TESIS – KI 42502

Ekstraksi Fitur Statistik untuk Deteksi Derau pada Dokumen Spesifikasi Kebutuhan Perangkat Lunak

Ahmad Mustofa
NRP. 5116201054

DOSEN PEMBIMBING

Daniel Oranova Siahaan, S.Kom, M.Sc, P.D.Eng

NIP. 197411232006041001

PROGRAM MAGISTER

DEPARTEMEN TEKNIK INFORMATIKA

FAKULTAS TEKNOLOGI INFORMASI & KOMUNIKASI

INSTITUT TEKNOLOGI SEPULUH NOPEMBER

SURABAYA

2018

[Halaman ini sengaja dikosongkan]

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar Magister Komputer (M.Kom.)

di

Institut Teknologi Sepuluh Nopember Surabaya

oleh:

Ahmad Mustofa

Nrp. 5116201054

Dengan judul :

Ekstraksi Fitur Statistik untuk Deteksi Derau pada Dokumen Spesifikasi Kebutuhan Perangkat Lunak

Tanggal Ujian : 20-7-2018

Periode Wisuda : 2017 Genap

Disetujui oleh :

Daniel Oranova Siahaan, S.Kom., M.Sc., PD.Eng. (Pembimbing 1)
NIP. 197411232006041001

Prof. Ir. Drs. Ec. Riyanarto Sarno, M.Sc., Ph.D. (Penguji 1)
NIP. 195908031986011001

Dr. Ir. Raden Venantius Hari Ginardi, M.Sc. (Penguji 2)
NIP. 196505181992031003

Dr.Eng. Darlis Herumurti, S.Kom., M.Kom (Penguji 3)
NIP. 197712172003121001

Dekan Fakultas Teknologi Informasi & Komunikasi,

Dr. Agus Zainal Arifin, S.Kom., M.Kom
NIP. 197208091995121001

[Halaman ini sengaja dikosongkan]

Nama Mahasiswa : Ahmad Mustofa
NRP : 5116201054
Pembimbing : Daniel Oranovora Siahaan, S.Kom. PD.Eng

ABSTRAK

Tahap spesifikasi kebutuhan adalah tahap pertama yang dilakukan dalam proses pengembangan perangkat lunak. Sehingga jika terjadi kesalahan pada tahap ini, secara otomatis akan terjadi kesalahan pada tahap-tahap selanjutnya. Kesalahan dalam pernyataan kebutuhan perangkat lunak diantaranya berupa *noise* (derau), ambigu, konflik, serta inkonsistensi. Beberapa penelitian sebelumnya telah berhasil mendeteksi ambigu, konflik, dan inkonsistensi dalam pernyataan kebutuhan secara otomatis. Akan tetapi pada saat laporan ini ditulis, belum ada penelitian yang dilakukan untuk mendeteksi derau dalam pernyataan kebutuhan perangkat lunak secara otomatis.

Penelitian ini mengajukan suatu metode untuk deteksi derau dalam pernyataan kebutuhan perangkat lunak secara otomatis. Metode yang diajukan berfokus pada bagaimana melakukan ekstraksi fitur lokal dari masing-masing pernyataan kebutuhan dalam sebuah dokumen Spesifikasi Kebutuhan Perangkat Lunak (SKPL) sehingga kemudian bisa digunakan secara bersamaan dengan pernyataan kebutuhan dari dokumen SKPL yang lain dalam membangun model klasifikasi. Ekstraksi fitur lokal dari sebuah pernyataan kebutuhan dilakukan dengan memanfaatkan fitur statistik dari kemiripan pernyataan kebutuhan tersebut dengan pernyataan kebutuhan yang lain dalam dokumen SKPL yang sama.

Metode yang diajukan kemudian diuji dengan menggunakan data pernyataan kebutuhan yang telah dilabeli secara manual yang kemudian akan diukur performanya dengan menggunakan metode *k-fold cross validation*. Hasil uji coba menunjukkan bahwa metode yang diajukan memiliki performa yang cukup baik yaitu akurasi 82.96%, true positive rate 73.17%, true negative rate 84.07%, serta f1 measure 46.51%.

Kata kunci*:* pernyataan kebutuhan, deteksi derau, fitur statistik

[Halaman ini sengaja dikosongkan]

Nama Mahasiswa : Ahmad Mustofa
NRP : 5116201054
Pembimbing : Daniel Oranovora Siahaan, S.Kom. PD.Eng

ABSTRACT

Requirement specification is the first step of a software development cycle. If errors occured in this step, errors will automatically occur in the next step. Errors in software requirements consist of noise, ambiguous, conflict, and inconsistency. Some research has been successfully detecting ambiguous, conflict, and inconsistency in software requirements automatically. But when this document is written, there was no research done to detect noise in software requirements automatically.

This research propose a method to detect noise in software requirements automatically. Proposed method focuses on how to extract local features of a requirement statement in a Software Requirement Specification (SRS) document so that this feature can be used globally with other requirement statements from another SRS document to build a classification model. Local feature extraction of a requirement statement is done by using statistical feature of the requirement statement's similarities with other requirement statement in the same SRS document.

Proposed method will be validated by using requirement statements data that have been labeled manually and it's performance will be measured later by using k-fold cross validation method. The testing result indicates that the proposed method has a good performance such as 82.96% accuracy, 73.17% true positive rate, 84.07% true negative rate, and 46.51% f1 measure.

***Keywords**:* requirements, noise detection, statistical feature

[Halaman ini sengaja dikosongkan]

KATA PENGANTAR

Segala puji bagi Allah SWT yang telah melimpahkan rahmat dan anugerah-Nya sehingga penulis dapat menyelesaikan tesis yang berjudul "*Ekstraksi Fitur Statistik untuk Deteksi Derau pada Dokumen Spesifikasi Perangkat Lunak*" dengan tepat waktu. Tesis ini dibuat sebagai salah satu syarat dalam menyelesaikan Program Studi Magister di Institut Teknologi Sepuluh November Surabaya. Penulis menyadari bahwa Tesis ini dapat diselesaikan karena dukungan dari berbagai pihak, baik dalam bentuk dukungan moral dan material. Melalui kesempatan ini dengan kerendahan hati penulis mengucapkan terima kasih dan penghargaan setinggi-tingginya kepada semua orang untuk semua bantuan yang telah diberikan, antara lain kepada:

1. Allah SWT, karena atas limpahan rahmat-Nya, penulis diberikan kemudahan dan kelancaran dalam mengerjakan tesis ini.
2. Orang tua dan keluarga penulis yang senantiasa memberikan doa dan dukungan kepada penulis untuk menyelesaikan pengerjaan tesis ini.
3. Bapak Daniel Oranova Siahaan, S.Kom., M.Sc., PD.Eng sebagai dosen pembimbing yang telah memberikan banyak arahan dan bantuan sehingga penulis dapat menyelesaikan tesis ini.
4. Bapak Prof. Ir. Drs. Ec. Riyanarto Sarno, M.Sc., Ph.D., bapak Dr. Ir. Raden Venantius Hari Ginardi, M.Sc., dan bapak Dr.Eng. Darlis Herumurti, S.Kom., M.Kom sebagai dosen penguji yang telah membantu penulis untuk menyelesaikan tesis ini.
5. Bapak Dr. Eng. Radityo Anggoro, S.Kom., M.Sc selaku dosen wali yang telah memberikan bimbingan dan nasihat selama masa perkuliahan.
6. Teman-teman satu angkatan terutama teman satu bimbingan yang telah memberikan bantuan dalam penyelesaian tesis ini.
7. Semua pihak yang telah membantu terselesaikannya tesis ini.

Penulis menyadari adanya banyak kekurangan dalam pengrejaan tesis baik dari segi program maupun laporan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun untuk penyempurnaan tesis ini. Akhir kata, penulis meminta maaf bila terdapat kesalahan dalam penulisan laporan tesis ini. Semoga hasil dari tesis ini dapat memberikan manfaat bagi pembaca pada umumnya dan penulis pada khususnya.

Surabaya,

Ahmad Mustofa

[Halaman ini sengaja dikosongkan]

DAFTAR ISI

DAFTAR GAMBAR

[Halaman ini sengaja dikosongkan]

DAFTAR TABEL

[Halaman ini sengaja dikosongkan]

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Tahap spesifikasi kebutuhan adalah tahap pertama yang dilakukan dalam proses pengembangan perangkat lunak. Sehingga jika terjadi kesalahan pada tahap ini, secara otomatis akan terjadi kesalahan pada tahap-tahap selanjutnya. Bertrand meyer mengelompokkan kesalahan dalam spesifikasi kebutuhan menjadi tujuh kelompok yang kemudian dikenal dengan istilah *Meyer's seven sins* [1]. Beberapa peneliti mengajukan metode untuk mendeteksi ambiguitas (salah satu kesalahan dalam *Meyer's seven* sins) dalam pernyataan kebutuhan perangkat lunak [2]. Noise (derau) adalah salah satu kesalahan dalam *Meyer's seven sins* yang disebabkan oleh adanya suatu elemen dalam teks yang memberikan informasi yang tidak relevan dengan domain masalah yang hendak diselesaikan [1]. Meyer membagi derau menjadi dua jenis, yaitu *remorse* dan *redundancy*. *Remorse* dapat didefinisikan sebagai pernyataan kebutuhan yang memberikan informasi yang tidak relevan dengan domain masalah, sedangkan *redundancy* adalah pernyataan kebutuhan yang memberikan informasi yang sama dengan pernyataan kebutuhan yang lain.

Di sisi lain, perkembangan yang pesat pada bidang *machine learning* telah mampu melakukan klasifikasi teks secara otomatis. Hal yang paling penting dalam pengolahan teks terletak pada tahap pembobotan. Hao Xu dan Bo Yu menggunakan metode pembobotan term frequency – invers document frequency (tf-idf) untuk melakukan klasifikasi email spam [3]. Bruno Trstenjak, Sasa Mikac, dan Dzenana Donko juga menggunakan tf-idf sebagai metode pembobotan untuk melakukan pengelompokan dokumen [4].

Derau dalam dokumen Spesifikasi Kebutuhan Perangkat Lunak (SKPL) hanya dapat dikenali secara lokal. Pernyataan kebutuhan yang tergolong derau dalam sebuah dokumen SKPL akan memiliki kriteria yang berbeda dengan peryataan kebutuhan yang tergolong derau di dokumen SKPL yang lain. Hal ini bertolak belakang dengan metode tf-idf yang melakukan pembobotan teks secara global. Karenanya dibutuhkan sebuah

metode untuk ekstraksi fitur lokal dari masing-masing pernyataan kebutuhan dalam sebuah dokumen SKPL.

Di sisi lain, B. Chandra, dan Manish Gupta menggunakan pendekatan statisitik untuk seleksi fitur yang akan digunakan dalam klasifikasi gen [5]. Wujie Zhou, Lu Yu, Weiwei Qiu, Yang Zhou, dan Mingwei Wu juga menggunakan nilai statistik untuk mendapatkan fitur lokal dari sebuah citra yang kemudian digunakan untuk menentukan kualitas dari citra tersebut [6]. Pendekatan secara statistik inilah yang kemudian akan digunakan oleh penulis untuk ekstraksi fitur lokal dari masing-masing pernyataan kebutuhan dalam sebuah dokumen SKPL. Setelah didapatkan fitur dari masing-masing pernyataan kebutuhan, deteksi pernyataan kebutuhan yang termasuk derau dalam sebuah dokumen SKPL dapat dicapai dengan melakukan klasifikasi masing-masing pernyataan kebutuhan dengan memanfaatkan fitur-fitur yang berhasil didapat.

1.2 Perumusan Masalah

Rumusan masalah yang diangkat dalam penelitian ini adalah:

1. Fitur apa saja yang tepat untuk klasifikasi derau dalam pernyataan kebutuhan perangkat lunak
2. Bagaimana mengekstraksi fitur-fitur dari pernyataan kebutuhan dalam dokumen spesifikasi kebutuhan perangkat lunak

1.3 Tujuan

Tujuan yang akan dicapai dalam pembuatan tesis ini adalah membangun sebuah metode yang dapat mendeteksi pernyataan kebutuhan yang termasuk derau dalam sebuah dokumen spesifikasi kebutuhan perangkat lunak dengan akurat.

1.4 Manfaat

Manfaat dari penelitian ini adalah untuk membantu perekayasa kebutuhan meningkatkan kualitas dokumen spesifikasi kebutuhan perangkat lunak yang akan dibuat dengan mengurangi pernyataan kebutuhan yang termasuk derau.

1.5 Kontribusi Penelitian

Kontribusi yang diharapkan dari penelitian ini adalah mengembangkan sebuah metode ekstraksi fitur untuk klasifikasi pernyataan kebutuhan yang termasuk derau dalam dokumen spesifikasi kebutuhan perangkat lunak.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Dokumen spesifikasi kebutuhan perangkat lunak yang digunakan hanya menggunakan Bahasa Inggris.
2. Pernyataan kebutuhan diekstrak secara manual dari dokumen spesifikasi kebutuhan perangkat lunak

[Halaman ini sengaja dikosongkan]

BAB 2
KAJIAN PUSTAKA

2.1 Kebutuhan Perangkat Lunak

Kebutuhan dalam perangkat lunak adalah kumpulan deskripsi tentang bagaimana perangkat lunak tersebut berperilaku sehingga memiliki nilai guna bagi penggunanya. Kebutuhan perangkat lunak terbagi menjadi 2 jenis yaitu kebutuhan fungsional dan kebutuhan non-fungsional. Kebutuhan fungsional adalah kebutuhan yang dilihat dari sudut pandang pengguna, sedangkan kebutuhan non-fungsional adalah kebutuhan yang dilihat dari sudut pandang sistem [7].

Daur hidup sebuah perangkat lunak terbagi menjadi dua siklus utama, yaitu daur hidup pengembangan dan daur hidup pengoperasian. Kedua daur ini dihubungkan oleh studi kelayakan dan proses *launching* (peluncuran). Tahap spesifikasi kebutuhan merupakan tahap pertama dari daur hidup pengembangan sebuah perangkat lunak. Pada tahap ini, semua kebutuhan-kebutuhan pengguna dispesifikasikan. Hasil dari tahap spesifikasi kebutuhan adalah sebuah dokumen Spesifikasi Kebutuhan Perangkat Lunak (SKPL). Dokumen ini berisi tentang kumpulan kebutuhan-kebutuhan perangkat lunak yang akan dibuat. Dokumen SKPL ini berfungsi sebagai landasan proses pengembangan perangkat lunak yang telah disetujui oleh pihak pengembang dan klien. Terdapat empat pendekatan dalam menyusun dokumen SKPL yaitu :

1. Bahasa Alamiah

   Bahasa alamiah adalah bahasa sehari-hari yang digunakan oleh manusia.
2. Bahasa Alamiah Terstruktur

   Bahasa alamiah terstruktur adalah bahasa sehari-hari yang ditata sesuai dengan struktur/aturan tertentu.
3. Bahasa Semi-formal

   Bahasa semi-formal adalah bahasa grafikal yang dilengkapi dengan penjelasan teks.
4. Bahasa Formal

   Bahasa formal adalah bahasa yang terdiri dari notasi-notasi matematika.

Diantara empat pendekatan bahasa tersebut, yang paling sering digunakan adalah pendekatan bahasa alamiah.

2.2 Derau pada Dokumen SKPL

Penggunaan bahasa alamiah dalam dokumen SKPL dapat menimbulkan beberapa kesalahan. Bertrand Meyer mengelompokkan kesalahan yang biasa terjadi pada proses spesifikasi kebutuhan perangkat lunak menjadi 7 bagian yang terkenal dengan sebutan *meyer's seven sin*. Satu diantara kesalahan itu adalah derau. Dera adalah salah satu kesalahan dalam *Meyer's seven sins* yang disebabkan oleh adanya suatu elemen dalam teks yang memberikan informasi yang tidak relevan dengan domain masalah yang hendak diselesaikan. Derau terdiri dari dua jenis, yaitu *remorse* dan *redundancy*. *Remorse* dapat didefinisikan sebagai pernyataan kebutuhan yang memberikan informasi yang tidak relevan dengan domain masalah, sedangkan *redundancy* adalah pernyataan kebutuhan yang memberikan informasi yang sama dengan pernyataan kebutuhan yang lain [1].

Jenis derau yang menjadi fokus pada penelitian ini adalah derau jenis *remorse*, yaitu pernyataan kebutuhan yang memberikan informasu yang tidak relevan dengan domain masalah yang hendak diselesaikan. Beikut adalah contoh kumpulan pernyataan kebutuhan beserta label(derau atau tidak)nya.

Tabel 2.1 Contoh Pernyataan Kebutuhan

ID	**Pernyataan Kebutuhan**	**Label**
R-01	submit jobs with the associated deadline, cost, and execution time	Normal
R-02	query the cluster to establish the current cost per unit time for submitting new jobs	Normal
R-03	monitor the status of submitted jobs	Normal
R-04	cancel jobs submitted by him	Normal
R-05	check his credit balance	Derau
R-06	check his usage history	Normal
R-07	check the status of each node of the cluster	Normal
R-08	check the usage pattern history of the cluster	Normal
R-09	check the status of all submitted jobs	Normal
R-10	check the load on each node of the cluster	Normal
R-11	alter the cost structure of the cluster	Normal
R-12	alter the scheduling policy of the cluster	Normal
R-13	cancel, suspend, and resume any job	Normal

2.3 TF-IDF

Term frequency – invers document frequency adalah metode pembobotan dalam pengolahan teks [4]. Nilai term frequency (tf) didapat dengan menghitung jumlah kemunculan sebuah term dalam sebuah dokumen. Sedangkan nilai invers document frequency didapat dengan rumus

$$\mathrm{idf_i} = \log_2 \frac{N}{\mathrm{df_i}} \tag{2.1}$$

dimana $\mathrm{df_i}$ adalah jumlah dokumen yang mengandung term i. Nilai tf-idf dari term i kemudian dapat dihitung dengan mengalikan nilai tf dari term i dengan nilai idf dari term i [4].

Sebagai contoh jika sebuah pernyataan kebutuhan adalah satu dokumen dan sebuah dokumen SKPL sebagai corpus, maka sebuah kebutuhan R-05 dari sebuah dokumen SKPL pada Tabel 2.1 yang memiliki term unik [alter, associ, balance, cancel, check, cluster, cost, credit, current, deadlin, establish, execut, histori, job, load, monitor, new, node, pattern, per, polici, queri, resum, schedule, statu, structure, submit, suspend, time, unit, usag] memiliki nilai vektor tf [0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]. Untuk mendapatkan nilai tf-idf dari pernyataan kebutuhan kebutuhan R05 maka kalikan vektor tf dari pernyataan kebutuhan tersebut dengan vektor idf dari term unik dalam corpus [0, 0, 0.65, 0, 0.38, 0, 0, 0.65, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]. Jadi hasil tf-idf dari pernyataan R-05 adalah [0, 0, 0.65, 0, 0.38, 0, 0, 0.65, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0].

2.4 Cosine Similarity

Dalam *text processing*, sebuah dokumen biasanya direpresentasikan sebagai sebuah vektor. Cosine similarity adalah sebuah metode yang digunakan untuk menghitung kemiripan antara dua vektor [8]. Nilai cosine similarity dari vektor A dan vektor B didapat dengan menggunakan rumus

$$\mathrm{sim}(A,\ B) = \frac{\sum_{\mathrm{k}=1}^{\mathrm{N}} \mathrm{A_k} \ * \ \mathrm{B_k}}{\sqrt{\sum_{k=1}^{N} \ \mathrm{A_k}^{\ 2}} \ * \sqrt{\sum_{k=1}^{N} \ \mathrm{B_k}^{\ 2}}} \tag{2.2}$$

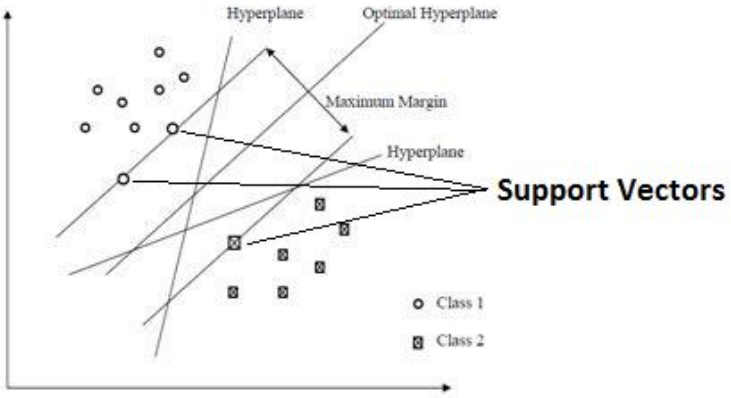
dimana $\mathrm{A_k}$ dan $\mathrm{B_k}$ adalah komponen dari vektor A dan B [8].

Sebagai contoh, dua buah vektor A : [2, 1, 3, 4, 2] dan B : [3, 2, 1, 2, 4] memiliki nilai cosine similarity

$$\text{sim(A,B)} = \frac{2*3 + 1*2 + 3*1 + 4*2 + 2*4}{\sqrt{2^2+1^2+3^2+4^2+2^2} * \sqrt{3^2+2^2+1^2+2^2+4^2}} = \frac{6+2+3+8+8}{\sqrt{4+1+9+16+4} * \sqrt{9+4+1+4+16}} = 0.79$$

2.5 Support Vector Machine

Support vector machine (SVM) adalah metode klasifikasi *supervised* (data latih sudah diketaui kelasnya) yang mengklasifikasikan dua kelas [9]. Pada metode ini, setiap data akan di-*plotting* ke dalam ruang *n*-dimensi (dimana *n* adalah jumlah fitur) dengan nilai masing-masing fitur menjadi nilai tertentu pada koordinat. Kemudian akan dilakukan klasifikasi dengan mencari *hyperplane* (bidang pembatas) untuk membedakan antara dua kelas sebaik mungkin. Gambar 2.1 merupakan beberapa contoh dari kemungkinan *hyperplane* yang digunakan untuk memisahkan kelas satu dengan lainnya. Metode SVM akan mencari h*yperplane* yang memiliki *margin* yang maksimal. *Support vector* adalah data-data yang dijadikan pembatas dengan kelas lain.



Sumber gambar : www.sine.ni.com

Gambar 2.1 Contoh alternatif hyperplane

Diberikan data masukan $(x_1, x_2, x_3, \ldots, x_n)$ dan masing-masing kelas dinotasikan $y_i \in \{-1, +1\}$ untuk $i = 1,2,3, \ldots, n$ dimana *n* adalah banyaknya data. Fungsi *hyperplane* dibuat dengan persamaan

$$w \,.\, x + b = 0 \qquad (2.3)$$

dengan batasan yang ditulis dalam persamaan

$$y_i(wx_i + b) + t_i \geq 1 \tag{2.4}$$

untuk mencari nilai *w* dan *b* yang optimal digunakan persamaan

$$\min \frac{1}{2} ||w||^2 + C \sum_{i=1}^{l} t_i \tag{2.5}$$

dimana nilai *C* adalah nilai pinalti dari kesalahan klasifikasi. Fungsi tujuan persamaan di atas berbentuk kuadrat, sehingga untuk menyelesaikannya, bentuk tersebut ditransformasikan ke dalam bentuk *dual space*. Persamaan *dual space* dapat ditulis menggunakan persamaan

$$\max \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j x_i x_j \tag{2.6}$$

dengan batasan dalam persamaan berikut.

$$\alpha_i \geq 0 , \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{2.7}$$

Untuk mencari nilai α_i, digunakan *sequential minimal optimization* (SMO). Persamaan *hyperplane* dilakukan dengan persamaan [9]:

$$f = w^T z + b = \sum_{i=1}^{s} \alpha_i y_i {x_i}^T z + b \tag{2.8}$$

dimana *z* adalah data masukan *support vector machine*. Pada banyak kasus, data yang diklasifikasikan tidak bisa langsung dipisahkan dengan garis yang linear. Oleh karena itu, digunakan metode kernel untuk mengatasi permasalahan tersebut. Dengan metode kernel, suatu data x di *input space* dipetakan ke fitur *space* F dengan dimensi yang lebih tinggi. Salah satu kernel yang biasa dipakai adalah kernel RBF dan *Polynomial*. Persamaan kernel RBF dan *Polynomial* berurutan dapat dilihat pada persamaan di bawah ini [9]:

$$k(x, y) = \exp(-\gamma|x - y|^2) \tag{2.9}$$

$$k(x, y) = (\gamma\langle x^T y\rangle + r)^p \tag{2.10}$$

Penggunaan fungsi kernel mengubah persamaan *dual space* menjadi

$$\max \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j k(x_i, x_j) \tag{2.11}$$

dan juga mengubah persamaan *hyperplane* menjadi

$$f = \sum_{i=1}^{s} \alpha_i y_i \, k(x_i, z) + b \tag{2.12}$$

2.6 Synthetic Minority Over-sampling Technique (SMOTE)

Untuk mengatasi adanya *imbalanced* pada data latih, SMOTE dapat digunakan untuk melakukan sintesa data sehingga data yang digunakan untuk melatih *classifier* dapat menjadi seimbang. SMOTE merupakan sebuah algoritma *oversampling* (menambahkan data baru dengan kelas minoritas) yang memanfaatkan kemiripan dari masing-masing data [10]. Berikut adalah pseudocode dari algoritma SMOTE

```
1  SMOTE(T, N, k)
2  #   Input: Number of minority class samples T; Amount of SMOTE N%;
       Number of nearest neighbors k
3  #   Output: (N/100)* T synthetic minority class samples
4      if N <100
5         then Randomize the T minority class samples
6         T = (N/100) ∗ T
7         N = 100
8      endif
9      N = (int)(N/100)
10     k = Number of nearest neighbors
11     numattrs = Number of attributes
12     Sample[ ][ ]: array for original minority class samples
13     newindex: 0
```

```
14      Synthetic[ ][ ]: array for synthetic samples
15      for i ← 1 to T
16          Compute k nearest neighbors for i, and save the indices in the nnarray
17          Populate(N, i, nnarray)
18      endfor
19  return
20  Populate(N, i, nnarray)
21      while N != 0
22          for attr ← 1 to numattrs
23              Compute: dif = Sample[nnarray[nn]][attr] − Sample[i][attr]
24              Compute: gap = random number between 0 and 1
25              Synthetic[newindex][attr] = Sample[i][attr] + gap ∗ dif
26          endfor
27          newindex++
28          N = N − 1
29      endwhile
30  return
```

Sebagai contoh 3 buah data minority A : [0,1 0.2], B : [0.2 0.2], dan C : [0.2 0.3] akan disintesa dengan menggunakan SMOTE dengan nilai k = 2 dan N = 100. Data A menghasilkan data sintesa baru A' : [0.15 0.2], data B menghasilkan data sintesa baru B' : [0.2 0.21], dan data C menghasilkan data sintesa baru C' : [0.2 0.3].

2.7 Sensitivity, Specificity, dan F1 Measure

Untuk menghitung tingkat performansi suatu sistem dapat digunakan perhitungan presisi, *recall*, dan F1 Measure. Secara matematis, rumus untuk menghitung sensitivity, specificity, dan F1 Measure dapat dilihat pada persamaan 2.14, 2.15, dan 2.16.

$$precision = \frac{TP}{TP + FP} \tag{2.13}$$

$$sensitivity = \frac{TP}{TP + FN} \tag{2.14}$$

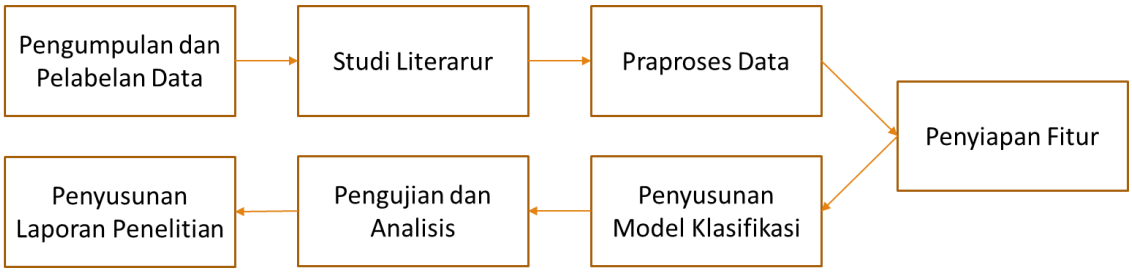
$$f1\ measure = 2 * \frac{precision * sensitivity}{precision + sensitivity} \tag{2.15}$$

$$specificity = \frac{TN}{TN + FP} \tag{2.16}$$

dimana TP adalah kebutuhan yang tergolong derau dan terklasifikasi sebagai derau, FP adalah kebutuhan yang bukan tergolong derau akan tetapi terklasifikasi sebagai derau, serta FN adalah kebutuhan yang tergolong derau akan tetapi tidak terklasifikasi sebagai derau [11].

BAB 3
METODOLOGI PENELITIAN

Bab ini akan memaparkan tentang metodologi penelitian yang digunakan pada penelitian ini, yang terdiri dari (1) pengumpulan dan pelabelan data, (2) studi literatur, (3) praproses data, (4) ekstraksi fitur, (5) penyusunan model klasifikasi, (6) pengujian dan analisis, dan (7) penyusunan laporan penelitian. Ilustrasi alur metodologi penelitian dapat dilihat pada berikut.



Gambar 3.1 Alur Penelitian

3.1 Pengumpulan dan Pelabelan Data

Penelitian diawali dengan proses pengumpulan data pernyataan kebutuhan yang kemudian dilakukan pelabelan secara manual untuk menentukan pernyataan kebutuhan yang tergolong derau atau tidak. Sebagai contoh pada sistem informasi penemuan barang hilang dan sisterm informasi ketersediaan dosen, F adalah singkatan dari Fungsionalitas seperti yang ditunjukkan pada tabel berikut.

Tabel 3.1 Pernyataan kebutuhan SI Penemuan Barang Hilang

Kode	Pernyataan Kebutuhan	Label
F01	Register a new user account	0
F02	Update user profile	0
F03	Report on finding thing	0
F04	Report on lost thing	0
F05	Asking help in form of question	1
F06	Delete user account	0
F07	Give reward	1

Tabel 3.2 Pernyataan kebutuhan SI Ketersediaan Dosen

Kode	Pernyataan kebutuhan	Label
F11	Register a new lecturer's account	0
F12	Delete lecturer's account	0
F13	Update lecturer's account information	0
F14	Add lecturer's availability status	0
F15	Update lecturer's availability status	0
F16	Delete lecturer's availability status	0
F17	Show lecturer's information and availability status	0
F18	Find lecturer's account	0
F19	Lecturer can approve student's plan	1

Dari total 405 data yang berhasil dikumpulkan dan dilabeli oleh 3 orang annotator, total persentase data derau hanya sekitar 10%. Hal ini menunjukkan bahwa persebaran data derau tidaklah seimbang. Karenanya dibutuhkan sebuah proses untuk melakukan penyeimbangan data. Proses penyeimbangan data dilakukan dengan menggunakan algoritma SMOTE.

Tabel 3.3 Persebaran derau dalam SKPL yang digunakan

Nama File	Jumlah Derau	Total	Persentase
1_libra_srs-uji1	1	13	7.6
24_srsv6-uji24	3	24	12.5
18_SoftwareRequirementsSpecification	9	86	10.46
7_16hlmnExemplu_cerinte_software	1	6	16.67
12_15hlmnSRS4.0_uji12	14	106	13.2
4.argos_urd-uji4	1	39	2.56
13_T1-Req-uji13	8	64	12.5
15_bpms-uji15	2	17	11.76
2_SRS_sample V1.2	1	17	5.88
14_yh-rr-ss-jw1-uji14	1	33	3.03

3.2 Studi Literatur

Proses kedua adalah pengkajian yang berkaitan dengan topik penelitian yang diambil. Pada penelitian ini, referensi yang digunakan adalah jurnal-jurnal yang berkaitan dengan metode-metode yang digunakan.

3.3 Praproses Data

Pada tahap ini akan dilakukan praproses pada data pernyataan kebutuhan yang telah berhasil dikumpulkan. Tahap ini terbagi menjadi 3 proses, yaitu *stopword removal*, *stemming*, dan tokenisasi.

3.3.1 Stopword Removal

Stopword removal adalah penghapusan kata penghubung dan tanda baca dari masing-masing data pernyataan kebutuhan. Sebagai contoh, sebuah pernyataan kebutuhan F17 “show lecturer’s information and availability status” akan menjadi “show lecturer information availability status” setelah melalui tahap ini.

3.3.2 Stemming

Stemming adalah proses pengubahan setiap kata dalam pernyataan kebutuhan menjadi kata dasar. Sebagai contoh, “show lecturer information availability status” akan menjadi “show lecture information available status” setelah melalui tahap ini.

3.3.3 Tokenisasi

Tokenisasi adalah proses pemecahan masing-masing pernyataan kebutuhan menjadi kumpulan kata yang unik. Sebagai contoh, “show information available lecture available student” akan menjadi {“show", “information”, “available”, “lecture”, “student”} setelah melalui tahap ini.

3.4 Penyiapan Fitur

Tahap ini terdiri dari dua proses utama, yaitu tahap pembobotan term dan tahap ekstraksi fitur.

3.4.1 Pembobotan Term

Masing-masing pernyataan kebutuhan yang telah melalui tahap praproses akan dihitung bobotnya dengan menggunakan rumus tf-idf. Setelah proses ini, masing-masing pernyataan kebutuhan akan direpresentasikan sebagai sebuah vektor.

Sebagai contoh, 2 pernyataan kebutuhan F17 “show lecturer’s information and availability status” dan F13 “update lecturer’s account information” akan menjadi {“show", “lecture”, “information”, “available”, “status”} dan {“update", “lecture”, “account”, “information”} setelah melalui tahap praproses. Token-token dari F17 dan F13 kemudian digabungkan menjadi {“show", “lecture”, “information”, “available”, “status”,

“update”, “account”} untuk kemudian digunakan sebagai fitur dalam penghitungan bobot tf-idf.

Untuk mendapatkan bobot tf-idf, pertama yang dilakukan adalah menghitung *term frequency* dari masing-masing kata fitur dalam masing-masing pernyataan kebutuhan sebagai berikut.

Tabel 3.4 Contoh Hasil Penghitungan Term Frequency

Kode	**Show**	**Lecture**	**Information**	**Available**	**Status**	**Update**	**Account**
F17	1	1	1	1	1	0	0
F13	0	1	1	0	0	1	1

Setelah itu, dilakukan penghitungan idf dari masing-masing kata fitur sebagai berikut.

Tabel 3.5 Contoh Hasil Penghitungan Invers Document Frequency

Kode	**Show**	**Lecture**	**Information**	**Available**	**Status**	**Update**	**Account**
Idf	0.3	0	0	0.3	0.3	0.3	0.3

Tahap akhir adalah mengalikan nilai tf dari masing-masing pernyataan kebutuhan dengan nilai idf.

Tabel 3.6 Contoh Hasil Penghitungan TF-iDF

Kode	**Show**	**Lecture**	**Information**	**Available**	**Status**	**Update**	**Account**
F17	0.3	0	0	0.3	0.3	0	0
F13	0	0	0	0	0	0.3	0.3

3.4.2 Ekstraksi Fitur

Untuk masing-masing vektor pernyataan kebutuhan, akan dihitung nilai similaritasnya dengan vektor pernyataan kebutuhan yang lain dalam dokumen SKPL yang sama dengan menggunakan *cosine similarity*. Setelah itu diambil tiga fitur yang kemudian akan digunakan dalam tahap klasifikasi. Tiga fitur tersebut adalah sebagai berikut.

a. Maximum of Similarity Measure

   Fitur ini adalah nilai maksimum dari nilai similaritas yang sudah didapatkan sebelumnya.

b. Mean of Similarity Measure

Fitur ini adalah nilai rata-rata dari nilai similaritas yang sudah didapatkan sebelumnya.

c. Standard Deviation of Similarity Measure

Fitur ini adalah nilai standar deviasi dari nilai similaritas yang sudah didapatkan sebelumnya.

Selain 3 fitur diatas, terdapat 2 fitur tambahan yang akan diuji coba untuk mendapatkan kombinasi fitur yang menghasilkan performa terbaik. Dua fitur tersebut adalah sebagai berikut.

a. Minimum of Similarity Measure

Fitur ini adalah nilai minimum dari nilai similaritas yang sudah didapatkan sebelumnya.

b. Variance of Similarity Measure

Fitur ini adalah nilai varian dari nilai similaritas yang sudah didapatkan sebelumnya.

Berikut adalah pseudocode dari proses ekstraksi fitur yang dilakukan untuk masing-masing pernyataan kebutuhan dalam sebuah dokumen SKPL.

```
1  ExtractFeature(DokumenSKPL)
2      feature_matrix = new matrix
3      foreach k = pernyataan_kebutuhan in DokumenSKPL
4          similarity = cosineSimilarity(k,
               rest_of_pernyataan_kebutuhan)
5          feature_matrix.add([mean(similarity), std(similarity),
               max(similarity)])
6      end foreach
7      return feature_matrix
```

Berikut adalah hasil ekstraksi fitur dari contoh kasus sistem informasi penemuan barang hilang dan sisterm informasi ketersediaan dosen.

Tabel 3.7 Hasil Ekstraksi Fitur pada Contoh Kasus

Kode	Mean	Standard Deviation	Maximum	Label
F01	0.062	0.119	0.3	0
F02	0.03	0.047	0.1	0
F03	0.076	0.185	0.5	0
F04	0.076	0.185	0.5	0
F05	0	0	0	1
F06	0.066	0.119	0.3	0

Kode	Mean	Standard Deviation	Maximum	Label
F07	0	0	0	1
F11	0.049394	0.075	0.197	0
F12	0.149962	0.219	0.642	0
F13	0.049394	0.075	0.197	0
F14	0.081858	0.117	0.281	0
F15	0.081858	0.117	0.281	0
F16	0.17665	0.228	0.642	0
F17	0.066294	0.093	0.21	0
F18	0.042397	0.063	0.164	0
F19	0	0	0	1

3.5 Penyusunan Model Klasifikasi

Pada tahap ini data-data pernyataan kebutuhan yang sudah direpresentasikan sebagai vektor fitur akan dibagi menjadi 2 kelompok, yaitu data latih dan data uji.

3.5.1 Pemisahan data

Pada penelitian ini, pengujian dilakukan dengan menggunakan *10-folds cross validation*. Dalam *10-folds cross validation*, dataset dibagi menjadi 10 bagian yang sama rata. Dengan melakukan iterasi sebanyak 10 kali, satu bagian dari data digunakan sebagai data uji, dan 9 bagian yang lain digunakan sebagai data latih. Satu bagian data yang digunakan sebagai data uji pada setiap iterasi adalah bagian data yang berbeda dari iterasi sebelumnya.

3.5.2 Penyeimbangan Data Menggunakan SMOTE

Dalam setiap iterasi pada *10-folds cross validation,* akan dilakukan penyeimbangan data derau pada data latih dengan menggunakan algoritma SMOTE. Berikut adalah hasil penyeimbangan data dari Tabel 3.7.

Tabel 3.8 Hasil Balancing Data Menggunakan Algoritma SMOTE dengan N=300

Kode	Mean	Standard Deviation	Maximum	Label
F01	0.062	0.119	0.3	0
F02	0.03	0.047	0.1	0
F03	0.076	0.185	0.5	0

Kode	Mean	Standard Deviation	Maximum	Label
F04	0.076	0.185	0.5	0
F05	0	0	0	1
F06	0.066	0.119	0.3	0
F07	0	0	0	1
F11	0.049394	0.075	0.197	0
F12	0.149962	0.219	0.642	0
F13	0.049394	0.075	0.197	0
F14	0.081858	0.117	0.281	0
F15	0.081858	0.117	0.281	0
F16	0.17665	0.228	0.642	0
F17	0.066294	0.093	0.21	0
F18	0.042397	0.063	0.164	0
F19	0	0	0	1
SMOTE 1	0	0	0	1
SMOTE 2	0	0	0	1
SMOTE 3	0	0	0	1
SMOTE 4	0	0	0	1
SMOTE 5	0	0	0	1
SMOTE 6	0	0	0	1
SMOTE 7	0	0	0	1
SMOTE 8	0	0	0	1
SMOTE 9	0	0	0	1

Dengan menggunakan data F05, F07, dan F19 serta nilai N=300 dan k = 2 dihasilkan 9 data sintesa baru. Data sintesa dengan kode SMOTE 1, SMOTE 2, dan SMOTE 3 didapatkan dengan memanfaatkan 2 data tetangga terdekat dari F05. Data sintesa dengan kode SMOTE 4, SMOTE 5, dan SMOTE 6 didapatkan dengan memanfaatkan 2 data tetangga terdekat dari F07. Data sintesa dengan kode SMOTE 7, SMOTE 8, dan SMOTE 9 didapatkan dengan memanfaatkan 2 data tetangga terdekat dari F19.

3.5.3 Training Klasifikasi

Setelah data latih mempunyai keseimbangan antara data derau dan data normal, maka proses *training* dilakukan untuk mendapatkan model klasifikasi. Model inilah yang nantinya akan dipakan untuk melakukan klasifikasi pada data uji.

3.6 Pengujian dan Analisis

Tujuan dari pengujian adalah untuk membuktikan bahwa metode yang diajukan dapat mendeteksi derau pada pernyataan kebutuhan dalam dokumen SKPL secara akurat. Data uji yang telah melalui tahap ekstraksi fitur seperti pada tabel 3.6 akan menjadi masukan yang akan diklasifikasi dengan menggunakan model klasifikasi yang sudah dilatih menggunakan data latih. Label output dari hasil klasifikasi ini kemudian akan dibandingkan dengan label hasil penilaian secara manual untuk kemudian dihitung nilai akurasi, sensitivitas, spesifisitas, dan f1 measure nya.

Skenario uji coba yang akan dilakukan dalam penelitian ini adalah sebagai berikut.

1. Untuk mendapatkan fitur statistik yang baik sangat tergantung pada bagaimana menghitung kemiripan antar pernyataan kebutuhan. Pada skenario ini akan dibandingkan 2 metode penghitungan kemiripan yaitu metode *Jaccard Similarity* dan metode *Cosine Similarity*. Hasil dari uji coba ini diharapkan mampu mendapatkan metode penghitungan kemiripan yang paling baik untuk kasus deteksi derau pada dokumen SKPL ini.
2. Untuk mendapatkan fitur yang terbaik untuk merepresentasikan sebuah kebutuhan, pada pengujian ini akan dilakukan variasi kombinasi dari fitur statistik yang disebutkan pada bab 3.4.2. Hasil dari uji coba ini diharapkan mampu mendapatkan kombinasi fitur statistik yang terbaik untuk merepresentasikan sebuah pernyataan kebutuhan untuk deteksi derau pada dokumen SKPL.
3. Dalam *text processing*, metode KNN dan *Neural Network* mampu menghasilkan performa yang bagus pada [3, 4] untuk mengkategorikan dokumen teks dan untuk menyaring surel spam. Uji coba ini dilakukan untuk mendapatkan metode klasifikasi mana diantara SVM, KNN, dan *Neural Network* yang paling cocok untuk digunakan pada sistem deteksi derau dalam dokumen SKPL ini.
4. Untuk mengetahui persentase data derau yang terbaik yang digunakan pada saat pelatihan model klasifikasi, dalam uji coba ini digunakan variasi persentase data derau

yang digunakan pada saat pelatihan model klasifikasi. Hasil dari uji coba ini diharapkan mampu mendapatkan persentase data derau yang paling baik untuk digunakan pada saat proses pelatihan sehingga akan menghasilkan sebuah model klasifikasi yang memiliki performa yang memuaskan.

[Halaman ini sengaja dikosongkan]

BAB 4
HASIL DAN PEMBAHASAN

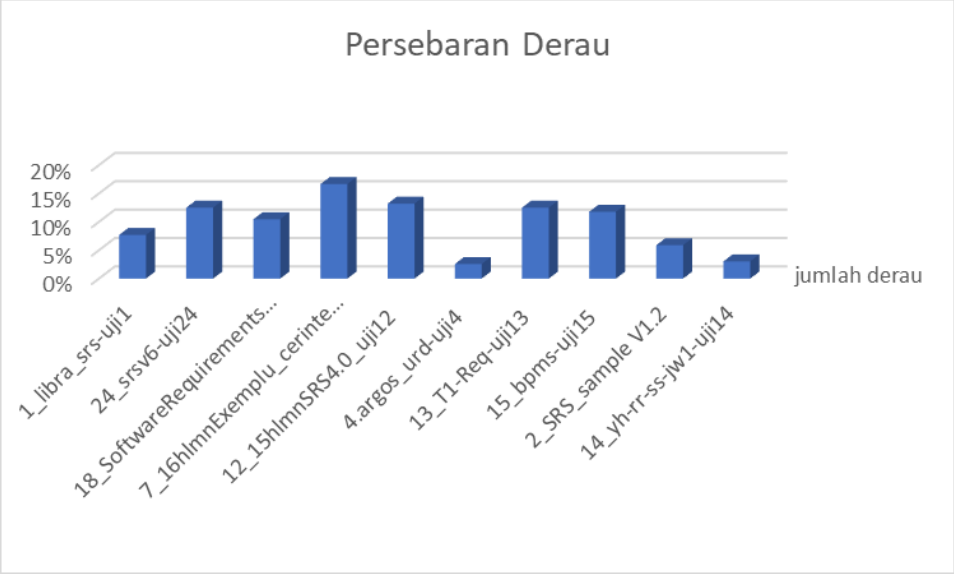
Bab ini akan memaparkan tentang hasil uji coba yang sudah dilakukan beserta pembahasan dan analisa dari masing-masing skenario uji coba yang telah dilakukan.

4.1 Implementasi Sistem

Penelitian ini diimplementasikan dengan menggunakan python dan matlab. Proses dari preproses data sampai ekstraksi fitur diimplementasi menggunakan python sedangkan proses balancing data dan proses klasifikasi diimplementasi menggunakan matlab.

4.1.1 Deskripsi Data Uji

Data yang digunakan dalam penelitian ini terdiri dari 10 dokumen SKPL dengan total 405 pernyataan kebutuhan. Dari 405 pernyataan kebutuhan, 41 diantaranya merupakan pernyataan kebutuhan yang tergolong derau. Berikut ini persebaran data derau di masing-masing dokumen SKPL.



Gambar 4.1 Persebaran Data Derau dalam Dokumen SKPL

Derau terbanyak terdapat pada dokumen 7_16hlmnExemplu_cerinte_software - uji7 dengan persentase derau sebesar 17% (1 dari 6). Hal ini menujukkan tidak seimbangnya data derau yang ada pada data yang digunakan. Tidak seimbangnya data saat *training* dapat menyebabkan rendahnya performa dari model klasifikasi yang

digunakan. Untuk mengatasi hal ini maka dalam penelitian ini akan digunakan metode SMOTE untuk *balancing* data saat *training* dilakukan.

4.1.2 Praproses Data

Masing-masing pernyataan kebutuhan akan melalui tahap praproses data terlebih dahulu. Praproses data terdiri dari 5 tahap yaitu menghapus karakter yang bukan alfabet, mengubah karakter menjadi *lowercase*, tokenisasi, menghapus *stopword*, dan *stemming*. Algoritma *stemming* yang digunakan adalah algoritma porter. Berikut adalah contoh hasil praproses data yang telah dilakukan pada dokumen 7_16hlmnExemplu_cerinte_software -uji7.

Tabel 4.1 Contoh Hasil Prapropes Data

Pernyataan Kebutuhan	**Hasil Praproses data**
User can see detailed information about SGP	user see detail inform sgp
User can apply for SGP	user appli sgp
User can application update form for SGP	user applic updat form sgp
User can update personal data	user updat person data
User can upload application related documents	user upload applic relat document
User can perform psychological test	user perform psycholog test

4.1.3 Ekstraksi Fitur

Hal pertama yang dilakukan untuk mendapatkan fitur statistik dari masing-masing pernyataan kebutuhan adalah dengan melakukan pembobotan term masing-masing kebutuhan dalam dokumen SKPL yang sama. Algoritma pembobotan yang digunakan pada penelitian ini adalah algoritma tf-idf. Berikut adalah hasil pembobotan term untuk dokumen 7_16hlmnExemplu_cerinte_software -uji7.

Tabel 4.2 Contoh Hasil Pembobotan Term

No	**[appli, applic, data, detail, document, form, inform, perform, person, psycholog, relat, see, sgp, test, updat, upload, user]**
1	[0,0,0,0.52,0,0,0.52,0,0,0,0,0.52,0.36,0,0,0,0.23]
2	[0.77,0,0,0,0,0,0,0,0,0,0,0,0.53,0,0,0,0.34]
3	[0,0.47,0,0,0,0.58,0,0,0,0,0,0,0.40,0,0.47,0,0.26]
4	[0,0,0.59,0,0,0,0,0,0.59,0,0,0,0,0,0.48,0,0.26]
5	[0,0.42,0,0,0.51,0,0,0,0,0,0.51,0,0,0,0,0.51,0.23]

No	[appli, applic, data, detail, document, form, inform, perform, person, psycholog, relat, see, sgp, test, updat, upload, user]
6	[0,0,0,0,0,0,0,0.56,0,0.56,0,0,0,0.56,0,0,0.25]

Setelah pembobotan term, maka akan dihitung kemiripan antar pernyataan kebutuhan dalam dokumen SKPL yang sama. Berikut adalah nilai kemiripan antar pernyataan kebutuhan dengan menggunakan algoritma *cosine similarity*.

Tabel 4.3 Kemiripan Masing-Masing Pernyataan Kebutuhan Pada Dokumen 7_16hlmnexemplu_Cerinte_Software -Uji7

Pernyataan Kebutuhan	1	2	3	4	5	6
1	-	0.272444	0.202941	0.060668	0.052244	0.057476
2	0.272444	-	0.300536	0.089844	0.077368	0.085116
3	0.202941	0.300536	-	0.295301	0.254296	0.063402
4	0.060668	0.089844	0.295301	-	0.059135	0.065057
5	0.052244	0.077368	0.254296	0.059135	-	0.056023
6	0.057476	0.085116	0.063402	0.065057	0.056023	-

Dari Tabel 4.3 kemudian dihitung fitur statistiknya untuk merepresentasikan masing-masing pernyataan kebutuhan.

Tabel 4.4 Hasil Eksraksi Fitur dari Dokumen 7_16hlmnexemplu_Cerinte_Software - Uji7

Pernyataan Kebutuhan	Mean	Maximum	Standard Deviation
1	0.129155	0.272444	0.091345
2	0.165062	0.300536	0.099623
3	0.223295	0.300536	0.087283
4	0.114001	0.295301	0.091327
5	0.099813	0.254296	0.077722
6	0.065415	0.085116	0.010426

4.1.4 Klasifikasi

Klasifikasi terdiri dari dua bagian utama yaitu proses *training* dan *testing*. Proses *training* dilakukan untuk mendapatkan model klasifikasi yang nantinya akan dipakai untuk mengklasifikasi data *testing*. Karena data yang digunakan memiliki kelas yang tidak seimbang, maka diperlukan proses penyeimbangan data sebelum proses *training* dilakukan. Dalam uji coba pada penelitian ini, digunakan metode validasi *10-fold cross*

validation. Berikut adalah sebagian hasil sintesa data dengan menggunakan algoritma SMOTE pada iterasi terakhir dari proses *10-fold cross validation*.

Tabel 4.5 Contoh Sintesa Data menggunakan SMOTE

Mean	Maximum	Standard Deviation	Label
0.024	0.173	0.039	1
0.062	0.169	0.074	1
0.043	0.193	0.059	1
0.020	0.154	0.033	1
0.027	0.169	0.040	1
0.030	0.190	0.046	1

4.2 Hasil Uji Coba

4.2.1 Skenario 1

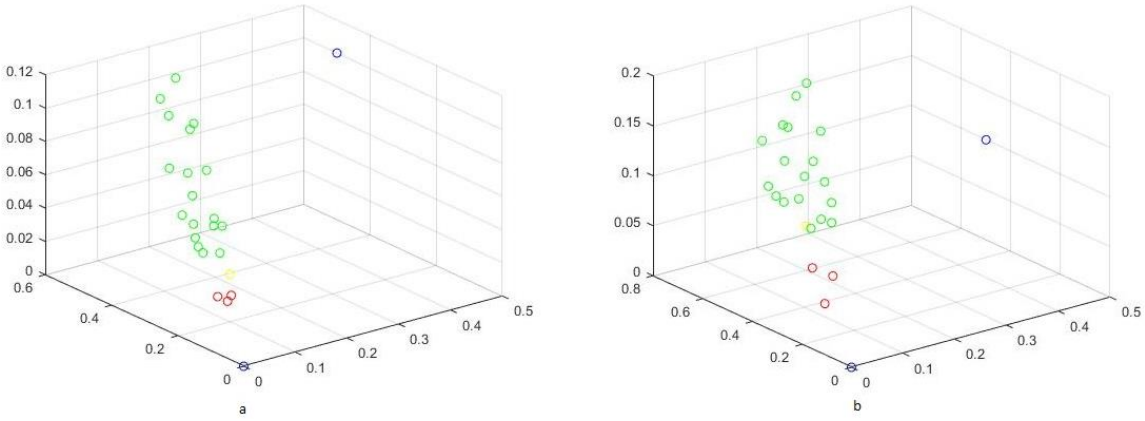
Skenario uji coba 1 dilakukan untuk mendapatkan metode penghitungan kemiripan yang paling cocok digunakan untuk ekstraksi fitur lokal dari masing-masing pernyataan kebutuhan dalam sebuah dokumen SKPL. Metode penghitungan kemiripan yang digunakan adalah *cosine similarity* dan *jaccard similarity*. Berikut adalah perbandingan performa yang didapatkan dari masing-masing metode.

Tabel 4.6 Hasil Uji Coba Skenario 1

Metode	Akurasi (%)	TP Rate (%)	TN Rate (%)	F1 Measure (%)
Jaccard similarity	81.23	63.41	83.24	40.63
Cosine similarity	82.22	68.29	83.79	43.75

Dari Tabel 4.6 terlihat bahwa metode *cosine similarity* memiliki performa yang lebih bagus dibandingkan dengan metode *jaccard similarity.* Keunggulan dari metode *cosine similarity* didapatkan karena metode ini menghitung kemiripan antara dua vektor berdasarkan nilai kosinus dari sudut yang dibentuk oleh kedua vektor tersebut, sedangkan *jaccard similarity* menghitung kemiripan dengan memanfaatkan perbandingan total nilai minimum dan total nilai maksimum dari kedua vektor tersebut. Berikut adalah contoh hasil perhitungan kemiripan masing-masing metode untuk penyataan kebutuhan “A table with login and password has been set up for each professor in the departments of the College of Engineering and is maintained in the local database” yang merupakan sebuah

pernyataan kebutuhan normal pada dokumen SKPL 24_srsv6-uji24 yang membahas tentang spesifikasi kebutuhan aplikasi sistem informasi akademik sebuah universitas.



Gambar 4.2 Contoh Hasil Metode (a) Jaccard Similarity dan (b) Cosine Similarity

Pada Gambar 4.2 (a) penyataan kebutuhan "*A table with login and password has been set up for each professor in the departments of the College of Engineering and is maintained in the local database*" yang diwakili oleh lingkaran kuning lebih dekat dengan lingkaran merah yang mewakili pernyataan kebutuhan yang tergolong derau dibandingkan dengan lingkaran hijau yang mewakili penyataan kebutuhan yang normal. Karenanya, model klasifikasi yang digunakan sistem salah mengklasifikasi pernyataan kebutuhan tersebut sebagai derau.

Disisi lain, pada Gambar 4.2 (b), pernyataan kebutuhan "*A table with login and password has been set up for each professor in the departments of the College of Engineering and is maintained in the local database*" yang diwakili oleh lingkaran warna kuning lebih dekat dengan lingkaran warna hijau yang mewakili penyataan kebutuhan yang normal dibandingkan dengan lingkaran merah yang mewakili pernyataan kebutuhan yang tergolong derau. Karenanya, model klasifikasi yang digunakan sistem berhasil melakukan klasifikasi pernyataan kebutuhan tersebut sebagai bukan derau (normal).

Dengan memperhatikan fakta-fakta diatas, maka dapat disimpulkan bahwa metode penghitungan kemiripan yang cocok untuk digunakan dalam sistem ini adalah metode *cosine similarity*.

4.2.2 Skenario 2

Skenario uji coba 2 dilakukan untuk mengetahui apakah fitur statistik yang digunakan sudah cukup untuk merepresentasikan fitur lokal dari masing-masing

pernyataan kebutuhan dalam dokumen SKPL. Uji coba dilakukan dengan menambahkan nilai minimum dan nilai varian dari vektor nilai kemiripan yang didapat secara bergantian dan secara bersamaan. Berikut adalah perbandingan performa sistem dengan variasi fitur yang digunakan.

Tabel 4.7 Hasil Uji Coba Skenario 2

Fitur	**Akurasi (%)**	**TP Rate (%)**	**TN Rate (%)**	**F1 Measure (%)**
Proposed	82.22	68.29	83.79	43.75
Proposed + Minimum	81.23	65.85	82.97	41.54
Proposed + Varian	81.48	68.29	82.97	42.75
Proposed + Varian dan Minimum	81.73	63.41	83.79	41.27

Dari Tabel 4.7 dapat disimpulkan bahwa fitur statistik yang diajukan (rata-rata, maksimum, dan standar deviasi) adalah kombinasi terbaik untuk merepresentasikan fitur lokal dari masing-masing pernyataan kebutuhan. Terlihat dari nilai akurasi, TP rate, TN rate, dan f1 measure yang lebih tinggi jika dibandingkan dengan kombinasi fitur yang lain.

Selain itu, jika dilihat dari nilai *information gain* dari kelima properti statistik yang diujikan pada skenario ini, 3 properti statistik yang menghasilkan performa klasifikasi tertinggi juga memiliki nilai *information gain* yang paling tinggi. *Information gain* menggambarkan berapa banyak informasi yang dibawa oleh masing-masing fitur dalam menentukan kelas atau label dari sebuah pernyataan kebutuhan. Semakin tinggi nilai *information gain* dari sebuah fitur, maka semakin tinggi pula peran fitur tersebut dalam menentukan hasil klasifikasinya. Berikut adalah nilai *information gain* dari masing-masing properti statistik yang digunakan pada skenario uji coba ini.

Tabel 4.8 Information Gain Semua Fitur

No	**Properti Statistik**	**Information Gain**
1	Mean of Similarity Measure	0.079
2	Maximum of Similarity Measure	0.060
3	Standar Deviation of Similarity Measure	0.078
4	Minumum of Similarity Measur	0.010
5	Variance of Similarity Measure	0.055

Dengan mempertimbangkan fakta diatas, maka kombinasi fitur yang akan digunakan oleh sistem ini adalah *Mean of Similarity Measure*, *Maximum of Similarity Measure*, serta *Standard Deviation of Similarity Measure*.

4.2.3 Skenario 3

Skenario uji coba 3 dilakukan untuk mengetahui metode klasifikasi yang paling cocok untuk digunakan pada kasus ini. Metode klasifikasi yang digunakan antara lain adalah SVM, neural network, dan kNN. Berikut adalah perbandingan performa yang didapat dari masing-masing metode.

Tabel 4.9 Hasil Uji Coba Skenario 3

Metode Klasifikasi	**Akurasi (%)**	**TP Rate (%)**	**TN Rate (%)**	**F1 Measure (%)**
SVM	82.22	68.29	83.79	43.75
Neural Netwok	89.63	9.76	98.63	16
K Nearest Neighbor	87.41	9.76	96.15	13.51

Dari Tabel 4.9 dapat dinyatakan bahwa metode neural network dan knn memiliki nilai akurasi dan presisi yang sedikit lebih baik jika dibandingkan dengan SVM. Namun karena data yang digunakan memiliki ketimpangan jumlah data antara data dengan kelas positif (data yang tergolong derau) dan kelas negatif (data yang bukan derau), perlu diperhatikan juga nilai dari *true positive rate* dan nilai f1 measurenya. Metode neural network dan knn memiliki performa yang bagus hanya pada data yang memiliki kelas negatif (normal) saja, sedangkan pada data yang memiliki kelas positif (derau) kedua metode ini memiliki performa yang sangat buruk. Hal ini terlihat pada ketimpangan nilai akurasi dan tn rate yang sangat tinggi jika dibandingkan dengan tp rate dan f1 measurenya. Disisi lain, svm memiliki nilai true positive rate dan f1 measure yang jauh lebih tinggi daripada metode neural network dan knn, sehingga dapat disimpulkan bahwa metode klasifikasi SVM memiliki keseimbangan yang lebih tinggi dalam memprediksi data kelas positif (derau) dan data kelas negatif (normal) dengan benar.

4.2.4 Skenario 4

Skenario uji coba 4 dilakukan untuk mengetahui komposisi perbandingan data derau yang tepat untuk digunakan dalam proses training sehingga mendapatkan model klasifikasi yang memiliki performa terbaik. Variasi perbandingan data derau diperoleh dengan memanfaatkan metode SMOTE untuk melakukan sintesis data derau. Berikut performa sistem dengan variasi perbandingan data derau.

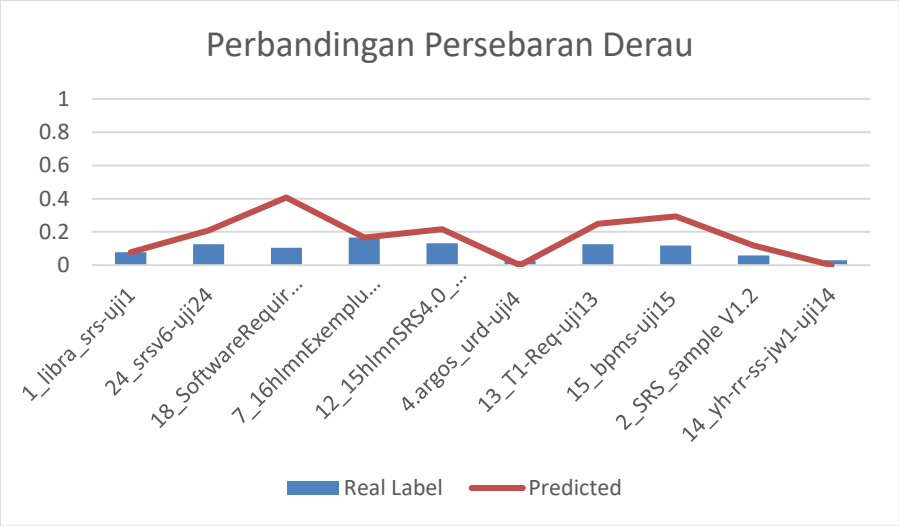
Tabel 4.10 Hasil Uji Coba Skenario 4

Persentase Data Derau (%)	**Akurasi (%)**	**TP Rate (%)**	**TN Rate (%)**	**F1 Measure (%)**
50	82.22	70.73	83.52	44.62
40	82.96	73.17	84.07	46.51
30	82.47	70.73	83.79	44.96
20	82.22	70.73	83.52	44.62
10	82.22	68.29	83.79	43.75

Dari Tabel 4.10 terlihat bahwa perubahan persentase data derau yang digunakan saat *training* memiliki pengaruh yang kecil pada performa klasifikasi. Hal ini disebabkan karena sintesa data yang dihasilkan oleh algoritma SMOTE hanya memiliki perbedaan yang sedikit dari data asli. Oleh sebab itulah data sintesa ini tidak memberikan dampak yang signifikan terhadap pembentukan model klasifikasi. Dengan mempertimbangkan nilai true positive rate yang relatif cukup tinggi dengan nilai performa lain yang tidak terlalu rendah, maka dapat disimpulkan bahwa persentase data derau yang baik untuk digunakan dalam penelitian ini adalah 40%.

4.3 Analisa Hasil Uji Coba

Dengan menggunakan algoritma dengan performa terbaik yang didapatkan dari semua skenario uji coba yang digunakan, maka didapatkan akurasi 82.96%, true positive rate 73.17%, true negative rate 84.07%, serta f1 measure 46.51%. Berikut adalah hasil pemetaan derau dari hasil klasifikasi yang dilakukan dengan menggunakan skema terbaik.



Gambar 4.3 Perbandingan Perserbaran Derau

Rata-rata perbedaan persentase derau hasil prediksi dengan data asli adalah 8.87% dengan perbedaan tertinggi terdapat pada dokumen 18_SoftwareRequirementsSpecification-uji18 yaitu sebesar 30.23%. Dokumen yang memiliki selisih persentase derau bernilai positif memiliki data false positive yang dominan, sedangkan dokumen yang memiliki selisih persentase derau bernilai negatif memiliki data false negative yang dominan. Pernyataan kebutuhan yang terdeteksi sebagai derau memiliki properti statistik kemiripan yang relatif rendah yaitu mean = 0.03 ± 0.02, max = 0.27 ± 0.1, std = 0.05 ± 0.02. Sedangkan pernyataan kebutuhan yang terdeteksi sebagai bukan derau memiliki properti statistik kemiripan yang relatif tinggi yaitu mean = 0.11 ± 0.06, max = 0.49 ± 0.2, std = 0.11 ± 0.04.

Berikut ini rincian performa sistem saat mendeteksi derau dari masing-masing dokumen SKPL yang digunakan dengan menggunakan skema terbaik yang didapatkan dari tahap uji coba.

Tabel 4.11 Performa Akhir Sistem

Nama File	**Akurasi (%)**	**TP Rate (%)**	**TN Rate (%)**	**F1 Measure (%)**
1_libra_srs-uji1	100	100	100	100
24_srsv6-uji24	91.67	100	90.48	75
18_SoftwareRequirementsSpecification-uji18	66.28	87.5	64.1	32.56

Nama File	Akurasi (%)	TP Rate (%)	TN Rate (%)	F1 Measure (%)
7_16hlmnExemplu_cerinte_software -uji7	100	100	100	100
12_15hlmnSRS4.0_uji12	89.62	92.86	89.13	70.27
4.argos_urd-uji4	97.44	0	100	NaN
13_T1-Req-uji13	71.88	37.5	76.79	25
15_bpms-uji15	70.59	50	73.33	28.57
2_SRS_sample V1.2	94.12	100	93.75	66.67
14_yh-rr-ss-jw1-uji14	96.97	0	100	NaN

4.3.1 Analisa Dokumen SKPL 1_libra_srs-uji1

Dokumen SKPL ini terdiri dari 13 pernyataan kebutuhan dengan 1 pernyataan yang tergolong derau. Dokumen ini memiliki performa yang bagus karena pernyataan kebutuhan yang tegolong derau memiliki properti statistik kemiripan yang relatif rendah jika dibandingkan dengan pernyataan kebutuhan lain. Hal ini menunjukkan bahwa pernyataan kebutuhan yang tergolong derau pada dokumen ini memiliki kemiripan yang rendah dengan pernyataan kebutuhan yang lain jika ditinjau dari kata penyusun serta frekuensi kata penyusunnya.

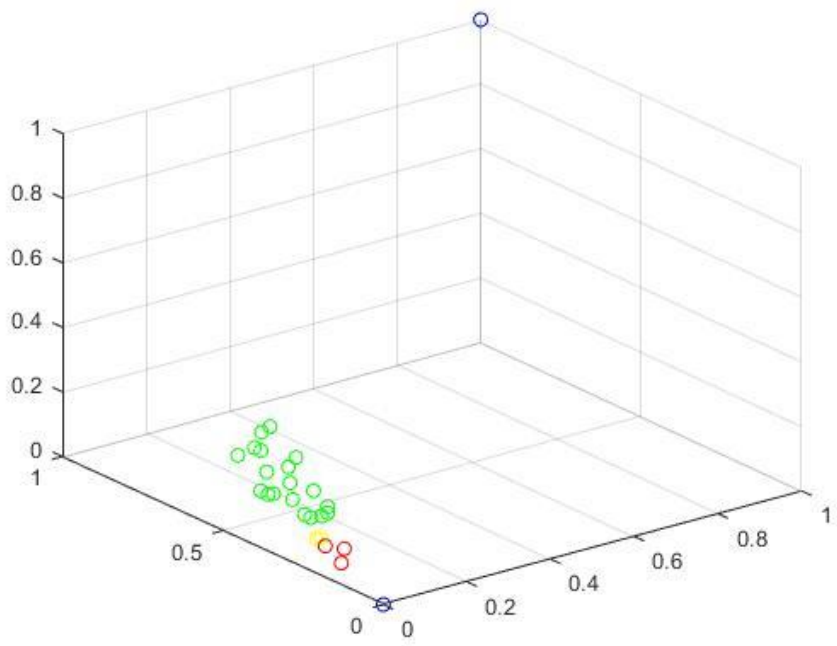
Pernyataan kebutuhan yang tergolong derau pada dokumen ini, "*check his credit balance*" hanya memiliki 1 kata yang sama dengan pernyataan kebutuhan "*check his usage history*", "*check the status of each node of the cluster", "check the usage pattern history of the cluster", "check the status of all submitted jobs",* dan "*check the load on each node of the cluster",* yaitu kata "*check*".

Di samping itu, jika ditinjau lebih mendalam dari sisi maknanya, pernyataan kebutuhan tersebut juga tidak memiliki keterkaitan dengan domain dokumen SKPL yang membahas tentang aplikasi penjadwalan pekerjaan.

4.3.2 Analisa Dokumen SKPL 24_srsv6_uji24

Dokumen SKPL ini terdiri dari 24 pernyataan kebutuhan dengan 3 pernyataan yang tergolong derau. 3 pernyataan yang tergolong derau pada dokumen ini, berhasil terdeteksi sebagai derau karena memiliki kemiripan yang rendah dengan pernyataan

kebutuhan yang lain jika ditinjau dari kata penyusun serta frekuensi kata penyusunnya. Selain itu, pada dokumen ini terdapat 2 pernyataan kebutuhan yang bukan derau akan tetapi terdeteksi sebagai derau (*false positive*). Pernyataan kebutuhan ini terdeteksi sebagai derau karena memiliki properti kemiripan yang hampir sama dengan pernyataan kebutuhan yang memang tergolong derau.



Gambar 4.4 Persebaran data Dokumen 24_srsv6-uji24

Dari Gambar 4.4 terlihat bahwa pernyataan kebutuhan yang bukan derau namun terdeteksi sebagai derau (lingkaran kuning) lebih dekat pada pernyataan kebutuhan yang merupakan derau (lingkaran merah) daripada pernyataan kebutuhan yang bukan merupakan derau (lingkaran hijau). Kedua pernyataan kebutuhan yang salah terdeteksi sebagai derau adalah "*The timeframe to upload data to the central database is the day that grades are due to the university and one week before the start of the new semester*" dan "*All data transmitted to the central database will be encrypted*".

Pernyataan kebutuhan "*The timeframe to upload data to the central database is the day that grades are due to the university and one week before the start of the new semester*" memiliki properti kemiripan mean = 0.05, max = 0.26, std = 0.07. Nilai properti kemiripan ini masih berada pada area properti kemiripan pernyataan kebutuhan yang tergolong derau (mean = 0.03 ± 0.02, max = 0.27 ± 0.1, std = 0.05 ± 0.02). Hal ini

menunjukkan bahwa pernyataan kebutuhan ini memiliki kemiripan yang rendah dengan pernyataan kebutuhan yang lain jika ditinjau dari kata penyusun serta frekuensi kata penyusunnya, karenanya sistem melabeli pernyataan kebutuhan ini sebagai derau.

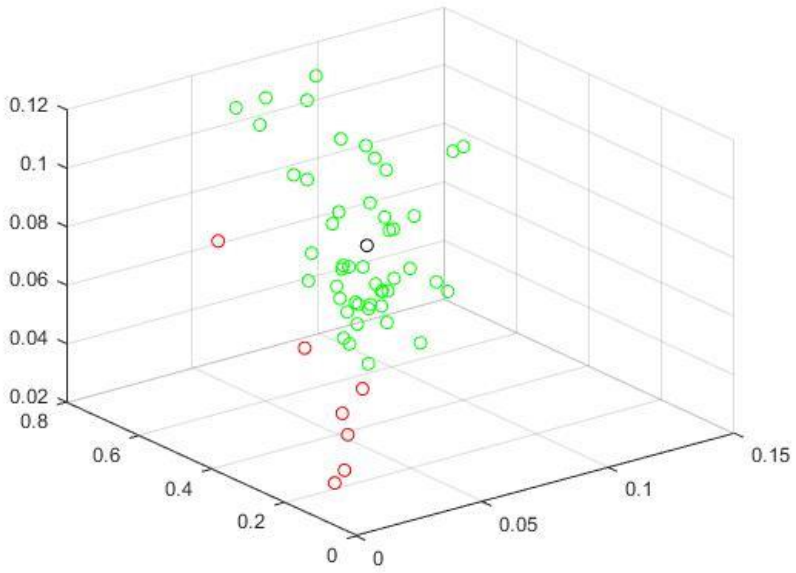
Akan tetapi jika ditinjau lebih dalam dari sisi maknanya, pernyataan kebutuhan ini masih berada dalam konteks yang dibahas dalam dokumen 24_srsv6_uji24, yaitu tentang Sistem Informasi Akademik sebuah Universitas, sehingga anotator melabeli pernyataan kebutuhan ini sebagai pernyataan kebutuhan yang normal.

Hal yang sama juga terjadi pada pernyataan kebutuhan "*All data transmitted to the central database will be encrypted.*". Pernyataan kebutuhan ini memiliki properti kemiripan mean = 0.04, max = 0.26, std = 0.07. Nilai properti kemiripan ini masih berada pada area properti kemiripan pernyataan kebutuhan yang tergolong derau (mean = 0.03 ± 0.02, max = 0.27 ± 0.1, std = 0.05 ± 0.02). Hal ini menunjukkan bahwa pernyataan kebutuhan ini memiliki kemiripan yang rendah dengan pernyataan kebutuhan yang lain jika ditinjau dari kata penyusun serta frekuensi kata penyusunnya, karenanya sistem melabeli pernyataan kebutuhan ini sebagai derau.

Akan tetapi jika ditinjau lebih dalam dari sisi maknanya, pernyataan kebutuhan ini masih berada dalam konteks yang dibahas dalam dokumen 24_srsv6_uji24, yaitu tentang Sistem Informasi Akademik sebuah Universitas, sehingga anotator melabeli pernyataan kebutuhan ini sebagai pernyataan kebutuhan yang normal.

4.3.3 Analisa Dokumen SKPL 18_SoftwareRequirementsSpecification-uji18

Pada dokumen yang membahas tentang spesifikasi kebutuhan perangkat lunak perpustakaan online pada sebuah universitas ini, terdapat satu pernyataan kebutuhan yang tergolong derau namun terdeteksi sebagai pernyataan kebutuhan yang bukan derau (*false negative*).



Gambar 4.5 Persebaran data *false negative* pada Dokumen 18_SoftwareRequirementsSpecification-uji18

Pada Gambar 4.5 terlihat dengan jelas bahwa pernyataan kebutuhan yang tergolong derau tapi terdeteksi sebagai normal (lingkaran hitam) memiliki kedekatan dengan data-data pernyataan kebutuhan yang tergolong normal (lingkaran hijau).

Pernyataan kebutuhan "*The university information security system must be compatible with the Internet applications*" memiliki properti kemiripan mean = 0.06, max = 0.38, std = 0.08. Nilai properti kemiripan ini masih berada pada area properti kemiripan pernyataan kebutuhan yang normal (mean = 0.11 ± 0.06, max = 0.49 ± 0.2, std = 0.11 ± 0.04). Hal ini menunjukkan bahwa pernyataan kebutuhan ini memiliki kemiripan yang tinggi dengan pernyataan kebutuhan yang lain jika ditinjau dari kata penyusun serta frekuensi kata penyusunnya, karenanya sistem melabeli pernyataan kebutuhan ini sebagai pernyataan kebutuhan yang normal.

Akan tetapi jika ditinjau lebih dalam dari sisi maknanya, pernyataan kebutuhan ini sudah berada diluar konteks yang dibahas dalam dokumen 18_SoftwareRequirementsSpecification-uji18, yaitu tentang perpustakaan online sebuah universitas, sehingga anotator melabeli pernyataan kebutuhan ini sebagai derau.

Dalam dokumen ini juga terdapat 28 pernyataan kebutuhan bukan derau yang terdeteksi sebagai derau (*false positive*). Berikut adalah 3 contoh dari 28 pernyataan kebutuhan yang tergolong *false positive* pada dokumen ini.

Tabel 4.12 Sampel *False Positive* pada Dokumen 18_SoftwareRequirementsSpecification-uji18

No	Pernyataan Kebutuhan	Mean	Max	Std
1	The system shall be comply with the TCP/IP protocol standards and shall be designed accordingly.	0.03	0.20	0.05
2	The product can access the university student database	0.03	0.33	0.06
3	The system can alert the Librarian or the administrator in case of any problems.	0.03	0.27	0.04

Pernyataan-pernyataan kebutuhan pada Tabel 4.12 memiliki nilai properti kemiripan yang masih berada pada area properti kemiripan pernyataan kebutuhan yang tergolong derau (mean = 0.03 ± 0.02, max = 0.27 ± 0.1, std = 0.05 ± 0.02). Hal ini menunjukkan bahwa pernyataan-pernyataan kebutuhan tersebut memiliki kemiripan yang rendah dengan pernyataan kebutuhan yang lain jika ditinjau dari kata penyusun serta frekuensi kata penyusunnya, karenanya sistem melabeli pernyataan-pernyataan kebutuhan tersebut sebagai derau.

Akan tetapi jika ditinjau lebih dalam dari sisi maknanya, pernyataan-pernyataan kebutuhan tersebut masih memiliki keterkaitan dengan konteks sistem perpustakaan online, sehingga annotator melabeli pernyataan-pernyataan kebutuhan tersebut sebagai pernyataan kebutuhan yang normal.

4.3.4 Analisa Dokumen SKPL 4.argos_urd-uji4

Pada dokumen yang terdiri dari 39 pernyataan kebutuhan ini, terdapat 1 pernyataan kebutuhan yang tergolong derau. Pernyataan kebutuhan yang tergolong derau ini dideteksi sebagai pernyataan kebutuhan yang normal oleh sistem (*false negative*). Pernyataan kebutuhan "*A role can chat with other person with the same role*" memiliki nilai properti kemiripan mean = 0.14, max = 0.44, std = 0.11. Nilai properti kemiripan ini masih berada pada area properti kemiripan pernyataan kebutuhan yang normal (mean = 0.11 ± 0.06, max = 0.49 ± 0.2, std = 0.11 ± 0.04). Hal ini menunjukkan bahwa pernyataan kebutuhan ini memiliki kemiripan yang tinggi dengan pernyataan kebutuhan yang lain jika ditinjau dari kata penyusun serta frekuensi kata penyusunnya, karenanya sistem melabeli pernyataan kebutuhan ini sebagai pernyataan kebutuhan yang normal.

Akan tetapi jika ditinjau lebih dalam dari sisi maknanya, pernyataan kebutuhan ini sudah berada diluar konteks yang dibahas dalam dokumen 4.argos_urd-uji4, yaitu tentang arsip digital sebuah universitas, sehingga anotator melabeli pernyataan kebutuhan ini sebagai derau.

4.3.5 Analisa Umum

Dari penjelasan sampel beberapa dokumen pada subbab sebelumnya, dapat disimpulkan bahwa kesalahan klasifikasi yang dilakukan oleh sistem terjadi karena gagalnya sistem dalam memperhatikan makna dari sebuah penyataan kebutuhan. Hal ini terjadi karena metode pembobotan TF-iDF tidak memperhatikan informasi semantik dari masing-masing term-nya.

Pada pernyataan kebutuhan yang tergolong *false positive*, sistem gagal mendeteksi kesamaan makna pernyataan kebutuhan tersebut dengan pernyataan kebutuhan yang lain yang bukan derau. Sebagai contoh, pernyataan kebutuhan "*The timeframe to upload data to the central database is the day that grades are due to the university and one week before the start of the new semester* " memiliki kemiripan makna dengan pernyataan kebutuhan lain yang bukan derau seperti "*Manage course information: the general user can use the system to enter and modify course information. The local database maintains the information.*" dan "*Manage student information: the general user can use the system to enter and modify student information, or to download student information from the university database. In addition, the user can delete a student from a course. The local database maintains the information.* " yang membahas tentang kebutuhan perangkat lunak sebuah sistem informasi akademik. Akan tetapi karena pernyataan kebutuhan "*The timeframe to upload data to the central database is the day that grades are due to the university and one week before the start of the new semester* " memiliki term-term yang unik, maka nilai kemiripan dari bobot TF-iDF nya dengan pernyataan kebutuhan yang lain tergolong rendah. Rendahnya nilai kemiripan inilah yang membuat sistem melabeli pernyataan kebutuhan ini sebagai pernyataan kebutuhan yang tergolong derau.

Pada pernyataan kebutuhan yang tergolong *false negative*, sistem gagal mendeteksi perbedaan makna pernyataan kebutuhan tersebut dengan pernyataan kebutuhan lain yang bukan derau. Sebagai contoh, pernyataan kebutuhan "*The university*

information security system must be compatible with the Internet applications" memiliki kata penyusun yang mirip dengan pernyataan kebutuhan "*The information of all the users must be stored in a database that is accessible by the Online Library System.*", "*The Online Library System is connected to the university computer and is running all 24 hours a day.*", dan "*The users access the Online Library System from any computer that has Internet browsing capabilities and an Internet connection.*". Akan tetapi jika diperhatikan dari maknanya, pernyataan kebutuhan "*The university information security system must be compatible with the Internet applications*" sama sekali tidak berhubungan dengan sistem perpustakaan online yang menjadi domain masalahnya.

BAB 5
PENUTUP

Berdasarkan skenario pengujian yang telah dilakukan, dapat ditarik beberapa kesimpulan dan saran penelitian yang akan dilakukan selanjutnya.

5.1 Kesimpulan

Berdasarkan hasil uji coba yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut:

a. Ekstraksi fitur lokal dari pernyataan kebutuhan dalam dokumen SKPL dapat dilakukan dengan cara menghitung kemiripan antar pernyataan kebutuhan dan kemudian dihitung fitur statistiknya.
b. Fitur statistik yang dapat digunakan untuk merepresentasikan masing-masing pernyataan kebutuhan untuk deteksi derau pada dokumen SKPL adalah nilai rata-rata, nilai maksimum, dan standar deviasi dari kemiripan masing-masing pernyataan kebutuhan dalam dokumen SKPL yang sama. Performa terbaik didapat dengan menggunakan metode kemiripan *Cosine Similarity* dan metode klasifikasi SVM.
c. Performa maksimal didapat jika data latih yang digunakan memiliki keseimbangan antara data yang tergolong derau dengan yang tidak.

5.2 Saran

Beberapa saran yang dapat dilakukan untuk penelitian selanjutnya adalah sebagai berikut:

a. Mempertimbangkan jarak antar kata dalam melakukan pembobotan agar bisa membedakan dua kalimat yang memiliki kata-kata penyusun yang sama dengan letak yang berbeda.
b. Memanfaakan informasi semantik kata untuk mengidentifikasi perbedaan konteks dari kalimat.
c. Memanfaatkan thesaurus untuk mengidentifikasi 2 kata yang berbeda akan tetapi memiliki makna yang sama.

[Halaman ini sengaja dikosongkan]

DAFTAR PUSTAKA

[1] B. Meyer, "On Formalism in Specification," *IEEE Software,* 1985.

[2] D. M. Berry and E. Kamsties, "Ambiguity in Requirements Specification," in *Perspectives on Software Requirements*, Kluwer Academic Publishers, 2004, pp. 7-44.

[3] H. Xu and B. Yu, "Automatic thesaurus construction for spam filtering using revised back propagation neural network," *Expert Systems with Applications,* no. 37, pp. 18-23, 2010.

[4] B. Trstenjak, S. Mikac and D. Donko, "KNN with TF-IDF Based Framework for Text Categorization," *Procedia Engineering,* no. 69, pp. 1356-1364, 2014.

[5] B. Chandra and M. Gupta, "An efficient statistical feature selection approach for classification of gene," *Journal of Biomedical Informatics,* vol. 44, pp. 529-535, 2011.

[6] W. Zhou, L. Yu, W. Qiu, Y. Zhou and M. Wu, " Local Gradient Patterns (LGP): An Effective Local-Statistical-Feature Extraction Scheme for No-Reference Image Quality Assessment," *Information Sciences,* Vols. 397-398, pp. 1-14, 2017.

[7] D. Siahaan, Analisa Kebutuhan Dalam Rekayasa Perangkat Lunak, Jogjakarta: Andi Publisher, 2012.

[8] A. Singhal, "Modern Information Retrieval: A Brief Overview," Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2001.

[9] J. C. Platt, "Sequential Minimal Optimization : A Fast Algorithm for Training Support Vector Machines," Technical Report MSR-TR-98-14, 1998.

[10] N. V. Chawla, K. W. Bowyer, L. O. Hall and P. W. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence,* no. 16, pp. 321-357, 2002.

[11] D. M. W. Powers, "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation," School of Informatics and Engineering Flinders University, Adelaide, 2007.

LAMPIRAN

Berikut ini adalah hasil deteksi derau dengan menggunakan skema terbaik yang didapatkan dari semua skenario uji coba yang dilakukan.

Tabel 7.1 Hasil Deteksi Derau dengan Skenario Terbaik

No Dokumen	Pernyataan kebutuhan	Mean	Maximum	Standard Deviation	Label	Prediction
1	submit jobs with the associated deadline, cost, and execution time	0.11	0.32	0.12	0	0
1	query the cluster to establish the current cost per unit time for submitting new jobs	0.12	0.32	0.09	0	0
1	monitor the status of submitted jobs	0.15	0.67	0.20	0	0
1	cancel jobs submitted by him	0.15	0.51	0.20	0	0
1	check his credit balance	0.06	0.17	0.07	1	1
1	check his usage history	0.12	0.76	0.21	0	0
1	check the status of each node of the cluster	0.20	0.64	0.19	0	0
1	check the usage pattern history of the cluster	0.15	0.76	0.20	0	0
1	check the status of all submitted jobs	0.24	0.67	0.20	0	0
1	check the load on each node of the cluster	0.14	0.64	0.17	0	0
1	alter the cost structure of the cluster	0.09	0.38	0.11	0	0
1	alter the scheduling policy of the cluster	0.07	0.38	0.11	0	0
1	cancel, suspend, and resume any job	0.08	0.51	0.14	0	0
2	Manage course information: the general user can use the system to enter and modify course information. The local database maintains the information.	0.22	0.64	0.18	0	0
2	Manage student information: the general user can use the system to enter and modify student information, or to download student information from the university database. In addition, the user can delete a student from a course. The local database maintains the information.	0.20	0.64	0.17	0	0
2	Enter outcomes: the general user can use the system to enter course outcomes. The user can assign a ranking to each course outcome. The local database maintains the information.	0.19	0.52	0.15	0	0
2	Enter targeted competency: the general user can use the system to select one of two types of targeted competencies that can be used to automatically check progress toward achieving course outcomes the user can enter values associated with the different types. The user can enter a description of a targeted competency that can be used for manual checking. The local database maintains the information.""	0.12	0.33	0.09	0	0
2	Map entities: the general user can use the system to map course outcomes to program outcomes to map AIUs to course outcomes and to map course outcomes to targeted competencies. The local database maintains the information.""	0.12	0.52	0.13	0	0
2	Define AIs: the general user can use the system to enter information related to an AI. For each instrument, the stored information includes such items as description, category, assessment instrument part (AIP) label and description, AIU	0.12	0.33	0.11	0	0

	label and description, AIU point value. For each student, the user can enter the points earned for each assessment unit. The local database maintains the information.					
2	Modify AIs: the general user can use the system to modify information related to an AI. The local database maintains the information.	0.16	0.59	0.15	0	0
2	Enter grades: the general user can use the system to enter student grades related to an AI and to enter formulae for calculating grades. The local database maintains the information.	0.16	0.61	0.15	0	0
2	Archive course and student information: the general user can use the system to archive course and student information to the central database from the local database at the end of the semester.	0.18	0.53	0.12	0	0
2	Access the system: the general user can access the system through a unique login name and password.	0.13	0.46	0.12	0	0
2	Open course: the general user can select a course from the list of courses that s/he owns.	0.09	0.31	0.10	0	0
2	Set password - The user can set the password for accessing and entering information on the course.	0.14	0.40	0.12	0	0
2	Set property list The user can set the property list that gives the login names for the users who can access a course.	0.09	0.40	0.11	0	0
2	Enter formula: the general user can define the formulas that will be used to calculate students grades on an AI, assessment category, or course.	0.12	0.61	0.14	0	0
2	The system will not be accessible to unauthorized users.	0.09	0.46	0.10	0	0
2	All data transmitted to the central database will be encrypted.	0.04	0.26	0.07	0	1
2	The system will be completed by the end of April 2001.	0.03	0.17	0.04	1	1
2	The university database is the Banner system.	0.06	0.20	0.06	1	1
2	Each professor has access to a desktop computer.	0.03	0.22	0.07	1	1
2	A table with login and password has been set up for each professor in the departments of the College of Engineering and is maintained in the local database.	0.08	0.35	0.09	0	0
2	The initial login and password for general users will be encrypted.	0.08	0.33	0.09	0	0
2	The timeframe to upload data to the central database is the day that grades are due to the university and one week before the start of the new semester.	0.05	0.26	0.07	0	1
2	Program outcomes that include ABET outcomes a-k and additional program outcomes are stored in the local database.	0.06	0.46	0.12	0	0
2	Once a student is dropped, the information associated with the student cannot be removed or modified.	0.07	0.45	0.11	0	0
3	The product provides the members with online blocking of books capabilities and the Online Library System is up and running all day.	0.10	0.45	0.10	0	0
3	The system provides logon facility to the users.	0.06	0.62	0.09	0	0
3	The system provides the members with the option to check their account and/or change their options like password of the account whenever needed all through the day during the library hours.	0.05	0.34	0.06	0	0
3	The system allows the members to block the books 24 hours a day and all the through the semester.	0.06	0.41	0.10	0	0
3	The system lets the library staff to check which all members have blocked the books and whether they can borrow any more books or not.	0.06	0.35	0.09	0	0

3	The system allows the Librarian to create the books catalog, add/delete books and maintain the books catalog.	0.06	0.72	0.11	0	0
3	The system updates the billing system as and when the member borrows or returns a book.	0.08	0.51	0.11	0	0
3	The book catalog is automated and the decision of offering the book based on the category of the book is automatically decided.	0.03	0.41	0.08	0	1
3	We also have an order department, which manages to add or remove a book from the Library.	0.04	0.25	0.05	0	1
3	The member should be provided with the updated information about the books catalog.	0.07	0.44	0.11	0	0
3	Provisions for the members to borrow the books they want, if all the other required rules hold good.	0.03	0.26	0.07	0	1
3	The member is given a provision to check his account information and change the account information any time in the given valid period.	0.03	0.34	0.06	1	1
3	The members are provided with the books available roster and allowed to choose the books, which they want to use in the coming up days.	0.07	0.37	0.09	0	0
3	The librarian can get the information about the members who have borrowed or returned the books.	0.05	0.51	0.10	0	0
3	The librarian is provided with interfaces to add/delete the books available in the book catalog.	0.07	0.72	0.11	0	0
3	The members when complete the book borrowing or returning process, the due to be paid by the member must be calculated and the information about the member and the due amount is sent to the university billing system.	0.06	0.46	0.09	0	0
3	The system uses the University information security requirements to provide the login facility to the users.	0.08	0.47	0.08	0	0
3	The information of all the users must be stored in a database that is accessible by the Online Library System.	0.09	0.44	0.08	0	0
3	The university information security system must be compatible with the Internet applications.	0.06	0.38	0.08	1	0
3	The Online Library System is connected to the university computer and is running all 24 hours a day.	0.08	0.45	0.10	0	0
3	The users access the Online Library System from any computer that has Internet browsing capabilities and an Internet connection.	0.08	0.71	0.11	1	0
3	The billing system is connected to the Online Library System and the database used by the billing system must be compatible with the interface of the Online Library System.	0.11	0.44	0.10	0	0
3	The users must have their correct usernames and passwords to enter into the Online Library System.	0.06	0.31	0.06	0	0
3	The users have sufficient knowledge of computers.	0.02	0.21	0.04	1	1
3	The University computer should have Internet connection and Internet server capabilities.	0.04	0.71	0.11	0	0
3	The users know the English language, as the user interface will be provided in English	0.04	0.31	0.06	0	0
3	The product can access the university student database	0.03	0.33	0.06	0	1
3	The system shall provide the users with logon capabilities.	0.08	0.62	0.10	0	0
3	The Online Library System is also supported on mobile devices such as cell phones.	0.05	0.34	0.07	0	0
3	The system can alert the Librarian or the administrator in case of any problems.	0.03	0.27	0.04	0	1

3	The system shall allow the users to access the system from the Internet using HTML or it&#65533;s derivative technologies.	0.08	0.34	0.07	0	0
3	The system uses a web browser as an interface.	0.05	0.42	0.07	0	0
3	Since all users are familiar with the general usage of browsers, no specific training is required.	0.02	0.18	0.04	1	1
3	The system is user friendly and self-explanatory.	0.03	0.12	0.03	0	1
3	The system has to be very reliable due to the importance of data and the damages incorrect or incomplete data can do.	0.02	0.22	0.04	0	1
3	The system is available 100% for the user and is used 24 hrs a day and 365 days a year.	0.06	0.43	0.07	0	0
3	The system shall be operational 24 hours a day and 7 days a week.	0.04	0.43	0.08	0	0
3	The system will be developed in such a way that it may fail once in a year.	0.03	0.21	0.04	0	1
3	Even if the system fails, the system will be recovered back up within an hour or less.	0.04	0.20	0.04	0	1
3	The accuracy of the system is limited by the accuracy of the speed at which the employees of the library and users of the library use the system.	0.06	0.26	0.06	0	0
3	The system shall provide 100% access reliability.	0.07	0.33	0.08	0	0
3	The Splash Page or Information page should be able to be downloaded within a minute using a 56K modem.	0.02	0.26	0.05	0	1
3	The information is refreshed every two minutes.	0.02	0.22	0.05	0	1
3	The access time for a mobile device should be less than a minute.	0.03	0.34	0.07	0	1
3	The system shall respond to the member in not less than two seconds from the time of the request submittal.	0.04	0.27	0.05	0	0
3	The system shall be allowed to take more time when doing large processing jobs.	0.04	0.32	0.06	0	0
3	The system shall take as less time as possible to provide service to the administrator or the librarian.	0.06	0.32	0.07	0	0
3	The number of transactions is directly dependent on the number of users, the users may be the Librarian, employees of the Library and also the people who use the Library for checking-out books, returning books and checking online library account.	0.08	0.28	0.07	0	0
3	The system is capable of handling 250 users at a time.	0.06	0.38	0.06	0	0
3	The resources are modified according the user requirements and also according to the books requested by the users.	0.04	0.16	0.05	0	1
3	The system designers shall take in to considerations the following supportability and technical limitations.	0.03	0.21	0.05	1	1
3	The system shall be comply with the TCP/IP protocol standards and shall be designed accordingly.	0.03	0.20	0.05	0	1
3	The system shall support the UHCL information security requirements and use the same standard as the UHCL information security requirements.	0.05	0.47	0.07	0	0
3	The member balance amount that will be calculated and sent to the billing system shall be compatible with the data types and design constraints of the billing system.	0.05	0.39	0.07	0	0
3	The maintenance of the system shall be done as per the maintenance contract.	0.03	0.31	0.04	0	1

3	The coding standards and naming conventions will be as per the American standards.	0.01	0.21	0.04	0	1
3	The languages that shall be used for coding the Online Library System are Active Server Pages (ASP), Java Servlets, Java Server Pages (JSP), HTML, JavaScript, and VBScript.	0.06	0.48	0.09	0	0
3	For working on the coding phase of the Online Library System, the Internet Information Services (IIS) Server needs to be installed.	0.07	0.36	0.07	0	0
3	Will make use of the available Java Development Tool kits for working with Java Beans and Java Server Pages.	0.03	0.48	0.08	0	1
3	Also will make use of the online references available for developing programs in ASP, HTML and the two scripting languages, JavaScript and VBScript.	0.05	0.60	0.08	0	1
3	Will make use of the existing Java libraries available for JSP and Servlets.	0.05	0.44	0.08	0	0
3	Also we need to develop some new libraries for the web-based application.	0.05	0.35	0.07	0	0
3	Also will develop new programs using ASP and scripting languages.	0.04	0.60	0.08	0	1
3	Online help is provided for each of the feature available with the Online Library System.	0.09	0.72	0.12	0	0
3	All the applications provide an on-line help system to assist the user.	0.06	0.35	0.07	0	0
3	The nature of these systems is unique to application development as they combine aspects of programming (hyperlinks, etc) with aspects of technical writing (organization, presentation).	0.02	0.15	0.03	1	1
3	Online help is provided for each and every feature provided by the system.	0.07	0.72	0.10	0	0
3	The User Manual describes the use of the system to Librarian and Employees.	0.06	0.46	0.07	0	0
3	It describes the use of the system on mobile systems.	0.06	0.46	0.07	0	0
3	The user manual should be available as a hard copy and also as online help.	0.05	0.39	0.07	0	0
3	An installation document will be provided that includes the installation instructions and configuration guidelines, which is important to a full solution offering.	0.02	0.33	0.05	0	1
3	Also, a Read Me file is typically included as a standard component.	0.02	0.30	0.05	0	1
3	The Read Me includes a ""What's New With This Release"" section	0.01	0.19	0.03	0	1
3	Most users also appreciate documentation defining any known bugs and workarounds in the Read Me file.	0.02	0.30	0.04	0	1
3	Since the installation of Online Library System is a complex process, our experts will do it.	0.05	0.21	0.06	0	1
3	So an installation Guide will not be provided to the user.	0.05	0.33	0.07	0	0
3	The System Administrator will need to purchase the license for IIS Server.	0.04	0.62	0.08	0	1
3	Mostly it is available with Windows Environment.	0.01	0.12	0.03	1	1
3	So the system need not purchase any licensing products.	0.04	0.62	0.08	1	1
3	Will make use of the existing Web Browsers such as Microsoft Internet Explorer or Netscape.	0.03	0.42	0.07	0	1
3	The user-interface of the system shall be designed as shown in the user-interface prototypes.	0.06	0.33	0.07	0	0
3	The existing Local Area Network (LAN) will be used for collecting data from the users and also for updating the Library Catalogue.	0.04	0.18	0.04	0	1

3	A firewall will be used with the server to prevent unauthorized access to the system.	0.04	0.20	0.05	0	1
3	The Online Library System will be connected to the World Wide Web.	0.06	0.33	0.07	0	0
3	The usage is restricted to only University of Houston-Clearlake Library who is purchasing the Online Library System from Library InfoSys and signs the maintenance contract.	0.06	0.31	0.07	0	0
3	Online Library System is a trademark of Library InfoSys and cannot be used without its consent.	0.05	0.31	0.07	0	0
4	User can see detailed information about SGP	0.13	0.27	0.09	0	0
4	User can apply for SGP	0.17	0.30	0.10	0	0
4	User can application update form for SGP	0.22	0.30	0.09	0	0
4	User can update personal data	0.11	0.30	0.09	0	0
4	User can upload application related documents	0.10	0.25	0.08	0	0
4	User can perform psychological test	0.07	0.09	0.01	1	1
5	The system shall display detailed product categorization to the user.	0.09	0.43	0.07	0	0
5	The system shall enable user to enter the search text on the screen.	0.07	0.53	0.10	0	0
5	The system shall enable user to select multiple options on the screen to search.	0.08	0.53	0.10	0	0
5	The system shall display all the matching products based on the search.	0.08	0.54	0.08	0	0
5	The system shall display only 10 matching result on the current screen.	0.06	0.37	0.08	0	0
5	The system shall enable user to navigate between the search results.	0.07	0.41	0.08	0	0
5	The system shall notify the user when no matching product is found on the search.	0.07	0.54	0.08	0	0
5	The system shall allow user to create profile and set his credential.	0.07	0.45	0.08	0	0
5	The system shall authenticate user credentials to view the profile.	0.06	0.45	0.07	0	0
5	The system shall allow user to update the profile information.	0.10	0.39	0.10	0	0
5	The system shall display both the active and completed order history in the customer profile.	0.08	0.44	0.07	0	0
5	The system shall allow user to select the order from the order history.	0.11	0.66	0.14	0	0
5	The system shall display the detailed information about the selected order.	0.11	0.55	0.12	0	0
5	The system shall display the most frequently searched items by the user in the profile.	0.08	0.28	0.07	0	0
5	The system shall allow user to register for newsletters and surveys in the profile.	0.06	0.34	0.07	0	0
5	The system shall provide online help, FAQ&#65533;s customer support, and sitemap options for customer support.	0.07	0.40	0.09	0	0
5	The system shall allow user to select the support type he wants.	0.09	0.39	0.09	0	0
5	The system shall allow user to enter the customer and product information for the support.	0.13	0.58	0.11	0	0
5	The system shall display the customer support contact numbers on the screen.	0.08	0.51	0.09	0	0
5	The system shall allow user to enter the contact number for support personnel to call.	0.07	0.51	0.09	0	0
5	The system shall display the online help upon request.	0.06	0.63	0.08	0	0
5	The system shall display the FAQ&#65533;s upon request.	0.06	0.63	0.08	0	0

5	The system shall maintain customer email information as a required part of customer profile.	0.06	0.31	0.07	0	0
5	The system shall send an order confirmation to the user through email.	0.07	0.37	0.07	0	0
5	The system shall display detailed invoice for current order once it is confirmed.	0.08	0.48	0.10	0	0
5	The system shall optionally allow user to print the invoice.	0.08	0.42	0.08	0	0
5	The system shall provide shopping cart during online purchase.	0.05	0.44	0.07	0	0
5	The system shall allow user to add/remove products in the shopping cart.	0.07	0.44	0.07	0	0
5	The system shall display different shipping options provided by shipping department.	0.07	0.46	0.08	0	0
5	The system shall enable user to select the shipping method during payment process.	0.09	0.60	0.13	0	0
5	The system shall display the shipping charges.	0.07	0.46	0.08	0	0
5	The system shall display tentative duration for shipping.	0.06	0.38	0.07	0	0
5	The system shall allow user to enter the order information for tracking.	0.11	0.62	0.13	0	0
5	The system shall display the current tracking information about the order.	0.09	0.62	0.11	0	0
5	The system shall calculate tax for the order.	0.07	0.72	0.10	0	0
5	The system shall display tax information for the order.	0.10	0.61	0.12	0	0
5	The system shall display available payment methods for payment.	0.08	0.73	0.12	0	0
5	The system shall allow user to select the payment method for order.	0.13	0.63	0.15	0	0
5	The system shall display the orders that are eligible to change.	0.08	0.51	0.09	0	0
5	The system shall allow user to select the order to be changed.	0.12	0.66	0.15	0	0
5	The system shall allow user to cancel the order.	0.10	0.49	0.10	0	0
5	The system shall allow user to change shipping, payment method.	0.10	0.63	0.13	0	0
5	The system shall notify the user about any changes made to the order.	0.08	0.52	0.09	0	0
5	The system shall display the reviews and ratings of each product, when it is selected.	0.08	0.61	0.08	0	0
5	The system shall enable the user to enter their reviews and ratings.	0.07	0.61	0.10	0	0
5	The system shall display all the available financing options.	0.08	0.62	0.10	0	0
5	The system shall allow user to select the financing option.	0.11	0.62	0.13	0	0
5	The system shall notify the use about the financing request.	0.06	0.33	0.07	0	0
5	The system shall allow user to view detailed sitemap.	0.08	0.36	0.08	0	0
5	The system shall display all the available promotions to the user.	0.10	0.76	0.10	0	0
5	The system shall allow user to select available promotion.	0.11	0.76	0.12	0	0
5	The system shall allow user to confirm the purchase.	0.09	0.37	0.09	0	0
5	The system shall enable user to enter the payment information.	0.10	0.51	0.12	0	0
5	The system shall provide a uniform look and feel between all the web pages.	0.04	0.23	0.04	0	1
5	The system shall provide a digital image for each product in the product catalog.	0.06	0.22	0.06	0	1

5	The system shall provide use of icons and toolbars.	0.05	0.25	0.05	0	0
5	The system shall provide handicap access.	0.06	0.40	0.06	0	0
5	The system shall provide multi language support.	0.05	0.33	0.06	0	0
5	The system shall provide storage of all databases on redundant computers with automatic switchover.	0.05	0.35	0.06	0	0
5	The system shall provide for replication of databases to off-site storage locations.	0.04	0.35	0.05	0	0
5	The system shall provide RAID V Disk Stripping on all database storage disks.	0.03	0.29	0.04	0	0
5	The system shall provide a contractual agreement with an internet service provider for T3 access with 99.9999% availability.	0.06	0.77	0.09	0	0
5	The system shall provide a contractual agreement with an internet service provider who can provide 99.999% availability through their network facilities onto the internet.	0.05	0.77	0.09	0	0
5	The product shall be based on web and has to be run from a web server.	0.04	0.33	0.06	0	0
5	The product shall take initial load time depending on internet connection strength which also depends on the media from which the product is run.	0.03	0.33	0.05	0	1
5	The performance shall depend upon hardware components of the client/customer.	0.03	0.21	0.05	0	1
5	The system shall use secure sockets in all transactions that include any confidential customer information.	0.05	0.32	0.06	0	0
5	The system shall automatically log out all customers after a period of inactivity.	0.04	0.20	0.04	0	1
5	The system shall confirm all transactions with the customer&#65533;s web browser.	0.06	0.31	0.08	0	0
5	The system shall not leave any cookies on the customer&#65533;s computer containing the user&#65533;s password.	0.06	0.79	0.09	0	0
5	The system shall not leave any cookies on the customer&#65533;s computer containing any of the user&#65533;s confidential information.	0.07	0.79	0.09	0	0
5	The customer&#65533;s web browser shall never display a customer&#65533;s password. It shall always be echoed with special characters representing typed characters.	0.05	0.37	0.06	0	0
5	The customer&#65533;s web browser shall never display a customer&#65533;s credit card number after retrieving from the database. It shall always be shown with just the last 4 digits of the credit card number.	0.05	0.30	0.06	0	0
5	The system&#65533;s back-end servers shall never display a customer&#65533;s password. The customer&#65533;s password may be reset but never shown.	0.05	0.37	0.07	0	0
5	The system&#65533;s back-end servers shall only be accessible to authenticated administrators.	0.04	0.41	0.06	0	1
5	The system&#65533;s back-end databases shall be encrypted.	0.04	0.41	0.05	0	0
5	The source code developed for this system shall be maintained in configuration management tool.	0.03	0.28	0.04	0	1
5	The system shall be built using a standard web page development tool that conforms to either IBM&#65533;s CUA standards or Microsoft&#65533;s GUI standards.	0.03	0.36	0.04	0	1
5	There are no memory requirements.	0.01	0.21	0.03	1	1
5	The computers must be equipped with web browsers such as Internet explorer.	0.02	0.30	0.06	1	1
5	The product must be stored in such a way that allows the client easy access to it.	0.04	0.19	0.05	1	1

5	Response time for loading the product should take no longer than five minutes.	0.01	0.33	0.04	1	1
5	A general knowledge of basic computer skills is required to use the product.	0.02	0.21	0.04	1	1
5	As the product is E-store, On-line help system becomes a critical component of the system which shall provide .	0.07	0.26	0.05	0	0
5	It shall provide specific guidelines to a user for using the E-Store system and within the system.	0.08	0.26	0.05	0	0
5	To implement online user help, link and search fields shall be provided.	0.05	0.34	0.06	0	0
5	There are many types of interfaces as such supported by the E-Store software system namely User Interface	0.03	0.36	0.06	1	1
5	The protocol used shall be HTTP.	0.03	0.61	0.07	1	1
5	The Port number used will be 80.	0.02	0.25	0.05	1	1
5	There shall be logical address of the system in IPv4 format.	0.02	0.05	0.01	1	1
5	The user interface for the software shall be compatible to any browser such as Internet Explorer, Mozilla or Netscape Navigator by which user can access to the system.	0.06	0.36	0.06	1	0
5	The user interface shall be implemented using any tool or software package like Java Applet, MS Front Page, EJB etc.	0.03	0.28	0.04	1	1
5	Since the application must run over the internet, all the hardware shall require to connect internet will be hardware interface for the system. As for e.g. Modem, WAN LAN, Ethernet Cross-Cable.	0.03	0.24	0.04	1	1
5	The e-store system shall communicate with the Configurator to identify all the available components to configure the product.	0.05	0.28	0.06	0	0
5	The e-store shall communicate with the content manager to get the product specifications, offerings and promotions.	0.04	0.34	0.06	0	0
5	The e-store system shall communicate with billPay system to identify available payment methods , validate the payments and process payment.	0.06	0.73	0.11	0	0
5	The e-store system shall communicate to credit management system for handling financing options.	0.07	0.45	0.08	0	0
5	The e-store system shall communicate with CRM system to provide support.	0.09	0.35	0.08	0	0
5	The e-store system shall communicate with Sales system for order management.	0.08	0.45	0.08	0	0
5	The e-store system shall communicate with shipping system for tracking orders and updating of shipping methods.	0.08	0.45	0.10	0	0
5	The e-store system shall communicate with external Tax system to calculate tax.	0.06	0.72	0.09	0	0
5	The e-store system shall communicate with export regulation system to validate export regulations.	0.04	0.20	0.04	0	1
5	The system shall be verisign like software which shall allow the users to complete secured transaction. This usually shall be the third party software system which is widely used for internet transaction.	0.06	0.32	0.05	0	0
5	The e-store system shall use the HTTP protocol for communication over the internet and for the intranet communication will be through TCP/IP protocol suite.	0.04	0.61	0.07	1	1
5	E-store should display the disclaimers, copyright, word mark, trademark and product warranties of the Marvel electronics and home entertainment.	0.02	0.11	0.03	0	1
5	It shall be as per the industry standard.	0.02	0.36	0.03	1	1

6	The archive administrator can create a new client space.	0.08	0.50	0.14	0	0
6	The client administrator can manage roles for its client.	0.14	0.63	0.15	0	0
6	The archive administrator cannot access documents.	0.09	0.45	0.12	0	0
6	The client administrator can view all roles.	0.15	0.63	0.14	0	0
6	The client administrator can�t read the contents of documents.	0.09	0.44	0.12	0	0
6	The client administrator cannot be deleted.	0.11	0.53	0.18	0	0
6	An end user can have multiple roles at the same time.	0.09	0.50	0.11	0	0
6	Each end user has a person role.	0.14	0.53	0.14	0	0
6	Only authenticated end users at the client server can access (as phrased in URCAR 45-48 and URCAR50) the document space.	0.09	0.84	0.14	0	0
6	End user cannot access (as phrased in URCAR 45-48 and URCAR50) the document spaces of other clients.	0.11	0.84	0.15	0	0
6	An end user can view assigned rights that are assigned to his role.	0.16	0.53	0.15	0	0
6	Each role is managed by at most one other role.	0.15	0.46	0.12	0	0
6	Each role inherits all static rights from the roles it manages.	0.14	0.40	0.11	0	0
6	Each role has a set (possibly empty) of person roles assigned to it.	0.13	0.55	0.12	0	0
6	Each role owns a set of documents in the SDA.	0.13	0.58	0.13	0	0
6	Each role has a set of rights assigned to it.	0.17	0.61	0.17	0	0
6	The archive administrator can limit the available client space.	0.07	0.50	0.13	0	0
6	A role can�t be deleted when there exists documents that are owned by that role.	0.18	0.48	0.13	0	0
6	An end user can for a predefined period of time delegate a role to another person role.	0.13	0.53	0.12	0	0
6	A role has a log in which the assignment of rights and the expiration of rights are logged.	0.13	0.65	0.14	0	0
6	A role has a log in which delegations and the expiration of delegations are logged.	0.05	0.65	0.11	0	0
6	A role can have the right to assign a role-right to a subordinate role.	0.28	0.74	0.23	0	0
6	A role can have the right to remove a role-right from a subordinate role.	0.23	0.74	0.20	0	0
6	A role can have the right to create subordinate roles.	0.21	0.64	0.18	0	0
6	Any role has the right to read (view) all roles.	0.18	0.48	0.13	0	0
6	A role can have the right to update subordinate roles.	0.22	0.69	0.20	0	0
6	A role can have the right to delete subordinate roles.	0.25	0.70	0.22	0	0
6	A role cannot be deleted when it manages subordinate roles.	0.19	0.66	0.16	0	0
6	A role can have the right to couple a person role to a subordinate role.	0.23	0.65	0.18	0	0
6	A role can have the right to decouple a person role from a subordinate role.	0.23	0.65	0.18	0	0
6	A role can have the right to assign a document right of a document he owns to another role.	0.25	0.82	0.18	0	0
6	A role can have the right to take a document right of a document he owns from another role.	0.21	0.82	0.16	0	0
6	A role can upload files into the document space associated with that role.	0.16	0.69	0.14	0	0

6	A role can download files from the document space associated with that role.	0.16	0.69	0.14	0	0
6	A role with the update right can update files in the document space associated with that role.	0.17	0.69	0.15	0	0
6	A role with the delete right can delete files in the document space associated with that role.	0.21	0.63	0.16	0	0
6	Each document in the SDA is owned by exactly one role.	0.11	0.58	0.12	0	0
6	A role that owns a document can hand over that ownership to another role.	0.14	0.52	0.11	0	0
6	A role can chat with other person with the same role	0.14	0.44	0.11	1	0
7	The only requirements besides having an iPhone will be having a Twitter account and internet acces	0.02	0.17	0.04	1	1
7	The Twitter client will use several different data objects while in operation.	0.07	0.36	0.09	1	0
7	The data objects that are while used inside the Application will be the main driving force of the client.	0.05	0.50	0.09	1	0
7	They will provide all the information the user wants to see based on what he or she inputs	0.03	0.29	0.05	0	1
7	The user is the main data object of the client's system.	0.09	0.50	0.11	0	0
7	The user will log into the client by supplying a username and password.	0.05	0.36	0.07	0	0
7	If the user does not already have a Twitter account, then they will have to create one in order to log into the client	0.06	0.29	0.06	0	0
7	After logging in, the user will then be able to use a search box for any anything they may want to search for, or they can enter a new Tweet for their own account.	0.08	0.52	0.09	0	0
7	The user will also be able to select who to follow, and whether or not to allow someone to follow them.	0.04	0.57	0.09	0	1
7	The user's Twitter page is an important data object in regards to the client.	0.09	0.54	0.10	1	0
7	This is where all the users Tweets are displayed and also the list of whom he or she is following or who is following them.	0.06	0.57	0.10	0	0
7	As the user enters new Tweets or someone tags his or her name in a Tweet, the page will change.	0.08	0.47	0.10	0	0
7	The new page must be updated every time a change occurs.	0.02	0.26	0.05	0	1
7	A Tweet is another data object.	0.06	0.41	0.10	1	0
7	A Tweet will be input by the user of the client.	0.09	0.30	0.08	0	0
7	A Tweet can be any sort of text as long as it is under the maximum number of characters allowed per Tweet.	0.03	0.19	0.06	0	1
7	A newly entered Tweet will appear on the user's Twitter page upon submission.	0.07	0.31	0.07	0	0
7	Other pages on Twitter are also data objects.	0.09	0.54	0.12	1	0
7	The user will be able to access other pages by either searching or clicking a linked page on their page.	0.06	0.44	0.09	0	0
7	The client would not be appealing if only the user's page could be accessed, so ability to access other pages is essential.	0.05	0.44	0.07	0	0
7	A trending topics page will be very useful to the user because it will keep them up to date on what is popular.	0.05	0.38	0.07	0	0
7	The user will enter a Tweet that can be seen by themselves and anyone who is following them.	0.06	0.43	0.09	0	0

7	If the user is tagged in a tweet it will immediately be viewable by the user.	0.07	0.48	0.09	0	0
7	The user has the ability to search for other Twitter pages by using a search box.	0.08	0.52	0.12	0	0
7	The Twitter client will store data only in one location.	0.06	0.29	0.08	0	0
7	The only data that will be stored will be the user's username and password to increase ease of access.	0.05	0.36	0.07	0	0
7	The tweets that the user enters and the tweets that the user is tagged in will all be handled by the Twitter API that we will be working with.	0.10	0.47	0.11	0	0
7	The searches that the user submits do not need to be saved, so they will not be.	0.04	0.29	0.07	0	0
7	A user will enter his/her credentials into the proper fields and the user will be logged into the Twitter system.	0.08	0.41	0.09	0	0
7	User will be given option to remember credentials so further input at startup is not necessary.	0.04	0.30	0.06	0	0
7	When the user clicks logout all personal data will be removed and all program data will be wiped.	0.04	0.19	0.06	0	1
7	The user's password will be removed and the user will be forced to re-enter their credentials at next startup.	0.05	0.30	0.07	0	0
7	Once the user has successfully logged into the Twitter system, the users Twitter feed will be displayed on the screen.	0.09	0.37	0.09	0	0
7	The number of previous data displayed will be preset by the user in the preferences pane.	0.04	0.29	0.05	0	0
7	A new view will display and show a composition window with various rich text and multimedia options.	0.04	0.26	0.06	0	1
7	The user can then enter his/her tweet and submit it to the Twitter system.	0.11	0.43	0.11	0	0
7	The tweet will be immediately displayed on the users feed.	0.07	0.48	0.09	0	0
7	When in the composition window, the user will be prompted with an image picker which will allow the user to insert an image from the users device.	0.06	0.36	0.07	0	0
7	When in the composition windows, the camera display will be shown and once the image is taken, the user will be given the option to keep re-taking.	0.06	0.49	0.09	0	0
7	If the user keeps the photo it will be used in the composition window, if the user selects re-take, the photo will be deleted and the camera display will be shown again.	0.06	0.49	0.08	0	0
7	The user will be prompted to insert existing voice memo or take a new voice memo.	0.06	0.68	0.13	0	0
7	The user will be prompted with a clip picker so the user may select a pre-existing clip on their device.	0.04	0.30	0.05	0	0
7	The user will be shown the voice memo window in which he/she may record a new voice memo.	0.07	0.77	0.15	0	0
7	The user will then have the option to hear the voice memo before submitting.	0.05	0.47	0.10	0	0
7	Once listened to, the user may select keep or re-record.	0.04	0.26	0.06	0	0
7	If the user decides to keep the memo, the memo will be inserted into the composition window.	0.06	0.62	0.12	0	0
7	If the user decided to re-record, the memo will be removed and the user will be shown the voice memo window again.	0.08	0.77	0.14	0	0
7	This function will display all mentions of the users handle (or username) in the Twitter system.	0.07	0.31	0.08	0	0
7	This function will display a search window in which the user can search for specific words in Tweets.	0.08	0.42	0.09	0	0

7	The window will also offer the ability to search for users on Twitter.	0.08	0.51	0.09	0	0
7	This window will display the users Direct Message (or DM) inbox.	0.04	0.30	0.05	0	0
7	This function will also allow the user to compose a new message, delete current messages, and read existing messages in greater detail.	0.03	0.30	0.05	0	1
7	This function will display topics which are currently trending on twitter.	0.05	0.38	0.07	0	0
7	The order of the topics will be determined by the items rise in popularity over a given timeline.	0.01	0.26	0.04	0	1
7	This function will display a preferences view which will contain various settings which the user may change based on preference.	0.04	0.29	0.06	0	0
7	The Twitter client will run natively on all iOS devices, including the iPod touch and iPad.	0.04	0.40	0.07	0	0
7	The client will be designed specifically for the iPhone/iPod Touch but it will retain the same functionality on the iPad.	0.03	0.40	0.06	0	1
7	The Twitter client will rely heavily on the Twitter API for retrieval of user data.	0.07	0.27	0.08	0	0
7	The client must work closely with the Twitter systems to authenticate and deliver content.	0.04	0.19	0.06	1	1
7	The client will be designed to work with the latest API available and compatibility is guaranteed with and iOS device running the latest version of iOS.	0.02	0.32	0.05	1	1
7	The client will utilize a common design theme throughout the entire system.	0.03	0.24	0.05	0	1
7	The application will contain multiple views, each of which will serve some primary function.	0.01	0.27	0.04	0	1
7	Each of these interfaces will differ only in content, but will maintain similar design themes.	0.01	0.24	0.05	0	1
7	The interface will utilize a fluid and robust interface, it will be easy to understand and will not require a tutorial for first-use (although one will be provided).	0.02	0.18	0.04	0	1
8	Items are bought from various vendors located in or outside United States.	0.05	0.27	0.10	0	0
8	The buyer get in contact with various vendors and create catalog (list of items) from where the Store Managers of every store buy items for their store.	0.12	0.30	0.11	0	0
8	The Online system will give the advantage of keeping track of all purchase orders at any given point of time thus eliminating paper work and in turn error caused by it	0.02	0.09	0.03	0	1
8	All The above users have similar rights.	0.03	0.49	0.12	1	1
8	They have total access the BPMS system.	0.10	0.43	0.14	0	0
8	They have the rights to generate new users in the system.	0.05	0.49	0.12	0	0
8	They will have the total access to all the information in the system being a people from the management team of the organization.	0.08	0.43	0.11	0	0
8	District manager may have more than 1 store under him.	0.08	0.39	0.10	0	0
8	He logs in he sees only stores under him i.e. a list of his stores and toggle between them.	0.09	0.33	0.11	0	0
8	He should have managerial type access only for those stores that are under him.	0.08	0.25	0.08	0	0
8	The Store Manager and Assistant Store Manager have limited access in the BPMS.	0.14	0.39	0.14	1	0
8	They will be able to see the details restricted to their Store only.	0.08	0.33	0.10	0	0
8	Cost of items bought will not be disclosed	0.04	0.27	0.08	0	1

8	The comparison with other vendor items will not shown	0.04	0.25	0.09	0	1
8	Enter the details of Sales for his store registers	0.06	0.28	0.08	0	0
8	The Usage of the system will be restricted only to add cash registers and Changing his password	0.04	0.16	0.06	0	1
8	The Stores will have limited Internet access that is the usage of BPMS system online rest of the Internet will not be accessible.	0.11	0.39	0.12	0	0
9	The system administrator must be able to deactivate and reactivate student account login	0.11	0.50	0.13	0	0
9	The system administrator must be able to force the sending of a new password to a student via email	0.11	0.44	0.12	0	0
9	The system administrator must be able to change any of a student's details	0.19	0.50	0.14	0	0
9	Student can be register on the system and fill in all detail and forward to choose project/supervisor	0.07	0.20	0.06	0	1
9	Student can change detail if information is incorrect such as telephone number	0.11	0.55	0.14	0	0
9	Student can change his login password at any time for security reason	0.09	0.23	0.07	0	0
9	Student can request his password if he/she forgotten the password	0.08	0.30	0.10	0	0
9	All staff can view the details of any student	0.12	0.29	0.09	0	0
9	Certain staff may be designated as Unit or Cohort Co-ordinators and can change the details of any student doing their unit or project cohort	0.18	0.57	0.17	0	0
9	Unit Cohort co-ordinator can change student detail for incorrect information	0.21	0.57	0.21	0	0
9	Unit Cohort co-ordinator can view student information and monitor their progress	0.14	0.54	0.18	0	0
9	Unit Cohort co-ordinator can list all students in different period of different group	0.19	0.78	0.25	0	0
9	Unit Cohort co-ordinator can reset the student&#65533;s password if required	0.19	0.64	0.20	0	0
9	System Administrator can list all students in different period of different group to check any error	0.15	0.78	0.23	0	0
9	System Administrator can reset the student&#65533;s password if required	0.16	0.64	0.16	0	0
9	Administration Staff can list all students in different period of different group	0.16	0.78	0.24	0	0
9	Administration Staff can make a school announcement	0.05	0.20	0.06	1	1
10	A ""meeting initiator"" shall initiate a meeting by deciding on a ""meeting topic""	0.25	0.60	0.11	0	0
10	A ""meeting initiator"" or ""potential meeting participant"" shall provide the ability where a person may ""delegate"" the ability to initiate or accept (or decline) a meeting to another system user.""	0.17	0.32	0.07	0	0
10	A ""meeting initiator"" shall be one of the ""potential meeting participants"" by default but may opt to remove himself as a ""potential meeting participant"".""	0.21	0.42	0.09	0	0
10	A ""meeting initiator"" shall confirm the meeting and the system shall change the ""time slots"" of accepting ""meeting participants"" from a temporary reservation to a scheduled meeting	0.33	0.51	0.09	0	0
10	A ""meeting initiator"" shall cancel the meeting and the system shall change the ""time slots"" from being temporarily reserved to be freed once the meeting is canceled.""	0.23	0.54	0.12	0	0
10	A ""meeting initiator"" shall reschedule the meeting and the system reschedule the meeting by	0.16	0.43	0.10	0	0

	releasing the temporary reservations and selecting a different ""data range""					
10	A ""meeting initiator"" may send a ""meeting proposal"" for a ""virtual meeting"" for the available ""time slot"" if the ""date range"" and ""duration"" is acceptable but no location for the meeting is available.""	0.24	0.63	0.13	0	0
10	A ""meeting initiator"" may cancel the meeting or reschedule the meeting at any time prior to the start of the meeting.""	0.23	0.43	0.09	0	0
10	A meeting scheduler may (optionally) automatically propose another meeting if current meeting is canceled by an important participant.	0.23	0.98	0.16	0	0
10	A meeting scheduler may provide the ""meeting initiator"" a summary of the scan of ""potential meeting participants"" showing available ""time slots"" and schedule conflicts as a means of informing the ""meeting initiator"" of the overall results of the system.""	0.27	0.99	0.17	0	0
10	The ""meeting initiator"" may designate one or more ""potential meeting participants"" as ""important"" meaning that their attendance at the meeting is required in order to have the meeting.""	0.22	0.42	0.08	0	0
10	The ""meeting proposal"" may include an ""agenda"" or list of topics for discussion during the meeting and may include a list of ""required equipments"".""	0.15	0.71	0.14	0	0
10	A meeting scheduler will scan all the list of ""potential meeting participants"" to determine a ""time slot"" of the required ""duration"" exists among all ""potential meeting participants"" once a ""meeting proposal"" is entered to the system.""	0.26	0.47	0.10	0	0
10	A meeting scheduler will inform the ""meeting initiator"" that no ""time slot"" exists for all ""potential meeting participants"" and may optionally suggest an alternative ""date range""	0.26	0.56	0.13	0	0
10	A meeting scheduler will temporarily reserve the ""time slots"" for the proposed meeting and inform the ""potential meeting participant"" of the meeting and request input as to ""will attend"" or ""will not attend""	0.23	0.62	0.13	0	0
10	A ""potential meeting participant"" or their ""delegate"" may accept or refuse the meeting. If accepting	0.18	0.40	0.10	0	0
10	A ""confirmed meeting participant"" may request special equipment.""	0.12	0.36	0.07	0	0
10	A ""potential meeting participant"" or ""confirmed meeting participant"" may confirm or cancel their attendance at the meeting subject to the administrative rules and practices of the participant&#65533;s.""	0.18	0.37	0.10	0	0
10	Any physical changes to the ""location"" and its ""required equipment"" shall be kept up-to-date.""	0.09	0.60	0.12	0	0
10	If any physical changes to the ""location"" and its ""required equipments"" shall occur after a ""meeting proposal"" and before the meeting date	0.20	0.60	0.10	0	0
10	The meeting scheduler system shall be able to schedule a meeting with a meeting topic, date range, duration, and location for a list of participants.	0.28	0.60	0.11	0	0
10	The meeting scheduler system shall monitor meetings	0.20	0.38	0.07	0	0
10	The meeting scheduler system shall be able to select a participant as an important participant.	0.22	0.45	0.11	0	0

10	The meeting scheduler system shall cancel a meeting due to canceling of an important participant.	0.21	0.52	0.12	0	0
10	The meeting scheduler system shall reschedule a meeting to support conflict resolutions.	0.16	0.31	0.07	0	0
10	The meeting scheduler shall be accessed from the Web.	0.11	0.21	0.05	1	0
10	The meeting scheduler system may (optionally) automatically propose another meeting if current meeting is canceled by an important participant.	0.25	0.98	0.16	0	0
10	The meeting scheduler system may provide the ""meeting initiator"" a summary of the scan of ""potential meeting participants"" showing available ""time slots"" and schedule conflicts as a means of informing the ""meeting initiator"" of the overall results of the system.""	0.27	0.99	0.17	0	0
10	The meeting scheduler system may be able to include an agenda for a meeting proposal.	0.22	0.68	0.13	0	0
10	The meeting scheduler system may suggest a ""virtual meeting"" for available ""time slots"" if no location is available or feasible for the meeting.""	0.21	0.63	0.12	0	0
10	The meeting scheduler system may be able to include a list of required equipment for a meeting proposal.	0.24	0.71	0.14	0	0
10	A meeting scheduler system will temporarily reserve the ""time slots"" for the proposed meeting.""	0.23	0.62	0.14	0	0
10	A meeting scheduler system will inform the ""potential meeting participant"" of the meeting.""	0.31	0.52	0.12	0	0

BIODATA PENULIS

Ahmad Mustofa, penulis dari buku tesis ini lahir di kabupaten Sumenep tanggal 3 Juli 1994. Penulis telah menempuh pendidikan di SD Negeri Pangarangan IV (2000-2006), SMP Negeri 1 Sumenep (2006-2009), SMA Negeri 1 Sumenep (2009-2012) dan S1 Teknik Informatika ITS Surabaya (2012-2016). Penulis memiliki peran akfif dalam dunia pengembangan perangkat lunak, karena selain menjalani perkuliahan selama masa perkuliahan paska sarjana, penulis juga aktif sebagai *co-founder* YukBelajar, sebuah perusahaan *startup* yang bergerak di bidang Pendidikan, dan sebagai *Android Developer* di sebuah *Software House* di Surabaya. Selain itu penulis juga tertarik pada topik *image processing*, *signal processing*, dan *text processing*. Penulis dapat dihubungi melalui email mustofa.ahmad12@mhs.if.its.ac.id dan mustofaahmad327@gmail.com