

# ANALISIS NILAI SENSITIVITAS UNTUK MENGUKUR AKURASI KOMBINASI KATA YANG DIGUNAKAN DALAM PREDIKSI KATEGORI TUGAS/PEKERJAAN

Eko Wahyu Tyas Darmaningrat<sup>1)</sup>, Shu-Chiang Lin<sup>2)</sup>

<sup>1)</sup>Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember  
Jl. Raya ITS, Kampus ITS Sukolilo Surabaya, 60111

Telp: (031) 5999944, Fax: (031) 5964965

<sup>2)</sup>Department of Industrial Management

National Taiwan University of Science and Technology, Taiwan, R.O.C

E-mail: [tyas.darmaningrat@gmail.com](mailto:tyas.darmaningrat@gmail.com)

---

## Abstract

*Broadly speaking, task analysis is a technique that determines the inputs, tools, and skills or knowledge necessary for successful task performance. Questions about what tasks the users will perform with the system often arise in task analyst's mind. Unfortunately, a task analysis of users' current activities is insufficient to guess what tasks the users will do following the previous tasks. The purpose of this research is to examine and further analyze the primitive results of Bayesian based task analysis model obtained by Lin and Lehto (2009) by comparing the prediction accuracy among all the word combinations from fuzzy Bayes model. This Bayesian based semi-automated task analysis tool was developed to help task analysts predict categories of tasks/subtasks performed by knowledge agents from telephone conversations where agents were trying to help customers to troubleshoot their problems. This study investigates two factors, subtask category and word predictor, based on three responses: hit rate, false alarm rate, and sensitivity value. The ANOVA test results reveal that both factors have significant effect to the three responses. Our findings imply that predictors contained single word always have higher hit rate than others. Moreover, predictors which have high hit rate tend to have high false alarm rate as well. Therefore, predictors with high hit rate do not always have high sensitivity value. Although the ANOVA test results cannot specify the most accurate predictor, they provide information on group of predictors with the best and the worst performance. These findings support our hypothesis that the word combination affects the accuracy of predicted subtask category.*

## Abstrak

Berbagai buku dan penelitian telah membahas aplikasi dari konsep analisis tugas (task analysis) dalam berbagai bidang. Secara garis besar, analisis tugas adalah teknik yang digunakan untuk menentukan masukan, alat, dan keterampilan atau pengetahuan yang diperlukan untuk menyelesaikan suatu tugas atau pekerjaan dengan baik. Pertanyaan tentang tugas apa yang akan pengguna selesaikan dengan menggunakan sistem sering muncul dalam pikiran seorang analis tugas. Sayangnya, analisis dari aktifitas pengguna saat ini tidak cukup untuk menebak tugas apa yang akan pengguna lakukan setelah melakukan tugas sebelumnya. Tujuan dari penelitian ini adalah untuk menguji dan menganalisis lebih lanjut hasil dari model analisis tugas dengan metode Bayesian yang dikembangkan oleh Lin dan Lehto (2009) dengan membandingkan akurasi prediksi antara semua kombinasi kata yang dihasilkan dari model Fuzzy Bayes. Alat analisis tugas semi-otomatis dengan metode Bayesian ini dikembangkan guna membantu analis tugas memprediksi kategori tugas/sub-tugas yang dilakukan oleh agen pengetahuan (informan) dari percakapan telepon di mana informan berusaha untuk membantu pelanggan memecahkan masalah mereka. Penelitian ini fokus pada dua faktor, yakni kategori sub tugas dan prediktor kata, berdasarkan tiga parameter: hit rate, false alarm rate, dan nilai sensitivitas. Hasil uji ANOVA menunjukkan bahwa kedua faktor berpengaruh signifikan terhadap ketiga parameter tersebut. Hasil penelitian ini menunjukkan bahwa prediktor yang mengandung kata tunggal selalu memiliki hit rate yang lebih tinggi. Selain itu, prediktor yang memiliki hit rate yang tinggi cenderung memiliki tingkat false alarm yang tinggi juga. Oleh karena itu, prediktor dengan hit rate yang tinggi tidak selalu memiliki nilai sensitivitas yang tinggi. Meskipun hasil tes ANOVA tidak dapat menentukan prediktor mana yang paling akurat, tetapi dapat memberikan informasi tentang kelompok prediktor dengan kinerja terbaik dan terburuk. Temuan ini mendukung hipotesis kami bahwa kombinasi kata mempengaruhi akurasi kategori sub tugas yang diprediksi.

**Kata kunci:** analisis tugas, model fuzzy bayes, hit rate, false alarm, nilai sensitivitas

## 1. PENDAHULUAN

Sejalan dengan perkembangan sistem komputer dan antarmukanya, pertanyaan tentang tugas apa yang akan pengguna selesaikan dengan menggunakan sistem sering muncul dalam pikiran seorang analis tugas. Sayangnya, analisis dari aktifitas pengguna saat ini tidak cukup untuk menebak tugas apa yang akan pengguna lakukan setelah melakukan tugas sebelumnya. Lin dan Lehto (2009) mengajukan metode analisis tugas baru yang menggabungkan pendekatan statistik, model fuzzy bayes, dan metode analisis tugas klasik. Dalam penelitian Lin dan Lehto (2009) tersebut, alat analisis tugas semi-otomatis berbasis inferensi Bayesian dikembangkan untuk membantu analis tugas memprediksi kategori tugas/sub-tugas yang dilakukan oleh agen-agen pengetahuan (informan) dari percakapan telepon dimana agen berusaha untuk membantu pelanggan memecahkan masalah mereka. Inferensi Bayesian merupakan pendekatan statistik dimana semua bentuk ketidakpastian dinyatakan dalam probabilitas. Inferensi Bayesian menggunakan beberapa bukti atau pengamatan untuk menghitung kemungkinan hipotesis menjadi kenyataan (probabilitas posterior), atau sebaliknya untuk memperbarui probabilitas yang dihitung sebelumnya (probabilitas prior) (Neal, 1998). Di antara berbagai aplikasi Model Bayesian, dua inferensi Bayesian sering diajukan: model Bayesian klasik (Naïve) dan model Fuzzy Bayesian. Jika beberapa bukti diperoleh, model Bayesian klasik mencoba untuk menggabungkan bukti-bukti tersebut dan menganggap bahwa mereka bebas dengan syarat hipotesis yang diberikan benar. Berbeda dengan Naïve Bayesian, model Fuzzy Bayesian menghindari membangun asumsi independensi yang kuat dari bukti-bukti yang diperoleh (Lin dan Lehto, 2009).

Evans dan Karwowski (1987) menyarankan Fuzzy Bayesian merupakan pendekatan yang tepat untuk menangani data yang tidak akurat dan tidak jelas. Keuntungan utama dari pendekatan Fuzzy Bayesian adalah dapat memberikan performa kinerja yang sama tetapi membutuhkan upaya komputasi yang jauh lebih sedikit dibandingkan dengan pendekatan lain yang serupa seperti neural networks, natural language processing methods dan model linguistik (Lehto dan Sorock, 1996). Hasil penelitian sebelumnya oleh Lin dan Lehto (2009) menginformasikan bahwa alat yang dikembangkan mampu belajar dan memprediksi kategori sub tugas dari percakapan telepon antara pelanggan dan informan. Hasil ini merupakan titik awal yang memungkinkan penelitian lanjutan tentang teori aplikasi Bayesian dalam bidang analisis tugas. Dengan banyaknya hasil yang diperoleh dari penelitian tersebut, menemukan hasil yang paling akurat

menjadi hal yang utama. Oleh karena itu, dalam penelitian ini, akan dilakukan analisis yang lebih mendalam berdasarkan hasil yang diperoleh sebelumnya untuk membandingkan akurasi prediksi antara semua kombinasi kata yang diperoleh dari model Fuzzy Bayes.

## 2. METODOLOGI

Menurut Annett et al. (1971) proses menganalisis tugas adalah proses mendiagnosa rencana yang diperlukan untuk mencapai tujuan yang telah ditentukan. Swezey dan Pearlstein (2001) menyatakan bahwa analisis tugas adalah teknik untuk menentukan masukan, peralatan, dan keterampilan atau pengetahuan yang diperlukan untuk menyelesaikan suatu tugas atau pekerjaan dengan baik. Analisis tugas bisa dalam bentuk menggambarkan pekerjaan dalam hal tindakan fisik individu (misalnya, menekan tombol) atau menjelaskan tindakan dalam hal tingkatan yang lebih tinggi (seperti tujuan, konsep domain, dll) (Hoffman dan Militello, 2008). Analisis tugas menggambarkan pemahaman kita tentang faktor yang mempengaruhi kinerja manusia dan kebutuhan informasi dari perancang sistem. Dalam beberapa tahun terakhir, teknologi modern telah banyak mengubah sifat pekerjaan manusia serta konsep dan teknik analisis, dan akan terus berubah untuk memenuhi kebutuhan yang baru (Annett dan Stanton, 2000a). Berbagai penelitian menunjukkan bahwa penggunaan analisis tugas telah diperluas dari spesialis tugas seperti ergonomists, desainer tugas, dan analis tugas, menjadi pekerja yang terkait dengan suatu tugas seperti operator, manajer, dan supervisor. Setiap individu telah memainkan peran dalam mengintegrasikan elemen manusia ke dalam desain sistem dan operasi secara efektif dan efisien (Lin dan Lehto, 2009).

### 2.1 Hierarchical Task Analysis (HTA)

Teknik Analisis Tugas Hirarkis (*Hierarchical Task Analysis/HTA*) membagi tugas atau pekerjaan ke dalam sub-pekerjaan dan operasi atau tindakan. Metode ini memerlukan identifikasi tugas/pekerjaan, mengkategorikannya, mengidentifikasi subtugas, dan memeriksa akurasi keseluruhan model (kristal dan Ellington, 2004). Menurut kristal dan Ellington (2004), HTA berguna untuk desainer antarmuka karena memberikan model untuk eksekusi tugas, memungkinkan desainer untuk memvisualisasikan tujuan, tugas, sub-tugas, operasi, dan rencana-rencana yang diperlukan untuk aktifitas pengguna. HTA berguna untuk mendekomposisi tugas-tugas kompleks, tetapi memiliki pandangan yang sempit terhadap tugas, dan biasanya digunakan bersama dengan metode analisis tugas lainnya untuk meningkatkan efektivitasnya. Stanton dan Young (1999) menyebutkan

beberapa keuntungan dari HTA seperti kemudahannya untuk diimplementasikan setelah konsep awal dipahami. Selain itu, eksekusi yang cepat memberikan kepuasan bagi pengguna. Namun, HTA juga memiliki beberapa kelemahan seperti memberikan informasi deskriptif bukan informasi analitis; biasanya informasi yang dihasilkan perlu untuk disimpulkan terlebih dahulu sebelum digunakan sebagai solusi desain; HTA tidak menangani komponen kognitif dari tugas (misalnya pengambilan keputusan), hanya unsur-unsur yang dapat diamati.

## 2.2 Cognitive Task Analysis (CTA)

Dibandingkan dengan HTA, Analisis Tugas Kognitif (Cognitive Task Analysis/CTA) menyajikan tantangan analisis yang sangat berbeda. CTA membutuhkan hubungan yang mendalam dengan domain pengetahuan tertentu, dan bekerja sama dengan ahli bidang tertentu untuk memperoleh pengetahuan mereka tentang berbagai tugas (Chipman et al., 2000). CTA merupakan upaya untuk memperoleh keahlian tertentu, karena keahlian biasanya tidak berwujud sehingga lebih sulit untuk menganalisisnya dibandingkan dengan tindakan eksplisit seperti yang biasanya dianalisis dalam HTA. Bahkan, CTA memerlukan untuk "membuat pengetahuan dan pekerjaan implisit menjadi eksplisit" (Dubois dan Shalin, 2000). CTA telah meningkatkan pemahaman tentang berbagai aspek kognitif dari lingkungan kerja modern. Namun, tidak jelas seberapa efektif teknik CTA dalam merepresentasikan aspek-aspek ini secara sistematis (Shepherd, 2001). Masalah lain yang signifikan dengan banyak teknik CTA adalah bahwa mempelajari fungsi kognitif tingkat tinggi dalam lingkungan pekerjaan yang sebenarnya sangat sulit. Penelitian tentang hal ini bisa memerlukan waktu berbulan-bulan bahkan bertahun-tahun. Akibatnya, beberapa praktisi mengembangkan pendekatan sederhana yang dapat membentuk "toolkit" bagi para praktisi (Militello dan Hutton, 2001). Ketika user interface perlu dirancang untuk pekerjaan yang bergantung pada integrasi informasi, misalnya navigasi, teoriti yang tersedia sangat terbatas. Kombinasi CTA dan HTA merupakan dasar dari definisi antarmuka (Westrenen, 2010).

## 2.3 Metode Bayesian

Di antara sejumlah pendekatan statistik, aplikasi model Bayesian terbukti telah berhasil di berbagai bidang. Zhu dan Lehto (1999) mengemukakan bahwa aturan inferensi Bayesian merupakan model statistik yang tepat untuk menemukan hubungan ketergantungan antara indeks istilah dan kata-kata dalam teks. Lehto dan Sorock (1996) menilai inferensi Bayesian sebagai teknik machine learning dan

menyimpulkan bahwa metode ini berhasil mengetahui dan memprediksi kategori dan subkategori kecelakaan kendaraan bermotor dari data naratif yang ada. Lin dan Lehto (2009) menerapkan model naive (klasik) bayes, fuzzy bayes, dan hybrid bayes untuk mengembangkan alat analisis tugas semi-otomatis untuk memprediksi kategori sub-tugas. Penulis mencatat dialog yang digunakan pelanggan dan informan di call center sebagai sumber data untuk melatih dan menguji alat mereka usulkan. Sebanyak dua belas hasil prediksi diperoleh dan dicatat dalam dua belas tabel prediksi. Setiap tabel berisi 5.184 hasil prediksi, prediksi kategori sub-tugas untuk berbagai kombinasi kata. Setiap record mengungkapkan prediksi kategori sub tugas, kata kunci yang digunakan untuk membuat prediksi, dan kekuatan (tingkat kepercayaan) kategori sub-tugas untuk prediksi tertentu.

## 2.4 Hit Rate dan False Alarm Rate

Definisi paling sederhana dari *hit rate* dan *false alarm rate* adalah bahwa jika alat yang dikembangkan mengidentifikasi atau memprediksi kategori sub tugas yang cocok dengan kategori sub tugas yang telah diberikan oleh seorang pakar untuk kata yang sama, maka ini adalah "*hit*" untuk kategori yang diprediksi; jika alat mengidentifikasi atau memprediksi kategori sub tugas yang tidak sesuai dengan kategori sub tugas yang diberikan pakar, maka ini adalah "*false alarm*" untuk kategori yang diprediksi dan *miss* untuk kategori aktual (Lin dan Lehto, 2009). Secara umum, hasil dari *hit rate* dan *false alarm rate* dari suatu percobaan adalah sebagai berikut:

Kategori Sub Tugas		Ditentukan oleh pakar	
		True	False
Hasil prediksi alat	True	Hit	False Alarm
	False	Miss	Correct Rejection

## 2.5 Signal Detection Theory (SDT)

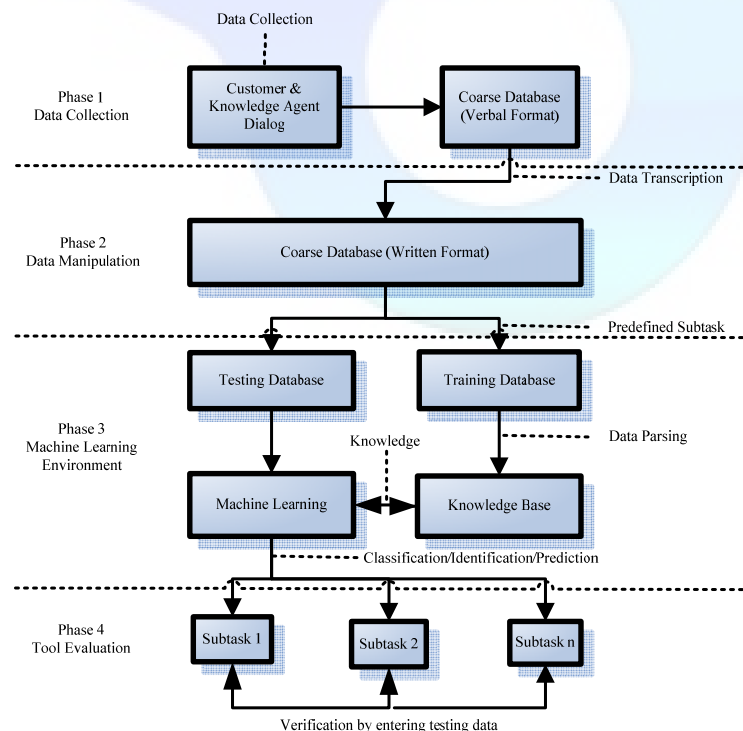
*Signal Detection Theory* (SDT) sepenuhnya dikembangkan sejak tahun 1960-an. *Signal Detection Theory* menyediakan kerangka kerja umum untuk menggambarkan dan mempelajari keputusan yang dibuat dalam situasi yang tidak pasti atau ambigu. Teori ini paling banyak digunakan dalam bidang *psychophysics* -domain dari penelitian yang menyelidiki hubungan antara stimulus fisik dan efek subjektif atau psikologisnya- tetapi teori ini memiliki implikasi tentang bagaimanakah jenis keputusan di bawah ketidakpastian dibuat (Wickens, 2002). SDT digunakan untuk menganalisis data yang berasal dari percobaan dimana tugasnya adalah untuk mengkategorikan stimulus ambigu yang dapat dihasilkan

baik oleh sebuah proses yang dikenal (disebut sinyal) atau yang diperoleh secara kebetulan (disebut *noise* dalam kerangka kerja SDT). SDT umumnya digunakan untuk menganalisis eksperimen yang memerlukan jawaban biner (misalnya 'Ya' atau 'Tidak') (Abdi, 2009).

## 2.6 Penyusunan Model

Penelitian ini menggunakan data rekaman dialog percakapan telepon di *call center* yang direkam saat agen pengetahuan (informan) melakukan proses *troubleshooting*. Seorang pelanggan menelepon *call center* untuk melaporkan masalah yang berhubungan dengan printernya dan dihubungkan dengan agen pengetahuan. Agen pengetahuan bertanggung jawab untuk menentukan, mengidentifikasi, dan/atau menilai masalah serta untuk memecahkan masalah, mendiagnosis, dan/atau menganalisis masalah dan menyusun rencana untuk menjelaskan dan/atau memberikan instruksi untuk memandu klien melalui resolusi jarak jauh secara tepat. Para agen pengetahuan sering memanfaatkan beberapa jenis sumber daya untuk membantu pelanggan menemukan masalahnya karena tingkat kompleksitas yang berbeda dari permasalahan seputar printer. Misalnya mencari berbagai tips dan/atau alat sistem pakar on-line, mengacu pada layanan manual berbasis kertas dan dokumen, melakukan pengujian produk fisik, berkomunikasi dan berkonsultasi dengan teknisi, supervisor, atau agen pengetahuan lain-

nya yang lebih tinggi tingkat pemahamannya, dll. Sehingga, dari sudut pandang pakar, pekerjaan ini tidak hanya melibatkan aktivitas fisik tetapi juga aktivitas psikologis seperti pengambilan keputusan. Gambar 1 menggambarkan seluruh proses eksperimental untuk mengembangkan alat analisis semi-otomatis berbasis Bayesian tugas yang diusulkan oleh Lin dan Lehto (2009). Panah dalam gambar menggambarkan aliran data. Data percakapan telepon antara pelanggan dan agen pengetahuan dalam bentuk verbal diubah menjadi bentuk tertulis dan digunakan sebagai input dalam penelitian ini. Kategori sub tugas didefinisikan dan secara manual diberikan kepada data yang sudah dalam format tertulis untuk membentuk *training set*. *Training set* akan menjadi masukan untuk melatih alat analisis tugas semi-otomatis yang dikembangkan. Kata kunci di-*parsing* dari *training set* dan disimpan sebagai basis pengetahuan (*knowledge base*). Alat analisis tugas semi-otomatis akan belajar pengetahuan (kata kunci) dari basis pengetahuan yang dihasilkan dari *training set* dan mengembangkan hasil kategori sub tugas yang diidentifikasi/diklasifikasikan sebagai output. *Test case* diberikan kepada alat tersebut untuk menguji hasil prediksi yang dihasilkan, yaitu untuk memverifikasi keakuratan hasil yang dihasilkan. Tabel 1 memberikan contoh data kategori sub tugas (2 dari 12 kategori) yang digunakan dalam penelitian ini.



Gambar 1 Empat Fase Pengembangan dari alat Analisis Tugas



Tabel 1. Dekomposisi Kategori Subtugas Agen dan Pelanggan

Main Category (Node)	Coarse Subcategory 1 (Node)	Coarse Subcategory 2 (Node)	Node <sup>1</sup>	Detailed Subcategory Knowledge Agent	Node <sup>2</sup>	Detailed Subcategory Customer
Greeting (1)	(1)		111	Greet the customer with HP customer care and provide the agent's name and support group	1111	Greet
Opening, gather and verify customer information (2)	(21-25)	(211-212)	211	Ask (obtain) the customer's first name/last name	2111	Provide name or Provide the agent's name and support group if customer is an agent
			212	Spell back and confirm the customer's first name/last name	2121	Verify the customer information
		(215-216)	215	Ask the customer's Phone number	2151	Provide phone number
			216	Spell back and confirm the customer's phone number		
		(221-222)	221	Ask company name	2211	Provide company name
			222	Spell back and confirm company address	2221	Verify the company information
		(223-224)	223	Ask company address		
			224	Spell back and confirm company address		
		(225)	225	Ask alternate contact information	2251	Provide alternate contact information

Node<sup>1</sup>: subtugas untuk aktifitas agenNode<sup>2</sup>: subtugas untuk aktifitas pelanggan

Tabel 2. Contoh Hasil Prediksi dengan Prediktor Single Word

Cat <sup>1</sup>	SW Prediction <sup>2</sup>	SW Pred Acc <sup>3</sup>	SW strength <sup>4</sup>	SWpredictor <sup>5</sup>	Cat <sub>1</sub>	SW Prediction <sup>2</sup>	SW Pred Acc <sup>3</sup>	SW strength <sup>4</sup>	SW predictor <sup>5</sup>
111	111	1	1.00000	HPCustomerCare	382	382	1	0.50000	default
111	111	1	1.00000	LaserJetContract Support	382	382	1	0.50000	features
111	111	1	1.00000	LaserJetsupport	382	382	1	0.50000	load
111	111	1	0.90909	Randy	382	382	1	0.50000	manager
111	111	1	0.80000	Sheila	382	382	1	0.50000	NINEFIVE
111	111	1	0.79577	hello	383	383	1	1.00000	roughly
111	111	1	0.57798	myname	383	383	1	1.00000	specify
211	211	1	1.00000	Stephanie	383	383	1	1.00000	valid
211	211	1	0.92308	firstname	391	391	1	1.00000	TWOminutes
211	211	1	0.89474	yourname	391	391	1	0.80000	FOURminutes
211	211	1	0.83333	FOURNESS	391	391	1	0.56250	warmup

Cat<sup>1</sup>: Kategori subtugas yang diberikan oleh pakarSWPrediction<sup>2</sup>: Kategori subtugas hasil prediksi alat dengan predictor single wordSWPredAcc<sup>3</sup>: Akurasi prediksi untuk dialog tersebut; 1 = *correct (hit)*; 0 = *incorrect (miss)*SWStrength<sup>4</sup>: Probabilitas kategori subtugas untuk prediktor single word tersebutSWPredictor<sup>5</sup>: Kata kunci yang digunakan untuk prediksi

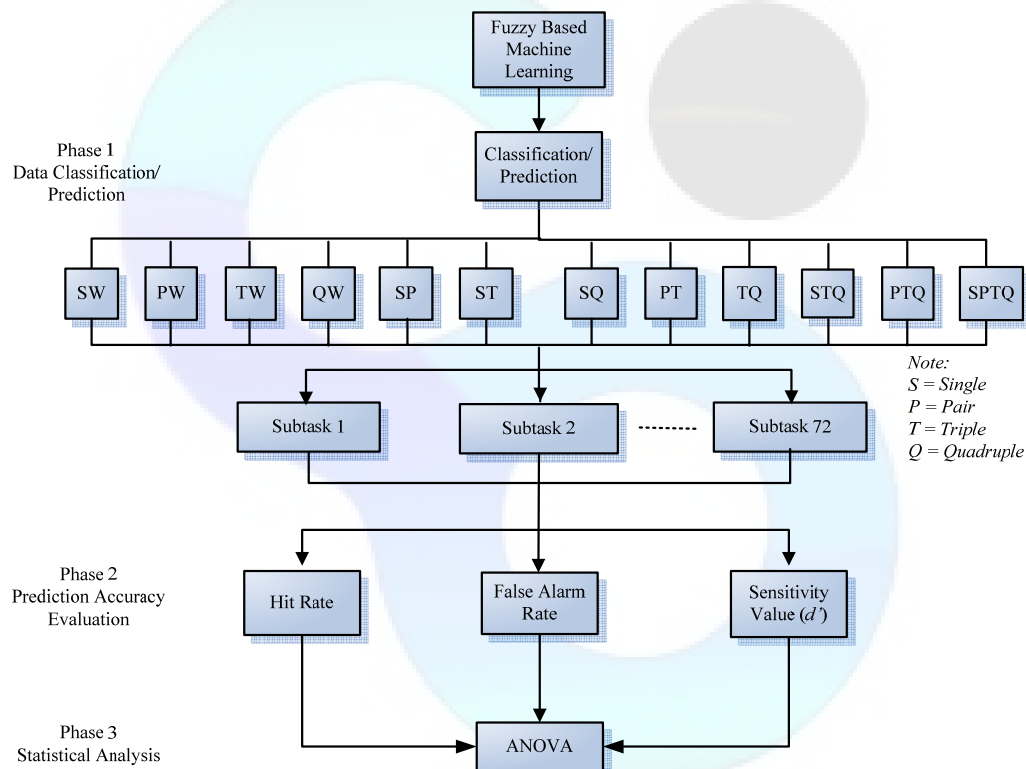
Hasil luaran dari alat analisis tugas semi-otomatis yang dikembangkan berupa hasil prediksi yang dicatat dalam dua belas tabel prediksi. Tabel 2 adalah contoh tabel prediksi untuk *single word*. Penelitian ini ditekankan pada tahap evaluasi alat, yaitu untuk mengevaluasi hasil dari proses pembelajaran yang dilakukan oleh alat yang dikembangkan dalam mengklasifikasi, mengidentifikasi, atau memprediksi kategori sub tugas dari percakapan telepon antara agen dan pelanggan. Gambar 2 mengilustrasikan ruang lingkup utama dari penelitian ini.

Dari model Fuzzy Bayes, dihasilkan total dua belas hasil prediksi yang dicatat dalam dua belas tabel prediksi. Tabel tersebut, masing-masing dengan 5.184 hasil prediksi, memprediksi kate-

gori sub tugas untuk berbagai kombinasi kata, mulai dari *single word* (kata tunggal), *pair-word*, *triple-word*, *quadruple-word*, hingga kombinasi *single-pair*, *single-triple*, *single-quadruple*, *pair-triple*, *pair-quadruple*, *triple-quadruple*, *single-triple-quadruple*, *pair-triple-quadruple*, dan kombinasi *single-pair-triple-quadruple*.

### 3. HASIL DAN PEMBAHASAN

Penelitian ini mengkaji dua faktor, yaitu kategori sub tugas dan kata yang digunakan untuk memprediksi kategori tersebut, berdasarkan tiga parameter: *hit rate*, *false alarm rate*, dan nilai sensitivitas. Tabel 3 menampilkan nilai *hit rate*, *false alarm rate*, dan nilai sensitivitas dari 12 prediktor kata yang dihasilkan dalam dua proses iterasi.



Gambar 2 Ruang Lingkup

Tabel 3. Hit Rate, False Alarm Rate, dan Nilai Sensitivity dari 12 Prediktor

Predictor	% Hit Rate Replicate 1	% Hit Rate Replicate 2	% False Alarm Rate Replicate 1	% False Alarm Rate Replicate 2	% Sensitivity Value Replicate 1	%Sensitivity Value Replicate 2
single word (SW)	49.56	50.17	0.731	0.721	2.43073	2.45074
pair-word (PW)	49.77	49.23	0.688	0.694	2.45766	2.44257
triple-word (TW)	36.97	35.11	0.628	0.685	2.16363	2.08283
quadruple-word (QW)	26.34	23.42	0.547	0.588	1.91190	1.79410
single-pair (SP)	54.15	54.17	0.666	0.667	2.57964	2.57932
single-triple (ST)	51.53	50.51	0.695	0.719	2.49832	2.46062
single-quadruple (SQ)	49.29	46.72	0.736	0.769	2.42172	2.34139
pair-triple (PT)	50.68	49.14	0.681	0.699	2.48423	2.43630

triple-quadruple (TQ)	37.89	35.19	0.670	0.709	2.16488	2.07256
single-triple-quadruple (STQ)	51.82	49.87	0.701	0.725	2.50267	2.44133
pair-triple-quadruple (PTQ)	50.70	48.87	0.679	0.690	2.48558	2.43388
single-pair-triple-quadruple (SPTQ)	54.60	53.21	0.659	0.677	2.59440	2.55004

Tabel 4. *General Linear Model: Hit Rate versus Predictor, Category* ( $\alpha = 0.05$ )

Source of Variation	Degree of Freedom	Sum of Squares	Mean Squares	F	P-value
Predictor	11	9.0632	0.8239	184.73	0.000
Category	71	121.2132	1.7072	382.78	0.000
Predictor*Category	781	16.5400	0.0212	4.75	0.000
Error	864	3.8535	0.0045		
Total	1727	150.6700			

Hasil analisis awal yang dirangkum pada Tabel 1 menunjukkan bahwa pada iterasi pertama kombinasi *single-pair-triple-quadruple* (SPTQ) memiliki *hit rate* tertinggi 54,60%, sedangkan kombinasi kata *quadruple* (QW) memiliki tingkat paling hit dari 26,34 % antara lain. Di sisi lain, dalam replikasi *single* kedua-pair (SP) kombinasi memiliki *hit rate* tertinggi 54,17% dan kombinasi *quadruple word* (QW) memiliki *hit rate* terendah 23,42% dibandingkan dengan kombinasi lainnya. Hasil ini menunjukkan bahwa penyelidikan lebih lanjut diperlukan untuk menyimpulkan prediktor kata mana yang memiliki akurasi tertinggi dalam memprediksi kategori sub tugas.

Hasil uji ANOVA pada *hit rate* vs prediktor kata dan kategori seperti yang ditunjukkan pada Tabel 4 menunjukkan bahwa kedua perlakuan (kategori dan prediktor) dan interaksi keduanya memiliki pengaruh yang signifikan terhadap *hit rate*. Ketika hasil ANOVA menunjukkan bahwa mean dari baris dan kolom berbeda, biasanya menarik untuk membuat perbandingan antara mean dari baris individu dan kolom untuk menemukan perbedaan yang spesifik. Dari hasil uji Tukey pada *hit rate* dibandingkan prediktor, diketahui bahwa kombinasi kata *quadruple* (QW) memiliki kinerja terburuk dibandingkan dengan kombinasi lainnya. Namun, kita tidak bisa menentukan prediktor mana yang paling akurat karena mean dari semua prediktor hanya sedikit berbeda antara satu dengan yang lain. Akan tetapi, prediktor kata dapat dikelompokkan menjadi dua kelompok utama berdasarkan perbedaan kemampuan mereka. Kelompok pertama terdiri dari *single-pair* prediktor (SP), *single-pair-triple-quadruple* (SPTQ), *single word* (SW), *single-triple* (ST), *single-triple-quadruple* (STQ), *single-quadruple* (SQ), *pair-triple* (PT), *pair-triple-quadruple* (PTQ), dan *pair word* (PW). Kelompok kedua terdiri dari *triple-quadruple* prediktor (TQ), *triple word* (TW), and *quadruple word* (QW). Semua anggota kelompok pertama

berbeda secara signifikan dengan QW, dan tiga anggota kelompok pertama yaitu PT, PTQ, dan PW sebenarnya tidak berbeda signifikan dengan prediktor TQ dan TW. Salah satu pendekatan alternatif untuk menyelidiki akurasi prediksi kata adalah dengan memeriksa *false alarm rate* dan nilai sensitivitas ( $d'$ ). Hasil uji ANOVA pada *false alarm rate* dan nilai sensitivitas terhadap prediktor dan kategori juga mengungkapkan bahwa kedua perlakuan (kategori dan prediktor) dan interaksi mereka memiliki pengaruh yang signifikan terhadap respon (*false alarm rate* dan nilai sensitivitas). Hasil yang ditunjukkan pada Tabel 1 menggariskan bahwa secara keseluruhan *hit rate* secara signifikan lebih besar daripada *false alarm rate*, dan nilai  $d'$  lebih besar dari nol. Meskipun dari hasil uji ANOVA kita tidak bisa menentukan prediktor yang paling akurat, tetapi hasilnya memberikan informasi tentang kelompok prediktor dengan kinerja terbaik dan kinerja terburuk. Temuan ini mendukung hipotesis kami bahwa kombinasi kata mempengaruhi akurasi kategori sub tugas yang diprediksi. Umumnya, subyek terbaik akan memaksimalkan *hit rate* (dan meminimalkan *miss rate*) dan meminimalkan *false alarm rate* (dan memaksimalkan *correct rejection rate*); oleh karena itu semakin besar perbedaan antara *hit rate* dan *false alarm rate*, semakin baik sensitivitas subyek. Dalam penelitian ini, prediktor yang memiliki *hit rate* yang tinggi cenderung memiliki *false alarm rate* yang tinggi juga. Oleh karena itu, prediktor dengan *hit rate* yang tinggi tidak selalu memiliki nilai sensitivitas yang tinggi. Tabel 5 merangkum kinerja dari dua belas prediktor berdasarkan *hit rate*, *false alarm rate*, dan nilai sensitivitasnya. Prediktor yang mengandung *single word* (SW) selalu memiliki *hit rate* yang lebih tinggi daripada yang lain. Selain itu, *triple-quadruple* (TQ), *triple word* (TW), dan *quadruple word* (QW) memiliki *hit rate* yang rendah, *false alarm rate* rendah, dan nilai sensitivitas rendah.

Tabel 5 Perbandingan Kinerja dari Prediktor Kata

Order (high to low)	Hit Rate	False Alarm Rate	Sensitivity Value ( $d'$ )
1	SP	SW	SPTQ
2	SPTQ	SP	STQ
3	SW	SQ	SP
4	ST	ST	ST
5	STQ	SPTQ	PTQ
6	SQ	STQ	PT
7	PT	PW	SW
8	PTQ	PT	SQ
9	PW	PTQ	PW
10	TQ	TQ	TQ
11	TW	TW	TW
12	QW	QW	QW

#### 4. SIMPULAN dan SARAN

Penelitian ini mengkaji dua faktor, yaitu kategori sub tugas dan prediktor kata, berdasarkan tiga parameter respon: yaitu *hit rate*, *false alarm rate*, dan nilai sensitivitas. Hasil uji ANOVA menggambarkan bahwa kedua faktor dan interaksinya berpengaruh signifikan terhadap ketiga parameter respon. Temuan kami menunjukkan bahwa prediktor yang mengandung *single word* (SW) selalu memiliki *hit rate* yang lebih tinggi daripada yang lain. Selain itu, *triple-quadruple* (TQ) *triple word* (TW), dan *quadruple word* (QW) memiliki *hit rate* rendah, *false alarm rate* rendah, dan nilai sensitivitas rendah. Prediktor yang memiliki *hit rate* yang tinggi cenderung memiliki *false alarm rate* yang tinggi juga. Oleh karena itu, prediktor dengan *hit rate* yang tinggi tidak selalu memiliki nilai sensitivitas yang tinggi. Meskipun kami tidak bisa menentukan prediktor yang paling akurat, tetapi hasil analisis kami dapat memberikan informasi tentang kelompok prediktor dengan kinerja terbaik dan kinerja terburuk, hal ini mendukung hipotesis kami bahwa kombinasi kata mempengaruhi akurasi kategori sub tugas yang diprediksi. Dalam penelitian selanjutnya, analisis yang menarik untuk dilakukan adalah menemukan hubungan antara prediktor kata dan kategori sub tugas.

#### 5. DAFTAR RUJUKAN

- Abdi, Hervé. (2009). Signal Detection Theory. In McGaw, B., Peterson, P.L., Baker, E. (Eds.): *Encyclopedia of Education* (3rd Ed). New York: Elsevier.
- Annett, J., Duncan, K.D., Stammers, R.B. and Gray, M.J. (1971). *Task Analysis*. London: HMSO.
- Annett, J. and Stanton, N.A. (2000a). *Research and Developments in Task Analysis*. In Annett and Stanton. (2000): Chapter 1 (pp. 1-8).
- Annett, J. and Stanton, N.A. (2000b). *Task Analysis*. London: Taylor and Francis.
- Chatterjee, S. (1998). *A Connectionist Approach for Classifying Accident Narratives*. Unpublished Ph.D. Dissertation. Purdue University, West Lafayette, IN. Retrieved October 19, 2011, from <http://search.proquest.com/docview/304450940>
- Chipman, S.F., Schraagen, J.M., and Shalin, V.L. (2000). *Introduction to Cognitive Task Analysis*. In Chipman, et al. (2000), Part I (page 3-24).
- Crystal, A. and Ellington, B. (2004). *Task Analysis and Human-Computer Interaction: Approaches, Techniques, and Levels of Analysis. Proceedings of the Tenth Americas Conference on Information Systems, New York*.
- Dubois, D. and Shalin, V. (2000). *Describing Job Expertise using Cognitively Oriented Task Analyses (COTA)*. In Chipman, et al. (2000), Part II (page 41-56).
- Evans, G.W., Wilhelm, M.R., and Karwowski, W. (1987). A layout design heuristic employing the theory of fuzzy sets. *International Journal of Production Research*, 25,1431-1450.
- Hoffman, R.R and Militello, L.G. (2008). *Perspective on Cognitive Task Analysis: Historical Origins and Modern Communities of Practice*. New York: Taylor and Francis.
- Lehto, M.R., Maruci-Wellman, H., and Corns, H., (2009). Bayesian Methods: A Useful Tool for Classifying Injury Narratives into Cause Groups. *Injury Prevention Module 2009;000:0-7*. doi:10.1136/ip.2008.021337.
- Lehto, M.R., and Sorock, G.S. (1996). Machine learning of motor vehicle accident categories from narrative data. *Meth. Inform. Med.*, 35 (4/5), 309-316.
- Lin, S. and Lehto, M.R. (2009). *A Bayesian Based Machine Learning Application to Task Analysis, Encyclopedia of Data Warehousing and Mining, Classification B*, 1-7, Wang, John (Ed., 2nd Edition).



- Militello, L. and Hutton, R. (2000). *Applied Cognitive Task Analysis (ACTA): A Practitioner's Toolkit for Understanding Cognitive Task Demands*. In Annett and Stanton (2000): page 90-113.
- Neal, R.M. (1998). *Philosophy of Bayesian Inference*. Retrieved May 25, 2011, from <http://www.cs.toronto.edu/~radford/res-bayes-ex.html>
- Shepherd, A. (2001). *Hierarchical Task Analysis*. New York: Taylor & Francis.
- Stanton, N.A. and Young, M.S. (1999). *A Guide to Methodology in Ergonomics: Designing for Human Use*. London: Taylor and Francis.
- Swezey, R.W. and Pearlstein, R.B. (2001). Selection, Training, and Development of Personnel. In Salvendy, G. (Eds). *Handbook of Industrial Engineering: Technology and Operations Management (3<sup>rd</sup> Ed)*. Chapter 35 (pp. 920-947). New York: John Wiley & Sons.
- Westrenen, F. (2010). *Cognitive Work Analysis and The Design of User Interfaces*. Cogn Tech Work (2011) 13:31–42. Springer.
- Wickens, T.D., (2002). *Elementary Signal Detection Theory*. New York: Oxford University Press.
- Zhu, W. and Lehto, M.R. (1999). Decision support for indexing and retrieval of information in hypertext system. *International Journal of Human Computer Interaction*, 11, 349-371.

