

Ekstraksi Informasi Pada Dokumen Surat Masuk Menggunakan Algoritma *Fuzzy K-Nearest Neighbour* (*Fuzzy K-NN*)

Adam Sulaiman¹, Nelly Indriani²

Teknik Informatika - Universitas Komputer Indonesia

Jl. Dipati Ukur 114 Bandung

E-mail : sulaimanadam@email.unikom.ac.id¹, nelly.indriani@email.unikom.ac.id²

ABSTRAK

Ekstraksi informasi merupakan salah satu dari cabang text mining yang bertujuan untuk mengambil informasi dari teks yang tidak terstruktur menjadi data yang terstruktur. Ekstraksi informasi dokumen yang menggunakan klasifikasi rule-based memiliki masalah penurunan akurasi pada karakteristik dokumen yang tidak identik. Oleh karena itu, penelitian ini menggunakan machine learning *Fuzzy K-Nearest Neighbour* (*Fuzzy K-NN*) sebagai klasifikasi ekstraksi informasi dokumen. Dokumen yang digunakan dalam penelitian ini adalah surat masuk. Tujuan penelitian ini adalah untuk menganalisis bagaimana tingkat performa klasifikasi *Fuzzy K-NN* dalam ekstraksi informasi surat masuk yang memiliki karakteristik data tidak identik dan membandingkan tingkat akurasi dengan klasifikasi *K-NN*. Penelitian ini menggunakan praproses *converting*, *filtering*, *tagging* dan *tokenization*, serta menggunakan *boolean weighting* sebagai metode pembobotan. Ekstraksi fitur yang digunakan memiliki 11 parameter. Pengujian yang dilakukan menggunakan pengujian *black box* dan perhitungan akurasi dengan *confusion matrix*. Hasil analisis pengujian akurasi menunjukkan bahwa nilai akurasi klasifikasi *Fuzzy K-NN* lebih baik daripada klasifikasi *K-NN* dengan 87,52 % berbanding 85,17%. Berdasarkan hasil tersebut, disimpulkan algoritma *Fuzzy K-NN* dapat diterapkan menjadi klasifikasi dalam ekstraksi informasi dokumen surat masuk.

Kata kunci: Ekstraksi Informasi, Machine Learning, Logika Fuzzy, *K-Nearest Neighbour*, Surat Masuk.

1. PENDAHULUAN

1.1 Latar Belakang

Ekstraksi informasi adalah cabang ruang lingkup dari text mining yang bertujuan untuk mengubah hasil proses text mining menjadi akar yang sama dengan dunia data yang terstruktur dalam data

mining[1] yang didalamnya terdapat metode klasifikasi sebagai penentu kelas data tersebut[2]. Dalam penelitian ini ekstraksi informasi diterapkan pada dokumen surat masuk. Dalam penelitian ini yang dimaksud dengan surat masuk adalah seluruh surat resmi atau dinas yang diterima oleh instansi maupun perseorangan[3].

Penelitian mengenai ekstraksi informasi telah dilakukan sebelumnya oleh Agny Ismaya. Pada penelitian tersebut dibuat sebuah tools dengan rule-based classification untuk melakukan ekstraksi informasi pada dokumen Laporan Hasil Pemeriksaan (LHP) atas Laporan Keuangan Pemerintah Daerah (LKPD)[4]. Dari hasil penelitian diketahui bahwa hasil klasifikasi pada data uji mengalami penurunan. Hal tersebut terjadi karena penggunaan metode *wordmatch* (kecocokan kata), sehingga apabila terjadi sedikit perbedaan pada dokumen akan terjadi kesalahan dalam klasifikasi. Permasalahan ini kemungkinan dapat diselesaikan jika metode klasifikasi dapat menentukan pola dokumen dari segi pembelajaran kemiripan antar dokumennya atau machine learning.

Dokumen yang digunakan dalam ekstraksi informasi pada penelitian ini adalah dokumen surat masuk. Hermawan S. melakukan penelitian administrasi di Kel. Jeruk Kab Sragen, dengan pembuatan aplikasi yang bertujuan untuk menangani proses surat menyurat[5]. Namun, dalam hasil penelitian tersebut penyimpanan data surat masuk masih diinputkan secara manual sehingga prosesnya masih sama dengan pencatatan data surat masuk ke buku agenda atau pengetikan manual ke data arsip komputer. Dengan ekstraksi informasi pada dokumen surat masuk maka pengambilan informasi dapat dilakukan secara otomatis sehingga akan meringkas proses input data secara manual. Namun ide ekstraksi informasi menggunakan rule-based kurang mampu menangani karakteristik data yang tidak identik seperti surat masuk, sehingga tetap dibutuhkannya metode

lain untuk dapat menangani klasifikasi karakteristik data yang tidak identik seperti dokumen surat masuk.

Metode yang akan digunakan untuk proses klasifikasi adalah Fuzzy K-Nearest Neighbour. Pada penelitian yang dilakukan oleh Zhang dkk[6] serta pada penelitian yang dilakukan Satria N. dkk[7], dikatakan bahwa kelebihan dari metode fuzzy K-NN terletak pada tingkat akurasi yang dihasilkan lebih baik daripada metode klasifikasi lain seperti Support Vektor Machine (SVM), dan K-Nearest Neighbour (K-NN). Namun Fuzzy K-NN belum diterapkan dalam ekstraksi informasi dokumen, sehingga belum diketahui apakah bila diterapkan tingkat akurasi Fuzzy K-NN tetap baik atau tidak.

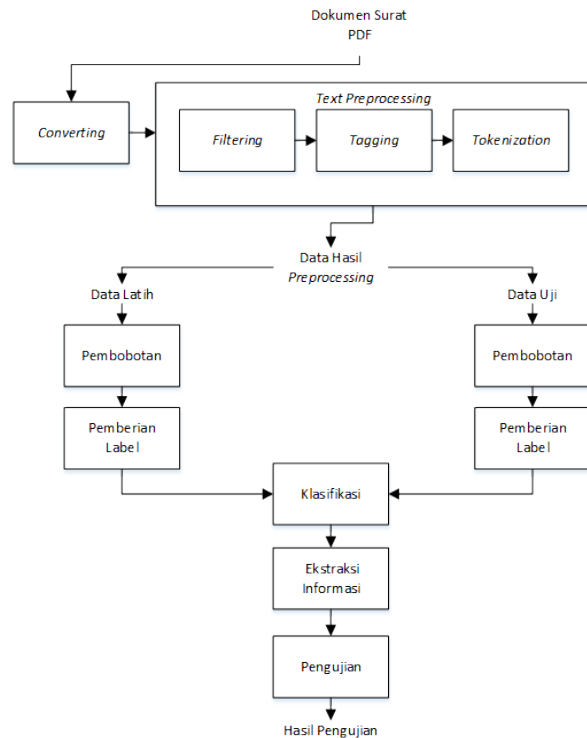
Berdasarkan paparan sebelumnya, maka dalam penelitian ini digunakan metode Fuzzy K-Nearest Neighbour sebagai klasifikasi dalam ekstraksi informasi. Hal ini dimaksudkan untuk mengolah dokumen surat masuk menjadi informasi-informasi yang terstruktur dan mengukur tingkat keakuratan algoritma Fuzzy K-NN dalam ekstraksi informasi pada dokumen surat masuk.

2. ISI PENELITIAN

2.1 Analisis Sistem

2.1.1 Deskripsi Proses

Dalam perancangan sistem akan dibangun ekstraksi informasi pada dokumen surat masuk menggunakan klasifikasi Fuzzy K-NN.. Sebelum dilakukan pembobotan dan klasifikasi, data masukan dilakukan proses converting menggunakan API pdftables, dimana hal ini dilakukan agar pembacaan teks menjadi lebih mudah dan kemudian melalui tahap preprocessing yaitu filtering, tagging, dan tokenization. Setelah pembobotan dan klasifikasi data uji menggunakan Fuzzy K-NN dilakukan ekstraksi informasi dan pengujian terhadap hasil klasifikasi. Adapun gambaran alur prosesnya dapat dilihat pada Gambar 1.



Gambar 1. Deskripsi Proses

2.1.2 Converting

Konversi pdf ke xml adalah proses mengubah file surat yang berformat pdf menjadi xml. Hal ini dilakukan supaya teks surat dapat lebih mudah dilakukan pembacaan oleh sistem. Proses konversi ini dilakukan oleh tools dari API yang berada di domain pdftables.com. Sebagai gambaran konversi pdf ke xml ditampilkan pada Gambar 2. dan Gambar 3.



Gambar 2. File Input Surat

2.1.4 Pembobotan dan Pelabelan

Pembobotan menggunakan boolean weighting dengan 11 fitur yang digunakan dapat dilihat pada Tabel 4. Bernilai 1 jika terdapat fitur tersebut dan bernilai 0 jika tidak[8].

Tabel 4. Fitur yang Digunakan

| FITUR | | |
|-------|-----------------|----------------------------|
| X1 | EMAIL | Memiliki email/web |
| X2 | ALLCAPS | Seluruhnya huruf kapital |
| X3 | DIGIT | Memiliki angka |
| X4 | CONTAINDASH | Memiliki tanda garis |
| X5 | CONTAINSLASH | Memiliki garis miring |
| X6 | INITIAL_KEYWORD | Memiliki kata kunci |
| X7 | LOCATION | Memiliki identitas tempat |
| X8 | CONTAINCOLON | Memiliki titik dua |
| X9 | STRING_LENGTH | Panjang Karakter > 250 |
| X10 | CONTAINCOMMA | Memiliki tanda koma |
| X11 | DATE | Memiliki identitas tanggal |

Setelah dilakukan pembobotan selanjutnya adalah proses pemberian label. Untuk label data latih digunakan dalam proses klasifikasi sementara label data uji digunakan pada proses pengujian akurasi dengan confusion matrix. Adapun label kelas yang digunakan dapat dilihat pada Tabel 5.

Tabel 5. Label yang Digunakan

| No. | KELAS | LABEL KELAS |
|-----|-------|--------------------|
| 1. | 1 | 1. Nama Instansi |
| 2. | 2 | 2. Alamat Instansi |
| 3. | 3 | 3. Tanggal Surat |
| 4. | 4 | 4. Nomor Surat |
| 5. | 5 | 5. Lampiran |
| 6. | 6 | 6. Perihal |
| 7. | 7 | 7. Tujuan |
| 8. | 8 | 8. Isi Surat |

2.1.5 Analisis Klasifikasi Fuzzy K-NN

Tahap klasifikasi dilakukan setelah proses pembobotan dan pelabelan data latih selesai, maka akan dilakukan klasifikasi data uji dengan menggunakan metode Fuzzy K-NN. Berikut penjelasan tahapan yang ada pada algoritma Fuzzy K-NN[9] untuk setiap data uji :

- 1 Menentukan nilai k
- 2.Menghitung jarak dokumen uji dan latih menggunakan metode euclidean distance (1)

$$\|x - x_j\| = \sqrt{\sum_{b=1}^p (x_b - x_{j,b})^2} \quad (1)$$

Keterangan :

$\|x - x_j\|$ = Jarak *euclidean* data uji x dengan data latih ke-j

p = Nilai maks. dari indeks bobot yang dihitung

b = Indeks bobot yang dihitung

x_b = Nilai bobot indeks ke-b di data uji x

$x_{j,b}$ = Nilai bobot indeks ke-b di data latih x_j

contoh menghitung jarak *euclidean distance* :

menghitung jarak data uji h1 dengan data latih d_1

$$\|h1 - d_1\|$$

$$= \sqrt{(h1_1 - d_{1,1})^2 + (h1_2 - d_{1,2})^2 + \dots + (h1_{11} - d_{1,11})^2}$$

$$\|h1 - d_1\| = \sqrt{(0 - 0)^2 + (1 - 1)^2 + \dots + (0 - 0)^2}$$

$$\|h1 - d_1\| = \sqrt{1} = 1$$

3. Ambil jarak tetangga terdekat sejumlah nilai k
4. Perbedaan Fuzzy K-NN dari K-NN adalah adanya perhitungan nilai keanggotaan setiap kelas dengan persamaan (2).

$$\mu_i(x) = \frac{\sum_{j=1}^k \mu_{ij} \left(\frac{1}{2} \right)}{\sum_{j=1}^k (1/\|x - x_j\|^{2/(m-1)})} \quad (2)$$

Keterangan :

$\mu_i(x)$ = Nilai keanggotaan kelas ke-i pada data uji x.

μ_{ij} = Nilai keanggotaan tetangga k (dari data latih x_j) pada kelas ke-i

(bernilai 1 jika kelas tetangga k dari data latih x_j sama dengan kelas ke-i, dan bernilai 0 jika tidak)[9]

j = Indeks data latih

k = Banyaknya nilai ketetanggaan terdekat yang diambil

m = Berat pangkat bobot yang sebanding dengan jarak antara x dan x_j

Nilai keanggotaan kelas dengan logika fuzzy tidak menghitung secara tegas kelas yang dipilih[10] tetapi memperhitungkan seluruh parameter pada setiap kelas, sehingga akurasi diharapkan lebih baik saat penentuan kelas. Contoh menghitung nilai keanggotaan data uji h1 pada kelas 1:

$$\mu_1(h1) = \frac{\mu_{1,9} \|h1 - d_9\|^{-2} + \mu_{1,1} \|h1 - d_1\|^{-2} + \mu_{1,14} \|h1 - d_{14}\|^{-2}}{\|h1 - d_9\|^{-2} + \|h1 - d_1\|^{-2} + \|h1 - d_{14}\|^{-2}}$$

$$\mu_6(h1) = \frac{1 \cdot (0,5)^{-2} + 1 \cdot (1)^{-2} + 0 \cdot (1,73)^{-2}}{(0,5)^{-2} + (1)^{-2} + (1,73)^{-2}} = \frac{5}{5,33} = 0,938$$

5. Ambil nilai keanggotaan kelas terbesar
6. Berikan kelas dengan nilai keanggotaan terbesar sebagai hasil klasifikasi data uji

2.1.6 Ekstraksi Informasi

Setelah hasil klasifikasi didapat maka kalimat-kalimat pada tabel data uji akan disusun ke dalam tabel baru yaitu tabel ekstraksi sesuai dengan kelas hasil klasifikasi. Berikut contoh tabel hasil ekstraksi dapat dilihat pada Tabel 6.

Tabel 6. Tabel Hasil Ekstraksi

| NO | NAMA INSTANSI | ALAMAT INSTANSI | NOMOR SURAT | TANGGAL |
|----|---|--|------------------------|-------------------------------------|
| 1 | MUSYAWARA H GURU MATA PELAJARAN PENDIDIKAN AGAMA ISLAM (PAI) & BUDI PEKERTI SMP NEGERI KABUPATEN KEDIRI | Jalan Raya Turus No. 108 Desa Turus Kec. Gurah Kab. Kediri Email: mgmppaismpka@gmail.com Blog: mgmppaismpka.blogspot.com | Kediri, 4 Oktober 2016 | Nomor : 005/397/418.47.2.89.02/2016 |

2.2 Pengujian Sistem

Pada pengujian sistem ini, data yang digunakan sebanyak 400 data latih dengan 75 data uji yang diperoleh dari total hasil preprocessing 60 dokumen surat masuk yang digunakan. Terdapat dua pengujian yang dilakukan, yaitu :

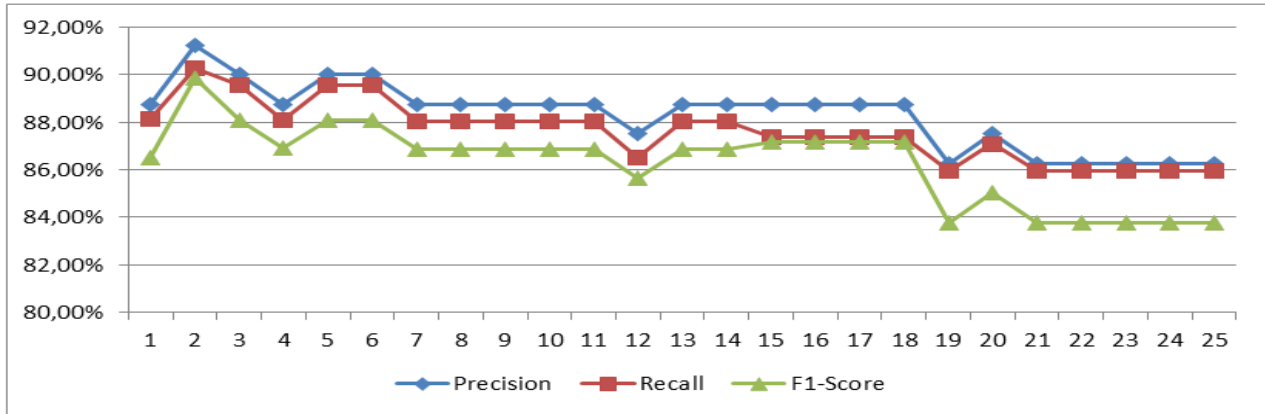
2.2.1 Pengujian Confusion Matrix

Jumlah data latih yang digunakan dalam pengujian ini adalah 400 data dengan sebaran yang sama rata setiap kelasnya berjumlah 50 data dan data uji sebanyak 75 data. Pengujian confusion matrix dilakukan untuk mengetahui tingkat akurasi dari klasifikasi Fuzzy K-NN[11] serta membandingkan dengan akurasi K-NN dan untuk mengetahui nilai k yang paling optimal. Hasil percobaan pengujian pengaruh nilai K terhadap akurasi dapat dilihat pada Tabel 7.

Tabel 7. Hasil Pengujian Confusion Matrix

| No. | k | FKNN | | | KNN | |
|-----------|----|-----------|--------|---------|---------|---------|
| | | Precision | Recall | F-Score | Akurasi | Akurasi |
| 1 | 1 | 88,75% | 88,14% | 86,50% | 88,00% | 88,00% |
| 2 | 2 | 91,25% | 90,28% | 89,84% | 90,67% | 89,33% |
| 3 | 3 | 90,00% | 89,57% | 88,10% | 89,33% | 85,33% |
| 4 | 4 | 88,75% | 88,10% | 86,89% | 88,00% | 85,33% |
| 5 | 5 | 90,00% | 89,57% | 88,10% | 89,33% | 88,00% |
| 6 | 6 | 90,00% | 89,57% | 88,10% | 89,33% | 88,00% |
| 7 | 7 | 88,75% | 88,01% | 86,87% | 88,00% | 86,67% |
| 8 | 8 | 88,75% | 88,01% | 86,87% | 88,00% | 85,33% |
| 9 | 9 | 88,75% | 88,01% | 86,87% | 88,00% | 85,33% |
| 10 | 10 | 88,75% | 88,01% | 86,87% | 88,00% | 84,00% |
| 11 | 11 | 88,75% | 88,01% | 86,87% | 88,00% | 84,00% |
| 12 | 12 | 87,50% | 86,53% | 85,66% | 86,67% | 84,00% |
| 13 | 13 | 88,75% | 88,01% | 86,87% | 88,00% | 84,00% |
| 14 | 14 | 88,75% | 88,01% | 86,87% | 88,00% | 84,00% |
| 15 | 15 | 88,75% | 87,39% | 87,18% | 88,00% | 85,33% |
| 16 | 16 | 88,75% | 87,39% | 87,18% | 88,00% | 84,00% |
| 17 | 17 | 88,75% | 87,39% | 87,18% | 88,00% | 85,33% |
| 18 | 18 | 88,75% | 87,39% | 87,18% | 88,00% | 84,00% |
| 19 | 19 | 86,25% | 85,95% | 83,77% | 85,33% | 84,00% |
| 20 | 20 | 87,50% | 87,09% | 85,02% | 86,67% | 84,00% |
| 21 | 21 | 86,25% | 85,95% | 83,77% | 85,33% | 84,00% |
| 22 | 22 | 86,25% | 85,95% | 83,77% | 85,33% | 85,33% |
| 23 | 23 | 86,25% | 85,95% | 83,77% | 85,33% | 85,33% |
| 24 | 24 | 86,25% | 85,95% | 83,77% | 85,33% | 85,33% |
| 25 | 25 | 86,25% | 85,95% | 83,77% | 85,33% | 85,33% |
| Rata-rata | | 88,30% | 87,61% | 86,31% | 87,52% | 85,17% |

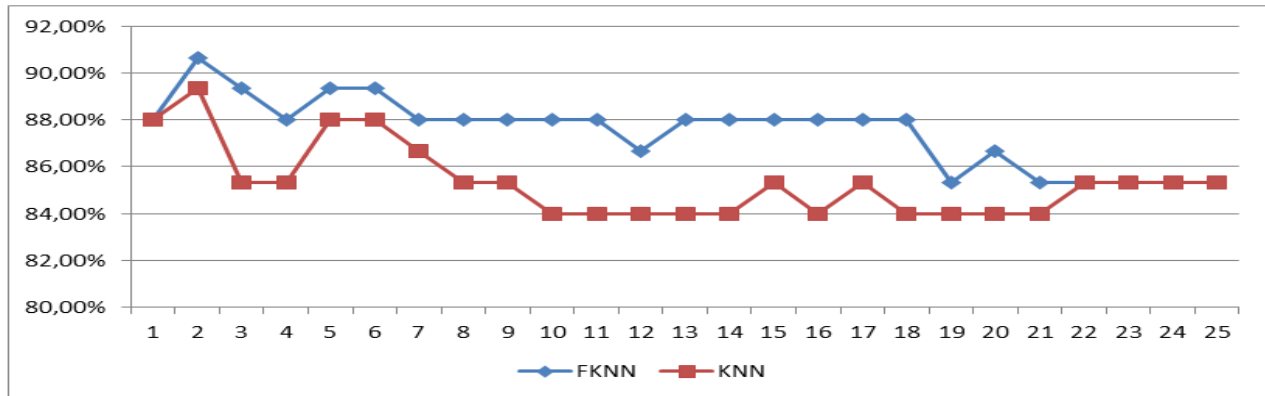
Dari Tabel 7. didapat grafik hasil perhitungan precision, recall dan f1-score yang dapat dilihat pada Gambar 9.



Gambar 9. Grafik Perhitungan Precision, Recall dan F1-score

Dari Tabel 7. pula didapat grafik hasil perbandingan akurasi Fuzzy K-NN dan K-NN pada Gambar 10.

berbeda-beda, akurasi yang dihasilkan pada penelitian ini akan berpengaruh atau tidak. Adapun 9 komposisi yang akan diujikan dapat dilihat pada Tabel 8.



Gambar 10. Grafik Perbandingan FK-NN dan K-NN

Dilihat dari hasil pada Tabel 7. maupun pada grafik di Gambar 9. dan Gambar 10. setelah nilai $k=21$ sampai nilai $k=25$ sudah tidak terjadi perubahan yang signifikan sehingga penggunaan nilai k berhenti di angka 25. Kemudian didapatkan nilai rata-rata precision, recall, f1-score, dan akurasi untuk klasifikasi Fuzzy K-NN adalah bernilai 88,30%, 87,61%, 86,31%, dan 87,52% dengan nilai precision, recall, f1-score dan akurasi tertinggi didapat pada saat nilai $k=2$ yaitu 91,25%, 90,28%, 89,84%, dan 90,67%. Sedangkan rata-rata akurasi dari KNN adalah 85,17% dengan nilai akurasi tertinggi didapat pada saat nilai $k=2$ yaitu 89,33%. Sehingga nilai k yang paling optimal adalah 2.

2.2.2 Pengujian Komposisi Data Latih

Pada pengujian pengaruh komposisi data latih terhadap akurasi, data latih yang digunakan adalah sebanyak 240 data latih dengan komposisi yang berbeda-beda. Pengujian ini bertujuan untuk mengetahui apakah dengan komposisi data latih yang

Tabel 8 Komposisi Data Latih

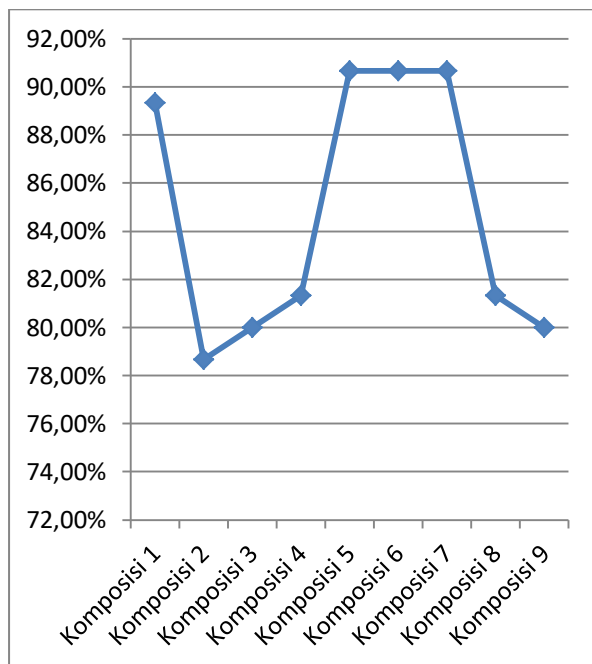
| | Jumlah Data Latih Kelas | | | | | | | | JML |
|----------------------|-------------------------|----|----|----|----|----|----|----|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| Komposisi Data Latih | 1 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 240 |
| | 2 | 40 | 35 | 35 | 30 | 30 | 25 | 25 | 240 |
| | 3 | 20 | 40 | 35 | 35 | 30 | 30 | 25 | 240 |
| | 4 | 25 | 20 | 40 | 35 | 35 | 30 | 30 | 240 |
| | 5 | 25 | 25 | 20 | 40 | 35 | 35 | 30 | 240 |
| | 6 | 30 | 25 | 25 | 20 | 40 | 35 | 35 | 240 |
| | 7 | 30 | 30 | 25 | 25 | 20 | 40 | 35 | 240 |
| | 8 | 35 | 30 | 30 | 25 | 25 | 20 | 40 | 240 |
| | 9 | 35 | 35 | 30 | 30 | 25 | 25 | 20 | 40 |

Hasil dari pengujian komposisi data latih dapat dilihat pada Tabel 9. dan grafik pada Gambar 11.

Tabel 9. Hasil Perhitungan Akurasi Sesuai Komposisi Data Latih

| No. | Komposisi Data Latih | Akurasi |
|-----------|----------------------|---------|
| 1 | Komposisi 1 | 89,33% |
| 2 | Komposisi 2 | 78,67% |
| 3 | Komposisi 3 | 80,00% |
| 4 | Komposisi 4 | 81,33% |
| 5 | Komposisi 5 | 90,67% |
| 6 | Komposisi 6 | 90,67% |
| 7 | Komposisi 7 | 90,67% |
| 8 | Komposisi 8 | 81,33% |
| 9 | Komposisi 9 | 80,00% |
| Rata-rata | | 84,74% |

Gambar 11. Grafik Pengujian Komposisi Data Latih



Berdasarkan Tabel 9. dan Gambar 11., akurasi yang dihasilkan komposisi data latih yang memiliki kelas dominannya adalah kelas 4,5 dan 6 didapati akurasi yang lebih tinggi dari komposisi lain, sedangkan yang tidak memiliki kelas 4,5, dan 6 sebagai kelas dominan dalam komposisinya mengalami penurunan akurasi. Hal tersebut dikarenakan kemampuan sistem untuk mengklasifikasi data yang nilai bobotnya hampir sama (kelas 4,5, dan 6) mengalami penurunan akibat berkurangnya training kelas tersebut dalam data uji. Oleh karena itu perlunya penambahan fitur agar nilai pembobotan dapat lebih

spesifik dan menghindari nilai pembobotan antar kelas yang hampir sama, sehingga dapat meminimalisir penurunan akurasi akibat berkurangnya data latih pada kelas tertentu.

3. PENUTUP

3.1 Kesimpulan

Kesimpulan yang didapat dari hasil penelitian adalah bahwa metode klasifikasi Fuzzy K-NN dapat diterapkan dalam ekstraksi informasi dokumen dan cukup mampu menangani karakteristik data yang tidak identik seperti surat masuk dibuktikan dengan hasil tingkat akurasi dan f1-score yang baik. Selain itu Metode Fuzzy K-NN memiliki akurasi yang lebih baik dibandingkan dengan metode klasifikasi K-NN. Pengujian ekstraksi informasi dokumen surat masuk dengan menggunakan algoritma Fuzzy K-NN sebagai metode klasifikasi menghasilkan akurasi terbesar 90,67% pada k=2.

3.2 Saran

Dalam keterbatasan yang ada, beberapa saran yang dapat diberikan adalah menggunakan media konversi yang lebih baik. karena dalam penelitian ini masih ditemukan noise pada hasil ekstraksi sehingga pembacaan informasi dapat terjadi kesalahan. Hal tersebut disebabkan kesalahan dalam proses konversi pada penelitian ini yang menggunakan API pdftables. Kemudian disarankan untuk menggunakan fitur yang lebih banyak sehingga nilai pembobotan dapat lebih spesifik. Tujuannya meminimalisir terjadinya penurunan akurasi klasifikasi Fuzzy K-NN akibat berkurangnya data latih yang memiliki nilai bobot yang hampir sama.

DAFTAR PUSTAKA

- [1] S. M. Weiss, N. Indurkha, F. Damerou, and T. Zhang, Text Mining: Methods for Analyzing Unstructured Information, Springer, 2004.
- [2] J. Han, M. Kamber, and J. Pei, Data Mining Concepts and Techniques Third Edition, Morgan Kaufmann Publishers, 2012.
- [3] Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia, Pedoman Tata Naskah Dinas Instansi Pemerintah, Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia, 2012.
- [4] Ismaya, Agny, "Algoritma Ekstraksi Informasi Berbasis Aturan", Jurnal Nasional Teknik Elektro dan Teknologi Informasi, Vol. 03, No. 04, pp. 242–247, 2014.
- [5] Hermawan, "Aplikasi Pendataan Arsip Dan Administrasi Di Kelurahan Jeruk Menggunakan Java Netbeans", Universitas Muhammadiyah Surakarta, 2013.

- [6] J. Zhang, Y. Niu, and H. Nie, "Web Document Classification Based on Fuzzy k-NN Algorithm", *Int. Conf. Comput. Intell. Secur.*, pp. 193–196, 2009.
- [7] S. D. Nugraha, R. R. M. Putri, and R. C. Wihandika, "Penerapan Fuzzy K-Nearest Neighbor (FK-NN) Dalam Menentukan Status Gizi Balita", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 9, pp. 925–932, 2017.
- [8] J. Han and M. Kamber, *Data Mining Concepts and Techniques Second Edition*, Morgan Kaufmann Publishers, 2006.
- [9] J. M. Keller and M. R. Gray, "A Fuzzy K-Nearest Neighbor Algorithm", *IEEE Trans. Syst. Man Cybern.*, vol. SMC-15, no. 4, pp. 580–585, 1985.
- [10] I. Nelly, "Model Perilaku Berjalan Agen-Agen Menggunakan Fuzzy Logic", *Jurnal Komputer dan Informatika (KOMPUTA)*, vol. Volume. 1, no. Edisi. I, pp. 37–43, 2012.
- [11] M. Sokolova and G. Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks", *Information Processing & Management.*, vol. 45, no. 4, pp. 427–437, 2009.