

IMPLEMENTASI TEXT MINING DALAM KLASIFIKASI BUKU DENGAN METODE NAÏVE BAYES CLASSIFIER STUDI KASUS PADA PERPUSTAKAAN STMIK PPKIA TARAKANITA RAHMAWATI

Dikky Praseptian M.¹⁾, Aida Indriani²⁾

¹⁾ Sistem Informasi, STMIK PPKIA Tarakanita Rahmawati
Jl. Yos Sudarso No.8 Tarakan Barat
email: dikkypraseptian@gmail.com

²⁾ Teknik Informatika, STMIK PPKIA Tarakanita Rahmawati
Jl. Yos Sudarso No.8 Tarakan Barat
email: ditamaniez2014@gmail.com

Abstrak- Pada implementasi metode naive bayes classifier terhadap data buku dibutuhkan informasi yang tepat dan akurat. Dalam hal ini metode naive bayes classifier mengklasifikasikan data berdasarkan probabilitas kata pada sinopsis yang telah di ambil sampelnya, naive bayes classifier menggunakan prinsip bahwa probabilitas maksimum pada kategori yang telah di tentukan sebelumnya itulah hasil keputusan klasifikasi.

Aplikasi yang dibuat menggunakan metode naive bayes classifier ini, diawali mengambil 100 data sampel buku. Setelah itu diambil lagi 75 data untuk dijadikan data training dan 25 untuk data testing. Langkah berikutnya memecah sinopsis menjadi kata kata yang akan dihitung. Kemudian mencari probabilitas data training yang nantinya dijadikan untuk pembandingan data tes. Selanjutnya, mencocokkan perhitungan data manual klasifikasi data tes dengan klasifikasi naive bayes classifier. Setelah selesai hitung akurasi klasifikasi naive bayes classifier. Setelah mendapatkan akurasi yang cukup besar, berarti metode ini dapat digunakan untuk klasifikasi buku.

Kata Kunci: Implementasi, Text Mining, Klasifikasi, Naive Bayes Classifier.

I. PENDAHULUAN

Perpustakaan adalah bagian yang penting dari suatu Sekolah ataupun Universitas karena menyediakan buku-buku referensi. Kesulitan yang terjadi adalah ketika perpustakaan harus mengklasifikasi buku-buku referensi tersebut sesuai dengan kategori atau jenisnya.

Ada beberapa buku yang memiliki kemiripan isi dengan jenis buku yang berbeda. Ada juga buku-buku yang memang mutlak memiliki satu jenis kategori, selama ini proses klasifikasi itu hanya dilakukan manual oleh petugas perpustakaan dengan melihat berdasarkan sinopsis buku.

Seiring berkembangnya teknologi terutama komputer. *Software* dapat membantu melakukan proses klasifikasi buku baik berdasarkan sinopsis ataupun judul. Namun belum banyak perpustakaan yang menggunakan *software* untuk melakukan proses klasifikasi tersebut.

Alasan penulis menggunakan *Text Mining* dengan memanfaatkan metode *Naive Bayes Classifier* karena dianggap mampu membantu menyelesaikan permasalahan yang ada dalam hal pengklasifikasian. Terutama memberikan keputusan yang lebih tepat dan cepat pada pengklasifikasian buku berdasarkan sinopsisnya. Berbeda dengan *Clustering*, klasifikasi digunakan ketika sasaran kategori sudah jelas

kategoriannya sedangkan *clustering* mengelompokkan data terlebih dahulu baru kemudian menentukan kategoriannya. Sehingga klasifikasi lebih sering digunakan untuk teks sedangkan *clustering* untuk data utuh.

Dibutuhkan waktu yang lama untuk mengklasifikasikan buku karena masih dilakukan secara manual oleh petugas perpustakaan belum adanya sistem komputerisasi serta dalam proses klasifikasi buku yang memiliki kandungan isi lebih dari satu kategori buku itu akan membutuhkan waktu yang lebih lama untuk klasifikasinya.

II. LANDASAN TEORI

Dengan adanya teknologi informasi dan komunikasi dapat memudahkan kita untuk belajar dan mendapatkan informasi yang kita butuhkan dari mana saja, kapan saja, dan dari siapa saja. Perkembangan zaman di era globalisasi ini juga telah banyak mendorong keingintahuan manusia yang sangat besar terhadap segala sesuatu yang mendorong untuk terciptanya suatu penemuan-penemuan baru yang sangat berguna dan bermanfaat bagi kehidupan manusia. Dimana dalam setiap kehidupan tidak lepas dari kehidupan akan informasi. Informasi dalam dunia informatika sangat penting perannya, hal ini karena

dalam sebuah pengambilan keputusan sangat diperlukan suatu informasi. Sistem informasi memiliki komponen berupa subsistem yang merupakan elemen-elemen yang lebih kecil yang membentuk sistem informasi tersebut misalnya bagian input, proses, output. Pada penelitian ini akan di bahas salah satu pengklaifikasian menggunakan Text Mining dengan metode naïve bayes classifier yang diharapkan dapat menjadi salah satu media untuk pengklasifikasian.

Dengan melihat semakin berkembangnya teknologi, maka keberadaan sistem komputerisasi merupakan salah satu cara untuk meningkatkan pelayanan dan keakuratan informasi. Namun sebelum melangkah pada implementasi Text Mining dengan metode naïve bayes classifier, perlu diketahui terlebih dahulu apa yang dimaksud dengan implementasi, Text Mining dan Naive bayes classifier.

2.1. Implementasi

Implementasi menurut beberapa sumber bahwa implementasi adalah sebagai berikut :

1. Implementasi adalah proses interaksi antara tujuan dan tindakan untuk mencapainya. Implementasi memerlukan jaringan pelaksana, birokrasi yang efektif. Efektivitas implementasi ditentukan oleh kemampuan untuk membuat hubungan dan sebab-sebab yang logis antara tindakan dan tujuan.[2]
2. Menurut Majone dan Wildavsky, mengartikan implementasi sebagai evaluasi.[4]
3. Browne dan Wildavsky, mengemukakan bahwa Implementasi adalah perluasan aktivitas yang saling menyesuaikan.[4]
4. Menurut Mclaughlin implementasi sebagai aktivitas yang saling menyesuaikan. [4]
5. Menurut Schubert mengemukakan bahwa implementasi adalah sistem rekayasa. [4]

Ungkapan-ungkapan tersebut mengandung arti bahwa implementasi bukan sekedar aktivitas, tetapi suatu kegiatan yang terencana dan dilakukan secara sungguh-sungguh berdasarkan acuan norma tertentu untuk mencapai tujuan kegiatan. [3]

2.2. Text Mining

Text mining adalah proses menambang data berupa teks dengan sumber data biasanya dari dokumen dan tujuannya adalah mencari kata - kata yang mewakili dalam dokumen sehingga dapat dilakukan analisa keterhubungan dalam dokumen. Data teks akan diproses menjadi data numerik agar dapat dilakukan proses lebih lanjut. Sehingga dalam *text mining* ada istilah *preprocessing data*, yaitu proses pendahulu yang diterapkan terhadap data teks yang bertujuan untuk menghasilkan data numerik. [1]

2.3. Naive Bayes Classifier

Naïve Bayes merupakan salah satu metode machine learning yang menggunakan perhitungan probabilitas. Untuk klasifikasi Bayes sederhana yang lebih dikenal sebagai naïve Bayesian Classifier dapat diasumsikan bahwa efek dari suatu nilai atribut sebuah kelas yang diberikan adalah bebas dari atribut-atribut

lain.[1] Asumsi ini disebut *class conditional independence* yang dibuat untuk memudahkan perhitungan-perhitungan pengertian ini dianggap “naive”, dalam bahasa lebih sederhana naïve itu mengasumsikan bahwa kemunculan suatu term kata dalam suatu kalimat tidak dipengaruhi kemungkinan kata-kata yang lain dalam kalimat padahal dalam kenyataannya bahwa kemungkinan kata dalam kalimat sangat dipengaruhi kemungkinan keberadaan kata-kata yang dalam kalimat. Naive Bayes classifier dapat dilatih untuk mengklasifikasi pola-pola yang melibatkan ribuan atribut dan diterapkan untuk ribuan pola. Akibatnya, Naive Bayes merupakan algoritma yang banyak digunakan untuk text mining dan masalah lain klasifikasi besar. Teori dari *Naive Bayes* sendiri adalah sebagai berikut:

$$P(A|B) = (P(B|A) * P(A))/P(B) \dots\dots (1)$$

Rumus di atas dapat dibaca sebagai peluang kejadian A sebagai B ditentukan dari peluang B saat A, Peluang A, dan Peluang B. Agar lebih jelas, dalam kasus klasifikasi emosi dari teks maka rumus di atas akan dirubah menjadi:

$$P(K_i|T) = (P(T|K_i)*P(K_i)) / P(T) \dots\dots (2)$$

Dimana P(K_i|D) adalah peluang dokumen teks T pada Kategori K_i

III. PEMBAHASAN

Ada 100 sampel data buku yang diambil dari perpustakaan yang akan diklasifikasi menjadi 5 klasifikasi yaitu Office for Bussines (Perkantoran), Pemrograman Visual, Pemrograman Web, Desain Grafis dan Jaringan Komputer. Yang dimana dari 100 sampel tadi telah ada klasifikasi manualnya.

3.1 Studi Kasus

Menampilkan Data Training 75 dari 100 sampel tadi. Kemudian menghitung jumlah kata yang ada dari keseluruhan data training melalui pruses *text mining*.

Tabel 1. Tabel Kata

Kata	D1	D2	D3	D4	D5	Dst
Fitur	2	0	0	0	0	
Utuh	1	0	0	0	0	..
Microsoft	3	2	1	0	0	..
Office	2	0	1	0	0	..
Uasa	3	0	0	1	0	..
Bel	1	0	0	0	0	..
Pintar	1	0	0	0	0	..
Bahas	1	1	1	0	0	..
Lengkap	1	1	0	0	0	..
Excel	1	4	1	0	6	..

powerpoint	1	0	1	0	0	..
Saji	1	1	0	0	2	..
Step	2	0	0	0	0	..
By	1	0	0	0	0	..
Rta	1	0	0	0	0	..
Cd	1	0	0	0	0	..
File	1	2	0	0	0	..
Dst

Setelah menentukan probabilitas dari data training tersebut dari 75 sampel masing memiliki 15 buku dalam satu kategori manual sehingga probabilitas yang dihasilkan yaitu 15/75 atau 0.2.

Mencari jumlah kata dari setiap dokumen buku.

Tabel 2. Tabel Jumlah Kata

Kode Buku	Jumlah Kata
D1	26
D2	71
D3	9
D4	66
D5	53
D6	65
D7	64
D8	33
D9	92
D10	47
D11	45
D12	70
D13	51
D14	37
D15	65
Dst	..

Setelah itu menghitung nilai klasifikasi naïve bayes classifier pada masing masing data tes satu persatu sebanyak 25 sampel data tes tadi. Mulai dari buku yang ke 76 sampai dengan 100. Dengan menghitung keterkaitan kata dengan data training

Tabel 3. Tabel Jumlah Kata

Kata	D1	D2	D3	D4	D5	Dst
Microsoft	4	3	2	1	1	..
bahas	2	2	2	1	1	..
konsep	1	1	1	2	1	..
Acu	1	1	1	2	1	..
standar	1	1	1	1	1	..
rogram	1	1	1	1	1	..
unix	1	1	1	1	1	..

Setelah itu menghitung jumlah kata-kata yang ada pada data tes yang diuji pada setiap klasifikasinya.

Tabel 4. Tabel Jumlah Kata Data Tes

Kata	K1	K2	K3	K4	K5
microsoft	90	34	21	21	21
Bahas	53	44	42	38	43
Konsep	25	23	26	25	26
Acu	22	20	21	22	20
Standar	22	21	22	23	24
Rogram	21	22	24	21	20
Unix	21	20	21	21	21

Melakukan perhitungan naïve bayes sesuai dengan data yang didapatkan sebelumnya. Data yang digunakan yaitu probabilitas, jumlah kata keseluruhan, jumlah kata pada data tes dan membuat semacam tabel model probabilistik untuk mempermudah perhitungan.

Dimana Probabilitas kategori dikalikan dengan probabilitas kata di data tes.

$$P_{K1} = \frac{15}{5} * \frac{91}{255} * \frac{54}{255} * \frac{25}{255} * \frac{23}{255} * \frac{23}{255} * \frac{22}{255} * \frac{22}{255} \\ = 1,34591E-06$$

$$P_{K2} = \frac{15}{5} * \frac{35}{185} * \frac{45}{185} * \frac{24}{185} * \frac{21}{185} * \frac{22}{185} * \frac{23}{185} * \frac{21}{185} \\ = 3,41193E-06$$

$$P_{K3} = \frac{15}{5} * \frac{22}{178} * \frac{43}{178} * \frac{27}{178} * \frac{22}{178} * \frac{23}{178} * \frac{25}{178} * \frac{22}{178} \\ = 2,91502E-05$$

$$P_{K4} = \frac{15}{5} * \frac{22}{174} * \frac{39}{174} * \frac{26}{174} * \frac{23}{174} * \frac{24}{174} * \frac{22}{174} * \frac{22}{174} \\ = 2,68448E-05$$

$$P_{K5} = \frac{15}{5} * \frac{22}{176} * \frac{44}{176} * \frac{27}{176} * \frac{21}{176} * \frac{25}{176} * \frac{21}{176} * \frac{22}{176} \\ = 3,63557E-06$$

Hasil klasifikasi didapatkan dari nilai maksimal untuk perhitungan setiap kategori klasifikasi. Nilai maksimal menandakan probabilitas atau kemungkinan paling tinggi ada pada kategori atau klasifikasi tersebut, jadi pada perhitungan diatas kemungkinan terbesar klasifikasi berada pada K5.

Yang terakhir menentukan akurasi perhitungan naïve bayes classifier, dengan cara mencocokkan data klasifikasi manual dengan data klasifikasi yang dilakukan dengan naïve bayes classifier. Dari data sampel diambil 25 data tes. Dari data tes tersebut hanya terjadi 4 perbedaan klasifikasi manual dan klasifikasi naïve bayes classifier sehingga

DAFTAR REFERENSI

- [1] Masayu Leylia Khodra, "Text Mining Klasifikasi NB" Bandung : ITB, 2013.
- [2] Roger, S, Pressman, Phd: Rekayasa Perangkat Lunak Pendekatan Praktisi, Yogyakarta: Andi Offset, 1997.
- [3] Gunadarma, Pengertian Tentang Implementasi, Wartawarga, 2012.
- [4] Usman, Nurdin, Implementasi Berbasis krikulum, PT Raja Grafindo Persada, 2002.

Biodata Penulis

Dikky Praseptian M, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK PPKIA Tarakanita Rahmawati Tarakan, lulus tahun 2013. Saat ini menjadi Dosen di STMIK PPKIA Tarakanita Rahmawati Tarakan.

Aida Indriani, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika STMIK PPKIA Tarakanita Rahmawati Tarakan, lulus tahun 2009. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Teknologi Informasi Sekolah Tinggi Teknik Surabaya, lulus tahun 2013. Saat ini menjadi Dosen di STMIK PPKIA Tarakanita Rahmawati Tarakan.