INFORMATION RETRIEVAL DALAM PENCARIAN CERDAS MENGENAI MUAMALAH

I'lman Teguh Prasetya[1], Galih Hermawan, S.Kom., M.T.[2]

[1,2] Universitas Komputer Indonesia
Jl. Dipatiukur No. 112-116, Bandung 40132
E-mail : ilmanfrasetya@gmail.com [1], galih.hermawan@gmail.com [2]

ABSTRAK

Information Retrieval (IR) adalah ilmu yang mempelajari sebuah metode untuk menemukan kembali informasi relevan yang tersimpan disebuah koleksi sumber informasi. Dalam proses pencarian informasi dibutuhkun sebuah metode pencarian. Algoritma KNN menurut beberapa jurnal algoritma KNN dinilai baik dalam penyelesaian kasus recognition dan klasifikasi, pada penelitian ini akan dilakukan implementasi Algoritma KNN dalam kasus pencarian buku. Didalam penelitian ini melakukan analisis terhadap penerapan algoritma KNN dalam pencarian buku pada koleksi buku muamalah. Adapun tahapan yang di analisis tahap *preprocessing* data. Dalam penelitian ini juga dilakukan analisis pada setiap tahap algoritma KNN yaitu penentuan K, perhitungan jarak dokumen testing menuju dokumen training, urutkan data dan menentukan kelompok data testing. Berdasarkan hasil pengujian, algoritma KNN mampu menyelesaikan permasalahan pencarian dari koleksi informasi. Proses pencarian yang berdasarkan kesamaan kata dari data training dan data testing sangat membantu sekali dalam peningkatan kualitas dari hasil pencarian. Pencarian menggunakan KNN menggunakan 20 data *training* memiliki nilai akurasi 95%, Proses pencarian menggunakan algoritma KNN juga dapat memberikan informasi klasifikasi mengenai permasalah user yang membuat pengguna menjadi tidak kebingungan mengenai materi.

Kata kunci : Muamalah, Pencarian menggunakan KNN, Sistem temu balik informasi, Klasifikasi, akurasi, *preprocessing* data.

1. PENDAHULUAN

Information Retrieval (IR) adalah ilmu yang mempelajari konsep, prosedur–prosedur dan metode – metode untuk menemukan kembali infomasi relevan yang tersimpan dari berbagai *resources* atau koleksi sumber informasi yang dicari atau dibutuhkan [2]. Pada proses pencarian informasi disebuah koleksi sumber, *Information Retrieval* akan memerlukan metode pencarian. Algoritma K-*Nearest Neighbour* Menurut Ramdhan [2] memiliki akurasi 80% pada kasus klasifikasi kategori tugas besar [2] dan menurut Ilyas [17] metode K-*Nearest Neighbour* memiliki akurasi 100% pada kasus pengenalan wajah. Pada penelitian ini akan mengimplementasikan Algoritma K-*Nearest Neighbour* untuk menyelesaikan permasalahan pencarian.

Muamalah adalah bagian dari hukum islam yang mengatur hubungan antara sesama manusia. Pada kegiatan pembelajaran muamalah membutuhkan buku – buku sebagai referensi, Pada saat ini koleksi buku sumber mengenai mumalah sudah tersedia, tetapi akan sulit ketika mencari sebuah buku – buku yang sesuai dengan keinginan dan ketika melakukan pencarian disebuah koleksi buku yang memiliki koleksi buku yang biasanya lebih dari satu akan membutuhkan banyak waktu [3]. Masalah ini menimbulkan sedikitnya buku referensi yang dipakai atau bahkan tidak ditemukan buku yang dicari.

Dengan melihat masalah diatas, pencarian buku sesuai kebutuhan pengguna harus melewati tahap *preprocessing* terlebih dahulu, lalu data yang telah diolah tersebut dicari nilai kemirimpannya dan dibuat ranking yang tujuannya tersebut menunjukkan buku tersebut relevan dengan masalah yang dimasukan pengguna. setelah melakukan *ranking* data tersebut dijadikan sebuah daftar yang nantinya akan diberikan ke pengguna. Pada penelitian ini menggunakan proses text minning untuk mengenali masukan pengguna dan database buku, dimana pada proses pencarian kata dasarnya (*stemming*) dan menggunakan Information Retrieval untuk penarikan informasinya. Setelah tahap *preprocesing* selesai dilanjutkan dengan proses pencarian buku dan data masukan pengguna akan menggunakan Algoritma K-*Nearest Neighbour*. Pada penelitian ini akan dibuat sistem yang dapat membantu pengguna untuk mencari buku referensi muamalah yang sesuai masukan pengguna.

1.1 Identifikasi Masalah

Berdasarkan uraian pada latar belakang masalah, maka rumusan masalahnya adalah Bagaimana bagaimana cara untuk mencari buku sesuai keinginan pengguna dan memberikan daftar buku yang dibutuhkan oleh pengguna dalam permasalahan pencarian buku muamalah.

1.2 Maksud dan Tujuan

Berdasarkan permasalahan yang diteliti, maka maksud dari penelitian ini adalah

mengimplementasikan Algoritma K-Nearest Neighbour dalam kasus pencarian buku. Sedangkan tujuan yang ingin dicapai dari penelitian ini mengetahui akurasi Algoritma K-Nearest Neighbour dalam kasus pencarian.

1.3 Information Retrieval

Information Retrieval menangani proses yang berkaitan dengan representasi dokumen penyimpanan, penelusuran dan penemuan informasi yan relevan terhadap suatu sistem yang dapat menjembatani informasi dengan pengguna yang membutuhkan informasi [7]. Saat melakukan penelusuran pengguna menginput permintaan. Permintaan ini sama dengan analisis isi informasi dari suatu dokumen, dokumen dianalisis isinya untuk mengetahui topik atau pokok bahasan.

1.4 Preprocessing

Preprocessing data adalah Tahap yang mencakup semua rutinitas, dan proses untuk mempersiapkan data yang akan digunakan pada operasi *knowledge discovery* sistem *text mining* [2]. Tujuan *Preprocessing* tersebut adalah untuk mempermudah dalam proses *Minning*. Pada penelitian ini tahap *Preprocessing* yang dilakukan adalah *Case Folding, Cleaning Data, Tokenizing, Filtering* dan *Stemming, Stemming* menggunakan Algoritma Nazief dan Andriani, yaitu algoritma yang dikembangkan untuk stemming dokumen Bahasa Indonesia. Yaitu suatu proses pemotongan imbuhan-imbuhan suatu kata menjadi kata dasarnya yang merupakan suatu tahap penting dalam *Information Retrieval* [1] dan *Filtering* merupakan tahap pengambilan kata-kata yang penting dari hasil Tokenizing [8].perhitungan menggunakan persamaan dibawah.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{True\ Positives + False\ Positives + True\ Positives + False\ Negatives}$$

(1)

1.5 Pembobotan

Pembobotan dilakukan untuk mendapatkan nilai dari kata/term yang berhasil diekstrak. Perhitungan bobot dengan Term Frequency – Inverse Document Frequency (TF-IDF) menggunakan kombinasi dari dua nilai, yaitu frekuensi kata dan inverse frekuensi dokumen yang didapat dari membagi jumlah dokumen secara keseluruan dengan jumlah dokumen dimana kata tersebut muncul [3].

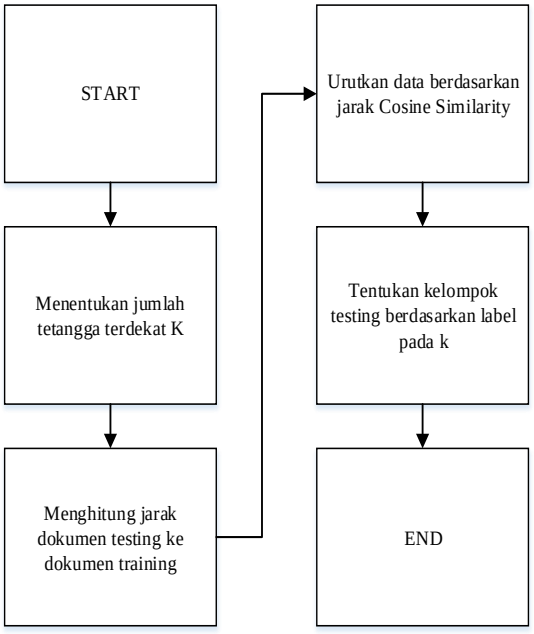
$$IDF = \log(\frac{n}{df}) \quad (2)$$

$$W_{dt} = tf_{df} * IDF_t \quad (3)$$

Deskripsi :
n : banyak dokumen koleksi termasuk dokumen testing
d : dokumen ke-d
t : kata ke-t dari kata kunci
W : bobot dokumen ke-d terhadap kata ke-t
tf : banyaknya kata yang dicari pada sebuah dokumen
IDF : Inversed Document Frequency Nilai IDF didapatkan dari IDF : log2 (D/df) dimana
D : total dokumen
df : banyak dokumen yang mengandung kata yang dicari

1.6 Algoritma *K-Nearest Neighbor*

K-Nearest Neighbor (KNN) adalah metode *machine learning* yang melakukan klasifikasi terhadap suatu objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Cara kerja dari KNN perlu adanya penetuan inputan berupa data latih, data uji dan nilai k. Kemudian mengurutkan data latih berdasarkan kedekatan jaraknya berdasarkan hitungan dari jarak data yang diuji dengan data latih. Setelah itu diambil dari k data latih teratas untuk menentukan kelas klasifikasi untuk kelas yang dominan dari k data latih yang diambil [10].



Gambar 1. Contoh penulisan Gambar

1.7 *Cosine Similarty*

Cosine similarity adalah ukuran kesamaan yang lebih umum digunakan dalam information retrieval dan merupakan ukuran sudut antara vektor dokumen dtest (titik (i,k)) dan dtraining (titik (j,k)) . Tiap vektor tersebut merepresentasikan setiap kata dalam setiap dokumen (teks) . Rumus Cosine similarity yang digunakan adalah persamaan 2.3 [10].

$$\cos(\Theta_{ij}) = \frac{\sum_k (d_{ik} d_{jk})}{\sqrt{\sum_k d_{ik}^2} \sqrt{\sum_k d_{jk}^2}} \quad (4)$$

1.8 *Confusion Matrix*

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Metode yang menggunakan tabel matriks [6].

Tabel 1. Tabel *Confusion Matrix*

klasifikasi yang benar	Diklasifikasikan sebagai	
	+	-
+	True Positives	False negatives
-	False Positives	True Negatives

True positives adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positives* adalah jumlah record negatif yang diklasifikasikan sebagai positif, *false negatives* adalah jumlah record positif yang diklasifikasikan sebagai negatif, true negatives adalah jumlah record negatif yang diklasifikasikan sebagai negative, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam confusion matrix, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity (recall), specificity, precision dan accuracy. Sensitivity* digunakan untuk membandingkan jumlah TP terhadap jumlah *record* yang positif sedangkan *specificity* adalah perbandingan jumlah TN terhadap jumlah record yang negative [12]. Untuk menghitung digunakan persamaan

2. ISI PENELITIAN

Pada tahap ini akan membahas mengenai masalah yang akan dianalisis. Beberapa hal yang terdapat dalam tahap ini yaitu mengenai analisis masalah, analisis sistem yang akan dibangun serta metode yang digunakan untuk menyelesaikan masalah.

2.1 Analisis Masalah

Analisis Masalah adalah tahap mengindentifikasi masalah yang akan muncul pada saat penerapan metode. Masalah yang akan diangkat dalam penelitian ini yaitu bagaimana penerapan metode K-NN untuk pencarian buku yang sesuai dengan text yang di input user. Masalah pencarian merupakan permasalahan yang kompleks karena untuk menemukan buku yang sesuai memerlukan metode pendekatan yang cocok untuk menemukan buku yang sesuai dengan text input user.

2.2 Analisis Sistem

Analisis sistem adalah tahap penjelasan dan penguraian dari sistem utama kedalam beberapa bagian sistem dengan tujuan untuk dapat melihat kebutuhan-kebutuhan yang diperlukan agar dapat menerapkan metode yang digunakan untuk menyelesaikan suatu masalah yang tujuannya membangun sebuah aplikasi untuk melakukan pengujian. Aplikasi yang akan dibangun memiliki beberapa tahapan. Setelah setiap tahapan dilakukan maka output yang didapatkan adalah hasil pencarian dari inputan user dan token – token yang sudah diberi bobot. Adapun tahapannya dari memasukan data test dan data training buku oleh pengguna. Lalu sistem akan memproses data-data tersebut. Langkah pertama melakukan proses Case-folding, proses Cleaning Data, proses Tokenisasi, proses Filtering, proses Stemming dan proses pencarian menggunakan Algoritma K-NN.

2.3 Analisis Data Masukan

Data masukan yang akan dianalisis yaitu data training dan data test, data test merupakan data masukan user yang berupa kalimat bahasa Indonesia yang baik dan benar

Tabel 2. Tabel Data *Test*

Term
Apakah asuransi dari kantor ABC itu haram ?

Tabel 3. Tabel Contoh Data *Training*

judul buku	klasifikasi
Bisnis dan Muamalah Kontemporer	Kontemporer

Tabel 4. Tabel Data *Training* yang diuji

NO	JUDUL	KLASIFIKASI
d1	Bisnis dan Muamalah Kontemporer	kontemporer
d2	Bisnis Islami & Kritik atas Praktik Bisnis ala Kapitalis	kontemporer
d3	Buku Pintar Ekonomi Syariah	jual-beli
d4	Syariah Marketing (Kertajaya dan Sula)	kontemporer
d5	Ekonomi Islam Umar bin Khattab	jual-beli

2.4 Preprocessing

Merupakan proses persiapan data *test* dan data *training* sebelum data diproses menggunakan Algoritma KNN. *Preprocessing* sendiri melewati beberapa tahap diantaranya :

Case Folding merupakan tahapan merubah semua data masukan huruf menjadi lower case.

Tabel 4. Tabel implementasi *Case folding* Data *Test*

sebelum	**sesudah**
Apakah asuransi dari kantor ABC itu haram ?	apakah asuransi dari kantor abc itu haram ?

Tabel 5. Tabel implementasi case folding Data *Test*

sebelum	**sesudah**
Bisnis dan mu'amalah kontemporer tidak bisa dipungkiri telah berkembang, dan menjadi bagian dari kehidupan umat Islam. Namun, tidak semua perkembangan tersebut sesuai dengan syariah. Ketika perkembangan ini menyangkut substansi yang bertentangan dengan syariah. Seperti PT, koperasi, asuransi dan saham. Di sinilah dibutuhkan kecermatan, dan ketepatan pandangan dan sikap. Karena ada juga perkembangan bisnis dan muamalah kontemporer yang dibenarkan oleh syariah. Seperti bisnis online, e-commerce, dan lain-lain. Semuanya bisa ditelaah dalam buku ini. Selamat membaca.	bisnis dan mu'amalah kontemporer tidak bisa dipungkiri telah berkembang, dan menjadi bagian dari kehidupan umat islam. namun, tidak semua perkembangan tersebut sesuai dengan syariah. ketika perkembangan ini menyangkut substansi yang bertentangan dengan syariah. seperti pt, koperasi, asuransi dan saham. di sinilah dibutuhkan kecermatan, dan ketepatan pandangan dan sikap. karena ada juga perkembangan bisnis dan muamalah kontemporer yang dibenarkan oleh syariah. seperti bisnis online, e-commerce, dan lain-lain. semuanya bisa ditelaah dalam buku ini. selamat membaca.

Cleaning Data merupakan tahapan yang digunakan untuk melakukan pembersihan text dari karakter-karakter selain huruf .

Tabel 6. Tabel Data *Test*

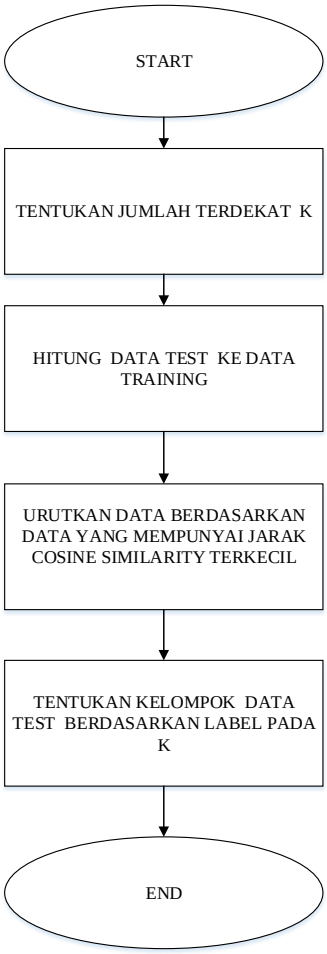
sebelum	sesudah
apakah asuransi dari kantor abc itu haram ?	apakah asuransi dari kantor abc itu haram

Tabel 7. Tabel Data *Test*

sebelum	sesudah
bisnis dan mu'amalah kontemporer tidak bisa dipungkiri telah berkembang, dan menjadi bagian dari kehidupan umat islam. namun, tidak semua perkembangan tersebut sesuai dengan syariah. ketika perkembangan ini menyangkut substansi yang bertentangan dengan syariah. seperti pt, koperasi, asuransi dan saham. di sinilah dibutuhkan kecermatan, dan ketepatan pandangan dan sikap. karena ada juga perkembangan bisnis dan muamalah kontemporer yang dibenarkan oleh syariah. seperti bisnis online, e-commerce, dan lain-lain. semuanya bisa ditelaah dalam buku ini. selamat membaca.	bisnis dan muamalah kontemporer tidak bisa dipungkiri telah berkembang dan menjadi bagian dari kehidupan umat islam namun tidak semua perkembangan tersebut sesuai dengan syariah ketika perkembangan ini menyangkut substansi yang bertentangan dengan syariah seperti pt koperasi asuransi dan saham di sinilah dibutuhkan kecermatan dan ketepatan pandangan dan sikap karena ada juga perkembangan bisnis dan muamalah kontemporer yang dibenarkan oleh syariah seperti bisnis online e-commerce dan lain lain semuanya bisa ditelaah dalam buku ini selamat membaca

2.4 Analisis Algoritma K-*Nearest Neighbor*

Analisis algoritma adalah proses menjabarkan atau menjelaskan tahapan-tahapan dari Algoritma K-*Nearest Neighbor* (K-NN atau KNN). tahapannya dari Menentukan parameter k (jumlah tetangga paling dekat), Menghitung kuadrat jarak *eucliden* objek terhadap data training yang diberikan, Mengurutkan hasil perhitungan eucliden secara ascending, Mengumpulkan kategori Y (Klasifikasi *nearest neighbor* berdasarkan nilai k), Dengan menggunakan kategori *nearest neighbor* yang paling mayoritas maka dapat dipredisikan kategori objek . Setelah semua tahapan maka hasil dari semua proses yaitu berupa klasifikasi kategori text tersebut masuk ke kategori yang mana.



Gambar 2. Tahapan Algoritma K-Nearest Neighbor

2.5 Menentukan Parameter K

Parmeter K adalah parameter data tetangga yang paling dekat. Pada proses ini adalah memasukan jumlah k yang dimana jumlah ini akan menentukan banyak data yang nanti akan dipilih pada proses pengelompokan dan pemberian label. K disni telah di tentukan sebanyak 3 .

2.6 Menghitung kemiripan data test dan data training

Pada tahap ini melakukan proses perhitung Cosine Similiarity . perhitungan menggunakan data testing dan data training yang telah diboboti menggunakan persamaan 3.

Tabel 8. Perkalian Skalar Antara Q dan dokumen lain

WDQ.WDI				
D1	**D2**	**D3**	**D4**	**D5**
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
…	…	…	…	…
0,0062697	0	0,0062697	0,0062697	0,0062697

WDQ.WDI				
D1	**D2**	**D3**	**D4**	**D5**
…	…	…	…	…
0	0,1812381	0	0	0
0,00627	**0,181238**	**0,00627**	**0,00627**	**0,00627**

Kedua Hitung panjang setiap dokumen, termasuk Q. Dengan cara kuadratkan bobot setiap term dalam setiap dokumen lalu jumlahkan nilai kuadrat dan terakhir akarkan

Tabel 9. Hitung panjang Vektor setiap dokumen

panjang vektor					
Q	**D1**	**D2**	**D3**	**D4**	**D5**
0	0	0	0,605519	0	0
0	0	0	0	0,227644	0,227644
0	0	0	0	0,227644	0,227644
0	0	0,764874	0,015609	0	0
0	0	0	0	0,227644	0,227644
0	0	0	0,910578	0	0
0	0	0,605519	0	0	0
0	0	0	0	0,605519	0
0	0	0	0	0,605519	0
0	0	0,605519	0	0	0
0	0	0	0	0,056427	0,025078
0	0	0	0	0,605519	0
0	0	0	0,605519	0	0
0	0	0	0	0,605519	0
0	0	0	0	0	0

	panjang vektor					
	Q	**D1**	**D2**	**D3**	**D4**	**D5**
	…	…	…	…	…	…
	0	0	0	0	0,227644 7	0,227644 7
	0	0	0	0,090619 1	0,090619 1	0,090619 1
total	0,702408 1	7,209008	42,74507 7	56,93837 9	115,84414	23,26673 3
akar	0,838097 9	2,684959 6	6,537971 9	7,545752 4	10,76309 2	4,823560 2

Menggunakan persamaan 3 rumus cosine similarity.hitung kemiripan Q dengan D1, D2, D3, D4 dan D5

Tabel 10. Perhitungan *Cosine Similarty*

Cosine Similarty				
D1	D2	D3	D4	D5
0,020086	0,048909	0,056448 5	0,080517	0,036084

Tabel 11. Perhitungan detail *Cosine Similarty*

Cosine similarty	**no**	**JUDUL**	**klasifikasi**
0,020086	D1	Bisnis dan Muamalah Kontemporer	kontemporer
0,036084	D5	Ekonomi Islam Umar bin Khattab	jual-beli
0,048909	D2	Bisnis Islami & Kritik atas Praktik Bisnis ala Kapitalis	kontemporer
0,056449	D3	Buku Pintar Ekonomi Syariah	jual-beli
0,080517	D4	Syariah Marketing (Kertajaya dan Sula)	kontemporer

2.7 Urutkan data berdasarkan jarak

Proses ini merupakan proses pengurutan data berdasarkan nilai Cosine Similarty secara Descending . untuk proses bisa dilihat pada tabel berikut:

Tabel 12. Tabel implimentasi Urut

Cosine similarty	**no**	**JUDUL**	**klasifikasi**
0,080517	D4	Syariah Marketing (Kertajaya dan Sula)	kontemporer
0,056449	D3	Buku Pintar Ekonomi Syariah	jual-beli
0,048909	D2	Bisnis Islami & Kritik atas Praktik Bisnis ala Kapitalis	kontemporer
0,036084	D5	Ekonomi Islam Umar bin Khattab	jual-beli
0,020086	D1	Bisnis dan Muamalah Kontemporer	kontemporer

2.8 Rank

Proses ini merupakan proses pemberian ranking berdasarkan nilai nilai Cosine Similarty yang telah diurutkan secara descending .

Tabel 13. Implemenasi *Rank*

Ranking	**Cosine similarty**	**no**	**JUDUL**	**klasifikasi**
1	0,080517	D4	Syariah Marketing (Kertajaya dan Sula)	kontemporer
2	0,056449	D3	Buku Pintar Ekonomi Syariah	jual-beli
3	0,048909	D2	Bisnis Islami & Kritik atas Praktik Bisnis ala Kapitalis	kontemporer
4	0,036084	D5	Ekonomi Islam Umar bin Khattab	jual-beli
5	0,020086	D1	Bisnis dan Muamalah Kontemporer	kontemporer

2.9 Pengelompokan

Dengan jumlah k = 3, maka diambil 3 urutan teratas. Dimana 3 urutan teratas D4 memiliki klasfikasi kontemporer, D3 memiliki klasifikasi jual-beli, D2 klasifikasi kontemporer.

Tabel 13. Implemenasi Pengenlompokan

Ranking	**Cosine similarty**	**no**	**JUDUL**	**klasifikasi**
1	0,080517	D4	Syariah Marketing (Kertajaya dan Sula)	kontemporer

Ranking	Cosine similarity	no	JUDUL	klasifikasi
2	0,056449	D3	Buku Pintar Ekonomi Syariah	jual-beli
3	0,048909	D2	Bisnis Islami & Kritik atas Praktik Bisnis ala Kapitalis	kontemporer

Setelah data testing (Q) di uji, k objek terdekat dengan data testing lebih banyak dari klasifikasi kontemporer. jadi Q diklasifikasikan sebagai klasifikasi kontemporer dan buku yang disarankan untuk pengguna adalah D4 dengan Syariah Marketing (Kertajaya dan Sula), D3 Buku Pintar Ekonomi Syariah Kertajaya dan Sula dan D2 Bisnis Islami & Kritik atas Praktik Bisnis ala Kapitalis.

Tabel 14. Implemenasi *detail* pengelompokan

Ranking	Cosine similarity	no	JUDUL	klasifikasi
1	0,080517	D4	Syariah Marketing (Kertajaya dan Sula)	kontemporer
2	0,056449	D3	Buku Pintar Ekonomi Syariah	jual-beli
3	0,048909	D2	Bisnis Islami & Kritik atas Praktik Bisnis ala Kapitalis	kontemporer

2.9 Pengujian Pengukuran Akurasi

Confucion matrix digunakan untuk pengukuran efektifitas klasifikasi. Dengan menggunakan persamaan 3, dilakukan proses perhitungan akurasi untuk 20 data uji.

$$Accuracy = \frac{3+16}{3+1+16+0} \times 100\% = 95\%$$

Berdasarkan hasil pengujian didapatkan akurasi untuk menyelesaikan permasalahan pencarian dengan data sebanyak 20 data sebesar 0.95 dan dapat disimpulkan akurasi algoritma KNN akan memiliki nilai tinggi atau rendahnya akurasi dari sebuah data training tergantung dari kualitas dari data training yang dibuat.

3. PENUTUP

3.1 Kesimpulan

Berdasarkan hasil yang didapat pada penelitian ini, diperoleh bahwa fungsional pencarian berjalan dengan baik selain itu pengujian akurasi KNN tingkat akurasi algoritmanya 95%, Tetapi dari hasil pencarian buku yang disarankan untuk pengguna masih sesuai rank.

3.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, masih perlu dilakukan beberapa kajian lebih lanjut. Adapun saran untuk penelitian lebih lanjut yaitu:

1. Menggunakan data training yang memiliki kategori selain jual beli dan kontemporer, sehingga mencapai solusi yang kaya akan informasi dan mendapatkan solusi yang optimal.
2. Mengoptimalkan proses pencarian agar lebih cepat dengan menggunakan system DBMS agar menyelesaikan permasalahan pencarian dengan data banyak.
3. Melakukan penanggulangan pencarian, agar buku yang diusulkan benar-benar optimal sepenuhnya.

DAFTAR PUSTAKA

[1] Abdurrasyid, S. A., & Suyanto. (2012). " Implementasi dan Optimasi Algoritma Nazief dan Andriani Untuk Stemming Dokumen Bahasa Indonesia ". Skripsi S1, Fakultas Teknik Informatika, Telkom University.

[2] Dr. Eng R.H. Sianipar, S. M. (2013). " JAVA teori, algoritma, dan aplikasi ". Yogyakarta: ANDI..

[3] Hamilton, K., & Miles, R. (2006). " Learning UML 2.0 ". O'Reilly.

[4] Haryanto, & Priyatmoko, B. (2015). " Sistem Informasi Pengolahan Data Pasien Rawat Inap Berbasis SMS gateway ". ISSN 2337 - 5213 , INFORMATIKA Vol 2. No 1 .

[5] Indriani, A. (2014). " Klasifikasi Data Forum dengan menggunakan ". ISSN: 1907 - 5022, Seminar Nasional Aplikasi Teknologi Informasi (SNATI).

[6] Leidiyana, H. (2013). " Penerapan Algoritma K-Nearst Neighbor untuk penentukan resiko kredit kepemilikan kendaraan bermotor ". Jurnal Penelitian Ilmu Komputer, System Embedded & Logic, 65-76.

[7] Pattah , S. H. (2013). " Peranan Intermediary dalam Sistem Temu Balik Informasi ". ISSN : 2353-9629, Vol. 1 No. 2.

[8] Putri, P. A., Ridok, A., & Indriati. (2016). " Implementasi Metode Improved K-Nearest Neighbor pada Analisis Sentimen Twitter berbahasa Indonesia ". Skripsi S1, Program Teknologi Informasi dan Ilmu Komputer, Universitas Brawijaya.

[9] Rustiana, D., & Rahayu, N. (2017). " Analisis Sentimen Pasar Otomotif Mobil : Tweet Twitter Menggunakan Naive Bayes ". ISSN: 2252-4983, Jurnal SIMETRIS, Vol 8 No 1.

[10] Sani, R. R., Zeniarja, J., & Luthfiarta, A. (2016). " Penerapan Algoritma K-nearest Neighbor pada Information Retrieval dalam penentuan Referensi Tugas Akhir ". Journal of Applied Intelligent System, Vol. 1, No. 2,.

[11] Susandi, D., & Sholahudin, U. (2016). " Pemanfaatan Vector Space Model pada Penerapan ". ISSN: 2406-7741, Jurnal ProTekInfo Vol. 3 No. 1.