

## Klasifikasi Kepakaran *Reviewer* Menggunakan Algoritma *K-Nearest Neighbor* (KNN)

Intan Putri Nur Fajarini, Imam Much Ibnu Subroto, Andi Riansyah  
Fakultas Teknologi Industri, Universitas Islam Sultan Agung

Correspondence Author: [intanfajar272@std.unissula.ac.id](mailto:intanfajar272@std.unissula.ac.id), [imam@unissula.ac.id](mailto:imam@unissula.ac.id),  
[andi@unissula.ac.id](mailto:andi@unissula.ac.id)

### Abstract

Ada banyak hal yang bisa dilakukan dalam meningkatkan kualitas pendidikan di negara Indonesia salah satunya yaitu Kemdikbud membuat sebuah program untuk meningkatkan jumlah publikasi jurnal di Indonesia yang bereputasi internasional. Jika dilakukan upaya peningkatan publikasi jurnal maka diperlukan *reviewer* jurnal, yang mana bidang keilmuan *reviewer* harus sesuai dengan bidang ilmu dari jurnal yang akan *direview*, data *reviewer* ini diperoleh dari database SINTA (*Science and Technology Index*). Bidang ilmu dari jurnal tersebut dikelompokkan ke dalam 5 bidang yaitu *Arts & Humanities*, *Engineering & Technology*, *Life Sciences & Medicine*, *Natural Science*, dan *Social Science & Management*. Saat ini belum ada penelitian yang menerapkan suatu metode untuk mengklasifikasikan *reviewer* dengan 5 bidang ilmu kepakaran, untuk memudahkan hal tersebut bisa memanfaatkan teknik yang ada saat ini yaitu menggunakan metode data *mining* untuk mengolah data tersebut, dalam penerapan data *mining* untuk klasifikasi akan digunakan algoritma KNN. Dalam penelitian ini tahap yang dilakukan adalah pengumpulan data, kemudian data tersebut akan diolah menggunakan metode *preprocessing* agar diperoleh data yang bersih dan memiliki kualitas yang baik, setelah itu akan diproses lagi dengan menerapkan metode klasifikasi menggunakan algoritma KNN. Setelah melakukan beberapa pengujian jumlah K dan menggunakan 3 skenario pengujian, pada penelitian ini menghasilkan akurasi 0,751 atau 75 %, *precision* 0.628, *recall* 0.54, dan *f-measure* 0.528 dengan K = 30 pada pengujian skenario uji 80% : 20%.

Kata Kunci : *reviewer* , data *mining*, klasifikasi, KNN

### 1. PENDAHULUAN

Kementerian Pendidikan dan Kebudayaan Republik Indonesia (Kemdikbud) merupakan kementerian pemerintah negara Indonesia yang mengurus segala urusan yang ada di bidang pendidikan, meliputi pendidikan anak usia dini, pendidikan dasar, pendidikan menengah, pembinaan masyarakat, dan budaya. Dalam hal ini Kemdikbud memiliki dan mengelola banyak data terkait pendidikan dan disiplin ilmu. Data-data tersebut dapat digunakan untuk meningkatkan kualitas pendidikan di Indonesia, jika diolah dengan baik dan benar maka akan menghasilkan sebuah informasi yang dapat berguna bagi kepentingan pendidikan.

Saat ini, Kemdikbud membuat sebuah program untuk meningkatkan jumlah publikasi jurnal di Indonesia yang bereputasi internasional. Untuk itu, diperlukan tenaga *reviewer* jurnal sesuai dengan bidang ilmu para peneliti dan dosen yang akan dipublikasi. Yang mana jumlah *reviewer* jurnal di Indonesia sangat banyak yaitu lebih dari 100 ribu dengan bidang ilmu yang berbeda-beda dari sejumlah 14.086 jurnal yang terindeks di Garuda Kemdikbud, dalam pembagiannya bidang ilmu yang berbeda-beda tersebut dikelompokkan menjadi 5 bidang keilmuan yaitu *Arts & Humanities*, *Engineering & Technology*, *Life Sciences & Medicine*, *Natural Sciences*, dan *Social Sciences & Management*. Oleh karena itu, diperlukan suatu metode untuk mengklasifikasikan secara otomatis sehingga bidang ilmu kepakaran dengan *reviewer* dapat sesuai. Selain itu belum adanya sebuah kebijakan untuk mengatur metode apa yang dapat digunakan untuk melakukan klasifikasi terhadap data *reviewer* tersebut, sehingga dari data-data yang ada masih sulit untuk dikelompokkan ke dalam 5 bidang ilmu.

Seiring berjalannya waktu, laju perkembangan teknologi semakin pesat, berbagai data dengan jumlah yang banyak dapat diolah menggunakan proses yang dikenal sebagai data *mining*. Data *mining* merupakan proses menambang informasi penting yang diperoleh dari sekumpulan data. Informasi yang penting tersebut diperoleh melalui proses yang sangat kompleks seperti penggunaan kecerdasan buatan, teknik statistik, dan lain-lain. Teknik yang kompleks ini mengidentifikasi dan mengekstraksi informasi yang berguna dari *database* besar [1]. Ada beberapa metode di dalam data *mining* salah satunya adalah klasifikasi. Klasifikasi adalah proses mengklasifikasikan fitur atau kata yang mirip ke dalam sebuah kelas menggunakan data latih (data *training*) sebagai acuan pengelompokan dan data uji (data *testing*) sebagai pengujian. Hal tersebut bertujuan agar objek kelas yang belum memiliki label dan terklasifikasi dapat diprediksi [2]. Ada beberapa algoritma yang dapat digunakan dalam metode klasifikasi salah satunya adalah KNN, dimana objek metode diklasifikasikan berdasarkan data pembelajaran dengan cara menghitung kedekatan jarak satu data terhadap data yang lainnya, yang mana KNN ini mempunyai beberapa keunggulan

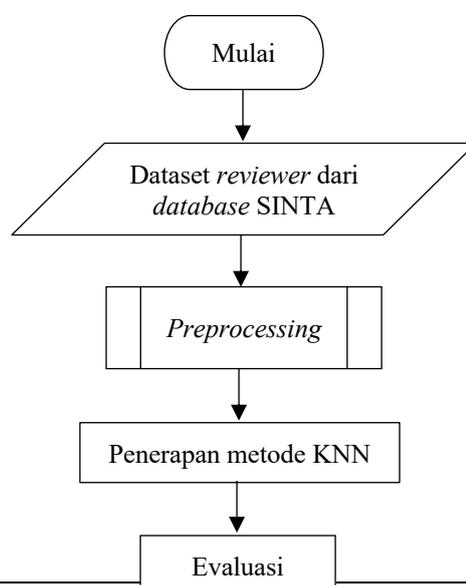
diantaranya memiliki kemampuan untuk melakukan pelatihan dengan sangat cepat, termasuk algoritma yang sederhana dan mudah untuk dipelajari, tahan pada data yang memiliki derau, dan efektif pada data pelatihan yang besar. Namun algoritma KNN ini juga mempunyai kelemahan diantaranya jumlah nilai  $k$  dapat menjadi bias, komputasinya kompleks, memori terbatas, serta mudah dikelabui dengan fitur yang tidak relevan [3].

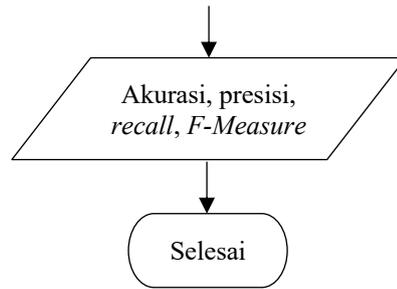
Adapun penelitian sebelumnya yang membahas mengenai klasifikasi laporan masyarakat menggunakan algoritma KNN, dalam penelitian tersebut menyebutkan bahwa algoritma KNN menghasilkan akurasi yang baik dalam mengklasifikasikan data laporan masyarakat ke dalam tiga kategori yaitu pengaduan, permintaan informasi, dan aspirasi. Data yang digunakan adalah 2235 data yang telah dikumpulkan oleh pemerintah kota Semarang selama tahun 2017. Setelah dilakukan pengolahan data diperoleh 930 data latih, masing-masing 310 data latih untuk setiap kelas laporan masyarakat dan 100 data uji, dengan uji evaluasi dan validasi menggunakan *confusion matrix* diketahui bahwa pada nilai  $k = 11$  memperoleh akurasi tertinggi yaitu 82% dalam penentuan kategori kelas laporan masyarakat [4]. Penelitian berikutnya mengenai penerapan algoritma KNN dalam mengklasifikasikan penentuan pengajuan kartu kredit, penelitian ini bertujuan untuk mempermudah pihak bank dalam penentuan kategori kartu kredit yang sesuai bagi nasabah, yang mana terdapat 4 kategori kartu kredit yaitu tidak diterima, *classic*, *gold*, dan platinum. Data yang digunakan adalah data nasabah Bank BNI Syariah Surabaya sebanyak 250 data, dengan rincian 250 data latih dan 100 data uji dengan mengambil sampel dari data latih yang diperoleh sebelumnya. Dalam penelitian ini menggunakan beberapa atribut untuk faktor pendukung diantaranya jenis kelamin, status rumah, status, jumlah tanggungan berupa anak, pekerjaan, serta penghasilan tiap tahun. Dalam penelitian ini diperoleh hasil nilai akurasi dengan rata-rata 93%, presisi 92%, dan *recall* 83% [5]. Penelitian berikutnya yaitu mengenai klasifikasi untuk pemberian sanksi pajak menggunakan metode KNN di KPP Pratama Cirebon Dua yang mana dapat digunakan untuk memudahkan dalam mengetahui perbedaan denda wajib pajak, penelitian tersebut menggunakan 229 data. Dari hasil pengujian akhir penerapan algoritma KNN dengan data *training* sebesar 80% yaitu 183 data serta jumlah data *testing* sebesar 20% yaitu 46 data menghasilkan nilai akurasi 97,78%, nilai presisi 97,78%, dan *recall* 100,00% [6]. Penelitian berikutnya mengenai penerapan algoritma KNN untuk mengklasifikasikan judul berita hoax di internet berdasarkan kategori berita hoax atau tidak. Dalam penelitian ini menggunakan data yang telah dikumpulkan secara manual oleh peneliti yaitu sebanyak 75 judul berita fakta yang diperoleh dari situs detikcom dan cnnindonesia, kemudian 75 judul berita hoax yang diperoleh dari situs turnbackhoax, pengumpulan data tersebut dilakukan pada tanggal 24 November 2021 sampai 22 Januari 2022. Pada penelitian ini menggunakan 3 jenis skenario pengujian yaitu 90% : 10%, 80% : 20%, dan 70% : 30%. Sehingga diperoleh akurasi terbaik pada skenario 90 : 10 dengan akurasi 93,33%, *precision* 100%, *recall* 80%, dan *f1-score* 88,89% [7].

Dari permasalahan tersebut penelitian ini mengusulkan metode KNN untuk klasifikasi kepakaran *reviewer*. Pada penelitian ini menggunakan data judul publikasi jurnal terindeks scopus sebagai data *training* dan data judul publikasi jurnal *reviewer* sebagai data *testing*. Sehingga bidang ilmu kepakaran dengan *reviewer* dapat sesuai. Dalam proses klasifikasi data *reviewer* ini akan menggunakan algoritma KNN dengan metode perhitungan jarak *euclidean distance*.

## 2. METODE PENELITIAN

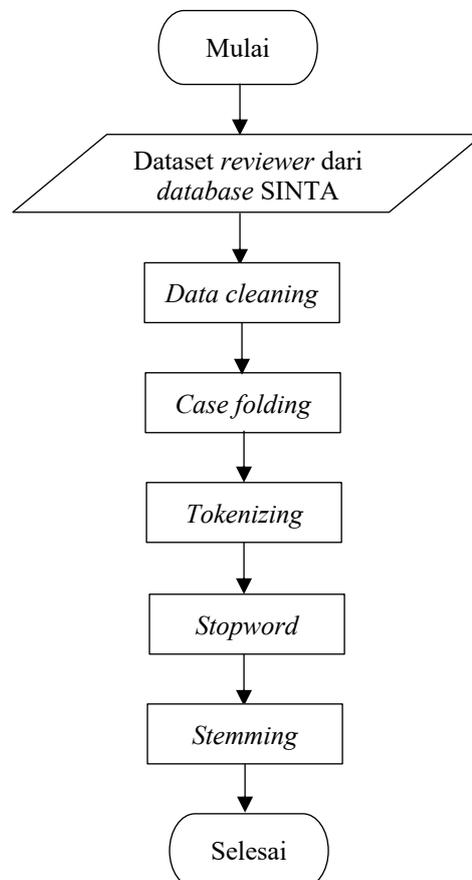
Dalam penelitian ini meliputi beberapa tahapan yang dilakukan diantaranya yaitu pengumpulan data, kemudian dilakukan pengolahan data yang disebut dengan *preprocessing*, data yang sudah melalui *preprocessing* akan diolah menggunakan metode klasifikasi data *mining* dengan menerapkan algoritma KNN, kemudian dari penerapan metode tersebut akan dihasilkan sebuah evaluasi akhir yaitu tingkat akurasi, presisi, *recall*, dan *F-Measure*. Adapun tahapan-tahapan dalam penelitian ditunjukkan pada gambar 1.





Gambar 1 Tahap-tahap Penelitian

Pada *preprocessing* data terdapat beberapa tahapan untuk mengolah data diantaranya yaitu data *cleaning* untuk menghilangkan karakter-karakter yang tidak memiliki pengaruh, *case folding* untuk mengubah karakter huruf menjadi besar atau kecil semua, *tokenizing* untuk memisahkan teks menjadi potongan kata yang disebut token, *stopword* untuk menghilangkan kosa kata atau simbol, dan terakhir yaitu *stemming* untuk memperoleh kata dasar dari suatu token. Adapun tahap-tahap *preprocessing* tersebut ditunjukkan pada gambar 2.



Gambar 2 Flowchart tahap *Preprocessing*

## 2. 1 Dataset yang digunakan

Pada penelitian ini menggunakan data judul publikasi jurnal terindeks scopus sebagai data *training* dan data judul publikasi jurnal *reviewer* sebagai data *testing* yang diperoleh dari *database* SINTA, SINTA adalah sistem informasi penelitian berbasis web yang memudahkan untuk mengukur kinerja peneliti, lembaga dan jurnal serta memberikan akses kutipan dan keahlian di Indonesia. Yang mana data tersebut telah divalidasi terlebih dahulu oleh pihak yang memiliki izin untuk mengakses data tersebut.

## 2. 2 Studi kasus

Studi kasus dalam penelitian tugas akhir ini adalah klasifikasi bidang ilmu dengan *reviewer* jurnal yang dikelompokkan ke dalam 5 bidang ilmu yaitu *Arts & Humanities, Engineering & Tecnology, Life Sciences & Medicine, Natural Sciences, dan Social Sciences & Management* menggunakan metode klasifikasi data *mining* dengan menerapkan algoritma KNN.

## 2. 3 Koleksi data

Pada proses koleksi data akan dilakukan labelisasi yaitu klasifikasi kepakaran *reviewer* ke dalam 5 bidang ilmu utama sebagaimana telah ditetapkan sebelumnya yang akan dijadikan sebagai label. Dalam penelitian ini menggunakan data sebanyak 500 dataset dengan rincian 100 data untuk setiap pengelompokan bidang ilmu. Adapun contoh datanya ditunjukkan pada tabel 1.

Tabel 1 Contoh data

No.	Area	Bidang Ilmu Kepakaran
1.	<i>Political Mediation in Ireland: Campaigning between Traditional and Tabloid Markets.</i>	<i>Social Science and Management</i>
2.	<i>Electrical Engineering Education in Indonesia.</i>	<i>Engineering and Tecnology</i>
3.	<i>Interpreting the atmospheric neutrino anomaly.</i>	<i>Natural Science</i>
4.	<i>Small doses of subcutaneous heparin preventing deep-vein thrombosis in major surgery.</i>	<i>Life Science and Medicine</i>
5.	<i>Audiovisual patterns in Wayang Kulit.</i>	<i>Arts &amp; Humanity</i>

2. 4 *Preprocessing*

*Preprocessing* yaitu tahapan-tahapan untuk memproses data mentah yang nantinya akan dimasukkan ke dalam sebuah metode klasifikasi data *mining*, pengolahan data ini berfungsi untuk mengoptimalkan kualitas suatu data, dan membuat proses penambahan data menjadi lebih efisien [8]. Pada tahap *preprocessing* akan dilakukan pengolahan data *reviewer* untuk memperoleh kualitas data yang baik dan bersih agar diperoleh data yang bersih dan siap untuk dilakukan proses klasifikasi dengan menerapkan algoritma KNN. Pada *preprocessing* akan dilakukan beberapa tahapan untuk mengolah data yang masih mentah diantaranya yaitu data *cleaning, case folding, tokenizing, stopword, dan stemming*.

- Cleaning* merupakan tahapan untuk menghapus karakter yang tidak berpengaruh terhadap pengolahan data seperti tanda baca, angka, link, dan lain-lain.
- Case folding* merupakan tahapan untuk mengubah semua huruf menjadi huruf besar atau kecil semua.
- Tokenizing* merupakan tahapan untuk memisahkan teks yang berupa kalimat, paragraf maupun dokumen menjadi potongan kata yang disebut token, tujuannya adalah untuk memperoleh potongan kata yang dijadikan entitas dan mempunyai nilai dalam matriks dokumen teks yang akan dianalisis.
- Stopword* merupakan tahapan untuk menghilangkan kosa kata atau simbol yang tidak memiliki informasi yang berguna seperti kata penghubung dan kata keterangan.
- Stemming* merupakan tahapan untuk mendapatkan kata dasar dari sebuah token yaitu dengan menghapus kata awalan, akhiran, sisipan, dan *confixes* (kombinasi kata awalan dan akhiran) [4][8][9].

## 2. 5 Penerapan metode KNN

Dalam penelitian ini menerapkan metode data *mining* yaitu klasifikasi menggunakan algoritma KNN, Klasifikasi yaitu sebuah bentuk analisis data yang mengekstraksi pola-pola yang mendeskripsikan kelas pada data, algoritma klasifikasi memprediksi label kelas kategorikal terhadap suatu data untuk mengklasifikasikannya ke dalam salah satu kelas yang telah teridentifikasi sebelumnya [10]. Klasifikasi adalah proses untuk menemukan pola yang menggambarkan dan membedakan antara kelas yang satu dengan yang lain dan berfungsi untuk menunjukkan suatu objek termasuk dalam label tertentu yang telah teridentifikasi sebelumnya. Klasifikasi terbagi menjadi 2 proses, pertama yaitu proses pembelajaran (*training*) yang mana sebuah algoritma klasifikasi dibangun untuk menganalisis data latih kemudian diungkapkan kedalam bentuk aturan klasifikasi, kedua yaitu proses klasifikasi yang mana data uji akan digunakan untuk mengevaluasi tingkat akurasi dari aturan klasifikasi tersebut [11].

Algoritma KNN yaitu suatu metode yang dimanfaatkan untuk mengklasifikasikan suatu objek berdasarkan data pembelajaran yang memiliki jarak terdekat dengan objek tersebut, tujuannya adalah untuk melakukan klasifikasi sebuah objek baru yang sesuai dengan atribut dan pola pelatihan [9]. KNN merupakan salah satu metode yang digunakan untuk mengklasifikasikan data, cara kerja dari algoritma KNN ini adalah dengan mengklasifikasikan data berdasarkan jarak terdekat satu data dengan data yang lain, dekat atau jauhnya jarak tersebut bisa dihitung menggunakan besaran jarak [8]. Algoritma KNN merupakan algoritma *supervised learning*, yang mana *supervised learning* adalah sebuah pembelajaran yang terawasi apabila *output* yang diinginkan telah teridentifikasi sebelumnya, pada umumnya pembelajaran tersebut digunakan pada data yang telah mempunyai label. Algoritma KNN melakukan proses klasifikasi objek dengan mengacu pada atribut dan sampel dari data *training* (data latih), jika algoritma ini diberikan titik *query* maka algoritma ini akan menemukan sekumpulan k objek atau titik latih yang memiliki jarak paling dekat dengan titik *query*. Klasifikasi menggunakan jumlah suara terbanyak dari k klasifikasi objek, dan algoritma KNN menggunakan klasifikasi tetangga untuk nilai prediksi dari *query instance* baru. Terdapat metode perhitungan jarak pada algoritma KNN salah satunya yaitu *Euclidean Distance* yang merupakan perhitungan jarak paling dekat, perhitungan jarak paling dekat ini digunakan untuk memperoleh jumlah kesamaan yang dihitung berdasarkan kesamaan kata yang terdapat dalam suatu paragraf [12].

Adapun tahapan implementasi dari algoritma KNN yaitu yang pertama dengan mendefinisikan jumlah nilai k sebagai ketetanggaan yang merupakan jarak terdekat dari data *training* ke data *testing*, tahap yang kedua yaitu menghitung jarak antara nilai vektor data *testing* dengan semua nilai vektor data *training* menggunakan perhitungan jarak *Euclidean Distance* sehingga dapat diketahui labelnya berdasarkan 5 bidang ilmu yang telah ditentukan, rumus algoritma KNN ditunjukkan pada persamaan (1).

$$d_{(i,j)} = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2)} \quad (1)$$

Keterangan :

- $d_{(i,j)}$  = Jarak dokumen i ke j.  
 $x_{i(n)}$  = Kata ke-n pada dokumen i.  
 $x_{j(n)}$  = Kata ke-n pada dokumen j.

Kemudian pada tahap ketiga yaitu mengambil sejumlah nilai parameter k dari data latih yang terdekat, dan tahap terakhir yaitu menentukan kategori yang dominan sebagai kategori untuk data uji[4].

## 2. 6 Evaluasi

Tahap akhir yaitu evaluasi yang mana pada tahap ini akan menghasilkan seberapa besar akurasi penggunaan algoritma KNN dalam mengklasifikasikan data judul artikel *reviewer* ke dalam 5 bidang ilmu yang telah ditentukan dengan melihat hasil akurasi, presisi, *recall*, dan *F-Measure*, yang mana penelitian ini akan menunjukkan apakah metode KNN tersebut layak digunakan dalam proses klasifikasi data *reviewer* terhadap 5 bidang ilmu keparakan. Tahap evaluasi algoritma KNN menggunakan *confusion matrix* terhadap akurasi, presisi, *recall*, dan *F-Measure*, *confusion matrix* merupakan sebuah bentuk perhitungan suatu perbandingan yang mendeskripsikan nilai positif dan nilai negatif dari hasil pengklasifikasian data. Hasil dari perhitungan evaluasi dan validasi akan digunakan untuk menghitung nilai dari *precision*, *recall*, *f-measure* dan *accuracy* sebagai mana yang ditunjukkan pada tabel 2.

Tabel 2 Tabel *confusion matrix*

		Nilai Prediksi	
		<i>Positive</i>	<i>Negative</i>
Nilai Aktual	<i>Positive</i>	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
	<i>Negative</i>	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

Keterangan :

- True Positive* (TP) : Nilai aktual positif yang diprediksi positif pada hasil klasifikasi  
*False Positive* (FP) : Nilai aktual negatif yang diprediksi positif pada hasil klasifikasi  
*True Negative* (TN) : Nilai aktual negatif yang diprediksi negatif pada hasil klasifikasi  
*False Negative* (FN) : Nilai aktual positif yang diprediksi negatif pada hasil klasifikasi

- a. *Precision* adalah perbandingan dari hasil *True Positive* (TP) dengan seluruh dokumen yang diprediksi sebagai *positive* oleh model klasifikasi, ditunjukkan oleh persamaan (2).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

b. *Recall* adalah perbandingan hasil *True Positive* (TP) dengan seluruh dokumen yang memiliki nilai *positive*, ditunjukkan oleh persamaan (3).

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

c. *F-Measure* adalah gabungan dari hasil perhitungan *precision* dan *recall* yang digunakan untuk melihat estimasi kinerja klasifikasi, ditunjukkan oleh persamaan (4).

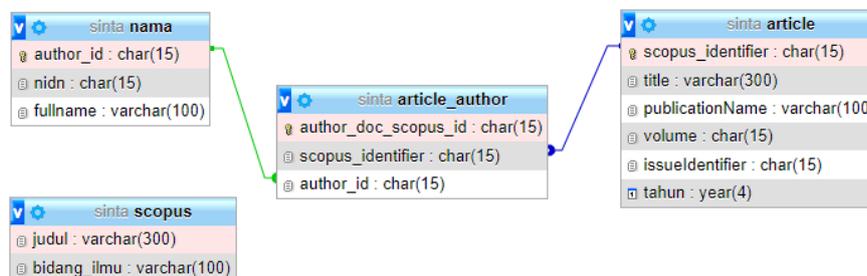
$$f - measure = \frac{2TP}{2TP + FP + FN} \quad (4)$$

d. *Accuracy* adalah hasil dari persentase keseluruhan dokumen yang diprediksi benar, ditunjukkan oleh persamaan (5) [4].

$$Accuracy = \frac{\sum TP + \sum TN}{\sum Data Uji} \quad (5)$$

## 2.7 Desain Database

Pada penelitian ini menggunakan data dengan format *file* CSV dan sql yang kemudian data-data tersebut diimport ke dalam *database* PhpMyAdmin, sehingga dibuatlah 1 *database* dengan nama *sinta*. Dalam merancang *database* diperlukan sebuah diagram yang berfungsi untuk menunjukkan hubungan atau relasi antara entitas dengan entitas yang lain, dan setiap entitas memiliki atribut. *Database relation design* pada penelitian ini ditunjukkan pada gambar 3.



Gambar 3 Database relation design

Gambar 3.4 merupakan tampilan *database relation design* pada penelitian. Terdapat 4 entitas berupa *nama*, *article*, *article\_author*, dan *scopus*, kemudian setiap entitas memiliki atributnya masing-masing seperti yang ada pada gambar. Yang mana pada gambar *database relation design* tersebut menjelaskan sebuah relasi yaitu *one to many* dimana seorang *reviewer* atau penulis dapat menulis *article* lebih dari satu. Dan terdapat satu tabel yang tidak memiliki relasi atau berdiri sendiri yaitu tabel *scopus*, karena tabel *scopus* berisi data judul jurnal terindeks *scopus* dan label bidang ilmu yang digunakan sebagai data *training*.

## 3. HASIL DAN ANALISA

### 3.1 Dataset

Dataset yang digunakan sebagai data *training* adalah data judul jurnal terindeks di *scopus* yang berjumlah 500 data dengan rinciang 100 data *Arts & Humanities*, 100 data *Engineering & Tecnology*, 100 data *Life Sciences & Medicine*, 100 data *Natural Sciences*, dan 100 data *Social Sciences & Management*. Kemudian untuk data *testing* kepakaran menggunakan data *reviewer* yang berjumlah 171 data, dan data judul artikel dengan jumlah keseluruhan data 716 data.

### 3.2 Hasil Pengujian

Pada penelitian ini menerapkan 3 skenario uji split data yaitu 70% : 30%, 80% : 20%, dan 90% : 10%. Untuk Evaluasi hasil akhir pada penelitian ini menggunakan *confusion matrix* terhadap hasil akurasi, presisi, *recall*, dan *f*-

*measure* dengan melakukan berbagai percobaan penentuan jumlah tetangga terdekat ( $k$ ) yaitu  $K = 3, 5, 9, 20, 25, 30, 35$ . Yang mana dilakukan percobaan pengujian 3 skenario uji terhadap semua parameter nilai  $K$ . Adapun tabel hasil perhitungan nilai *accuracy*, *precision*, *recall*, dan *f-measure* dari setiap jumlah  $k$  menggunakan 3 skenario split data ditunjukkan pada tabel 3 sampai dengan 5.

Tabel 3 Tabel *Accuracy, Precision, Recall, F-Measure* menggunakan split data 70% : 30%

Jumlah K	Accuracy	Precision	Recall	F-Measure
3	0.539	0.452	0.306	0.265
5	0.545	0.590	0.3	0.265
9	0.602	0.637	0.346	0.304
20	0.606	0.484	0.373	0.346
25	0.682	0.490	0.459	0.448
30	0.658	0.432	0.433	0.426
35	0.667	0.447	0.440	0.413

Tabel 4 Tabel *Accuracy, Precision, Recall, F-Measure* menggunakan split data 80% : 20%

Jumlah K	Accuracy	Precision	Recall	F-Measure
3	0.580	0.356	0.34	0.314
5	0.535	0.386	0.29	0.270
9	0.612	0.668	0.36	0.312
20	0.698	0.487	0.48	0.473
25	0.711	0.588	0.490	0.481
<b>30</b>	<b>0.751</b>	<b>0.628</b>	<b>0.54</b>	<b>0.528</b>
35	0.619	0.610	0.37	0.359

Tabel 5 Tabel *Accuracy, Precision, Recall, F-Measure* menggunakan split data 90% : 10%

Jumlah K	Accuracy	Precision	Recall	F-Measure
3	0.584	0.484	0.339	0.368
5	0.496	0.382	0.260	0.230
9	0.544	0.531	0.28	0.267
20	0.597	0.517	0.36	0.322
25	0.612	0.384	0.38	0.354
30	0.652	0.464	0.420	0.410
35	0.716	0.514	0.5	0.483

Dari perhitungan *accuracy*, *precision*, *recall*, dan *f-measure* menerapkan 3 skenario split data dengan percobaan beberapa nilai parameter  $K$  yang telah dicatat pada tabel-tabel diatas, maka dapat diketahui bahwa hasil akurasi tertinggi diperoleh pada split data 80% : 20% yaitu 0,751, *precision* 0.628, *recall* 0.54, dan *f-measure* 0.528 pada percobaan menggunakan nilai  $K = 30$ .

### 3.3 Analisa Hasil Pengujian

Berdasarkan dari hasil pengujian terhadap beberapa percobaan yang telah dilakukan dalam penelitian, selain pentingnya dilakukan pembersihan data yang disebut sebagai *preprocessing* data, penentuan split data atau pembagian data antara data *training* dan *testing* dan penentuan parameter nilai  $K$  juga memiliki pengaruh dalam hasil akurasi. Adapun contoh klasifikasi dari hasil prediksi data *reviewer* ke dalam 5 bidang ilmu ditunjukkan pada tabel 6 dan 7.

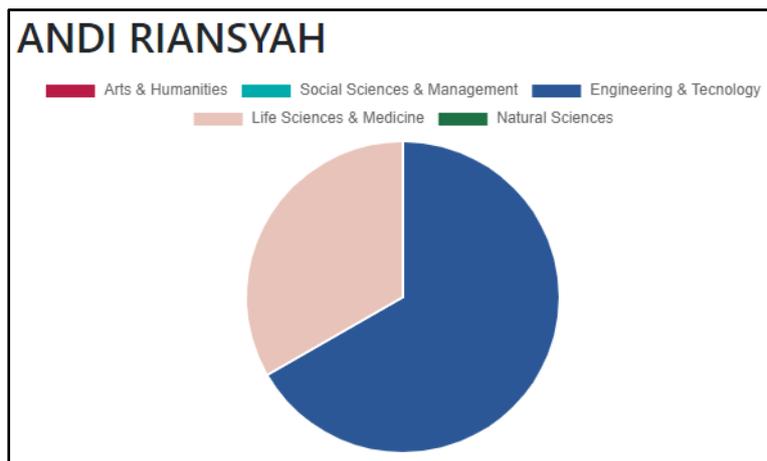
Tabel 6 Tabel data judul artikel *reviewer*

nama <i>reviewer</i> : ANDI RIANSYAH	
No.	Judul Publikasi
1	<i>Data acquisition in determining lab work assesment ranking using Fuzzy Analytic Hierarchy Process (FAHP)</i>
2	<i>Applying fuzzy proportional integral derivative on internet of things for figs greenhouse</i>

3 Fuzzy Implementation for Land Spatial Planning

Tabel 7 Tabel data judul artikel *reviewer* dan hasil prediksi

nama <i>reviewer</i> : ANDI RIANSYAH	
No.	Judul Publikasi
1	<i>Data acquisition in determining lab work assesment ranking using Fuzzy Analytic Hierarchy Process (FAHP)</i> <b>Engineering &amp; Technology</b>
2	<i>Applying fuzzy proportional integral derivative on internet of things for figs greenhouse</i> <b>LifeSciences_ &amp;_ Medicine</b>
3	<i>Fuzzy Implementation for Land Spatial Planning</i> <b>Engineering &amp; Technology</b>

Gambar 4 *Pie chart* dari hasil prediksi

Pada gambar 4 merupakan *pie chart* dari hasil prediksi kepakaran *reviewer*, dari gambar tersebut dapat diketahui bahwa kepakaran *reviewer* yang paling dominan adalah pada bidang ilmu *Engineering & Technology*.

#### 4. KESIMPULAN

Adapun kesimpulan yang dapat diambil dari penelitian ini adalah sistem klasifikasi yang dibangun dengan menerapkan metode KNN memperoleh hasil akurasi paling tinggi setelah dilakukan beberapa percobaan jumlah tetangga terdekat  $K = 30$  menggunakan skenario pengujian 80% : 20% yaitu sebesar 0,751 atau 75 %, serta diperoleh juga hasil *precision* sebesar 0.628, *recall* sebesar 0.54, dan *f-measure* sebesar 0.528.

#### ACKNOWLEDGEMENTS

Tugas akhir ini merupakan bagian dari penelitian dosen Teknik Informatika Universitas Islam Sultan Agung berjudul Pengembangan Sistem Klasifikasi Otomatis Publikasi Ilmiah pada *database SINTA Science and Technology Index* yang dibiayai dari Kementerian Pendidikan dan Kebudayaan, Pendidikan Tinggi, Riset dan Teknologi Republik Indonesia tahun 2022.

#### DAFTAR PUSTAKA

- [1] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner," *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 1, pp. 13–21, 2021, doi: 10.30813/jbase.v4i1.2729.
- [2] R. P. Fitrianti, A. Kurniawati, and D. Agustien, "Implementasi Algoritma K-Nearest Neighbor Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia," *Semin. Nas. Apl. Teknol. Inf.*, 2019.
- [3] A. Supriatman, "Pembobotan TF-IDF pada Judul Penelitian Dosen Sebagai Dasar Klasifikasi Menggunakan Algoritma K-NN ( Studi Kasus : Universitas Siliwangi )," vol. VI, no. 1, pp. 1573–1579, 2021.
- [4] H. P. Hadi and T. S. Sukamto, "Klasifikasi Jenis Laporan Masyarakat Dengan K-Nearest Neighbor Algorithm," *JOINS (Journal Inf. Syst.)*, vol. 5, no. 1, pp. 77–85, 2020, doi: 10.33633/joins.v5i1.3355.
- [5] Y. I. Kurniawan and T. I. Barokah, "Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor," *J. Ilm. Matrik*, vol. 22, no. 1, pp. 73–82, 2020, doi: 10.33557/jurnalmatrik.v22i1.843.
- [6] P. Saadah, O. Nurdiawan, D. A. Kurnia, D. R. Amalia, and I. Ali, "Klasifikasi Pemberian Sanksi Pajak Dengan Metode K-Nearest Neighbour Di KPP Pratama Cirebon Dua," *J. DATA Sci. Inform.*, vol. 1, no. 2, pp. 41–45, 2021.
- [7] M. D. Hendriyanto and B. N. Sari, "Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Judul Berita Hoax," *J. Ilm. Inform.*, vol. 10, no. 02, pp. 80–84, 2022, [Online]. Available:

*Title of manuscript is short and clear, implies research results (First Author)*

- <https://ejournal.upbatam.ac.id/index.php/jif/article/view/5477>
- [8] A. A. Irfa, A. Adiwijaya, and M. S. Mubarak, "Klasifikasi Topik Berita Berbahasa Indonesia Menggunakan k-Nearest Neighbor," *eProceedings Eng.*, vol. 5, no. 2, 2018.
- [9] J. J. A. Limbong, I. Sembiring, K. D. Hartomo, U. Kristen, S. Wacana, and P. Korespondensi, "ANALISIS KLASIFIKASI SENTIMEN ULASAN PADA E-COMMERCE SHOPEE BERBASIS WORD CLOUD DENGAN METODE NAIVE BAYES DAN K-NEAREST ANALYSIS OF REVIEW SENTIMENT CLASSIFICATION ON E-COMMERCE SHOPEE WORD CLOUD BASED WITH NAÏVE BAYES AND K-NEAREST NEIGHBOR METHODS," vol. 9, no. 2, pp. 347–356, 2022, doi: 10.25126/jtiik.202294960.
- [10] A. Nugroho and Y. Religia, "Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 504–510, 2021, doi: 10.29207/resti.v5i3.3067.
- [11] F. Maulana, M. Orisa, and H. Zulfia Zahro', "Klasifikasi Data Produk Mebel Aneka Jaya Menggunakan Metode K-Nearest Neighbor Berbasis Web," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 5, no. 2, pp. 460–466, 2021, doi: 10.36040/jati.v5i2.3782.
- [12] E. Hardiyanto and F. Rahutomo, "STUDI AWAL KLASIFIKASI ARTIKEL WIKIPEDIA BAHASA INDONESIA DENGAN MENGGUNAKAN METODA K NEAREST NEIGHBOR," *Semin. Nas. Terap. Ris. Inov.*, vol. 01, pp. 15–16, 2016.