

PERANCANGAN SISTEM KLASIFIKASI PENYAKIT DIABETES MENGUNAKAN ALGORITMA NAÏVE BAYES

Riyan Dwi Yulian Prakoso¹, Dhani Ratna Sari²,

^{1,2}. Program Bisnis Digital, Institut Teknologi Bisnis Muhammadiyah Purbalingga.

Jl. Letjen S Parman No.95, Bancar, Kec. Purbalingga, Kabupaten Purbalingga, Jawa Tengah, Indonesia

Email: riyandwwi@itbmp.ac.id

Penyakit Diabetes penyakit yang berlangsung lama atau kronis serta ditandai dengan kadar gula (glukosa) darah yang tinggi atau di atas nilai normal. Glukosa yang menumpuk di dalam darah akibat tidak diserap sel tubuh dengan baik dapat menimbulkan berbagai gangguan organ tubuh. Jika diabetes tidak dikontrol dengan baik, dapat timbul berbagai komplikasi yang membahayakan nyawa penderita. Di Indonesia, Penyakit Diabetes adalah kasus penyakit yang paling umum dijumpai pada orang dewasa khususnya lanjut usia (lansia). Tool yang di gunakan dalam mengklasifikasi penyakit diabetes adalah Naïve bayes. Metode ini menggunakan nilai probabilitas atau kemungkinan untuk menyelesaikan suatu masalah. Salah satu penerapan naïve bayes adalah untuk melakukan kalsifikasi serta mampu menyelesaikan permasalahan yang kompleks. Untuk menyelesaikan masalah naïve bayes memerlukan jumlah data yang banyak untuk mendapatkan nilai probabilitas. Data yang digunakan adalah adat penyakit Diabetes yang di ambil dari website UCI machine learning. Hasil dari penelitian ini adalah sebuah perancangan sistem klasifikasi dengan menggunakan naïve bayes untuk melakukan klasifikasi penyakit diabetes, hasil akhir penelitian ini adalah nilai dari akurasi, presisi, dan recall

Kata Kunci: Naive bayes, *Diabetes*, Kalsifikasi

1. PENDAHULUAN

Diabetes adalah salah satu penyakit dengan kematian pada orang yang cukup tinggi. Berdasarkan data kasus penyakit di Indonesia dari LITBANG KEMENKES terdapat jumlah kasus baru Penyakit Diabetes sebesar 56,82%. Pada tahun 2018 jumlah penderita penyakit Penyakit Diabetes sebesar 230.871 orang, dengan penderita terbanyak berdasarkan provinsi, yaitu pulau jawa dengan jumlah kasus sebesar 93.478 kasus [1] Penyakit Diabetes penyakit yang berlangsung lama atau kronis serta ditandai dengan kadar gula (glukosa) darah yang tinggi atau di atas nilai normal. Glukosa yang menumpuk di dalam darah akibat tidak diserap sel tubuh dengan baik dapat menimbulkan berbagai gangguan organ tubuh [2].

Klasifikasi adalah proses pengidentifikasian obyek ke dalam sebuah kategori, kelas atau kelompok berdasarkan prosedur, definisi dan karakteristik yang telah ditentukan sebelumnya [3] Klasifikasi bertujuan untuk menempatkan objek yang ditugaskan hanya ke salah satu kategori yang disebut kelas [4].

Tak hanya klasifikasi, proses pengelompokan obyek juga dapat dilakukan dengan menggunakan teknik clustering. Clustering merupakan pengelompokan obyek berdasarkan kemiripan antar obyek. Perbedaan antara klasifikasi dan clustering terletak pada proses pengelompokan obyek. Jika pada klasifikasi proses pengelompokan obyek dilakukan dengan membagi obyek berdasarkan kelompok / kategori yang telah didefinisikan sebelumnya, makaproses pengelompokan obyek pada clustering dilakukan dengan melihat kemiripan antar obyek, sehingga kategori belum terdefinisi sebelumnya. Salah satu metode klasifikasi yang sering digunakan adalah Naive Bayes.

Metode yang akan digunakan adalah metode Naïve Bayes Classifier, yang merupakan salah satu teknik pengklasifikasian dalam data mining [5]. Pada penelitian sebelumnya metode

Naive Bayes juga digunakan dalam memprediksi penyakit Dermatologi yang diabaikan tapi bahkan dapat menyebabkan kematian di mana metode Naive Bayes digunakan untuk mengenal pola data untuk mengungkap kemungkinan penyakit dermatologi[6]. Metode Naive Bayes juga dinilai berpotensi baik dalam mengklasifikasi dokumen dibandingkan metode pengklasifikasian yang lain dalam hal akurasi dan efisiensi komputasi [7].

Penelitian ini akan dibuat suatu perancangan sistem klasifikasi dengan judul “Perancangan Sistem Klasifikasi Untuk Penyakit Diabetes Menggunakan Algoritma Naive bayes ” yang dapat membantu petugas medis untuk memprediksi apakah pasien terjangkit diabetes atau tidak.

2. METODE

2.1. Metodologi Penelitian

Pada tahap ini peneliti akan melakukan studi literatur tentang metode prediksi partikel swam optimization dan naïve bayes. Kedua metode ini merupakan mesin pembelajaran yang sering digunakan untuk memprediksi atau meramalkan sesuatu. parameter yang digunakan pengujian dalam penelitian ini yaitu: Akurasi, presisi, recall pada hasil penerapan naïve bayes dalam klasifikasi penyakit diabetes.

2.2. Jenis, Sifat, Dan Pendekatan Penelitian

Adapun jenis, sifat dan pendekatan yang akan diimplementasikan pada penelitian ini sebagai berikut :

2.2.1 Jenis Penelitian Eksperimen

Penelitian ini merupakan penelitian eksperimen yaitu: meneliti perbandingan tingkat akurasi, Presisi, dan Recall pada metode algoritma Partikel swam optimization dan Naive bayes untuk klasifikasi penyakit diabetes.

2.2.2. Sifat Penelitian Deskriptif

Tujuan dari penelitian ini yaitu : mengetahui perbandingan tingkat Akurasi, Presisi, dan Recall untuk klasifikasi penyakit diabetes menggunakan algoritma Partikel Naive bayes dan mengetahui apakah jumlah record berpengaruh terhadap performa penggabungan kedua algoritma tersebut.

2.2.3. Pendekatan Penelitian Kuantitatif

Pada penelitian ini menggunakan pendekatan kuantitatif yang nantinya hasil dari penelitian ini merupakan informasi-informasi berupa angka dan diagram hasil dari eksperimen penggabungan dua metode yang dilakukan. Pengumpulan data dilakukan melalui hasil eksperimen yang kemudian data tersebut dilakukan penggabungan dan analisis seperti dibuatkannya tabel dan diagram untuk melihat penggabungan metode mana yang paling baik dalam klasifikasi penyakit diabetes.

2.3. Metode Pengumpulan Data

Pengumpulan data diperlukan dalam penelitian Adapun metode pengumpulan data antara lain:

1. Data Publik

Metode ini dimaksudkan untuk mendapatkan data publik yaitu data penyakit diabetes berjumlah 768 yang di ambil dari website UCI Machine Learning

2. Studi Pustaka

Melakukan penelusuran melalui dokumen-dokumen yang berhubungan baik dalam bentuk media cetak ataupun elektronik sebagai acuan dalam melakukan penelitian.

2.4. Metode Analisa Data

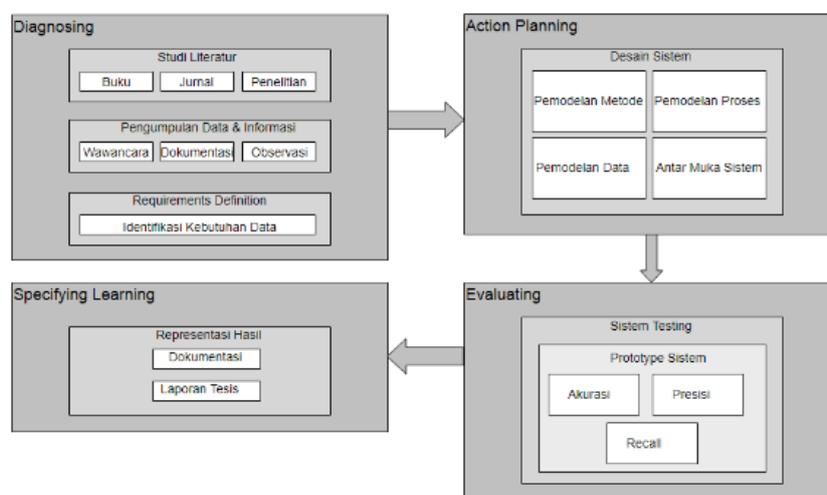
Tahap analisis data yang akan adalah membandingkan hasil eksperimen mulai dari awal sampai akhir. Eksperimen dimulai dari melakukan cleaning atau pembersihan pada data yang telah dikumpulkan. Tahap selanjutnya akan dilakukan transformasi data, dimana pada tahap ini akan dilakukan pengclusteran untuk dijadikan menjadi beberapa group atau kelompok data. Setelah data berhasil dikelompokkan maka proses selanjutnya akan dilakukan proses pemodelan data dimana data yang sudah dikelompokkan tersebut akan diubah menjadi nilai-nilai yang dipisahkan dengan tanda koma atau Comma Sparated Value (CSV) sebagai format data yang akan dimasukkan ke dalam database.

Setelah proses pemodelan data atau memasukan data di suatu database selesai dilakukan, maka proses yang dilakukan selanjutnya yaitu menerapkan model algoritma yaitu Naive bayes menggunakan data yang telah diperoleh. Kedua model algoritma ini akan digabungkan dengan menggunakan 5-FOLD Cross-Validation. Metode ini membagi secara acak 5 subset dan kemudian setiap subset $n=1$ akan menjadi data training sedangkan $n=2,3,\dots,5$ akan menjadi data testing. Kemudian dilakukan sebanyak 5 iterasi dengan data testing $n+1$ dan syarat data testing tidak sama dengan data training.

Penarikan kesimpulan dilakukan berdasarkan hasil evaluasi pada kedua model algoritma yang terpilih dengan menggunakan pengukuran nilai nilai performa Akurasi, presisi, recall dan waktu. Hasil pengukuran tersebut akan dijadikan acuan atau pedoman dalam menentukan hasil atau prediksi pada penelitian ini.

2.5. Alur Penelitian

Alur pada tahap ini berisi diagram proses urutan alur penelitian secara rinci dan detail yang dapat mencakup algoritma, rute, pemodelan-pemodelan, desain yang terdapat pada perancangan sistem. Alur penelitian pada sistem kelsifikasi penyakit diabetes ditunjukkan di Gambar 1.



Gambar 1
Metode Penelitian

Penjelasan singkat alur atau metode pada Gambar 1. seperti berikut:
Penjelasan mengenai metode penelitian pada Gambar 1. Adalah sebagai berikut:

1. *Diagnosing*

Pada *diagnosing* ini peneliti melakukan pengumpulan data melalui studi literatur pada media buku, jurnal ataupun artikel penelitian lain, selain itu peneliti juga mencari dataset melalui UCI machine learning untuk memperoleh informasi dataset dari penyakit diabetes untuk kemudian dilakukan identifikasi kebutuhan data yang sesuai dengan sistem yang akan dibangun.

2. *Action Planning*

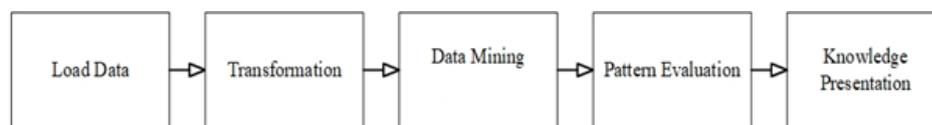
Pada *Action Planning* ini berisikan proses perancangan sistem yang akan dibangun. Perancangan tersebut berisikan pemodelan metode *Naïve Bayes* pemodelan proses dengan *usecase* dan *activity diagram* serta perancangan *User Interface* sebagai rancangan awal untuk dasar pembuatan sistem.

3. *Evaluating*

Pada *Evaluating* ini dilakukan pengujian sistem, dimana pengujian ini akan menilai seberapa besar tingkat akurasi, presisi dan recall melalui mekanisme validasi dengan 5-fold cross validation dari implementasi sistem dengan menggunakan metode *Naïve Naves* untuk klasifikasi penyakit diabetes .

4. *Specifying Learning*

Pada *Specifying Learning* ini dilakukan sebuah proses dokumentasi dan publikasi dari penelitian, dimana pada bagian ini berisikan hasil keseluruhan dari implementasi metode *Naïve bayes* untuk klasifikasi penyakit Diabetes.



Gambar 2
Alur Tahapan Data Mining (Klasifikasi)

Penjelasan secara singkat pada Gambar 2. alur tahapan data mining pada sistem klasifikasi penyakit diabetes seperti berikut:

1. Load Data

Proses awal dari data mining klasifikasi untuk diagnosa penyakit diabetes, dimana admin dapat meng-inputkan file excel dengan format CSV sebagai media untuk memasukan data.

2. Data Transformation

Pada tahapan penelitian ini, merubah data atribut yang berbentuk decimal kedalam bentuk bilangan bulat. Atribut-atribut yang dilakukan transformasi pada tabel data mining adalah 9 atribut yaitu Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, Diabetes Pedigree Function, Age, Outcome. serta Clasification.

3. Teknik Klasifikasi Dengan Metode Naïve Bayes

Teknik Klasifikasi Dengan Metode Naïve Bayes menggunakan algoritma Naïve bayes untuk klasifikasi Diabetes. Algoritma Naïve Bayes merupakan teknik data mining untuk melakukan klasifikasi data.

4. Pattern Evaluation

Pattern Evaluation dilakukan uji pemodelan pola untuk mengetahui nilai kesesuaian dan kesalahan atau error pada pola yang ditemukan. Pola yang sudah ditemukan kemudian dipresentasikan kepada pengguna agar mudah dimengerti. Pada penelitian ini pengujian sistem dilakukan menggunakan confusion matrix untuk mengetahui nilai akurasi.

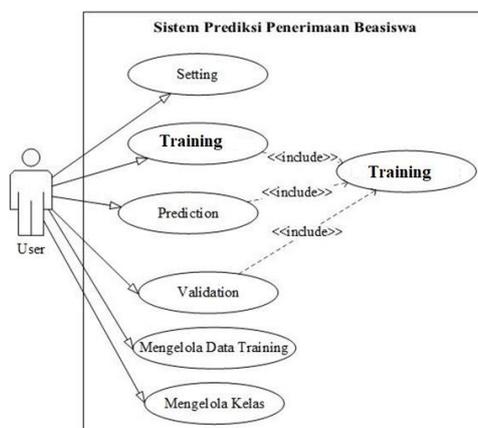
5. Representasi Pengetahuan (Knowledge Presentation)

Pada tahapan ini pola aturan algoritma Naïve Bayes yang telah ditemukan kemudian dipresentasikan kepada pengguna agar mudah dipahami.

3. HASIL DAN PEMBAHASAN

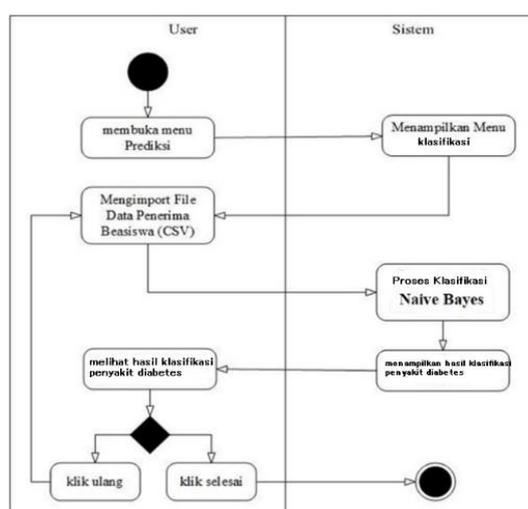
3.1. Perancangan Sistem

Pada proses perancangan sistem dibuat suatu perancangan dari sistem klasifikasi penyakit Diabetes melalui penerapan naïve bayes dengan menggunakan usecase pada Gambar 3 sebagai penjelasan kebutuhan fungsional dan activity diagram pada Gambar 4. sebagai penggambaran alur atau aktifitas pada sistem.



Gambar 3
Usecase Diagram

Gambar 2 menerangkan tentang fungsionalitas yang akan dibuat dan dirancang dengan pembuatan 5 menu yang diimplementasikan kedalam sistem, dimana terdapat menu pengaturan(setting) yang berfungsi untuk mengatur pembatasan jumlah dataset yang akan diolah dan dilakukan training, menu dataset untuk mengelola dataset seperti menambahkan dan menghapus dataset, menu prediction untuk melakukan kalidikasi penyakit diabetesan validation untuk melakukan validasi hasil berupa akurasi, presisi, dan recall, akan tetapi menu prediction dan validation hanya dapat dilakukan ketika proses pada menu Pada training dijalankan, menu training sendiri merupakan menu dimana dataset akan ditraining untuk dapat menemukan bobot yang akan digunakan untuk prediksi dan validasi.

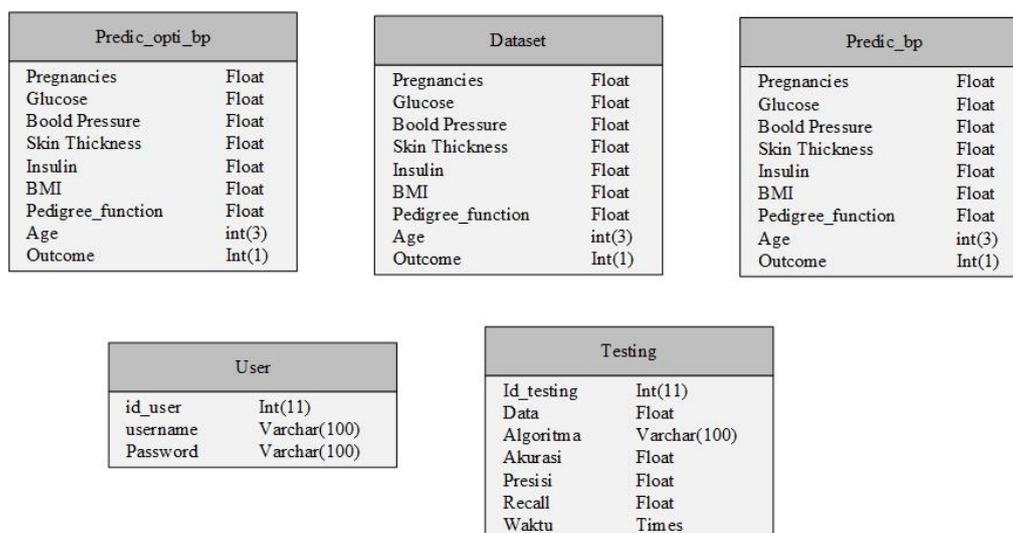


Gambar 4
Activity Diagram Prediksi

Pada gambar 3 menjelaskan tentang alur prediksi pada sistem dimana user membuka menu prediksi untuk memulai yang kemudian sistem akan merespon dan menampilkan menu selanjutnya user memasukkan data data pasien penyakit Diabetes dalam bentuk file Comma Sparated Value (CSV) dan sistem akan melakukan prediksi dan perhitungan dengan algoritma Naïve Bayes yang kemudian akan didapatkan hasil kalsifikasi penyakit diabetes yang ditampilkan ke user.

3.2. Analisa Model Data

Pada tahap ini akan dilakukan pemodelan data dengan menggunakan ERD atau Entity Relationship Diagram untuk membantu dalam pendeskripsian data pada sistem perangkat lunak yang akan digunakan dalam penelitian ini. Pemodelan ERD yang akan digunakan dalam penelitian ini dapat dilihat seperti pada gambar 5



Gambar 5
Entity Relationship Diagram

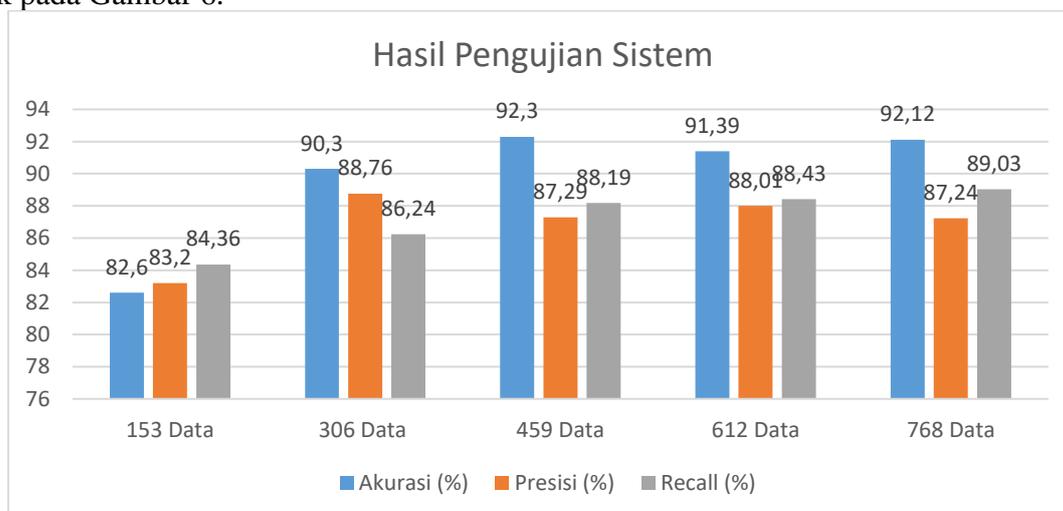
3.3. Pengujian Sistem

Model pengujian sistem dilakukan melalui beberapa percobaan dalam menentukan hasil nilai akurasi, presisi serta recall. Pengujian dilakukan dengan menggunakan 5-FOLD Cross-Validation, pembagian data dilakukan dengan membagi data menjadi 5 bagian seperti pada Tabel 2.

Tabel 1
Hasil Pengujian Sistem

No	Data	Akurasi (%)	Presisi (%)	Recall (%)
1	153	82,6	83,2	84,36
2	306	90,3	88,76	86,24
3	459	92,3	87,29	88,19
4	612	91,39	88,01	88,43
5	768	92,12	88,98	89,03
Rata-Rata		89,74	87,24	87,26

Berdasarkan hasil pada Tabel 2. Maka didapatkan hasil rata-rata akurasi sebesar 89,74%, presisi sebesar 87,44 %, recall sebesar 87,95%. Berdasarkan hasil tersebut terbentuk grafik pada Gambar 6.



Gambar 6
Hasil Pengujian Sistem

4. KESIMPULAN

Melalui hasil penelitian telah dijelaskan dan dilakukan pada pembuatan sistem klasifikasi penyakit diabetes menggunakan naïve bayes. Pembuatan sistem ini menyimpulkan nilai hasil akurasi rata-rata akurasi sebesar 89,74%, presisi sebesar 87,44 %, recall sebesar 87,95% dengan konfigurasi data yang terdapat pada data penyakit diabetes yang berisi 2 kelas dan 9 atribut dari 768 dataset

DAFTAR PUSTAKA

- [1] Kemenkes, R. (n.d.). Pusat Data Dan Informasi Kementerian Kesehatan RI. Retrieved Desember 19, 2018, from <http://www.depkes.go.id>
- [2] Aparimita Swain, Sachi Nandan Mohanty dan Ananta Chandra Das. 2016. Comparative Risk Analysis On Prediction Of Diabetes Mellitus Using Machine Learning Approach. International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT). Electronic ISBN: 978-1-4673-9939-5
- [3] Oana Geman, Luliana Chiuchisan dan Roxana Todorean. 2017. Inference System for diabetes classification and prediction. E-Health and Bioengineering Conference (EHB). Electronic ISBN: 78-1-5386-0358-1
- [4] Satwati Devi, Sanjay Kumar, dan Govind Singh Kushwaha. 2016. Adaptive Neuro Fuzzy Inference System For Prediction Of Anxiety Of Students. Eighth International Conference on Advanced Computational Intelligence ICACI). Electronic ISBN: 978-1-4673-7782-9
- [5] Mustafa , Muhammad Syukri . Simpen, I Wayan. (2014). Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus: Data Akademik Mahasiswa STMIK Dipanegara Makassar), ISSN: 2354-577
- [6] Manjusha K. K., Sankaranarayanan, K., Seena P., 2014, Prediction of Different Dermatological Conditions Using Naive Bayesian Classification, International Journal of Advanced Research in Computer Science and Software Engineering, Vol 4, No 1, Hal 864-868.

- [7] Ting, S. L., Ip, W. H., Tsang, A. H.C., 2011, Is Naive Bayes a Good Classifier for Document Classification?, International Journal of Software Engineering and Its Applications, Vol. 5, No. 3, Hal 37-46.