Jurnal ELTIKOM, Vol. 1 No.1, Juni 2017, Hal 23-30
ISSN 2598-3245 (Print), ISSN 2598-3288 (Online)
Tersedia Online di http://eltikom.poliban.ac.id

KLASIFIKASI PENYEBAB PENYALAHGUNAAN NARKOBA DARI BERITA ONLINE DENGAN MENGGUNAKAN NAIVE BAYES

Laili Wahyunita
Akademi Teknik Pembangunan Nasional
e-mail: laili.wahyunita@gmail.com

ABSTRACT

This research conducted the classification process by applying the method of classification of Naive Bayes. News article document is one form of text data that is not structured so that requires the process of cleaning data and pre-processing first. The Naive Bayes approach is an approach that refers to Bayes's Theorem, where it uses the principle of statistical opportunity to combine previous knowledge. The use of this technique is based on the need of the system to know the prob-ability value of the data to be classified. Waterfall method was used for built this classificaiton system. Accuracy rate up to 60 % with the highest precision and recall is 80% and 90%.

Keywords*: Naive Bayes, Text Document Represention, Precision, Recall.*

ABSTRAK

Penelitian ini melakukan proses klasifikasi dengan menerapkan metode klasifikasi Naive Bayes. Dokumen artikel berita merupakan salah satu bentuk data teks yang tidak terstruktur sehingga memerlukan proses pembersihan data dan pra pemrosesan terlebih dahulu. Pendekatan Naive Bayes merupakan pendekatan yang mengacu pada Teorema Bayes, dimana teorema ini menggunakan prinsip peluang statistika untuk mengkombinasikan pengetahuan sebelumnya. Penggunaan teknik ini didasari oleh keperluan dari sistem untuk mengetahui nilai probabilitas dari data yang akan diklasifikasi. Sistem klasifikasi dikembangkan dengan menggunakan metode waterfall. Akurasi yang dicapai mencapai 60 % dengan nilai precision dan recall tertinggi sebesar 80% dan 90%.

Kata Kunci*: Naive Bayes, Preprocessing Data, Waterfall, Representasi Dokumen Teks, Precision, Recall.*

I. PENDAHULUAN

Dokumen berita *online* merupakan sumber data yang banyak diteliti. Dokumen berita *online* terus bertambah jumlahnya seiring dengan perkembangan yang pesat dalam informasi digital[9]. Pengelompokan dokumen berita dibutuhkan untuk mempermudah pencarian informasi mengenai suatu event (kejadian) tertentu. Pencarian berita-berita lain yang berkaitan dengan kejadian tersebut tentu sulit dilakukan, bila hanya mengandalkan query biasa. Sebab pemilihan query yang kurang spesifik akan berakibat membanjirnya dokumen-dokumen yang tidak relevan [1].

Permasalahan yang timbul dari klasifikasi dokumen berita adalah perlunya algoritma klasifikasi dokumen. Selain itu, penentuan tingkat kemiripan (similarity) antar dokumen berdasarkan komposisi term. Ada banyak teknik yang dapat dilakukan untuk melakukan klasifikasi data, diantaranya Naïve Bayes *Classifier, Rule Based Classifier, Decision Tree* maupun *Support Vector Machine* [8]. Dalam penelitian ini digunakan algoritma Naive Bayes *Classifier* (NBC) dalam melakukan klasifikasi data. Pendekatan ini merupakan pendekatan yang mengacu pada Teorema Bayes, dimana teorema ini menggunakan prinsip peluang statistika untuk mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru. Prinsip ini kemudian digunakan untuk memecahkan masalah klasifikasi [17]. Penggunaan teknik ini didasari oleh keperluan dari sistem untuk mengetahui nilai probabilitas dari data yang akan diklasifikasi.

Di sisi lain Perkembangan kasus penyalahgunaan narkoba di Indonesia beberapa kurun waktu terakhir ini menunjukkan tingkat kenaikan yang sangat mengkhwatirkan. Jumlah pengguna narkoba di Indonesia pada tahun 2014 terhitung sebanyak 3,8 juta jiwa. Bahkan pada tahun 2015, diperkirakan jumlah pengguna narkoba di Indonesia mencapai 5,8 juta jiwa [7]. Pemerintah sudah menilai kejadian narkoba

termasuk dalam kategori darurat narkoba. Hal ini didasarkan pada banyak kasus yang terjadi, peredaran narkoba tidak hanya pada orang-orang dewasa saja, tetapi sudah di kalangan generasi muda anak-anak Indonesia.

II. KLASIFIKASI DATA

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek. Dengan kata lain, klasifikasi merupakan penempatan objek-objek ke salah satu dari beberapa kategori yang telah ditetapkan sebelumnya. Ada dua langkah dalam proses klasifikasi, yaitu [8] yaitu membangun suatu model dengan menganalisis data *training* dan menggunakan model untuk melakukan klasifikasi terhadap data yang belum diketahui label kelasnya.

Ada banyak teknik yang dapat dilakukan untuk mengklasifikasikan data, diantaranya adalah *decision tree*, *naive bayesian classifier*, *bayesian belief network* dan *rule based classifier* [8]. Setiap algoritma klasifikasi tersebut memiliki kelebihan dan kekurangan. Tetapi prinsip dari masing-masing algoritma tersebut sama, yaitu melakukan suatu pelatihan sehingga di akhir pelatihan, model dapat memprediksi setiap vektor masukan ke label kelas *output* dengan tepat [11]. Metode klasifikasi yang digunakan pada penelitian ini adalah Naive Bayes *Classifier*. Salah satu kelebihan Naive Bayes *Classifier* adalah sederhana tetapi memiliki akurasi yang tinggi [13].

III. NAIVE BAYES

Pendekatan naive bayes merupakan pendekatan yang mengacu pada teorema bayes yang merupakan prinsip peluang statistika untuk mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru. Prinsip ini kemudian digunakan untuk memecahkan masalah klasifikasi [17].

Hubungan hipotesis dan bukti dengan klasifikasi adalah, hipotesis dalam teorema bayes merupakan label kelas yang menjadi atribut target, sedangkan bukti merupakan himpunan atribut yang menjadi masukan dalam model klasifikasi. Klasifikasi naïve bayes berasumsi bahwa efek dari nilai atribut pada kelas tertentu bersifat independen dari nilai-nilai atribut lainnya. Asumsi ini disebut class *conditional probability*. Hal ini dilakukan untuk menyederhanakan perhitungan yang terlibat, dan dalam pengertian ini dianggap "naive" [8]. Maksud independensi yang kuat (naïve) artinya bahwa setiap atribut pada suatu data tidak memiliki hubungan atau ketergantungan antara satu atribut dengan atribut lainnya.

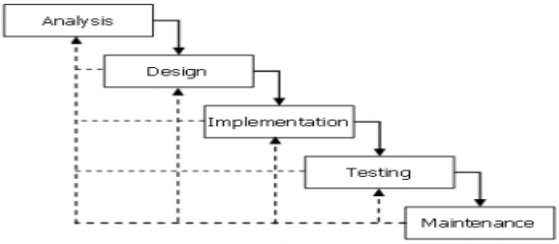
$$P\ (H|E) = \frac{P(E|H)x\ P(H)}{P\ (E)}$$

P(H|E) : Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis H terjadi jika diberikan bukti (*evidence*) E terjadi.

P(E|H) : Probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis H P(H) : Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun P(E) : Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis/ bukti yang lain.

IV. RESEARCH METHOD

Pengembangan sistem dalam penelitian ini menggunakan metode *waterfall*, pertama kali diperkenalkan oleh Windows W. Royce pada tahun 1970. Metode ini merupakan model klasik yang sederhana dengan aliran sistem yang linier, dan masih popular digunakan hingga saat ini. Output dari setiap tahap merupakan input bagi tahap berikutnya. Tahapan proses pada metode *waterfall* diperlihatkan seperti pada Gambar 1.



Gambar 1. Model waterfall

a. Analisis
Analisis dalam pengembangan sistem sering dikenal pula sebagai *Software Requirement Spesification* (SRS), merupakan deskripsi lengkap dan komprehensif perangkat lunak yang akan dikembangkan. Analisis ini digunakan untuk menentukan kebutuhan fungsional maupun non-fungsional. Biasanya, persyaratan fungsional didefinisikan dengan cara menggambarkan interaksi pengguna dengan perangkat lunak. Di dalamnya juga termasuk persyaratan seperti tujuan, ruang lingkup, perspektif, fungsi, atribut perangkat lunak, pengguna karakteristik, fungsi spesifikasi, antarmuka persyaratan, dan persyaratan database. Sebaliknya, persyaratan non-fungsional mengacu pada berbagai kriteria, kendala, keterbatasan, dan persyaratan yang diberlakukan pada desain dan pengoperasian perangkat lunak.

b. Desain
Tahap ini adalah proses perencanaan dan solusi pemecahan masalah perangkat lunak. Pengembang software dan desainer menentukan rencana yang meliputi desain algoritma, arsitektur perangkat lunak, skema konseptual *database*, diagram logis desain, desain konsep, desaingrafis antarmuka pengguna, dan definisi struktur data.

c. Implementasi
Tahap ini mengacu pada realisasi kebutuhan bisnis dan spesifikasi desain ke dalam eksekusi struktur program, *database*, *website*, atau komponen perangkat lunak. Pada fase ini kode ditulis dan dikompilasi menjadi sebuah aplikasi operasional, dengan kata lain, tahap ini adalah proses konversi seluruh persyaratan dan cetak biru ke dalam lingkungan produksi.

d. Pengujian
Tahap ini juga dikenal sebagai verifikasi dan validasi, berupa proses untuk memeriksa apakah perangkat lunak yang dibuat memenuhi persyaratan dan spesifikasi yang telah direncanakan. (IEEE-STD-610, 1991). Selain itu, tahap pengujian dilakukan untuk menemukan *bug* dan gangguan sistem, sehingga dapat dilakukan perbaikan dan penyempurnaan kembali.

e. Pemeliharaan
Tahap ini merupakan proses memodifikasi solusi perangkat lunak setelah selesai dan didistribusikan. Perbaikan output, kesalahan, serta meningkatkan kinerja dan kualitas. Kegiatan pemeliharaan tambahan dapat dilakukan dalam fase ini, termasuk menyesuaikan perangkat lunak untuk kebutuhan pengguna baru, maupun untuk meningkatkan keandalan perangkat lunak [12].

V. PROSES PEMBERSIHAN DATA (DATA CLEANING)

Data artikel berita online merupakan jenis data yang tidak terstruktur. Data yang tidak terstruktur memerlukan pembersihan terlebih dahulu sebelum diolah. Pembersihan data mencakup penghapusan data yang tidak diperlukan, data yang tidak konsisten, dan noise data. Di dalam sebuah sistem memungkinkan disediakan dan ditambahkan aturan-aturan yang bertujuan untuk membersihkan data.

VI. REPRESENTASI DOKUMEN TEKS

Dokumen teks direpresentasikan dengan mengambil kata-kata yang muncul dalam dokumen tersebut. Ada beberapa cara untuk melakukan konversi dokumen menjadi kata-kata yang akan digunakan dalam pelatihan data. Bigrams dan trigrams adalah contoh metode ekstraksi fitur kata yang digunakan dalam mereprentasikan dokumen teks. Kedua metode ini mengkombinasikan fitur menjadi dua atau tiga kata menjadi satu frase. Metode yang lain adalah bag-of-words yaitu representasi dari dokumen berdasarkan kata-kata yang muncul paling tidak sekali. Kata-kata yang dihasilkan akan diolah menjadi word dictionary untuk pelatihan data dalam proses klasifikasi [3].

VII. PRECISION & RECALL

Pengujian akurasi hasil klasifikasi dilakukan dengan membandingkan antara hasil klasifikasi dengan aktual data yang pada model klasifikasi.

ISSN 2598-3245 (Print)
ISSN 2598-3288 (Online)

TABEL 1
EVALUASI HASIL KLASIFIKASI

Nilai prediksi klasifikasi	Nilai sebenarnya	
	True	False
TRUE	TP (*True Positive*) Correct Result	FP (*False Positive*) *Unexpected result*
FALSE	FN (*False Negative*) *Missing result*	TN (*True Negative*) *Correct absence of result*

Pada Tabel 1 menyajikan evaluasi hasil klasifikasi dengan menggunakan dua kelas yang kemudian dapat dicari nilai precision, recall, dan accuracy. Precision menunjukkan keberhasilan klasifikasi yang dihasilkan untuk mengambil bagian dari dokumen yang relevan. Precision adalah persentase hubungan hasil ekstraksi yang benar dari total keseluruhan data yang diolah. Recall menunjukkan derajat kelengkapan dari nilai precision. Recall adalah sebagian kecil dari dokumen yang relevan yang akan diambil [10].

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (2)$$

Akurasi [10] adalah perhitungan alternatif untuk menilai tingkat kebenaran dalam klasifikasi. Akurasi dapat dihitung dengan menggunakan persamaan yang ditunjukkan pada Persamaan (3):

$$\text{Accuracy} = \frac{\text{TP+TN}}{(\text{TP+FN})+(\text{FP+TN}))} \quad (3)$$

Komponen evaluasi dari klasifikasi didefinisikan sebagai berikut [6].

a. *True Positive* (TP) adalah jumlah klasifikasi kelas yang benar yang berhasil sistem hasilkan. Artinya kelas yang ditentukan oleh sistem sesuai dengan kelas realnya.
b. *False Positive* (FP) adalah jumlah hasil klasifikasi yang pada aktualnya tidak masuk dalam suatu kelas, di dalam sistem dimasukkan dalam suatu kelas.
c. *False Negative* (FN) adalah kesalahan klasifikasi, dimana sistem tidak berhasil memasukkan objek ke suatu kelas, padahal aktualnya objek tersebut mempunyai kelas.
d. *True Negeative* (TN) adalah jumlah klasifikasi kelas yang salah yang berhasil sistem hasilkan. Artinya kelas yang ditentukan oleh sistem tidak sesuai dengan kelas realnya.

VIII. PENYALAHGUNAAN NARKOBA

Penyalahgunaan narkoba atau narkotika dan psikotrapika, merupakan kejahatan kemanusiaan yang berat, yang mempunyai dampak luar biasa, terutama pada generasi muda suatu bangsa yang beradab [15]. Penyalahgunaan narkoba merupakan penggunaan atau pemanfaatkan narkotika secara ilegal di luar untuk kepentingan pengobatan atau pelayanan kesehatan dan ilmu pengetahuan. Tindakan penyalahgunaan narkotika secara berlebihan akan membahayakan diri sendiri, baik secra fisik maupun psikis [14]. Kegiatan menjual belikan atau perdagangan narkoba untuk kepentingan bisnis juga termasuk dalam tindak penyalahgunaan narkoba [5].

Terdapat beberapa jenis narkotika yang cukup populer dan sering disalahgunakan yang dapat dilihat pada Tabel 2.

TABEL 2
DAFTAR JENIS NARKOBA

No	Nama Narkotika
1	Opium
2.	Morphin
3.	Ganja
4.	Cocaine
5.	Heroin
6.	Shabu-shabu
7.	Putaw
8.	Ectasy

Berbagai penelitian [14] mengemukakan bahwa faktor penyebab timbulnya penyalahgunaan narkoba yakni sebagai berikut:

1. Faktor individu.
   Terdiri dari aspek kepribadian, dan kecemasan atau depresi. Aspek kepribadian meliputi pribadi yang ingin tahu, mudah kecewa, sifat tidak sabar dan rendah diri. Sedangkan yang termasuk kecemasan atau depresi, karena tidak mampu menyelesaikan kesulitan hidup, sehingga melarikan diri dalam penggunaan narkoba.
2. Faktor sosial budaya
   Faktor ini terdiri dari kondisi keluarga dan pengaruh pergaulan. Keluarga dimaksudkan sebagai faktor disharmonis seperti orang tua yang bercerai, orang tua yang sibuk dan jarang dirumah, serta perekonomian keluarga yang serba kekurangan. Pengaruh pergaulan, dimaksudkan karena ingin diterima dalam pergaulan kelompok narkoba.
3. Faktor lingkungan
   Lingkungan yang tidak baik maupun tidak mendukung dalam menampung segala sesuatu yang menyangkut perkembangan psikologis anak dan kurangnya perhatian terhadap anak untuk menjadi pemakai narkotika.
4. Faktor narkoba
   Faktor ini ditandai dengan kemudahan dalam mendapatkan narkoba sehingga semakin mudah timbulnya penyalahgunaan narkoba. Meskipun peredaran narkoba dilarang, tetapi narkoba masih mudah diperoleh di kalangan masyarakat.

IX. HASIL DAN PEMBAHASAN

1. Spesifikasi Penggunaan *Hardware* dan *Software*

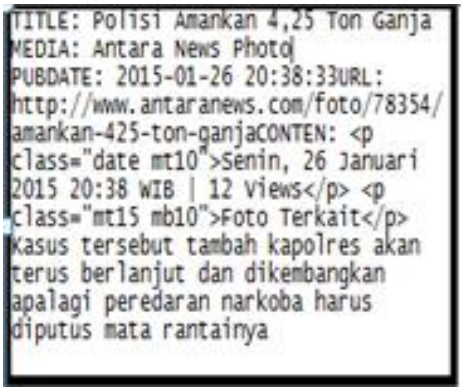
   Dalam pengembangan sistem ini menggunakan perangkat keras dan perangkat lunak dengan spesifikasi sebagai berikut:

   a. Perangkat keras:
      - *Processor* AMD A4-3330MX HDGraphic 2.30 GHz
      - RAM 4 GB
      - Tipe sistem 32-bit

   b. Perangkat Lunak:
      - XAMPP Control Panel
      - Sublime Text 2
      - Microsoft Windows 7

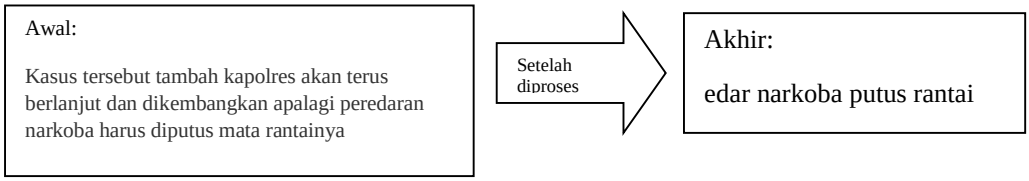
2. Pengembangan Sistem

   Analisa yang dilakukan dalam mengidentifikasi kebutuhan sistem menghasilkan diperlukan kamus data untuk proses pembersihan data. Pembersihan data bertujuan untuk menghilangkan bagian-bagian data yang tidak dibutuhkan. Sebagian contoh dari data-data yang dibutuhkan adalah kata “yang”, “untuk”, dan kata penghubung lainnya. Gambar 2 menunjukkan data mentah artikel yang akan diolah. Pada proses pembersihan data juga dilakukan penghapusan elemen data selain isi berita.



Gambar 2. Sampling data artikel yang diambil

Bagian isi berita akan dilakukan pembersihan data lagi sehingga dihasilkan data yang akan dimasukkan ke dalam data latih.



Gambar 3. Visualisasi proses pembersihan data

Kumpulan dokumen artikel berita yang digunakan sebagai data latih disimpan di dalam file data latih. Dokumen tersebut telah diberikan label kelas sesuai pembelajaran pemahaman dari manusia terhadap faktor penyebab kasus narkoba. Pemberian label dilakukan secara manual. Tabel 3 memperlihatkan sebaran penggunaan data latih untuk pembangunan model sistem klasifikasi.

TABEL 3
SEBARAN DATA LATIH

Kelas	Jumlah dokumen
Narkoba	119
Individu	80
Sosial budaya	60
Lingkungan	47

Dalam proses klasifikasi masing-masing kelas diberikan kata-kata kunci berdasarkan penelitian yang telah dilakukan terhadap penyebab penyalahgunaan narkoba. Adapun sebagian dari kata-kata kunci untuk masing-masing kelas faktor penyebab penyalahgunaan narkoba ditunjukkan pada tabel 4.

TABEL 4
KATA KUNCI KELAS PENYEBAB NARKOBA

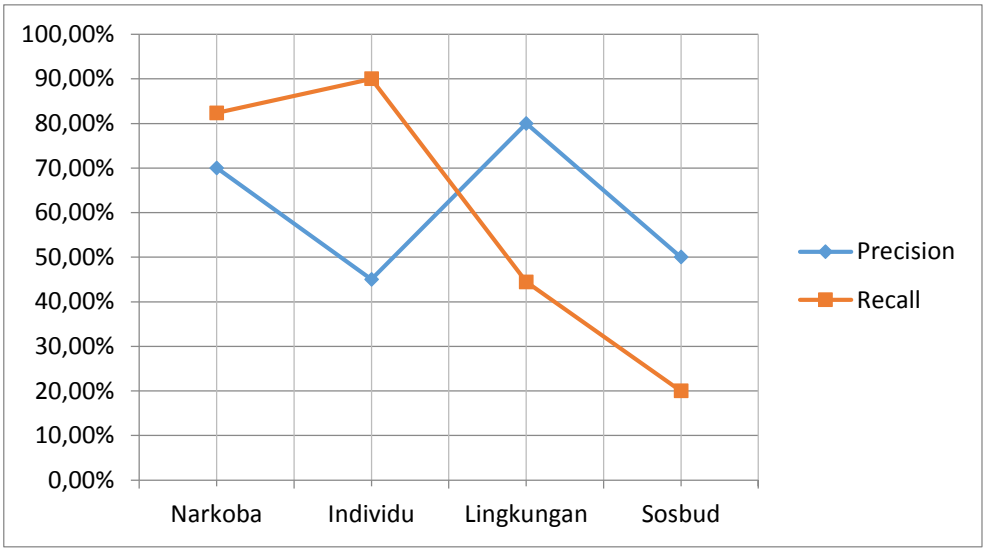
Kelas	Kata kunci
Narkoba	Murah mudah untung sindikat jangkau banyak
Individu	Coba penasaran senang bosan hibur diri gemuk badan ingin tahu tekanan
Sosial budaya	lembaga pemasyarakatan peredaran mengedarkan kuli bangunan masyarakat berbisnis sabu karena kebutuhan ekonomi hotel losmen diskotik tempat hiburan malam universitas mahasiswa pelajar kampus
Lingkungan	keluarga broken home ayah ibu orang tua kakak pencandu sekolah cafe tempat hiburan sekolah kurang pengawasan

Pengujian dilakukan dengan menggunakan 50 data uji. Nilai precision dan recall yang dicapat dapat dilihat pada tabel 5. Nilai precision tertinggi ada pada kelas Lingkungan dan terendah ada pada kelas Individu. Sedangkan untuk nilai recall yang tertinggi terdapat pada kelas Individu dan terendah pada kelas Sosbud.

TABEL 5
NILAI PRECISION DAN RECALL

Kelas	Precision	Recall
Narkoba	70,00%	82,35%
Individu	45,00%	90,00%
Lingkungan	80,00%	44,44%
Sosbud	50,00%	20,00%

ISSN 2598-3245 (Print)
ISSN 2598-3288 (Online)

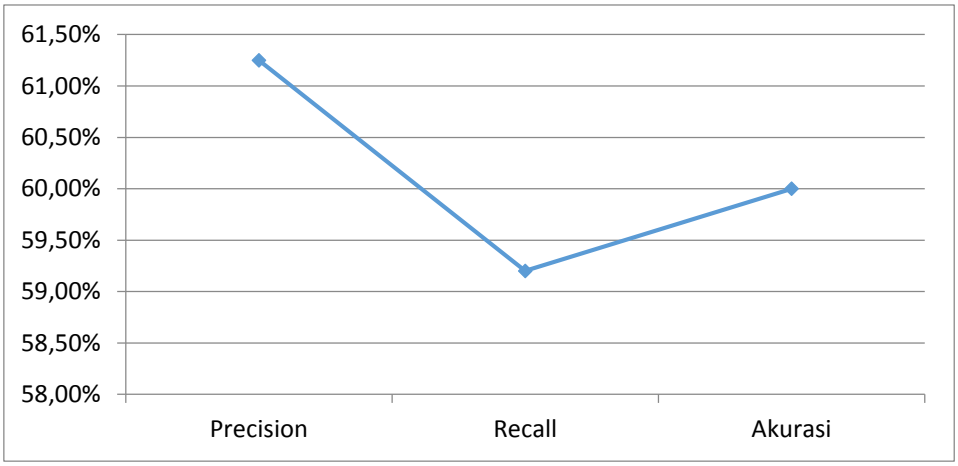


Gambar 4. Precesion dan Recall

Gambar 4 menunjukkan grafik perolehan precision dan recall untuk masing-masing kelas hasil klasifikasi. Sedangkan visualisasi dari Tabel 6 ditunjukkan melalui Gambar 5 yang menggambarkan nilai akurasi, presicion, dan recall untuk klasifikasi dengan menggunakan metode naive bayes.

TABEL 6
HASIL EVALUASI KLASIFIKASI

Item evaluasi	Perolehan
Precision	61,25%
Recall	59,20%
Akurasi	60%



Gambar 5. Hasil Evaluasi

DAFTAR PUSTAKA

[1] Arifin, A. Z., & Setiono, A. N. (2002). Klasifikasi dokumen berita kejadian berbahasa indonesia dengan algoritma single pass clustering. In *Prosiding Seminar on Intelligent Technology and its Applications (SITIA), Teknik Elektro, Institut Teknologi Sepuluh Nopember Surabaya*.
[2] Singhal, Anoop., "An Overview of Data Warehouse, OLAP and Data Mining"," Data Warehousing and Data Mining Techniques for Cyber Security", New York,2007
[3] Bramer, Max., "Principles of Data Mining",Springer, London, 2007.
[4] Afiatin, T., 2008, *Pencegahan penyalahgunaan narkoba,* Gadjah Mada University Press, Yogyakarta.
[5] Amir, M.P., dan Syahrul, B., 2007, *Narkoba ancaman generasi muda*, Gerpana, Kalimantan Timur.
[6] Candradewi, I. and Harjoko, A., 2015. *Pemrosesan Video Untuk* Klasifikasi *Jenis Kendaraan Menggunakan Algoritma Support Vector Machine* (Doctoral dissertation, Universitas Gadjah Mada).
[7] Damayanti, R., 2015, Laporan akhir, Survey Nasional Perkembangan Penyalahguna Narkoba Tahun Anggaran 2014, BNN,Jakarta.

[8] Han, J. dan Kamber, M., 2006, *Data Mining:Concepts and Technique 2nd Edition,* Morgon Kauffman Publisher, San Fransisco.
[9] Hamzah, A. (2012). KLASIFIKASI TEKS DENGAN NAÏVE BAYES CLASSIFIER (NBC). *Seminar Nasional Aplikasi* Sains & *Teknologi (SNAST) Periode III,* (p. 9). Yogyakarta.
[10] Junianto, E., 2014, Penerapan PSO untuk Seleksi Fitur pada Klasifikasi Dokumen Berita Menggunakan Naive Bayes Classifier, *Thesis*, Pascasarjana Magister Ilmu Komputer STIMIK Nusa Mandiri, Jakarta
[11] Prasetyo, E, 2012, *Data Mining Konsep dan Aplikasi Menggunakan MATLAB, 1st ed*., ANDI OFFSET, Yogyakarta.
[12] Pressman, R.S., 2002, *Rekayasa Perangkat Lunak*, (diterjemahkan oleh: LN *Harnaningrum*), Penerbit ANDI, Yogyakarta.
[13] Rish, I., 2001, *An Empirical Study Of The Naive Bayes Classifier*, *IBM Research Report*, Thomas J. Watson Research Center, Yorktown Heights, New York.
[14] Situmorang, Y. B., 2012, Perlindungan Hukum Bagi Korban Penyalahgunaan Narkotika (Studi Kasus Di Polresta Yogyakarta) (Doctoral dissertation, UAJY).
[15] Siswantoro, S., 2004, *Penegakan Hukum Psikotropika Dalam Kajian Sosiologis Hukum*, PT RajaGrafindo Persada, Jakarta. hlm.2.
[16] Sumanthi, S. dan Esakkirajan, S.PY., 2007, *Fundamentals of Relational Database Management Systems*, Springer Belin Heidelberg.
[17] Tan, P.N., Stelnbach, M., and Kumar,V., 2006, Introduction to Data Mining, Pearson Education, Boston