar atau event terbaru yang akan diselenggarakan.

5. Daftar Pustaka

- [1]. Prasetya, C. S. D. Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor. 2017.
- [2]. Oktora, R., & Susanty, W. Perancangan Aplikasi E-Commerce Dengan Sistem Rekomendasi Item-Based Collaborative Filtering. 2012.
- [3]. Melita, R., Amrizal, V., Suseno, H. B., & Dirjam, T. Penerapan Metode Term Frequency Inverse Document Frequency (TF-IDF) dan Cosine Similarity pada Sistem Temu Kembali Informasi untuk Mengetahui Syarah Hadits Berbasis Web. 2018.
- [4]. Wahyuni, R. T., Prastiyanto, D., & Suprpto, E. Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi. 2017.
- [5]. Dian Oktaviani, S. H. Rancang Bangun Portal Seminar Nasional. 2018.
- [6]. Sugiyamta. Sistem Deteksi Kemiripan Dokumen Dengan Algoritma Cosine Similarity Dan Single Pass Clustering. 2015.
- [7]. Rismayani. Sistem Rekomendasi Pencarian Jodoh Syariah Menggunakan. 2018.
- [8]. Kesuma, H. W. A. Penerapan Metode Tf-Idf Dan Cosine Similarity Dalam Aplikasi Kitab Undang- Undang Hukum Dagang. 2016.
- [9]. Ajiprayoga, H. P. Pemanfaatan Metode Cosine Similarity Dalam Menentukan Kemiripan Iklan Pada Situs Jual Beli Online. 2015.
- [10]. Arifin, M. F., & Fitriana, D. Penerapan Algoritma Klasifikasi C4.5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus: PT Atria Artha Persada. 2018.