KLASIFIKASI SURAT MENGGUNAKAN METODE NAÏVE BAYES PADA SISTEM INFORMASI MANAJEMEN SURAT

(Studi Kasus :Bagian Humas Setda Kabupaten Batang)

Mohamat Dodi Trisetiyo[1], Jati Sasongko Wibowo[2]
Teknik Informatika, Fakultas Teknologi Informasi, Universitas Stikubank Semarang
e-mail: [1]doditrisetiyo@gmail.com, [2]jatisw@edu.unisbank.ac.id

ABSTRAK

Dalam pengelolaan surat menyurat dibagian Humas Setda Kabupaten Batang, membutuhkan pendataan surat dan arsip surat yang begitu banyak. Selama ini proses yang dilakukan masih dengan cara manual, hal ini menimbulkan kesulitan didalam pencarian data surat jika suatu waktu diperlukan. Oleh karena itu, diperlukan adanya aplikasi sistem monitoring yang dapat membantu dalam proses pengarsipan surat, agar dapat mempercepat dalam proses pencarian surat, pengelolaan dan pengagendaan surat. Sistem ini dibuat dengan metode Naïve Bayes Classification (NBC). Metode NBC merupakan metode yang digunakan untuk klasifikasi, dalam penelitian ini data yang digunakan untuk klasifikasi adalah data yang tidak terstruktur berupa teks atau text mining.Sistem monitoring surat menggunakan metode Naïve Bayes yang dibangun berbasis web dengan bahasa pemrograman PHP dan basis data MySQL dengan pembatasan data input surat masuk yang berekstensi (.pdf) yang akan dikonversi dalam bentuk teks menggunakan OCR, ternyata menghasilkan nilai akurasi yang cukup rendah yaitu 33%. Hal itu disebabkan karena hasil konversi dari file PDF dengan OCR akurasinya juga cukup rendah dan menghasilkan teks yang tidak beraturan, sehingga banyak huruf yang tidak sesuai. Setelah melakukan pengujian ulang dengan cara semi manual hasil pengujian sistem didapatkan tingkat akurasi sebesar 83% dengan 6 data latih dan 6 data yang diujikan.

***Kata Kunci**: Monitoring Surat, Text Mining, Naïve Bayes*

1. PENDAHULUAN

Dalam pengelolaan surat menyurat dibagian Humas Setda Kabupaten Batang membutuhkan pendataan surat dan arsip surat yang begitu banyak. Selama ini proses yang dilakukan masih dengan cara manual, hal ini menimbulkan kesulitan didalam pencarian data surat jika suatu waktu diperlukan. Apalagi jika ada permasalahan lain yang terjadi karena kurangnya manajemen arsip surat. Permasalahan yang sering terjadi karena kurangnya manajemen surat misalnya sering tercecer atau hilangnya surat dalam pengagendaan dan pengarsipan surat, laporan surat bulanan atau tahunan yang masih ditulis tangan, timbulnya nomor surat yang *double* atau tumpang tindih, *track* surat dan status surat yang memerlukan waktu lama, kurangnya keamanan dalam menjaga kerahasiaan surat yang bersifat sensitif atau rahasia, serta penyalahgunaan nomor surat dalam instansi[1].

Oleh karena itu, diperlukan adanya aplikasi sistem monitoring yang dapat membantu dalam proses pengarsipan surat, agar dapat mempercepat dalam proses pencarian surat, pengelolaan dan pengagendaan surat.

Rumusan masalah yang diangkat dalam sistem ini adalah "Bagaimana membangun sistem monitoring surat yang dapat membantu dalam proses pengarsipan surat agar dapat mempercepat dalam proses pencarian, pengarsipan, dan pengelolaan surat di Humas Batang?" dan "Bagaimana mengimplementasikan metode *Naïve Bayes Classification* pada sistem monitoring?".

Batasan masalah dari penelitian ini diantaranya adalah

1. Sistem monitoring surat berbasis aplikasi web dengan bahasa pemrograman PHP dan basis data MySQL.
2. Metode yang digunakan dalam sistem monitoring adalah dengan metode *Naïve Bayes Classification* (NBC).
3. Pengambilan sampel berupa surat masuk yang terdiri dari 3 kategori, yaitu Publikasi, Komunikasi, dan Kerjasama yang berekstensi (.pdf).

Text mining adalah bidang khusus dari data mining, hanya yang membedakan adalah jenis datasetnya. Pada data mining dataset yang digunakan yaitudata terstruktur berupa angka, sementara pada text mining datasetyang dipergunakan adalah data yang tidak terstruktur berupa teks[2].

Salah satu implementasi dari text mining yaitu klasifikasi. Tanpa adanya klasifikasi, proses pencarian data akan memakan waktu yang lama dan memberikan hasil yang meluas dari topik pencarian yang dibutuhkan [3].

2. METODE PENELITIAN

Ada beberapa tahapan metode penelitian yang dilakukan dalam menyelesaikan penelitian inisebagai berikut.

2.1 Pengumpulan Data

Teknik pengumpulan data dilakukan dengan cara meneliti langsung, mengadakan pengamatan terhadap permasalahan yang diteliti dibagian Humas Batang dan melakukan wawancara langsung dengan pegawai

dibagian Humas Batang untuk mendapatkan data yang akurat dan mengetahui proses manajemen yang sedang berjalan.

Pengumpulan data juga dengan cara mencari dan mempelajari referensi dari berbagai sumber buku, modul, artikel dan jurnal yang terkait secara langsung maupun tidak langsung untuk mendukung dan mengetahui secara teoritis permasalahan yang dihadapi[4].

2.2 Analisis Sistem

Sistem ini terdiri dari tiga level user, yaitu Super Admin (Kepala Bagian Humas), Administrator (Staff Administrasi Humas), dan User Biasa (Masyarakat Umum). Setiap level user tersebut dibedakan berdasarkan hak akses didalam sistem dan untuk menggunakan sistem harus melakukan login terlebih dahulu kecuali User Biasa karena hak aksesnya hanya bisa membuka tampilan pengagendaan surat (E-Agenda) dari bagian Humas Batang.

2.3 Analisis Data

Pada penelitian ini data yang digunakan adalah data teks surat masuk dari Humas Batang. Untuk melakukan proses *text mining* perlu adanya tahapan pembobotan data teks atau pra-pemrosesan (*preprocessing).* Teks *preprocessing* merupakan suatu proses pengubahan bentuk data tekstual yang belum terstruktur menjadi data yang terstruktur[5]. Tahapan *preprocessing* terdapat beberapa langkah seperti tokenisasi, stopword, steming, dan pembobotan setiap kata dengan menggunakan skema TF *(Term Frequency)*[3]. Hasil dari *text preprocessing* berupa database yang akan digunakan untuk proses klasifikasi [5]. Bobot sebuah *term* pada sebuah teks ditunjukkan pada persamaan (1).

$$W(d,t) = TF(d,t) \tag{1}$$

2.4 Klasifikasi Naïve Bayes

Algoritma *Naïve Bayes Classification* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat [6]. Proses klasifikasi metode NBC dibagi menjadi dua tahap, yaitu tahap pelatihan dan tahap klasifikasi. Tahap pelatihan dilakukan proses analisis data untuk menghitung jumlah kemunculan kata terhadap sampel data yang sudah diketahui kategorinya. Data-data ini digunakan untuk bahan pembelajaran pada tahap proses klasifikasi untuk menentukan data uji termasuk dalam kategori mana[7].Teorema Bayes memiliki bentuk umum yang ditunjukkan pada persamaan (2)

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{2}$$

Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (Vmap). Dengan menerapkan teorema Bayes ditunjukan pada persamaan (3) dapat ditulis :

$$\text{Vmap} = \text{argmax}_{Vj \in V} \frac{P(X1,X2,X3,\ldots Xn|Vj)}{P(X1,X2,X3,\ldots Xn)} \tag{3}$$

Untuk $P(x_1, x_2, x_3)$ nilainya konstan untuk semua kategori *(Vj)* sehingga persamaan (4) menjadi:

$$\text{Vmap} = \text{argmax}_{Vj \in V} P(X1, X2, X3, \ldots Xn|Vj) \tag{4}$$

Persamaan diatas dapat disederhanakan yang ditunjukkan pada persamaan (5).

$$\text{Vmap} = \text{argmax}_{Vj \in V} \prod_i \overset{n}{=} 1 \, P(Wk|Vj)P(Vj) \tag{5}$$

Dimana *Vj* adalah kategori *j*, sedangkan *P(Vj)* adalah probabilitas *Vj*, dan *P*(*Wk*| *Vj*) adalah probabilitas *Wk* dalam kategori *Vj*.Untuk menghitung *P(Vj)* dan *P(Wk /Vj)* dihitung pada saat pelatihan dimana ditunjukkan pada persamaan (6) dan (7).

$$P(Vj) = \frac{|doc\ j|}{|contoh|} \tag{6}$$

Dimana $|doc\ j|$ adalah jumlah dokumen dari setiap kategori, sedangkan $|contoh|$ adalah jumlah dokumen dari semua kategori.

$$P(Wk|Vj) = \frac{nk + 1}{n + |kosakata|} \tag{7}$$

Dimana *nk* adalah jumlah frekuensi dari setiap kata *Wk* dalam dokumen yang berkategori *Vj*, sedangkan nilai *n* adalah jumlah kata dari dokumen berkategori*Vj*, dan |kosakata| adalah jumlah semua kata dari semua kategori.

2.5 Evaluasi

Proses evaluasi pada penelitian ini menggunakan perhitungan akurasi, metode akurasi digunakan untuk mengukur keakuratan hasil dari klasifikasi. Akurasi menghitung banyaknya data prediksi yang benar dari proses klasifikasi yang akan ditunjukkan pada persamaan (8)[7].

$$Accuracy = \frac{\text{jumlah dokumen terklasifikasi dengan benar}}{\text{jumlah dokumen Uji}} \times 100 \quad (8)$$

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini data surat diperoleh dari bagian Humas Batang. Pada proses klasifikasi data, surat yang sudah discan berupa file PDF diinputkan, kemudian file akan diubah kedalam bentuk teks dengan menggunakan OCR. Setelah itu masuk ketahap *preprocessing*, sebagai contoh data teks surat diambil dari perihal surat yang terdiri dari tiga kategori, yaitu Publikasi, Komunikasi, dan Kerjasama.

Contoh data latih :

Publikasi : Publikasi dan Dokumentasi
Komunikasi : Permohonan Sambutan Bupati
Kerjasama : Pemberitahuan Pemberhentian Kerjasama

Dari data tersebut kemudian dilakukan *reprocessing* dan hasilnya akan ditunjukan pada tabel 1.

Tabel 1. Himpunan Data Latih

Data	Kategori	Frekuensi Kemunculan Kata
D1	Publikasi	publikasi(1), dokumentasi(1)
D2	Komunikasi	mohon(1), sambut(1), bupati(1)
D3	Kerjasama	pemberitahuan(1), henti(1), kerjasama(1)

Jumlah kosakata yang dihasilkan dari data latih adalah 8. Kemudian menghitung *P(Vj)* dari setiap kategorinya berdasarkan rumus persamaan (6) yang ditunjukkan pada tabel 2.

Tabel 2. Nilai P(Vj)

Data	Kategori	P(Vj)
D1	Publikasi	1/3
D2	Komunikasi	1/3
D3	Kerjasama	1/3

Untuk setiap kata nk pada kelas vj ditetapkan perhitungan berdasarkan rumus 7. Sebagai contoh untuk menampilkan perhitungannya akan diambil satu kata, yaitu perhitungan terhadap kata "mohon" untuk setiap kategoriyang akan ditunjukkan pada tabel 3.

Tabel 3. Nilai P(mohon)

Vj	Publikasi		Komunikasi		Kerjasama	
	nk = 0	n = 2	nk = 1	n = 3	nk = 0	n = 3
P*(Wk|Vj)*	1/10		2/11		1/11	

Kemudian untuk contoh probabilitas setiap kata dalam setiap kategoriditunjukan pada tabel4.

Tabel 4. Model Probabilitas

Vj	P(Vj)	P(Wk|Vj)							
		publikasi	Dokumentasi	mohon	sambut	bupati	pemberitahuan	henti	kerjasama
D1	1/3	2/10	2/10	1/10	1/10	1/10	1/10	1/10	1/10
D2	1/3	1/11	1/11	2/11	2/11	2/11	1/11	1/11	1/11
D3	1/3	1/11	1/11	1/11	1/11	1/11	2/11	2/11	2/11

Kemudian dilakukan proses klasifikasi data uji dengan mengambil contoh data perihal surat "Bantuan Protokol dan Dokumentasi". Kemudian dilakukan proses *preprocessing* data uji dan hasilnya akan ditunjukkan pada tabel 5.

Tabel 5. Himpunan Data Uji

Frekuensi Kemunculan Kata		
bantu(1)	protokol(1)	dokumentasi(1)

Pada tahap klasifikasi dilakukan perhitungan nilai Vmap untuk setiap kategori berdasarkan rumus persamaan (5).

1. Nilai Vmap untuk kategori "Publikasi"
   Vmap(Publikasi) = P(bantu | Publikasi) P(protocol | Publikasi) P(dokumentasi | Publikasi) P(Vj)
   = 1/10 * 1/10 * 2/10 * 1/3
   = 0,00067
2. Nilai Vmap untuk kategori "Komunikasi"

Vmap(Komunikasi) = P(bantu | Komunikasi) P(protocol | Komunikasi) P(dokumentasi | Komunikasi) P(Vj)
= 1/11 * 1/11 * 1/11 * 1/3
= 0,00025

3. Nilai Vmap untuk kategori “Kerjasama”

Vmap(Komunikasi) = P(bantu | Komunikasi) P(protocol | Komunikasi) P(dokumentasi | Komunikasi) P(Vj)
= 1/11 * 1/11 * 1/11 * 1/3
= 0,00025

Kelas suatu kategori ditentukan berdasarkan Vmap terbesar, pada perhitungan diatas didapatkan bahwa nilai Vmap untuk kelas kategori Publikasi memiliki nilai tertinggi dibandingkan dengan nilai kelas yang lain. Jadi untuk data surat yang diujikan termasuk dalam kategori Publikasi.

Untuk menjalankan metode NBC pada sistem monitoringyang dilakukan pertama adalah mencari nilai dari setiap data yang dibutuhkan pada persamaan rumus NBC yang dijalankan menggunakan query. Hasil query yang dijalankan untuk mencari nilai P(Vj) dari persamaan (6)ditunjukkan pada gambar 1.

```
select *, docs/contoh pvj from (
select count(id) docs, kategori from (
select distinct(id),kategori from token_set where id<>'0' group by kategori,id) as l
group by kategori) as n
join
(select sum(docs) contoh from (
select count(id) docs,kategori from (
select distinct(id),kategori from token_set where id<>'0' group by kategori,id) as l group by kategori) as o) as p;
```

Gambar 1. Query Nilai P(Vj)

Hasil query yang dijalankan untuk mencari nilai P(Wk|Vj) dari persamaan (7)ditunjukkan pada gambar 2.

```
select id, kategori, round(exp(sum(log(nkplus/(n+kosakata)))),9) pwkvj from (
select f.id,term,kategori,nk,nkplus,n from (
select e.id, e.term, e.kategori, sum(nk) nk, sum(nk)+1 nkplus from (
select c.id, c.term, c.kategori, coalesce(d.frequency,0) nk  from (
select a.id, b.term, a.kategori from (
select id,term,kategori from token_set where id<>'0' order by id) as a
join
(select id,term,kategori from token_set where id='0') as b group by b.term,a.id order by a.id,b.term) as c
left join
(select a.id, a.term, a.frequency, a.kategori, coalesce(b.term+a.frequency,0) nk from (
select * from token_set where id<>'0' order by id) as a
left outer join
(select * from token_set where id='0' order by id) as b
on a.term=b.term order by a.id) as d
on c.term=d.term and c.id=d.id group by c.id, c.term order by c.id,c.term) as e
group by e.term, e.kategori order by id) as f
join
(select id,sum(frequency) n from token_set where id<>'0' group by kategori) as g
on f.id=g.id order by f.id,f.term) as h
join
(select sum(frequency) kosakata from token_set where id<>'0') as i group by id;
```

Gambar 2. Query Nilai P(Wk|Vj)

Setelah menemukan nilai P(Vj) dan P(Wk|Vj) adalah menghitung nilai Vmap dari persamaan (5) yang ditunjukkan pada gambar 3.

```
select *, pwkvj*pvj vmap from(
select id, kategori, round(exp(sum(log(nkplus/(n+kosakata)))),9) pwkvj from (
select f.id,term,kategori,nk,nkplus,n from (
select e.id, e.term, e.kategori, sum(nk) nk, sum(nk)+1 nkplus from (
select c.id, c.term, c.kategori, coalesce(d.frequency,0) nk  from (
select a.id, b.term, a.kategori from (
select id,term,kategori from token_set where id<>'0' order by id) as a
join
(select id,term,kategori from token_set where id='0') as b group by b.term,a.id order by a.id,b.term) as c
left join
(select a.id, a.term, a.frequency, a.kategori, coalesce(b.term+a.frequency,0) nk from (
select * from token_set where id<>'0' order by id) as a
left outer join
(select * from token_set where id='0' order by id) as b
on a.term=b.term order by a.id) as d
on c.term=d.term and c.id=d.id group by c.id, c.term order by c.id,c.term) as e
group by e.term, e.kategori order by id) as f
join
(select id,sum(frequency) n from token_set where id<>'0' group by kategori) as g
on f.id=g.id order by f.id,f.term) as h
join
(select sum(frequency) kosakata from token_set where id<>'0') as i group by id) as i

left join

(select kategori, docs/contoh pvj from (
select count(id) docs, kategori from (
select distinct(id),kategori from token_set where id<>'0' group by kategori,id) as l
group by kategori) as n
join
(select sum(docs) contoh from (
select count(id) docs,kategori from (
select distinct(id),kategori from token_set where id<>'0' group by kategori,id) as l group by kategori) as o) as p) as j
on i.kategori=j.kategori;
```

Gambar 3. Query Nilai Vmap

Hasil pengujian data surat yang di*scan* dalam bentuk PDF dan dikonversi kedalam bentuk teks denganmenggunakan metode *Naïve BayesClassification* dengan jumlah data latih sebanyak 6 dokumen dan 6 data dokumen surat yang diujikan ditunjukkan pada tabel 6.

Tabel 7. Hasil Pengujian PDF

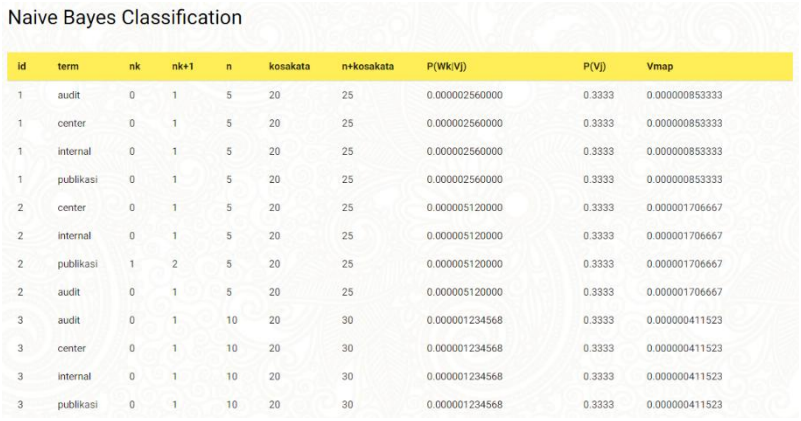
No	Dokumen	Nilai Vmap	Hasil Kategori
1.	Publikasi1	0.000000000001	Kerjasama
2.	Publikasi2	0.000000001924	Kerjasama
3.	Komunikasi1	0.000344352333	Kerjasama
4.	Komunikasi2	0.000003897992	Publikasi
5.	Kerjasama1	0.000000000001	Kerjasama
6.	Kerjasama2	0.000000001065	Kerjasama

Karena hasil yang didapatkan dari pengujian dengan file PDF baik data latih maupun data uji hanya menghasilkan tingkat akurasi yang cukup rendah yaitu 33%, maka dilakukan pengujian ulang dengan cara semi manual. Maksud dari semi manual yaitu isi dari surat masuk diinput manual baik data latih maupun data uji dan hanya diambil pada bagian perihalnya. Hasil pengujian dengan cara semi manual ditunjukkan pada tabel 7.

Tabel 7. Hasil Pengujian Semi Manual

No	Dokumen	Nilai Vmap	Hasil Kategori
1.	Publikasi1	0.000001706667	Publikasi
2.	Publikasi2	0.000085333333	Publikasi
3.	Komunikasi1	0.000000204800	Komunikasi
4.	Komunikasi2	0.004799999995	Komunikasi
5.	Kerjasama1	0.000000000046	Kerjasama
6.	Kerjasama2	0.000000853333	Komunikasi

Hasilpengujian sistemdengan cara semi manual menghasilkan tingkat akurasi yang cukup tinggi yaitu sebesar 83%. Tampilan hasil perhitungan ditunjukkan pada gambar 4.



Naive Bayes Classification

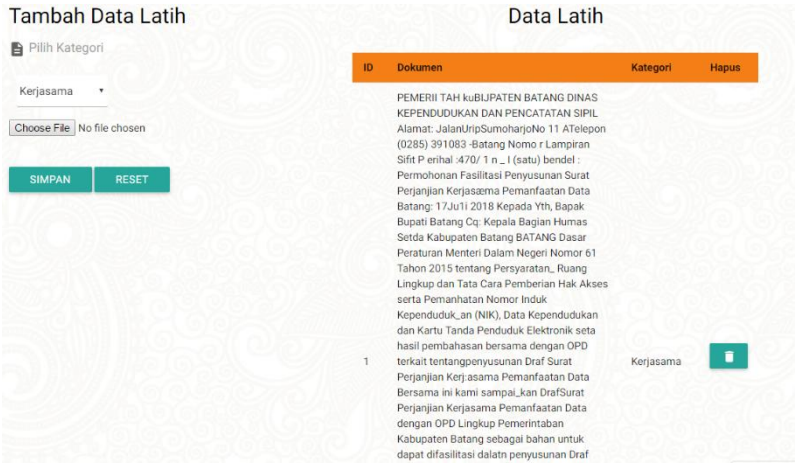
id	term	nk	nk+1	n	kosakata	n+kosakata	P(Wk|Vj)	P(Vj)	Vmap
1	audit	0	1	5	20	25	0.000002560000	0.3333	0.000000853333
1	center	0	1	5	20	25	0.000002560000	0.3333	0.000000853333
1	internal	0	1	5	20	25	0.000002560000	0.3333	0.000000853333
1	publikasi	0	1	5	20	25	0.000002560000	0.3333	0.000000853333
2	center	0	1	5	20	25	0.000005120000	0.3333	0.000001706667
2	internal	0	1	5	20	25	0.000005120000	0.3333	0.000001706667
2	publikasi	1	2	5	20	25	0.000005120000	0.3333	0.000001706667
2	audit	0	1	5	20	25	0.000005120000	0.3333	0.000001706667
3	audit	0	1	10	20	30	0.000001234568	0.3333	0.000000411523
3	center	0	1	10	20	30	0.000001234568	0.3333	0.000000411523
3	internal	0	1	10	20	30	0.000001234568	0.3333	0.000000411523
3	publikasi	0	1	10	20	30	0.000001234568	0.3333	0.000000411523

Gambar 4. Tampilan Hasil NBC

Pada gambar 4 dapat dilihat hasil dari setiap data yang dicari dan dapat dibandingkan hasil nilai Vmap yang terbesar. Hasil klasifikasi sistem ditunjukkan pada gambar 5.

Gambar 5. Tampilan Hasil Klasifikasi

Pada gambar 5 dapat dilihat hasil klasifikasi yang berada dibawah no. agenda setelah menjalankan proses klasifikasi. Tampilan proses input data latih dengan file PDF dapat dilihat pada gambar 6.



Gambar 6. Tampilan Data Latih PDF

Pada gambar 6 diatas dapat dilihat proses input file data yang akan mengkonversi file berekstensi (.pdf) dan hasilnya akan ditampilkan pada tabel data latih disebelahnya.

Gambar 7. Tampilan Data Latih Semi Manual

Pada gambar 7 diatas dapat dilihat proses input data semi manual dan hasilnya akan ditampilkan pada tabel data latih disebelahnya.

4. KESIMPULAN

Sistem monitoring surat menggunkan metode Naïve Bayes yang dibangun berbasis web dengan bahasa pemrograman PHP dan basis data MySQL dengan menggunakan surat masuk yang berekstensi (.pdf) yang akan dikonvert dalam bentuk teks menggunakan OCR ternyata menghasilkan nilai akurasi yang cukup rendah yaitu 33%. Hal itu disebabkan karena hasil konvert dari file PDF dengan OCR akurasinya juga cukup rendah dan hasilnya teks tidak beraturan sehingga banyak huruf yang tidak sesuai. Setelah melakukan pengujian ulang dengan cara semi manual hasil pengujian sistem didapatkan tingkat akurasi sebesar 83% dengan 6 data latih dan 6 data yang diujikan. Sehingga dari hasil tersebut dapat disimpulkan bahwa input data dengan cara semi manual dapat menghasilkan tingkat akurasi yang cukup tinggi dan mengurangi kesalahan pada data teks.

5. SARAN

1. Menggunakan teknik lain yang lebih baik dalam proses seleksi kata yang dianggap tidak perlu ada pada data latih untuk meningkatkan hasil klasifikasi.
2. Menambah jumlah data latih sehingga dapat digunakan sebagai kata kunci dan menampung semua data dari data uji.
3. Dalam pemilihan skema pembobotan pada penelitian ini hanya menggunakan skema TF (*Term Frequency)*, untuk penelitian selanjutnya bisa dikembangkan dengan skema pembobotan *TF-IDF*.
4. Pada penelitian ini menggunakan metode pengelompokan klasifikasi*Naïve Bayes Classification*, dalam penelitian selanjutnya dapat dikembangkan dengan metode pengelompokan lainnya seperti klastering atau pengklasifikasian lainnya.

DAFTAR PUSTAKA

[1] Rizkyanto, Hafidh dan Astuti, Hanum M., 2012, *Pembuatan Perangkat Lunak untuk Workflow Pengelolaan Surat Dinas Bagian Surat Keluar di Pemerintah Kabupaten Buton Utara"*, Jurnal Teknik ITS, Vol. 1, 2301-9271.

[2] Wijaya, Akhmad P., 2016. *Klasifikasi Dokumen dengan Naive Bayes Classifier (NBC) Untuk Mengetahui Konten EGoverment*, Journal of Applied Intelligent System, Vol.1, No. 1, pp. 48-55.

[3] Kolakasari, Dea H., Shofi, Imam M., dan Setyaningrum, Anif H., 2017, *Implementasi Algoritma Multinomial Naive Bayes Classifier Pada Sistem Klasifikasi Surat Keluar (Studi Kasus: Diskominfo Kabupaten Tangerang)*, Jurnal Informatika, Vol. 10, No. 2, Hal 109-118.

[4] Rachmatullah, Sholeh, dan Wijaya, Agung P., 2019, *Rekomendasi Disposisi Surat dengan Metode Naïve Bayes Pada Arsip Surat di Kantor Bakorwil Kabupaten Pamekasan*, Journal of Computer and Information Technology, vol. 2, No. 2, Hal 50-59.

[5] Rahman, Amelia, Wiranto, dan Doewes, Afrizal, 2017, *Online News Classification Using Multinomial Naive Bayes*, Jurnal Ilmiah Teknologi dan Informasi, No. 1, Vol. 6, 2301-7201.

[6] Feldman, R., dan Sanger, J., 2007, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press, New York.

[7] Hamzah, Amir, 2012, *Klasifikasi Teks Dengan Naïve Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita Dan Abstract Akademi*, Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III, Yogyakarta, 3 November.