

Analisis Kinerja Smoothing pada Naive Bayes untuk Pengkategorian Soal Ujian

Indah Listiowarni^[1], Eka Rahayu Setyaningsih^[2]

Teknologi Informasi, Sekolah Tinggi Teknik Surabaya
listioindah@gmail.com
Teknologi Informasi, Sekolah Tinggi Teknik Surabaya
eka@stts.edu

Abstract

Cognitive domain taxonomy bloom is a reference to classify examination based on difficulty levels on education world, the result of those categories will be used to compile a variety of exam questions. In this research, Naive bayes classifier will be used to categorize the text about biology exam for high school. Chi-square is feature selection method that will be used to remove an unuse features on examination, and increase speed of process text categorization. In addition, naive bayes classifier is a method that causing missclassification result of categorization text, this case will be happen if testing data is not found in training data, so we need another method to minimize it, that method called by smoothing method. In this research we will test performance and impact of smothing method for naive bayes classifier and chi-square as feature selection method. The smothing methods to be compared on this research are : Laplace, Dirichlet and Two Stage smothing.

Keywords: bloom's taxonomy, smoothing, naive bayes, chi-square, feature selection, examination

1. PENDAHULUAN

Taksonomi berasal dari dua kata dalam bahasa Yunani yaitu tassein yang berarti mengklasifikasi dan nomos yang berarti aturan. Jadi Taksonomi berarti hierarki klasifikasi atas prinsip dasar atau aturan. Taksonomi bloom pertama kali digagas oleh ilmuwan bernama Benjamin Samuel Bloom. Jadi, taksonomi bloom merupakan sebuah tingkatan skills yang digunakan dan dikenal dalam dunia pendidikan, pada taksonomi bloom terdapat 3 domain utama, yaitu : kognitive, afektif, dan psikomotorik. Setiap domain pada taksonomi bloom memiliki penilaian skill yang berbeda, afektif domain mencakup penilaian sikap, psikomotorik domain mencakup tentang ketrampilan, dan kognitif domain yang merupakan domain yang akan dipakai dalam penelitian ini merupakan sebuah domain yang penilaiannya mencakup intelektual, pengetahuan, dan ketrampilan berpikir. Cognitive domain taksonomi bloom memiliki 6 level, yaitu : Mengingat (remembering), Memahami (understanding), Menerapkan (applying), Menganalisis (analysing), Mengevaluasi (evaluating), Membuat (creating) dimana level terendahnya adalah mengingat (C1), kemudian diikuti level berikutnya, hingga level tertinggi atau tersulit adalah kategori membuat (C6).

Proses kategorisasi soal ujian sesuai dengan level cognitive domain pada taksonomi bloom, biasanya dilakukan secara manual oleh tenaga pendidik, dengan memperhatikan setiap term yang ada pada soal ujian, kemudian mencocokkannya dengan kata kerja operasional pada setiap level cognitive domain.

Dalam proses penyusunan soal ujian, biasanya digunakanlah pembagian level sesuai dengan 6 level yang ada pada taksonomi bloom. Hal ini merupakan salah satu uapaya untuk pemerataan jenis soal pada sebuah soal ujian. Hal yang biasa dilakukan untuk

mengetahui kategori dari sebuah soal dalam level taksonomi bloom adalah dengan memperhatikan secara manual setiap term yang ada, berusaha memahami maksud dari soal ujian dan mencocokkannya dengan setiap level cognitive domain, hal itu tentu saja menyita waktu dan kurang efektif. Untuk itulah diperlukan sebuah sistem yang dapat membantu mengkategorikan soal secara otomatis ke dalam salah satu level taksonomi bloom.

Setiap proses pada sebuah sistem tidak luput dengan penggunaan sebuah metode. Naive Bayes classifier merupakan metode yang paling banyak digunakan untuk proses klasifikasi. Naive Bayes classifier menggunakan teorema Bayes yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Ciri utama dari naive bayes classifier adalah asumsi yang kuat akan independensi dari masing-masing kondisi, artinya naive bayes classifier memiliki variabel yang independen. Dengan menerapkan naive bayes pada sebuah data terkadang juga menyebabkan missclassification jika data training hanya sedikit sehingga data testing tidak ditemukan pada data training, dan menyebabkan hasil probabilitas bernilai 0 [1] dan menyebabkan error pada proses klasifikasi.

Pada penelitian sebelumnya [2][3], kekurangan dari naive bayes dapat diminimalisir menggunakan metode smoothing. Sehingga, pada penelitian ini akan diuji dan dianalisis kinerja dan pengaruh metode smoothing terhadap metode naive bayesian dan feature selection Chi square yang juga akan digunakann pada penelitian ini untuk mengurangi banyaknya atribut yang digunakan pada data soal, dan membuang fitur-fitur yang tidak diperlukan, sehingga dapat mempercepat proses kategori teks. Metode smoothing yang akan diuji-cobakan adalah Laplace, Dirichlet dan Two-Stage

2. METODE PENELITIAN

Dataset

Data yang digunakan merupakan data soal mata pelajaran biologi berbahasa Indonesia tingkat SMA dari kelas X hingga kelas XII yang diambil dari berbagai sumber, diantaranya buku trainingan LKS, Ulangan Harian, dan modul biologi. Pada penelitian ini akan digunakan 2 macam dataset, yaitu dataset pertama dengan jumlah 600 data soal ujian dengan masing-masing jumlah soal pada class nya adalah 100 soal, yang akan diujikan pada 100 data soal ujian baru yang belum dikenali sebelumnya. Sedangkan pada dataset kedua data yang akan dijadikan data testing merupakan bagian dari data training. Kombinasi data testing dan data training dipilih secara acak oleh sistem.

Cognitive domain taksonomi bloom

Taksonomi bloom merupakan sebuah aturan klasifikasi/kategori yang digagas oleh ilmuwan yang bernama Benjamin Samuel Bloom yang terdiri level-level dari tingkat yang rendah hingga ke level yang tertinggi (hierarki). Bloom membagi kerangka konsep ini dalam tiga garis besar yaitu : kognitif, afektif dan psikomotorik.

Ranah kognitif (cognitive domain) mengurutkan keahlian berpikir sesuai dengan tujuan yang diharapkan. Proses berpikir menggambarkan tahap berpikir yang harus dikuasai oleh siswa agar mampu mengaplikasikan teori kedalam perbuatan. Ranah kognitif ini terdiri atas enam level, yaitu: Mengingat (remembering), Memahami (understanding), Menerapkan (applying), Menganalisis (analysing), Mengevaluasi (evaluating), Membuat (creating). Ke-enam level pada ranah kognitif dapat digambarkan dengan menggunakan piramida, pada gambar 1



Gambar 1 .piramida taksonomi bloom ranah kognitif

Piramida pada gambar 1 menjelaskan tentang level atau class pada ranah kognitif taksonomi bloom. Tiga level pertama (terbawah) merupakan Lower Order Thinking Skills, sedangkan tiga level berikutnya Higher Order Thinking Skill. Level tersebut yang akan digunakan sebagai class untuk mengelompokkan data soal yang ada.

Class-class untuk pengelompokan (pengkategorian) soal sesuai dengan taksonomi bloom ranah kognitif memiliki ciri khusus pada setiap levelnya. Pada tabel 1 akan dipaparkan tentang ke-6 class yang digunakan, dan kata kerja operasional yang

mencerminkan setiap class-nya, sekaligus memudahkan penulis mengetahui maksud dari setiap kategori dan menerapkannya pada pengkategorian soal ujian. Untuk memudahkan penyebutan class, penulis membuat inisial untuk setiap class sesuai dengan level. Level terendah, menjadi C1 dan level tertinggi menjadi C6 [4].

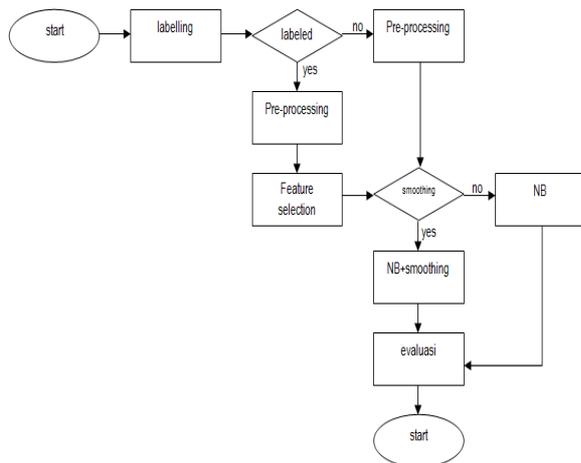
Tabel 1 : daftar class dan contoh kata kerja kunci

Kategori	Contoh Kata Kerja Kunci
C1	Mengutip, menyebutkan, mendaftar, menunjukkan, mengidentifikasi, melabeli, memasangkan, menamai, menandai, membaca, menyadari, menghafal, mencatat, mengulang memilih, menulis.
C2	Menerangkan, menjelaskan, menterjemahkan, menguraikan, mengartikan, menyatakan kembali, menafsirkan, menginterpretasikan, mendiskusikan, menyeleksi, mendeteksi, melaporkan, menduga, mengelompokkan, memberi
C3	Memilih, menerapkan, melaksanakan, mengubah, menggunakan, mendemonstrasikan, memodifikasi, menginterpretasikan, menunjukkan, membuktikan, menggambarkan, mengoperasikan, menjalankan, memprogramkan, mempraktekkan, memulai
C4	Mengkaji ulang, membedakan, membandingkan, mengkontraskan, memisahkan, menghubungkan, menunjukan hubungan antara variabel, memecah menjadi beberapa bagian, menyisahkan, menduga, mempertimbangkan
C5	Mengkaji ulang, mempertahankan, menyeleksi, mempertahankan, mengevaluasi, mendukung, menilai, menjustifikasi, mengecek, mengkritik, memprediksi, membenarkan, menyalahkan.
C6	Merakit, merancang, menemukan, menciptakan, memperoleh, mengembangkan, memformulasikan, membangun, membentuk, melengkapi, membuat, menyempurnakan, melakukan, inovasi, mendisain, menghasilkan karya.

Alur sistem

Pada penelitian ini, terdapat dua macam data yang akan digunakan, yaitu data soal yang diberi label (labeled) dan tidak terlabeli (unlabeled). Data yang terlabeli merupakan data yang akan digunakan sebagai data training. Seperti yang digambarkan pada gambar 2 mengenai alur sistem dari penelitian ini, kami akan menganalisis kinerja smoothing dan menguji pengaruh metode smoothing tersebut terhadap metode naive bayessian dan metode feature selection, yaitu chi-square (X^2). Metode feature selection akan diterapkan pada data

soal yang terlabeli (labeled) dan berfungsi untuk menghilangkan fitur-fitur yang tidak relevan.



Gambar 2 .Alur sistem

Naive Bayes

Naive bayes classifier, merupakan algoritma klasifikasi berdasarkan penerapan teorema bayes dengan asumsi independen yang kuat, output dari classifier berupa nilai probabilitas yang menunjukkan sebuah dokumen atau data harus diklasifikasikan atau dikelompokkan, sesuai nilai probabilitas tertinggi yang dihasilkan Naive Bayes Classifier. Jika terdapat variabel *d* yang harus diklasifikasikan, NBC (Naive Bayes Classifier) mengasumsikan bahwa fitur-fitur nya independen, dan variabel *c* merupakan variabel class, sehingga *p(c_i)* dan *p(d|c_i)* dapat dirumuskan dengan formula sebagai berikut :

$$p(c_i) = \frac{|c_i|}{|C|} \tag{1}$$

$$p(d|c_i) = \prod_{k=1}^{|d|} p(w_k|c_i) \tag{2}$$

Untuk meningkatkan performace naive bayes classifier, dan meminimalisir adanya missclassification pada hasil klasifikasi, dibutuhkan adanya metode smoothing. Smoothing yang akan dibandingkan pada penelitian ini adalah :

Laplace Smoothing

Laplace smoothing merupakan metode yang banyak digunakan, sekaligus smoothing yang disebut sebagai default smoothing dan smoothing tertua yang pernah di implementasikan pada Naive Bayes Classifier[5]. Laplace smoothing juga disebut dengan Add-one smoothing karena metode smoothing ini menambahkan angka 1 pada setiap frekuensi token yang didapat. Rumus naive bayessian dengan laplace smoothing, dapat dihitung menggunakan formula sebagai berikut :

$$P(t|c) = \frac{T_{ct}+1}{\sum_{t \in V} T_{ct'}+|V|} \tag{3}$$

Dimana *P(t|c)* adalah peluang munculnya sebuah kata dalam sebuah kelas, dan *|V|* merupakan jumlah kata unik pada semua kelas.

Dirichlet Smoothing

Dirichlet smoothing merupakan smoothing yang

menggunakan distribusi multinomial, berikut rumus naive bayessian dengan dirichlet smoothing [2]:

$$P_{\mu}(t|c) = \frac{T_{ct}+\mu P(t|C)}{\sum_{t \in V} T_{ct'}+\mu} \tag{4}$$

dimana μ merupakan koefisien control, dengan nilai di antara 0-1.

Two-Stage Smoothing

Two stage smoothing merupakan smoothing yang juga menggunakan koefisien kontrol seperti halnya dirichlet smoothing, yang dapat dipilih sesuai dengan nilai akurasi yang dihasilkan. Berikut merupakan rumus naive bayessian dengan two-stage smoothing [6]

$$P_{\lambda, \mu}(t|c) = (1 - \lambda) \frac{T_{ct} + \mu P(t|C)}{\sum_{t \in V} T_{ct'}} + \lambda P(t|C) \tag{5}$$

Dimana *P(t|c)* merupakan peluang munculnya sebuah kata dalam sebuah kelas, μ dan λ merupakan koefisien control, dan *T_{ct}* merupakan munculnya kata/term tertentu dalam sebuah kelas (kata *t* dalam kelas *c*). Two stage smoothing merupakan hasil kombinasi dari dua metode smoothing yang berbeda, yaitu dirichlet smoothing dan jelineck mercer smoothing [6], terlihat dari penggunaan koefisien control, yaitu μ merupakan koefisien control dari dirichlet smoothing dan λ merupakan koefisien control dari jelineck mercer smoothing

Chi square

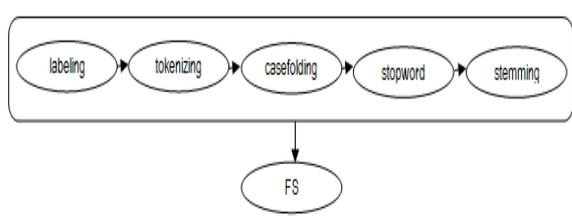
Chi-Square atau yang biasa disimbolkan sebagai χ^2 merupakan metode feature selection atau pemilihan fitur yang termasuk dalam pendekatan metode filter [7]. Pada umumnya metode feature selection ini, digunakan pada tahap pre-processing.

$$X^2(t, c) = \frac{N(A \times D - B \times C)^2}{(A+B) \times (C+D) \times (A+C) \times (B+D)} \tag{2.1}$$

variabel *t* merupakan kata yang sedang diujikan terhadap suatu kelas *c*, *N* merupakan jumlah dokumen training, *A* merupakan banyaknya dokumen pada kelas *c* yang memuat kata *t*, *B* merupakan banyaknya dokumen yang tidak berada di *c* namun memuat kata *t*, *C* merupakan banyaknya dokumen yang berada di kelas *c* namun tidak memiliki kata *t* di dalamnya, serta *D* merupakan banyaknya dokumen yang bukan merupakan dokumen kelas *c* dan tidak memuat kata *t*[8]

Preprocessing

Secara garis besar, tahapan pre-processing akan dilakukan berurutan seperti yang dijelaskan pada gambar 3. Tahap Tokenizing dilakukan dilakukan pertama kali setelah proses labeling pada data soal selesai dilakukan, dilanjutkan dengan tahapan casefolding, stopword removal dan berakhir dengan tahapan stemming sebagai hasil akhir dari rangkaian alur preprocessing.



Gambar 3 .Alur pre processing

Setelah data selesai melewati tahap persiapan data yang diakhiri dengan proses stemming, tahap selanjutnya yaitu memilih feature yang akan digunakan ke tahap selanjutnya, penghapusan dan pemilihan fitur pada data tergantung pada penentuan nilai ambang batas (threshold) chi-square. Fitur yang tidak memenuhi nilai threshold akan dihapus.

Evaluasi

Tahapan evaluasi merupakan salah satu tahapan penting dalam sebuah penelitian untuk mengetahui kinerja dan performance dari serangkaian metode yang dilakukan, yaitu dengan menggunakan F measure, untuk mendapatkan F measure, diperlukan perhitungan precision dan recall untuk setiap pengujian. Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Sedangkan recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{8}$$

$$F - \text{measure} = \frac{2PR}{P+R} \tag{9}$$

parameter TN,TP ,FP ,FN didapatkan dari tabel confusion matrix, seperti pada tabel 2. Menurut Han dan Kamber (2015) Confusion matrix merupakan alat yang digunakan untuk menganalisa seberapa baik klasifikasi mengenali tuple kelas yang berbeda.

Tabel 2. Tabel confusion matrix

		Prediksi	
		Ya	Tidak
Aktual	Ya	TP	TN
	Tidak	FP	FN

3. HASIL DAN PEMBAHASAN

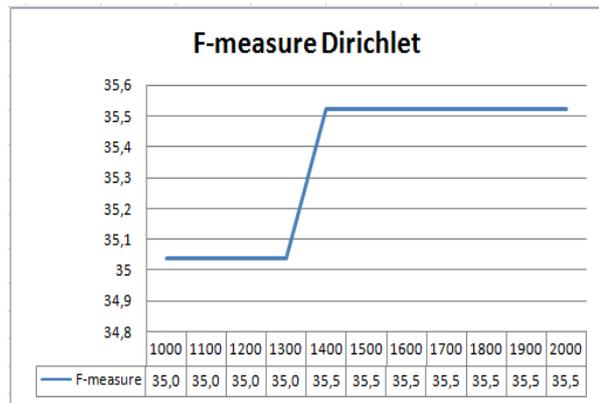
Penelitian ini bertujuan untuk mengetahui kinerja smoothing terhadap metode classifier Naive Bayessian dengan menggunakan metode feature selection Chi-Square untuk mengkategorikan data soal ujian ke dalam 6 class berdasarkan level cognitive domain taksonomi bloom. Metode smoothing yang akan diujikan terdiri dari metode Laplace, Dirichlet dan Two-Stage Smoothing. Sebelum dilakukan pengujian terhadap dataset yang sudah disediakan, perlu dilakukan penentuan parameter yang tepat untuk Dirichlet dan Two-Stage smoothing. Tabel 3

merupakan rentan parameter μ dan λ yang akan di uji cobakan, dan memilih parameter dengan hasil F-measure tertinggi untuk digunakan pada ujicoba selanjutnya pada penelitian ini.

Tabel 3 . parameter range

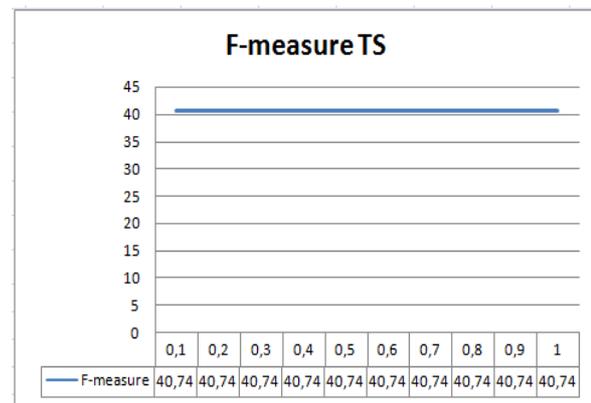
parameter	Minimum	Maximum	Interval
μ	1000	2000	100
λ	0,1	1	0,1

dengan menggunakan 100 data testing baru terhadap 600 data training, didapatkan hasil F-measure dirichlet smoothing pada gambar 4.



Gambar 4 .Parameter dirichlet smoothing

Pada gambar 4, nilai F-measure pada dirichlet smoothing mulai terlihat meningkat pada penggunaan parameter $\mu=1500$ hingga $\mu=2000$, sehingga penulis memilih $\mu=2000$ untuk penggunaan parameter dirichlet smoothing untuk ujicoba selanjutnya. Kemudian, parameter λ akan dipilih berdasarkan parameter range pada tabel 3 untuk digunakan pada Two-Stage smoothing.



Gambar 5. Parameter Two-Stage smoothing

Pada gambar 5, Two-stage tidak terlalu menunjukkan hasil yang signifikan terhadap pengaruh rentan parameter range 0,1-1 sesuai pada tabel 2. Sehingga parameter λ yang akan digunakan adalah $\lambda=0,5$ dan $\mu=2000$.

Setelah menemukan nilai yang tepat untuk setiap parameter yang akan digunakan pada metode smoothing, ujicoba selanjutnya adalah bertujuan untuk mengetahui

pengaruh jumlah kombinasi data testing dan data training yang digunakan, dan sekaligus mengetahui kinerja masing-masing metode smoothing dalam hal mengkategorikan data soal ujian baru dan mengenali kembali data yang sudah terlabeli dan menjadi bagian dari data training.

Sehingga, pada penelitian ini digunakan 2 jenis dataset. Dataset pertama dilakukan pengujian, dengan menggunakan 100 data testing baru yang belum dikenali sebelumnya terhadap kombinasi data training, rincian jumlah kombinasi data training dijelaskan pada tabel 4.

Tabel 4 : variasi jumlah data training

%	Jumlah data training
10%	60
20%	120
30%	180
40%	240
50%	300
60%	360
70%	420
80%	480
90%	540
100%	600

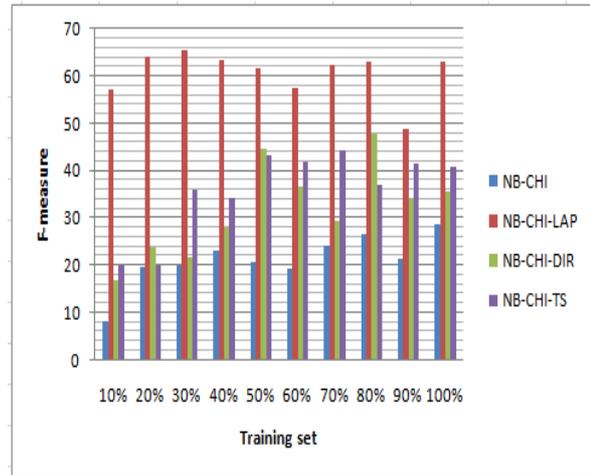
Sesuai dengan jumlah variasi data training yang digunakan, didapatkan nilai F-measure untuk setiap percobaan menggunakan metode smoothing pada tabel 5. Nilai threshold yang digunakan pada metode feature selection Chi Square adalah 0,5

Tabel 5 : hasil ujicoba dataset 1

Training set	F-measure			
	NB-CHI	NB-CHI-LAP	NB-CHI-DIR	NB-CHI-TS
10%	8,14	57,27	16,70	19,73
20%	19,61	64,09	23,55	19,70
30%	19,73	65,30	21,71	35,81
40%	22,91	63,21	27,80	34,15
50%	20,57	61,69	44,60	43,18
60%	18,99	57,51	36,41	41,94
70%	24,15	62,49	29,35	44,26
80%	26,60	63,18	47,81	36,96
90%	21,21	48,74	34,19	41,42
100%	28,44	62,88	35,52	40,74
Rata-rata	21,035	60,636	31,764	35,789

pada hasil ujicoba menggunakan dataset pertama, metode laplace smoothing berhasil mengkategorikan

data soal ujian dengan hasil rata-rata nilai F-measure mencapai 60,363 mengungguli 2 metode smoothing lainnya dan berhasil meningkatkan nilai F-measure Naive Bayes tanpa smoothing sebanyak 39,601. Kemudian, pada hasil percobaan di tabel 4 terlihat bahwa jumlah training sangat berpengaruh pada hasil F-measure. Gambar 6 merupakan bentuk diagram dari tabel 5.



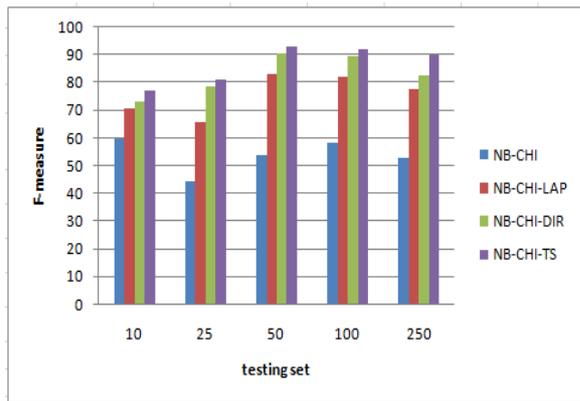
Gambar 6. Grafik Hasil Ujicoba Dataset 1

Ujicoba selanjutnya dilakukan dengan menggunakan dataset kedua, yaitu dilakukan pengujian sejumlah data testing, dimana data testing tersebut merupakan bagian data training, jumlah data testing yang akan diujikan sebanyak 25, 50, 100 dan 250. Ujicoba ini dilakukan untuk mengetahui sejauh mana metode smoothing mengenali kembali data testing yang merupakan bagian dari data training.

Tabel 5 : hasil ujicoba dataset 2

testing set	F-measure			
	NB-CHI	NB-CHI-LAP	NB-CHI-DIR	NB-CHI-TS
10	59,68	70,80	73,37	77,02
25	44,29	65,85	78,58	80,95
50	53,92	82,99	90,37	92,82
100	58,31	82,31	89,77	91,81
250	52,59	77,76	82,36	90,05
Rata-rata	53,758	75,942	82,89	86,53

pada ujicoba dataset kedua seperti yang dijelaskan pada tabel 6, diketahui metode smoothing dengan nilai rata-rata F-measure tertinggi adalah Two-stage smoothing yang berhasil mengungguli 2 metode smoothing lainnya, dan berhasil meningkatkan nilai F-measure Naive Bayes tanpa smoothing hingga 32,78. Gambar 7 merupakan bentuk diagram dari tabel 5



Gambar 7. Hasil Ujicoba Dataset 2

4. KESIMPULAN

Metode smoothing berhasil meningkatkan performance metode Naive Bayesian –Chi Square. Two stage smoothing berhasil mengungguli metode smoothing lainnya dalam hal mengenali kembali data testing yang sudah terlabeli dan merupakan bagian dari data training, TS smoothing berhasil meningkatkan metode naive bayes-chi square sebanyak 32,78.

Laplace smoothing memiliki rata-rata nilai F-measure tertinggi mengalahkan TS dan dirichlet smoothing dalam hal mengenali data soal ujian baru yang belum terlabeli/dikenali sebelumnya, dalam hal ini metode laplace smoothing dapat meningkatkan nilai rata-rata F-mesure pada naive bayessian-chi square tanpa smoothing hingga 39,601.

Metode smoothing terbukti dapat meminimalisir hasil akhir klasifikasi sama dengan 0, sehingga metode smoothing dapat memperbaiki hasil klasifikasi dan meningkatkan akurasi.

5. REFERENSI

- [1] F. Ramadhani, Rif'at Ahdi Indriani and D. T. Nugrahadi, "Comparison of Naive Bayes Smoothing Methods for Twitter Sentiment Analysis," *ICACSYS*, vol. 16, pp. 287–292, 2016.
- [2] S. Aggarwal, "Naïve Bayes Classifier with Various Smoothing for text document," *Int. J. Comput. Trends Technol.*, vol. 4, no. April, pp. 873–876, 2013.
- [3] E. I. W. M. X. L. Evmsyw, S. J. Sqtyxiv, R. Rxy, and I. H. Y. Wk, "Enhancing Naive Bayes with Various Smoothing Methods for short Text Classification," *WWW 2012 Companion*, vol. 4, no. April 16-20, pp. 645–646, 2012.
- [4] R. Utari, W. Madya, and Pusdiklat KNPk, "Taksonomi bloom Apa dan Bagaimana Menggunakannya?"
- [5] Z. H. Kilimci and M. C. Ganiz, "Evaluation of Classification Models for Language Processing," *Innov. Intell. Syst. Appl.*, 2015.
- [6] C. Zhai and J. Lafferty, "The Dual Role of Smoothing in the Language Modeling Approach," vol. 451326, pp. 5–10.
- [7] Y. Yang and J. Pedersen, "A comparative Study on Feature Selection in Text Categorization," 1997.
- [8] D. A. Abduljabbar and N. Omar, "Exam Questions Classification Based On Bloom's Taxonomy Cognitive Level Using Classifiers Combination," *J. Theor. Appl. Inf. Technol.*, vol. 78, no. 3, pp. 447–455, 2015.