Pengkategorian Data Angket Mahasiswa dengan Mutual Information dan K-Nearest Neighbor

Indra Tri Saputra

Sistem Informasi, STMIK PPKIA Tarakanita Rahmawati
Jl. Yos Sudarso No. 8 Tarakan Kalimantan Utara
Email : indratriputra@gmail.com

Abstrak. *Salah satu cara untuk memajukan kualitas Perguruan Tinggi yaitu dengan mengevaluasi data angket yang diisi oleh Mahasiswa. Angket biasanya berbentuk pernyataan dalam bentuk uraian yang dapat diisi oleh Mahasiswa terkait pelayanan maupun fasilitas. Pihak penganggung jawab data angket melakukan pengkategorian agar memudahkan dalam evaluasi. Pengkategorian dibedakan menjadi 3 (tiga) yaitu perpustakaan, laboratorium dan fasilitas. Pengkategorian data angket dapat dilakukan secara otomatis yaitu dengan menggunakan teknik text mining. Teknik pengklasifikasian sangat berperan penting dalam menentukan kategori yang cocok terhadap isi pernyataan angket. Tahapan pengklasifikasian teks dimulai dari melakukan pre-processing dokumen, perhitungan nilai pembobotan, perhitungan nilai cosine similarity, dan perhitungan akurasi. Pada penelitian ini, nilai bobot dihitung dengan menggunakan Mutual Information sebagai pengganti TF-IDF, cosine similarity dengan menggunakan metode klasifikasi k-Nearest Neighbor (k-NN), untuk nilai akurasi digunakan confusion matrix. Hasil akhir dari penelitian ini adalah memperoleh nilai akurasi kecocokan sebesar 67% untuk k=3 dan 78% untuk k=5.*

Kata kunci*: klasifikasi, mutual information, k-nn, confusion matrix*

1. Pendahuluan

Angket merupakan sebuah gambaran tentang baik/tidaknya sebuah objek. Angket biasanya terdiri atas pernyataan tentang objek yang akan dinilai. Perguruan tinggi tidak lepas dari sebuah penilaian Mahasiswa. Baik/buruknya sebuah perguruan tinggi tergantung bagaimana pihak perguruan tinggi dapat memberikan fasilitas dan pelayanan khususnya kepada Mahasiswa secara maksimal. Untuk dapat mengukur baik/buruknya diperlukan sebuah kepuasan user dalam hal ini adalah Mahasiswa tentang apa yang dirasakan terkait dengan fasilitas dan pelayanan yang diberikan oleh pihak perguruan tinggi.

Pengkategorian angket dibagi menjadi 3 (tiga) yaitu perpustakaan, laboratorium dan fasilitas. Pemberian kategori angket dapat dilakukan secara otomatis dengan melihat dari isi pernyataan angket. Pengkategorian angket sangat membantu pihak penanggung jawab untuk mempermudah dalam proses evaluasi. Banyak metode yang dapat digunakan untuk melakukan pengkategorian secara otomatis yaitu dengan metode k-NN, Rocchio, Naive Bayes Classifier, WA k-NN dan masih banyak lagi metode yang dapat digunakan.

Pada penelitian ini, penulis menggunakan teknik *Text Mining* (pengolahan kata). *Text mining* merupakan *varian* dari data *mining* yang bekerja dengan cara menemukan pola dari kumpulan data teks dengan jumlah yang besar [1]. Pembobotan kata yang dilakukan dengan menggunakan *Mutual Information* (MI). MI adalah salah satu metode yang dapat digunakan pada *differential cluster labeling* untuk menghitung nilai deskriftif suatu calon *cluster* label. Nilai deskriftif calon *cluster* label tertinggi kemudian dapat dipilih sebagai label dari *cluster* tersebut. MI mengukur derajat ketergantungan dari dua *variable* acak. Formula untuk *Weight Initialization Using Mutual Information* dapat dilihat pada Persamaan (1).

$$MI\ (W) = \sum_{c \in C}(P(c,w)\log\frac{P(c,w)}{P(c)P(w)} + P(c,\bar{w})\log\frac{P(c,\bar{w})}{P(c)p(\bar{w})}) \quad \ldots\ldots\ldots\ldots\ldots (1)$$

Dimana,
P(c) adalah probabilitas dari class c
P(w) adalah probabilitas dari adanya kata
P($\bar{w}$) adalah probabilitas dari tidak adanya kata
P(c,w) dan P(c,$\bar{w}$) adalah probabilitas gabungan [2].

Tahapan awal pengkategorian teks untuk pengolahan dokumen yaitu *pre-processing*. *Pre-processing* dilakukan dalam 3 (tiga) tahapan. Pertama, tokenisasi yang digunakan untuk merubah dokumen menjadi bentuk paling sederhana dalam bentuk kumpulan kata (*term*). Pada tahap ini juga dilakukan perubahan semua huruf menjadi huruf kecil. Kedua, *filtering* yang digunakan untuk menyaring kata dari tahapan tokenisasi yang tidak memiliki makna (*stopword*). *Stopword* yaitu kata yang sering muncul dalam dokumen dengan jumlah banyak tetapi tidak mempunyai kaitan dengan topik tertentu. Contoh *stopword* yaitu "apa", "adalah", "di", dll. Ketiga adalah stemming, merubah kata yang diperoleh dari hasil filtering menjadi kata dasar [3].

Salah satu algoritma yang dapat digunakan untuk melakukan klasifikasi terhadap suatu objek adalah algoritma K-Nearest Neighbor (K-NN). Pengklasifikasian dikakukan berdasarkan k buah data latih dengan melihat jarak yang paling dekat dengan objek tersebut. Untuk menentukan nilai k, ada beberapa syarat yaitu nilai k tidak lebih besar dari jumlah data latih, nilai k bernilai ganjil dan lebih dari satu. Dekat atau jauhnya jarak data latih dengan objek yang akan diklasifikasi dapat dihitung dengan menggunakan metode *cosine similarity*. *Cosine similarity* digunakan untuk menguji ukuran yang dapat digunakan sebagai interpretasi kedekatan jarak berdasarkan kemiripan dokumen [4].

Persamaan (2) berikut ini adalah rumus untuk menghitung jarak pada algoritma KNN dengan metode *cosine similarity*:

$$Cos(\theta_{QD}) = \frac{\sum_{i=1}^{n} Q_i D_i}{\sqrt{\sum_{i=1}^{n} (Q_i)^2} \cdot \sqrt{\sum_{i=1}^{n} (D_i)^2}} \quad \text{..... (2)}$$

Dimana,
Cos(Q,D) = Kemiripan Q terhadap dokumen D
Q = Data Uji
D = Data Latih
n = Banyaknya data

Pengukur akurasi kecocokan klasifikasi dengan menggunakan *confusion matrix*. *Confusion matrix* digambarkan dengan tabel yang berisi jumlah klasifikasi yang benar dan salah terhadap data uji [5]. Berikut adalah tabel *confusion matrix* yang dapat dilihat pada Tabel 1:

Tabel 1. Tabel Confusion Matrix

		Kelas Prediksi	
		+	-
Kelas Sebenarnya	+	TP	FN
	-	FP	TN

Berdasarkan Tabel 1, dapat dijelaskan sebagai berikut:
True Positives (TP) adalah jumlah record data positif yang diklasifikasikan sebagai nilai positif
False Positives (FP) adalah jumlah record data negatif yang diklasifikasikan sebagai nilai positif
False Negatives (FN) adalah jumlah record data positif yang diklasifikasikan sebagai nilai positif
True Negatives (TN) adalah jumlah record data negatif yang diklasifikasikan sebagai nilai negatif.

Nilai yang dihasilkan melalui metode *confusion matrix* adalah berupa *accuracy*, yaitu prosentase jumlah *record* data yang diklasifikasikan (prediksi) secara benar oleh algoritma dengan Persamaan (3).

$$Accuracy = \frac{(TP+TN)}{Total\ Data} \quad \text{..... (3)}$$

Dimana,
TP = *True Positives*
TN = *True Negatives*
Total Data = Jumlah Seluruh Dokumen

2. Pembahasan

Pada tahapan pembahasan, akan diuraikan secara tahap demi tahap yang dilakukan dalam pengkategorian data angket secara otomatis berdasarkan dengan isi angket. Diawali dengan tahapan *pre-processing*, pembobotan *term* dengan *mutual information*, perhitungan *cosine similarity* dan perhitungan akurasi kategori dengan *confusion matrix*.

2.1. *Pre-processing*

Tahap awal yang dilakukan yaitu *pre-processing*. *Pre-processing* dilakukan terhadap sekumpulan data latih yang telah mempunyai label kategori yaitu perpustakaan, laboratorium atau fasilitas. Berikut adalah contoh koleksi data latih pernyataan angket sesuai dengan kategori masing-masing seperti yang ditunjukkan pada Tabel 2.

Tabel 2. Data Latih Angket

No	Data Latih Angket	Kategori
D1	Sebaiknya ruangan Perpustakaan jangan sering tutup, setidaknya ada satu Petugas untuk melayani peminjaman buku	Perpustakaan
D2	Ruangan di lantai 2 sebaiknya diberi pendingin, ruangan terasa panas	Fasilitas
D3	Laptop / PC harus diupdate atau diganti	Laboratorium
D4	Perpustakaan terasa panas	Perpustakaan
D5	AC diruangan kelas lebih diperbanyak lagi	Fasilitas
D6	Asisten sebaiknya jangan berkuasa di Lab	Laboratorium
D7	Waktu peminjaman Buku ditambahkan lagi menjadi 1 minggu	Perpustakaan
D8	Kipas diruangan depan prodi sebaiknya dinyalakan, karena panas	Fasilitas
D9	Waktu praktikum untuk Mahasiswa malam lebih diperbanyak	Laboratorium

Pada Tabel 2, terdapat 9 (sembilan) data angket yang merupakan data latih dengan masing-masing kategori sebanyak 3 (tiga) data. Langkah selanjutnya yaitu melakukan *pre-processing* yaitu tokenisasi, *filtering* dan *stemming*. Hasil dari *pre-processing* dapat dilihat pada Tabel 3.

Tabel 3. Hasil *Pre-processing* Data Latih Angket

No	Data Latih Angket	Kategori
D1	ruang pustaka petugas layan pinjam buku	Perpustakaan
D2	ruang lantai dingin ruang panas	Fasilitas
D3	laptop pc update	Laboratorium
D4	pustaka panas	Perpustakaan
D5	ac ruang kelas	Fasilitas
D6	asisten lab	Laboratorium
D7	pinjam buku	Perpustakaan
D8	kipas ruang prodi panas	Fasilitas
D9	praktikum mahasiswa	Laboratorium

Tabel 3 merupakan hasil *pre-processing* data latih angket. Dari hasil *pre-processing* dibuatkan sebuah matriks antara *term* dan dokumen yang biasa disebut sebagai *inverted index*. Hasil *inverted index* yang terbentuk dari data latih angket dapat dilihat pada Tabel 4.

Tabel 4. *Inverted Index* Data Latih Angket

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9
Ruang	1	2	0	0	1	0	0	1	0
Pustaka	1	0	0	1	0	0	0	0	0
Petugas	1	0	0	0	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…
Buku	1	0	0	0	0	0	1	0	0
Update	0	0	1	0	0	0	0	0	0
Prodi	0	0	0	0	0	0	0	1	0
Praktikum	0	0	0	0	0	0	0	0	1
Mahasiswa	0	0	0	0	0	0	0	0	1

Inverted index pada Tabel 4 menyatakan, *term* "ruang" terdapat pada dokumen D1, D2, D5 dan D8. Sedangkan untuk *term* "mahasiswa" hanya terdapat pada dokumen D9.

2.2. *Mutual Information* (MI)

Pada penelitian sebelumnya dengan judul Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan *Cosine Similarity* yang dilakukan oleh Bening Herwijayanti dkk. Pembobotan *term* dilakukan dengan menggunakan metode pembobotan TF-IDF. Metode TF-IDF dilakukan dengan menggabungkan frekuensi kemunculan kata dan *inverse* frekuensi dokumen yang mengandung kata tersebut [6]. Pada penelitian ini, penulis menggunakan pembobotan *term* dengan MI. MI digunakan untuk memberikan label kategori pada setiap *term* yang terdapat pada tabel *inverted index*. Perhitungan label kategori dilakukan dengan menggunakan Persamaan (1). Berikut ini adalah contoh perhitungan nilai MI untuk *term* "ruang" dengan kategori "perpustakaan".

Berdasarkan tabel *inverted index*, dibuatkan tabel *contingency* dari *term* "ruang" pada kategori "perpustakaan" dapat dilihat pada Tabel 5.

Tabel 5. Tabel *Contingency* dari *Term* "ruang" pada kategori "perpustakaan"

	Dokumen pada kategori "perpustakaan"	Dokumen yang tidak pada kategori "perpustakaan"
Dokumen yang mengandung term "ruang"	1 + 1 (a)	3 + 1 (c)
Dokumen yang tidak mengandung term "ruang"	2 + 1 (b)	3 + 1 (d)

Setiap nilai yang terdapat pada Tabel 5 ditambah dengan 1, untuk menghindari perhitungan yang tidak terdefinisi jika ada jumlah dokumen pada tabel *contingency* bernilai 0.
Total dokumen = (a) + (b) + (c) + (d) = 2+3+4+4 = 13
Probabilitas gabungan kategori dan *term* dilakukan perhitungan sebagai berikut:
p (*term* "ruang" kategori "perpustakaan") = a/total dokumen = 2/13 = 0.154
p (bukan *term* "ruang" kategori "perpustakaan") = b/total dokumen = 3/13 = 0.231
p (*term* "ruang" selain kategori "perpustakaan") = c/total dokumen = 4/13 = 0.308
p (bukan *term* "ruang" selain kategori "perpustakaan") = d/total dokumen = 4/13 = 0.308
Probabilitas kategori dan *term* dilakukan perhitungan sebagai berikut:
p (kategori "perpustakaan") = (a) + (b) / total dokumen = (2 + 3)/13 = 0.385
p (*term* "ruang") = (a) + (c) / total dokumen = (2 + 4)/13 = 0.462
p (selain kategori "perpustakaan") = (c) + (d) / total dokumen = (4 + 4)/13 = 0.615
p (selain *term* "ruang") = (b) + (d) / total dokumen = (3 + 4)/13 = 0.538

Nilai MI *term* "ruang" = (0.154*log(0.154/(0.385*0.462))) + (0.231*log(0.231/(0.385*0.538))) + (0.308*log(0.308/(0.615*0.462))) + (0.308*log(0.308/(0.615*0.538))) = 0.007. Hasil dari pembobotan MI dapat dilihat pada Tabel 6.

Tabel 6. *Mutual Information Term*

Term	Perpustakaan	Fasilitas	Laboratorium	Kategori
Ruang	0.007	0.219	0.131	Fasilitas
Pustaka	0.183	0.025	0.025	Perpustakaan
Petugas	0.071	0.002	0.002	Perpustakaan
…	…	…	…	…
Buku	0.183	0.025	0.025	Perpustakaan
Update	0.002	0.002	0.071	Laboratorium
Prodi	0.002	0.071	0.002	Fasilitas
Praktikum	0.002	0.002	0.071	Laboratorium
mahasiswa	0.002	0.002	0.071	Laboratorium

Kategori pada *term* ditentukan dengan mengambil nilai MI pada masing-masing kategori yang mempunyai nilai tertinggi. Pada Tabel 6 menyatakan, *term* "ruang" termasuk kategori Fasilitas dengan nilai MI 0.219 lebih besar dari nilai MI kategori Perpustakaan dan Laboratorium.

2.3. K-Nearest Neighbor (k-NN)

Untuk melakukan pengkategorian pada data uji digunakan algotirma k-NN dengan menggunakan *cosine similarity* seperti pada Persamaan (2). Pada penelitian ini, terdapat 9 (sembilan) data angket yang dijadikan sebagai data uji. Contoh data uji ke-1 yang akan dikategorikan secara otomatis yaitu "buku – buku pustaka harap di*update* sesuai dengan kondisi sekarang". Data uji ke-1 yang telah mengalami *pre-processing* "buku buku pustaka *update*". Tabel *inverted index* data latih dan data uji dapat dilihat pada Tabel 7.

Tabel 7. *Inverted Index* Data Latih dan Data Uji dengan Nilai MI

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	Uji ke-1	MI
ruang	1	2	0	0	1	0	0	1	0	0	0.219
pustaka	1	0	0	1	0	0	0	0	0	1	0.183
petugas	1	0	0	0	0	0	0	0	0	0	0.071
…	…	…	…	…	…	…	…	…	…	…	…
buku	1	0	0	0	0	0	1	0	0	2	0.183
update	0	0	1	0	0	0	0	0	0	1	0.071
prodi	0	0	0	0	0	0	0	1	0	0	0.071
praktikum	0	0	0	0	0	0	0	0	1	0	0.071
mahasiswa	0	0	0	0	0	0	0	0	1	0	0.071

Untuk mendapatkan bobot *term*, dilakukan perkalian nilai MI terhadap masing-masing dokumen data latih dan uji yang diperoleh dari Tabel 7. Hasil pembobotan MI dapat dilihat pada Tabel 8.

Tabel 8. Nilai Bobot *Term*

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	Uji ke-1
ruang	0.219	0.438	0	0	0.219	0	0	0.219	0	0
pustaka	0.183	0	0	0.183	0	0	0	0	0	0.183
petugas	0.071	0	0	0	0	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…	…
buku	0.183	0	0	0	0	0	0.183	0	0	0.365
update	0	0	0.071	0	0	0	0	0	0	0.071
prodi	0	0	0	0	0	0	0	0.071	0	0
praktikum	0	0	0	0	0	0	0	0	0.071	0
mahasiswa	0	0	0	0	0	0	0	0	0.071	0

Pada Tabel 8 menyatakan *term* "ruang" pada dokumen D1 mempunyai bobot sebesar 0.219 yang diperoleh dari jumlah *term* "ruang" pada dokumen D1 dikalikan dengan nilai MI *term* "ruang". Langkah selanjutnya yaitu menghitung *cosine similarity* dengan menggunakan nilai-nilai pada Tabel 8, dengan tahapan sebagai berikut:

$\sum_{i=1}^{n} Q_i D_i$ yaitu menjumlahkan nilai bobot setiap term pada dokumen D1 dikalikan dengan bobot term data uji. Perhitungan dijabarkan sebagai berikut:
(0.219*0.000) + (0.183*0.183) + (0.071*0.000) + … + (0.183*0.365) + (0.000*0.071) + (0.000*0.000) + (0.000*0.000) + (0.000*0.000) = 0.100

$\sqrt{\sum_{i=1}^{n} (Q_i)^2}$ yaitu menjumlahkan nilai bobot setiap *term* pada dokumen data uji dipangkat 2 kemudian diakarkan. Perhitungan dijabarkan sebagai berikut:
√ (0.000^2) + (0.183^2) + (0.000^2) + … + (0.365^2) + (0.071^2) + (0.000^2) + (0.000^2) + (0.000^2) = 0.414

$\sqrt{\sum_{i=1}^{n} (D_i)^2}$ yaitu menjumlahkan nilai bobot setiap *term* pada dokumen D1 dipangkat 2 kemudian diakarkan. Perhitungan dijabarkan sebagai berikut:
√ (0.219^2) + (0.183^2) + (0.071^2) + … + (0.183^2) + (0.000^2) + (0.000^2) + (0.000^2) + (0.000^2) = 0.398

$Cos(\theta_{QD})$ yaitu menghitung jarak data uji dengan dokumen D1. Perhitungan dijabarkan sebagai berikut:
0.11 / (0.414*0.398) = 0.607

hasil perhitungan jarak data uji dengan seluruh dokumen dapat dilihat pada Tabel 9.

Tabel 9. Hasil *Cosine Similarity*

Dokumen	$\sum_{i=1}^{n} Q_i D_i$	$\sqrt{\sum_{i=1}^{n} (Q_i)^2}$	$\sqrt{\sum_{i=1}^{n} (D_i)^2}$	$Cos(\theta_{QD})$	Jarak Terdekat	Kategori Dokumen
D1	0.100	0.414	0.398	0.607	2	Perpustakaan
D2	0	0.414	0.458	0	5	Fasilitas
D3	0.005	0.414	0.124	0.099	4	Laboratorium
D4	0.033	0.414	0.203	0.396	3	Perpustakaan
D5	0	0.414	0.241	0	6	Fasilitas
D6	0	0.414	0.101	0	7	Laboratorium
D7	0.067	0.414	0.458	0.623	1	Perpustakaan
D8	0	0.414	0.257	0	8	Fasilitas
D9	0	0.414	0.101	0	9	Laboratorium

Pada Tabel 9 menyatakan, dokumen D1 memiliki urutan ke-2 dan D2 urutan ke-5 jarak terdekat dari data uji. Penentuan urutan jarak terdekat yaitu dengan mengambil nilai *cosinie similarity* dari yang terbesar sampai terkecil. Langkah selanjutnya yaitu menentukan label kategori pada data uji dengan cara melakukan pengurutan jarak terdekat dari 1 sampai dengan 9. Dikarenakan data latih yang digunakan sebanyak 9 dokumen, maka nilai k yang digunakan yaitu k=3 dan k=5. Nilai k=3 dengan hasil kategori perpustakaan, dan k=5 dengan hasil kategori perpustakaan. Hasil label kategori terhadap 9 (sembilan) data uji dapat dilihat pada Tabel 10.

Tabel 10. Label Kategori Data Uji

Dokumen Data Uji	Kategori Manual	Kategori Dokumen	
		k=3	k=5
Q1	Perpustakaan	Perpustakaan	Perpustakaan
Q2	Fasilitas	Fasilitas	Fasilitas
Q3	Laboratorium	Laboratorium	Laboratorium
Q4	Perpustakaan	Perpustakaan	Perpustakaan
Q5	Fasilitas	Fasilitas	Fasilitas
Q6	Laboratorium	Laboratorium	Perpustakaan
Q7	Perpustakaan	Laboratorium	Laboratorium
Q8	Laboratorium	Perpustakaan	Laboratorium
Q9	Fasilitas	Perpustakaan	Fasilitas

2.4. *Confusion Matrix*

Confusion matrix digunakan untuk mengukur prosentase akurasi dari pengkategorian dengan nilai k=3 dan k=5. Dengan menggunakan Persamaan (3), diperoleh nilai akurasi sebesar 67% untuk k=3 dan 78% untuk k=5.

3. Kesimpulan

Dari uraian hasil pembahasan, dapat disimpulkan menjadi beberapa poin sebagai berikut:
1. Pembobotan term dengan menggunakan Mutual Information dapat digunakan sebagai pengganti TF-IDF.
2. Penentuan urutan jarak terdekat antara data uji terhadap kumpulan data latih diambil dari nilai *cosine similarity* yang terbesar sampai terkecil.
3. Dari hasil uji coba terhadap 9 (sembilan) data angket diperoleh nilai akurasi terhadap k=3 sebesar 67% dan k=5 sebesar 78%.
4. Untuk pengembangan lebih lanjut dapat menggunakan algoritma WA k-NN yang merupakan pengembangan dari algoritma k-NN.

Ucapan Terima Kasih
Terima kasih penulis sampaikan yang sebesar-besarnya kepada orang-orang yang sangat berperan penting dalam pembuatan penelitian ini yang tidak bisa penulis sebut satu persatu. Sehingga penelitian ini dapat selesai dengan tepat waktu.

Daftar Pustaka

[1]. Indriani, A., dkk. 2018. *Implementasi Jaccard Index dan N-Gram pada Rekayasa Aplikasi Koreksi Kata Berbahasa Indonesia*. Jurnal Nasional SEBATIK ISSN 1410-3737 Vol. 22 No. 2, pp 95-101. STMIK Widya Cipta Dharma, Samarinda.

[2]. Indriani, A., dkk. 2013. *Weight Adjusted K-Nearest Neighbor dan Minimum Spanning Tree untuk Information Retrieval System di Perpustakaan STMIK PPKIA Tarakanita Rahmawati Tarakan*. Seminar Nasional Aplikasi Teknologi dan Informasi (SNAT) ISSN 1907-5022, pp F18-F22. Universitas Islam Indonesia, Yogyakarta.

[3]. Ruli, Riki A. Siregar., dkk. 2017. *Aplikasi Penentuan Dosen Penguji Skripsi menggunakan Metode TF-IDF dan Vector Space Model*. Journal of Computer Science and Information Systems ISSN 2549-2810 Vol. 1 No. 2, pp 171-186. Universitas Tarumanagara, Jakarta.

[4]. Rivki, M. dan Bachtiar, A.M. 2017. *Implementasi Algoritma K-Nearest Neighbor dalam Pengklasifikasian Follower Twitter yang menggunakan Bahasa Indonesia*. Jurnal Sistem Informasi ISSN 2088-7043 Vol. 13 No. 1, pp 31-37. Universitas Indonesia, Depok.

[5]. Rahman, M. Fadly., dkk. 2017. *Klasifikasi untuk Diagnosa Diabetes menggunakan Metode Bayesian Regularization Neural Network (RBNN)*. Jurnal Informatika ISSN 1978-0524 Vol. 11 No. 1, pp 36-45. Universitas Ahmad Dahlan, Yogyakarta.

[6]. Herwijayanti, B., dkk. 2018. *Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer ISSN 2548-964X Vol. 2 No. 1, pp 306-312. Universitas Brawijaya, Malang