



Forum Guru Besar
Institut Teknologi Bandung



Forum Guru Besar
Institut Teknologi Bandung

Orasi Ilmiah Guru Besar
Institut Teknologi Bandung

Profesor Dwi H. Widyantoro

**PEMBELAJARAN MESIN:
KONSEP DAN APLIKASINYA
PADA KLASIFIKASI, EKSTRAKSI
DAN PERINGKASAN INFORMASI**

18 Oktober 2017
Aula Timur Institut Teknologi Bandung

**Orasi Ilmiah Guru Besar
Institut Teknologi Bandung**
18 Oktober 2017

Profesor Dwi H. Widyantoro

**PEMBELAJARAN MESIN:
KONSEP DAN APLIKASINYA
PADA KLASIFIKASI, EKSTRAKSI DAN
PERINGKASAN INFORMASI**



Forum Guru Besar
Institut Teknologi Bandung

Hak cipta ada pada penulis

Judul: PEMBELAJARAN MESIN: KONSEP DAN APLIKASINYA PADA KLASIFIKASI, EKSTRAKSI DAN PERINGKASAN INFORMASI
Disampaikan pada sidang terbuka Forum Guru Besar ITB,
tanggal 18 Oktober 2017.

Hak Cipta dilindungi undang-undang.

Dilarang memperbanyak sebagian atau seluruh isi buku ini dalam bentuk apapun, baik secara elektronik maupun mekanik, termasuk memfotokopi, merekam atau dengan menggunakan sistem penyimpanan lainnya, tanpa izin tertulis dari Penulis.

UNDANG-UNDANG NOMOR 19 TAHUN 2002 TENTANG HAK CIPTA

1. Barang siapa dengan sengaja dan tanpa hak mengumumkan atau memperbanyak suatu ciptaan atau memberi izin untuk itu, dipidana dengan pidana penjara paling lama 7 (tujuh) tahun dan/atau denda paling banyak Rp 5.000.000.000,00 (lima miliar rupiah).
2. Barang siapa dengan sengaja menyiarkan, memamerkan, mengedarkan, atau menjual kepada umum suatu ciptaan atau barang hasil pelanggaran Hak Cipta atau Hak Terkait sebagaimana dimaksud pada ayat (1), dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/atau denda paling banyak Rp 500.000.000,00 (lima ratus juta rupiah).

Hak Cipta ada pada penulis
Data katalog dalam terbitan

Dwi H Widyantoro
PEMBELAJARAN MESIN: KONSEP DAN APLIKASINYA PADA KLASIFIKASI,
EKSTRAKSI DAN PERINGKASAN INFORMASI
Disunting oleh Dwi H Widyantoro

Bandung: Forum Guru Besar ITB, 2017
vi+52 h., 17,5 x 25 cm
ISBN 978-602-6624-04-8
1. Informatika 1. Dwi H Widyantoro

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah SWT, yang berkat rahmat dan karunia-Nya, naskah orasi ilmiah ini dapat diselesaikan. Penulis juga mengucapkan terima kasih kepada Forum Guru Besar Institut Teknologi Bandung yang telah memperkenankan penulis menyampaikan orasi ilmiah sebagai pertanggung-jawaban akademik Guru Besar. Naskah orasi ini pada dasarnya memperkenalkan serta menyoroti perjalanan penelitian penulis dalam menekuni bidang pembelajaran mesin.

Bagian awal tulisan memaparkan motivasi dan relevansi pembelajaran mesin dalam penyelesaian persoalan yang sering dihadapi pada masa sekarang ini, khususnya yang terkait dengan 'ledakan' informasi. Selanjutnya diperkenalkan secara singkat konsep dasar dan contoh-contoh pendekatan untuk memberikan gambaran ringkas terkait bidang ini. Rangkuman penelitian penulis dalam bidang pembelajaran mesin dijabarkan dengan contoh-contoh kegiatan pengembangan dan penerapan pembelajaran mesin untuk menyelesaikan persoalan klasifikasi, ekstraksi dan peringkasan informasi, khususnya dalam domain teks.

Semoga tulisan ini dapat menambah wawasan dan memperluas cakrawala bagi para pembaca.

Bandung, 2 Oktober 2017

Prof. Dwi H. Widyantoro

DAFTAR ISI

KATA PENGANTAR	iii
DAFTAR ISI	v
1. PENDAHULUAN	1
2. KATEGORI PEMBELAJARAN MESIN	2
3. PENDEKATAN DALAM PEMBELAJARAN MESIN	5
4. PEMBELAJARAN DINAMIKA PREFERENSI PENGGUNA	9
4.1 Pembelajaran dengan Representasi Tiga-Deskriptor	10
4.2 Pembelajaran dengan Masukan yang Terbatas	13
5. PEMBELAJARAN MULTI-STRATEGI BERBASIS <i>WINNER-TAKES-ALL (WTA)</i>	15
5.1 Penerapan Multi-Strategi berbasis WTA untuk Ekstraksi Informasi	17
5.2 Penerapan Multi-Strategi berbasis WTA untuk Klasifikasi Retorik Kalimat	20
6. PERINGKASAN OTOMATIS BERBASIS KEBUTUHAN PENGGUNA	21
6.1 Peringkasan Kumpulan Makalah Ilmiah berbasis Spesifikasi Ringkasan	23
6.2 Visualisasi Dinamis Ringkasan Makalah Ilmiah berbasis Jaringan Sitasi	27
6.3 Peringkasan Opini	29
7. PENUTUP	31
8. UCAPAN TERIMA KASIH	32

9. DAFTAR PUSTAKA	33
CURRICULUM VITAE	37

PEMBELAJARAN MESIN: KONSEP DAN APLIKASINYA PADA KLASIFIKASI, EKSTRAKSI DAN PERINGKASAN INFORMASI

1. PENDAHULUAN

We are drowning in information but starved for knowledge. ----- John Naisbitt

Saat ini kita hidup dalam era informasi yang berlimpah dan berlebihan. Sebagai contoh, jumlah halaman web yang diindeks oleh search engine terbesar saat ini, GOOGLE, mencapai 47 milyar. Total durasi video yang diunggah ke YOUTUBE setiap menitnya sekitar 300 jam. Jumlah transaksi yang ditangani oleh salah satu toko serba ada terbesar di Amerika Serikat, WALMART, sebanyak lebih dari 1 Juta per jamnya.

Kemudahan yang disediakan oleh teknologi informasi dan komunikasi saat ini sangat mempermudah sebuah organisasi atau individu untuk mendapatkan data dalam jumlah yang sangat besar. Agar data tersebut bernilai, maka data tersebut perlu dianalisa untuk mendapatkan wawasan baru, pengetahuan atau informasi yang berguna bagi organisasi tersebut dalam pengambilan keputusan strategis yang lebih baik. Namun apabila data tersebut sedemikian banyaknya sehingga melebihi kapasitas manusia untuk memprosesnya, maka proses analisis data menjadi persoalan tersendiri.

Fenomena 'kebanjiran' data ini memerlukan proses otomatis untuk analisis data, yang dapat diberikan oleh Pembelajaran Mesin (*Machine Learning*). Secara khusus, Pembelajaran mesin adalah sekumpulan metoda

yang dapat mendeteksi pola-pola (*patterns*) dalam data secara otomatis dan selanjutnya menggunakan pola yang ditemukan untuk memprediksi data-data baru yang belum pernah dijumpai sebelumnya. Sebagai contoh, Pembelajaran Mesin dapat diterapkan untuk mengidentifikasi pola atau karakteristik pemakai kartu kredit yang menguntungkan atau merugikan dari contoh-contoh ribuan bahkan jutaan profil pemakai kartu kredit. Pola yang diidentifikasi tersebut selanjutnya dapat dipakai untuk memprediksi apakah permintaan kartu kredit harus diterima (karena karakteristik calon nasabah memenuhi pola pemakai kartu kredit yang menguntungkan) atau harus ditolak (karena memenuhi pola pemakai kartu kredit yang merugikan). Saat ini, individu juga sudah banyak mendapatkan manfaat dari penerapan Pembelajaran Mesin untuk *Spam Filtering*, sehingga surel yang masuk secara otomatis dapat dikelompokkan ke berbagai kategori, termasuk yang tergolong *spam*.

Pada bagian berikutnya penulis akan memaparkan secara singkat berbagai konsep dasar dalam Pembelajaran Mesin serta pendekatan-pendekatan yang pernah dikembangkan oleh komunitas ini untuk menemukan pola secara otomatis, termasuk kelebihan dan kekurangannya. Selanjutnya akan diuraikan beberapa contoh kegiatan penelitian yang merepresentasikan kegiatan pengembangan keilmuan yang telah penulis lakukan serta penerapannya dalam klasifikasi, ekstraksi dan peringkasan informasi, khususnya dalam domain dokumen teks.

2. KATEGORI PEMBELAJARAN MESIN

Pembelajaran mesin secara inheren adalah bidang multi-disiplin.

Bidang ini banyak terinspirasi dan mengadopsi hasil-hasil yang telah dikembangkan dari disiplin ilmu-ilmu lainnya seperti kecerdasan buatan (*artificial intelligent*), probabilitas dan statistik, teori kompleksitas komputasional, teori informasi, teori kendali, psikologi, filosofi, neurobiologi serta bidang-bidang lainnya.

Terkait dengan jenis masukan yang tersedia untuk pembelajaran, ada tiga kategori utama dalam pembelajaran mesin. Kategori pertama sering dikenal dengan pendekatan **pembelajaran prediktif** atau **pembelajaran terawasi** (*Supervised Learning*). Tujuan dari pendekatan ini adalah untuk mendapatkan pemetaan dari masukan x ke luaran y , jika diberikan sekumpulan pasangan masukan-luaran $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$. Di sini \mathcal{D} disebut sebagai kumpulan data latih (*training set*) dan N adalah jumlah contoh data latih. Dalam bentuk yang paling sederhana, adalah vektor angka berdimensi D , yang merepresentasikan nilai masukan, misalnya *pendapatan* dan *umur* seseorang. Masukan ini sering disebut sebagai **fitur** (*features*) atau **atribut**, dan direpresentasikan dalam matriks X berukuran $N \times D$. Secara umum x_i dapat berupa objek berstruktur kompleks seperti gambar, kalimat, sekuen, graph dll.

Seperti halnya dengan masukan, luaran secara prinsip dapat berupa nilai apapun dan sering disebut sebagai **fitur target**. Meskipun demikian, y_i pada umumnya variabel kategoris atau nominal dari beberapa kumpulan berhingga $y_i = \{1, \dots, c\}$ (misal, *prospektif* atau *tidak prospektif*), atau berupa bilangan skalar bernilai riil (misal, *harga saham*). Pada saat y_i berupa variabel kategoris, maka persoalannya dikenal sebagai persoalan **klasifikasi** atau **pengenalan pola**. Misal, klasifikasi aplikasi kartu kredit

masuk kategori sebagai calon nasabah prospektif atau tidak. Sedangkan pada saat y_i bernilai riil, persoalannya dikenal sebagai **regresi**. Misal dalam memprediksi harga saham dari waktu ke waktu.

Kategori pembelajaran mesin yang kedua adalah **pembelajaran tidak terawasi** (*unsupervised learning*). Disini, yang ada hanya berupa masukan $\mathcal{D} = \{x_i\}_{i=1}^N$, kumpulan data tanpa label, dengan tujuan utamanya adalah mencari pola-pola yang 'menarik' dari data masukan tersebut. Persoalan dalam pengaturan ini sering dikenal dengan penemuan pengetahuan (*knowledge discovery*). Dibandingkan dengan pembelajaran terawasi, persoalan dalam penemuan pengetahuan relatif kurang terdefinisi dengan baik karena di awalnya pola yang 'menarik' itu tidak diketahui, yang tentunya berbeda dari satu domain ke domain lainnya. Salah satu contoh yang cukup populer dalam literatur adalah ditemukannya pola bahwa pria yang membeli diaper di toko serba ada di Amerika Serikat di awal malam hari juga cenderung membeli bir. Dengan pengetahuan ini, penempatan bir di dekat rak diaper membuat penjualan bir meningkat tajam.

Kategori ketiga dalam pembelajaran mesin adalah *reinforcement learning*. Pendekatan ini bertujuan untuk belajar menentukan aksi yang harus dilakukan pada lingkungan yang sesekali mendapatkan sinyal berupa penghargaan (*reward*) atau hukuman (*punishment*). Contoh sukses dari pendekatan ini adalah pada penerapannya untuk belajar memenangkan permainan (misal, *backgammon*, catur), dan belajar untuk melakukan navigasi robot otonom dalam lingkungan yang tidak dikenal.

Kategori lainnya adalah *semi-supervised learning*. Pembelajaran mesin

masuk dalam kategori ini apabila beberapa atau sebagian besar nilai luaran y_i dalam data latih $\mathcal{D} = \{x_i\}_{i=1}^N$ tidak diketahui. Situasi ini muncul pada jenis data yang jumlahnya sangat besar dan mudah di dapat, namun harga yang dibayar untuk memberi label (nilai luaran y_i) terlalu mahal. Sebagai contoh, memberi label jenis topik dari setiap halaman web yang ada di Internet.

3. PENDEKATAN DALAM PEMBELAJARAN MESIN

Saat ini sudah banyak metoda yang telah dikembangkan dalam pembelajaran mesin, baik yang diadopsi maupun terinspirasi atau dimodifikasi dari hasil-hasil dari bidang lainnya. Untuk memberikan gambaran, di sini akan dijabarkan sekilas metoda-metoda utama yang cukup populer dan sukses dalam penerapannya seperti *Decision Tree*, *Bayesian Learning*, *Neural Network*, *Support Vector Machine*, khususnya yang masuk dalam kategori *supervised learning*.

Dalam pembelajaran terawasi, diberikan data latih yang merupakan contoh masukan-keluaran $\mathcal{D} = \{x_i, y_i\}_{i=1}^N$. Tujuan pembelajaran adalah mencari hipotesis untuk estimasi nilai fungsi target (nilai y_i) untuk setiap x_i . Misal H merepresentasikan himpunan dari semua hipotesis yang mungkin untuk mengidentifikasi fungsi target. Jika $X = \{x_i\}_{i=1}^N, Y = \{y_i\}_{i=1}^N$, dan $h \in H$ merepresentasikan fungsi target $h: X \rightarrow Y$, maka tujuan dari pembelajaran terawasi adalah mencari suatu hipotesis h yang terbaik, yaitu hipotesis yang dapat memenuhi $h(x_i) = y_i$ untuk setiap $x_i \in X$. Asumsi dalam pembelajaran mesin ini adalah bahwa hipotesis yang dapat dipakai untuk perkiraan (*approximate*) fungsi target dengan baik pada himpunan

data latih dengan jumlah yang cukup besar akan juga dapat melakukan perkiraan fungsi target dengan baik pada data lainnya yang belum pernah terobservasi.

Hipotesis (fungsi pemetaan) yang dicari dalam proses pembelajaran mesin dapat direpresentasikan dalam berbagai macam bentuk. Sebagai contoh, hipotesis dapat berupa pohon keputusan, bobot-bobot pada jaringan syaraf tiruan, nilai probabilitas kemunculan atribut/fitur, aturan dengan sintaksis tertentu, atau nilai-nilai parameter pada beragam persamaan regresi atau pendekatan algoritmik untuk proses pembelajaran fungsi target. Proses pembelajaran untuk mendapat hipotesis yang terbaik dan selanjutnya pemanfaatannya untuk memprediksi data baru akan sangat bergantung pada jenis representasi hipotesis yang dipilih.

Inductive Decision Tree (ID3) Learning adalah metoda pembelajaran yang diterapkan untuk hipotesis yang berupa pohon keputusan. Simpul internal (*internal node*) dalam pohon keputusan berisi nama atribut data dan setiap cabang berisi nilai atribut. Simpul daun (*leaf node*) memuat nilai fungsi target (klasifikasi). Proses klasifikasi dilakukan dengan mensortir nilai-nilai atribut data baru mulai dari akar pohon sampai pada simpul daun, yang menginformasikan nilai target (hasil klasifikasi). Salah satu tujuan pembelajaran dalam *ID3 Learning* adalah mencari pohon terbaik yang paling pendek kedalamannya, sehingga mempercepat proses klasifikasi. Teori informasi banyak diadopsi dalam proses pencarian pohon keputusan yang terbaik.

Jaringan Saraf Tiruan (JST) merupakan metoda klasik dan sebagian terinspirasi berdasarkan pengamatan pada sistem pembelajaran biologis

yang terbangun dari jejaring kompleks neuron yang saling terhubung. JST memiliki satu lapisan masukan (untuk memasukkan data), satu atau lebih lapisan tersembunyi, dan satu lapisan luaran (untuk mengeluarkan hasil prediksi). Setiap lapisan terdiri dari satu atau lebih simpul-simpul unit/ Setiap unit pada suatu lapisan terhubung ke semua unit pada lapisan berikutnya dan kekuatan keterhubungannya ditandai dengan bobot koneksi, yang merupakan hipotesis yang dicari. Proses pelatihan pada JST dilakukan dengan mencari nilai setiap bobot koneksi, yang dilakukan secara iteratif, sehingga luarannya dapat memprediksi dengan benar nilai target dari contoh-contoh data latih. *BACKPROPAGATION* merupakan salah satu algoritma yang banyak dipergunakan untuk pelatihan JST, khususnya pada arsitektur JST dengan jumlah lapisan tersembunyi yang relatif sedikit.

Support Vector Machine (SVM) merupakan salah satu teknik terbaik pada saat ini untuk tugas klasifikasi. SVM pada prinsipnya mencari satu atau lebih *hyperplane* pada ruang berdimensi tinggi yang dapat memisahkan kelas-kelas data dengan margin yang maksimal. Hipotesis pada SVM adalah nilai koefisien (atau bobot) pada sekumpulan data yang merupakan *support vector*, yang menjadi basis untuk menentukan margin yang maksimal. SVM dikembangkan pertama kali oleh Vapnik pada tahun 1996, namun kepopulerannya mulai muncul pada era 1990-an pada saat diterapkan untuk persoalan klasifikasi. Berbagai variasi SVM saat ini telah banyak dikembangkan untuk mengatasi persoalan pada keragaman karakteristik data, misal data dengan label yang tidak seimbang, data yang sebagian tidak berlabel, dataset yang berukuran sangat besar atau data dengan dimensi sangat tinggi.

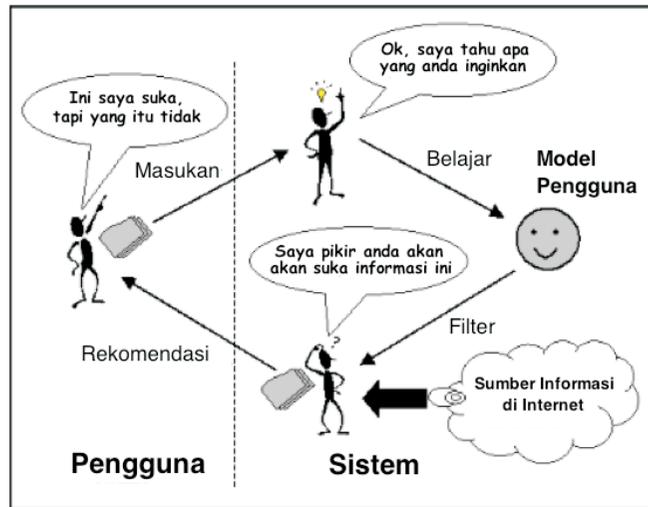
Bayesian Learning adalah pendekatan probabilistik yang juga banyak dipakai dan cukup efektif untuk tugas klasifikasi. Dengan mengadaptasi metoda Bayesian, *Bayes classifier* menghitung secara eksplisit probabilitas dari semua kategori yang akan diprediksi dan menentukan kelas fungsi target dari kategori dengan nilai probabilitas tertinggi. Hipotesis dari pendekatan ini berupa himpunan nilai probabilitas kelas dan kemunculan kombinasi atribut pada setiap kelas yang dikalkulasi dari data latih. Untuk persoalan nyata, banyak kombinasi nilai probabilitas yang tidak diketahui karena keterbatasan jumlah data latih. Kekurangan ini secara praktis dapat diatasi dengan menerapkan asumsi ketidakbergantungan antar atribut (*independent assumption*) sehingga semua nilai probabilitas yang dibutuhkan dapat dihitung. Pendekatan yang menerapkan asumsi ini dikenal dengan *Naïve Bayes Classifier*. Walaupun *independent assumption* tidak sepenuhnya benar sehingga nilai probabilitas yang dihitung tidak akurat, namun hasil klasifikasinya sangat efektif dan memberikan kinerja yang kompetitif dengan pendekatan lainnya yang lebih kompleks.

Keempat contoh di atas merepresentasikan pendekatan pembelajaran dengan teknik yang cukup bervariasi serta telah banyak diteliti dan digunakan baik di dunia akademik maupun industri. Pendekatan lainnya dalam pembelajaran terawasi adalah sekumpulan metoda dan variasinya yang masuk ke dalam kelompok *Ensemble Learning*, *Deep Learning*, *Instance-based Learning*, *Analytical Learning*, *Active Learning* dll. Setiap algoritma yang dikembangkan tersebut selalu didasarkan pada *inductive bias*, yaitu kumpulan asumsi yang mendasari kriteria pemilihan model klasifikasi dari suatu algoritma pembelajaran. *Inductive bias* ini dapat menginduksi model yang melakukan generalisasi sampai di luar data

latih. Jika kurang tepat, *Inductive bias* ini dapat menuntun kepada kesalahan prediksi karena model klasifikasi yang dihasilkan *underfitting* atau *overfitting*. *Underfitting* terjadi apabila model yang dihasilkan terlalu sederhana untuk merepresentasikan keterkaitan antara atribut masukan dengan fungsi target. Sedangkan *overfitting* terjadi pada saat algoritma menghasilkan model yang terlalu kompleks serta sangat spesifik untuk data latih sehingga kemampuan generalisasinya di luar data latih dikorbankan. Algoritma pembelajaran yang baik adalah algoritma yang dapat menghindari kedua permasalahan di atas, dan ini merupakan persoalan utama dalam pengembangan algoritma pembelajaran mesin.

4. PEMBELAJARAN DINAMIKA PREFERENSI PENGGUNA

Pesatnya perkembangan Internet di era 1990-an telah banyak mendorong perkembangan aplikasi Internet baru yang menggunakan informasi dari penggunaannya sebagai basis untuk meningkatkan fungsionalitas dari sistem. Salah satu di antaranya adalah sistem berbasis Web yang menawarkan layanan personal dalam penyediaan informasi berita terkini yang diminati. Dengan layanan ini, pengguna tidak perlu mencari-cari berita yang ingin dibaca karena sistem akan secara otomatis merekomendasikan berita-berita yang bakal diminati oleh pengguna. Untuk dapat memberikan layanan seperti itu, sistem harus selalu memutakhirkan informasi tentang preferensi penggunaannya sehingga akurat dalam merekomendasikan bacaan yang diminati.



Gambar 1. Ilustrasi sistem pemberian rekomendasi.

Gambar 1 memberikan ilustrasi cara kerja sistem pemberian rekomendasi yang dimaksud. Pertama sistem akan merekomendasikan sekumpulan berita yang diprediksi akan diminati penggunanya. Berdasarkan rekomendasi yang diberikan, pengguna akan memberikan masukan kepada sistem terkait berita-berita yang disukai atau tidak disukai. Sistem selanjutnya belajar dari masukan tersebut dan memutakhirkan model profil minat pengguna untuk basis pemberian rekomendasi selanjutnya. Proses ini berlanjut secara kontinyu sehingga informasi tentang preferensi minat pengguna terjaga kemutakhirannya.

4.1 Pembelajaran dengan Representasi Multi Tiga-Deskriptor (MTDR)

Dalam dunia nyata, topik yang diminati untuk dibaca tidak statis,

namun dapat mengalami perubahan. Terkait dengan penelitian ini, penulis berkontribusi dalam mengembangkan algoritma pembelajaran mesin yang dapat beradaptasi secara fleksibel terhadap minat baca pengguna yang dinamik, yang dapat berubah dari waktu ke waktu. Ini masuk kategori persoalan klasifikasi yang fungsi targetnya dapat berubah sewaktu-waktu.

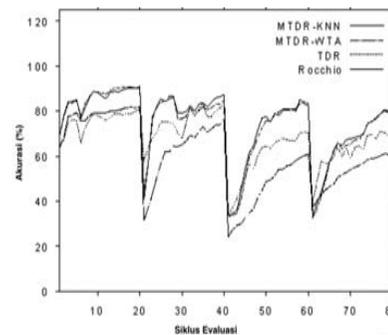
Agar dapat beradaptasi dengan perubahan yang cepat, penulis mengusulkan model representasi tiga-deskriptor (TDR) untuk merepresentasikan minat baca pengguna. Model yang diusulkan memiliki deskriptor jangka panjang untuk mengelola kategori minat secara umum, dan memiliki deskriptor jangka pendek untuk mengikuti perkembangan kategori minat yang terkini dan berubah lebih cepat. Deskriptor jangka panjang dibangun secara bertahap sehingga kinerjanya meningkat secara konsisten sejalan dengan banyaknya masukan yang diperoleh. Model minat jangka pendek terdiri dari deskriptor positif (untuk menyimpan deskriptor kategori yang disukai) dan deskriptor negatif (deskriptor kategori yang tidak disukai). Deskriptor jangka pendek ini dibangun secara progresif agar dapat bereaksi dengan cepat terhadap perubahan kategori minat yang sedang terjadi.

Sebuah algoritma pembelajaran selanjutnya dikembangkan berdasarkan representasi tiga-deskriptor tersebut sehingga mendapatkan akurasi yang tinggi untuk dapat mengenali kategori minat jangka panjang dan beradaptasi dengan cepat pada kategori minat yang berubah dalam jangka pendek. Model minat jangka-panjang dan jangka-pendek ini menawarkan manfaat yang saling melengkapi. Model representasi tiga-deskriptor ini mencoba mengkombinasikan manfaat dari kedua model

dan menghilangkan kekurangan masing-masing sebanyak mungkin. Maksud ini dapat dicapai dengan belajar minat jangka panjang dan minat jangka pendek secara terpisah, dan mengoptimalkan penggunaannya dalam memprediksi minat pada bacaan baru. Model ini selanjutnya diperluas menjadi representasi multi tiga-deskriptor (MTDR) agar dapat menangkap domain kategori minat yang lebih luas.

Tabel 1: Rangkuman Hasil Eksperimen Semua model untuk semua task.

	Rata-rata Akurasi (%)			
	Rocchio	TDR	MTDR-WTA	MTDR-KNN
T ₁	79.9	87.7	82.3	86.8
T ₂	75.5	81.5	81.9	83.7
T ₃	75.5	79.8	85.2	86.4
T ₄	59.7	68.0	74.0	74.5



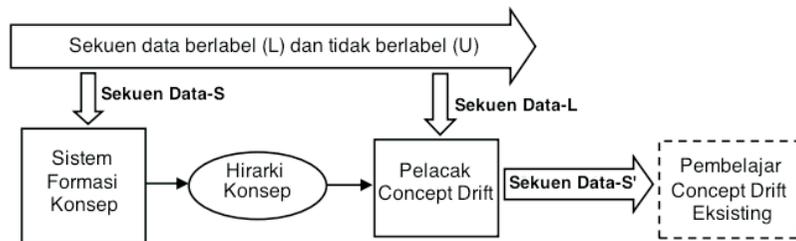
Gambar 2. Kinerja Pembelajaran dengan Perubahan Signifikan

Hasil eksperimen untuk pembelajaran minat yang berubah dari waktu ke waktu menunjukkan keunggulan algoritma yang berbasis MTDR, seperti yang terlihat pada Gambar 2. Kecepatan pemulihan akurasi akibat adanya perubahan pada pada model TDR masih baik, tetapi karena keterbatasan ruang representasi, akurasi akhirnya tidak setinggi model MTDR. Seperti yang diperlihatkan pada Tabel 1, kelebihan pendekatan berbasis MTDR terlihat lebih jelas pada nilai rata-rata akurasi ketika tingkat kesulitan persoalan semakin meningkat dari yang paling mudah (T₁) ke yang paling sulit (T₄).

4.2 Pembelajaran dengan Masukan yang Terbatas

Algoritma pembelajaran yang dikembangkan dari model representasi MTDR di atas telah berhasil beradaptasi mengikuti perubahan kategori minat, dengan asumsi jumlah masukan terkait dokumen yang disukai atau tidak disukai relatif cukup banyak. Asumsi ini secara realistis hanya dapat dipenuhi oleh sebagian kecil pengguna yang dapat meluangkan waktu untuk memberikan contoh-contoh positif dan negatif yang mencukupi. Sebagian besar pengguna masih cenderung enggan, dan walaupun mau, sangat sedikit contoh-contoh yang dapat diberikan. Kajian teoritis menunjukkan bahwa pembelajaran konsep yang berubah-ubah akan menjadi lebih sulit apabila contoh yang diberikan sangat sedikit.

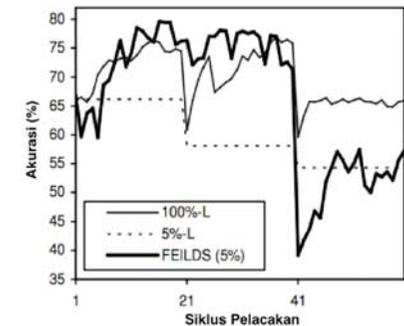
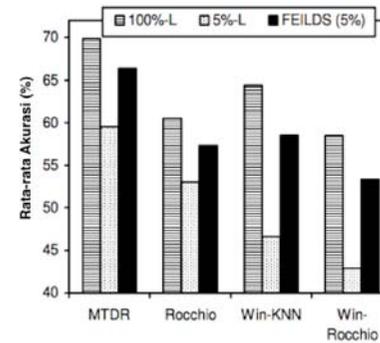
Kontribusi penulis berikutnya adalah mengatasi persoalan kurangnya masukan dari pengguna dalam proses pembelajaran konsep yang berubah-ubah. Penulis mengusulkan sebuah model komputasional yang disebut FEILDS (*Framework for Extending Incomplete Labeled Data Stream*). Salah satu masukan dari model ini adalah sekuen data berlabel yang diasumsikan dengan jumlah yang minimal (misal, masukan pengguna yang sangat terbatas) yang menjadi masukan standar dari sistem. Sistem selanjutnya membangkitkan sekuen data baru yang lebih panjang dengan memanfaatkan data-data tidak berlabel yang relevan. Algoritma pembelajaran yang sudah ada (seperti MTDR) selanjutnya belajar dari sekeuen data yang baru tersebut. Dengan demikian, FEILDS secara modular meningkatkan kemampuan algoritma pembelajaran yang sudah ada dengan memodifikasi masukan aslinya tanpa memodifikasi algoritma pembelajarannya.



Gambar 3. Arsitektur FEILDS.

Gambar 3 memperlihatkan model komputasional dari FEILDS. Secara umum, FEILDS menggunakan data tidak berlabel yang relevan untuk mengkompensasi kurangnya data berlabel (masukan pengguna), tetapi untuk pembelajaran konsep yang berubah-ubah secara dinamik. Masukan dari sistem adalah sekuen data S yang terdiri dari data berlabel L dan tidak berlabel U dengan jumlah data berlabel jauh sangat sedikit. Inti dari pendekatan ini adalah mengidentifikasi sekuen data berlabel $L_R \subseteq L$ yang masih relevan dan memperluas L_R dengan data tidak berlabel yang relevan. Sistem selanjutnya menambahkan data tidak berlabel yang relevan ke dalam L_R , atau $\{X_i \cup U \mid X_i \in L_R\}$, untuk menghasilkan sekuen data baru S' dengan memberikan label baru untuk setiap data tidak berlabel $X_i \in U$ dengan labelnya X_i , yang menjadi basis perluasan data tidak berlabel. Seperti yang diperlihatkan pada gambar, sistem ini memanfaatkan sebuah *concept formation system* (CFS) untuk pembentukan hirarki konsep (*concept hierarchy*) dalam rangka mengidentifikasi data tidak berlabel yang relevan. CFS pada dasarnya adalah algoritma untuk pembangunan cluster berhirarki secara incremental dari sekuen data, yang merupakan kontribusi lain dari penulis dalam *unsupervised learning*.

Tabel 2.: Rata-rata Kinerja FEILDS



Gambar 4. Kinerja pelacakan FEILDS

Tabel 2 memperlihatkan kinerja dari FEILDS dengan data yang jumlah labelnya direduksi menjadi 5%. Sistem dengan 100% data mendapatkan label data yang lengkap dan berfungsi sebagai batas atas. Semua algoritma dengan 5% data berlabel kinerjanya sangat berkurang saat mendapatkan data dengan label yang minimal. Seperti yang diperlihatkan pada gambar, FEILDS dapat meningkatkan kinerja secara signifikan dari semua algoritma yang mendapatkan jumlah data berlabel minimal (5%). Gambar 4 menunjukkan kinerja pelacakan FEILDS yang dalam beberapa kasus tingkat kecepatan pemulihannya lebih baik dari sistem yang mendapatkan 100% data berlabel.

5. PEMBELAJARAN MULTI-STRATEGI BERBASIS WINNER-TAKES-ALL (WTA)

Pembelajaran multi-strategi adalah sebuah pendekatan pembelajaran untuk menyelesaikan persoalan yang memiliki tingkat kompleksitas

bervariasi dengan menggunakan beragam bias dan paradigma komputasional dalam proses pembelajaran. Pendekatan multi-strategi ini pada umumnya bekerja dengan menjalankan dua atau lebih *classifier*, dan mengambil keputusan akhir dengan mengkombinasikan prediksi dari semua *classifier* yang dilibatkan. Dengan mengasumsikan keragaman dari paradigma pembelajaran, bagaimana derau (*noise*) pada data latih ditangani dan hasil prediksi dari setiap paradigma digabungkan, kinerja dari multi-*classifier* diharapkan akan lebih baik dari kinerja *classifier* tunggal. Pendekatan ini pada umumnya bekerja dengan baik pada saat semua atau sebagian besar *classifier* yang dipakai dapat dengan tepat mengidentifikasi konsep kuncinya. Namun, apabila persoalannya sangat sulit sehingga sebagian besar *classifier* gagal memprediksi luaran, maka hasil akhir juga cenderung salah.

Untuk mengatasi persoalan di atas, penulis menggunakan satu *classifier*, tetapi yang terbaik, pada saat proses prediksi. Prinsip utamanya adalah menyerahkan pekerjaan ke pakar/spesialis terbaik sesuai dengan bidangnya. Apabila sebuah *classifier* diketahui memberikan kinerja terbaik (dibandingkan *classifier-classifier* lainnya yang dipakai) untuk memprediksi sebuah kelas atau kategori selama fase pelatihan, maka *classifier* terbaik tersebut akan digunakan untuk memprediksi pada kategori tersebut (di sini istilah *Winner-takes-all* diambil). Dengan demikian, *classifier* lainnya yang kurang baik kinerjanya tidak akan mengganggu prediksi dari *classifier* terbaik.

Diberikan himpunan algoritma pembelajaran basis $\mathcal{L} = \{\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_n\}$ dan himpunan kategori $\mathcal{C} = \{c_1, c_2, \dots, c_n\}$ yang akan diprediksi. Pembelajaran multi-strategi dengan pendekatan *Winner-takes-*

all secara umum dapat diringkas sebagai berikut:

1. Pelatihan awal: cari algoritma pembelajaran terbaik $\mathcal{L}_i \in \mathcal{L}$ untuk klasifikasi setiap kategori tertentu C_i . Proses ini menggunakan *k-fold cross validation* untuk mengukur kinerja dari setiap algoritma.
2. Pelatihan akhir: lakukan pelatihan ulang *classifier* terbaik untuk setiap kategori C_i dengan data latih yang lengkap.
3. Saat prediksi data baru, gunakan *classifier* terbaik untuk kategori C_i pada saat mengidentifikasi apakah data tersebut masuk kategori C_i .

Dua sub-seksi berikut memperlihatkan efektivitas pembelajaran multi-strategi berbasis *Winner-takes-all* untuk persoalan ekstraksi informasi dan klasifikasi retorik kalimat.

5.1 Penerapan Multi-Strategi berbasis WTA untuk Ekstraksi Informasi

Ekstraksi informasi adalah proses otomatis untuk mendapatkan informasi spesifik yang sudah ditentukan sebelumnya dari dokumen yang tidak terstruktur seperti teks dan halaman Web. Informasi yang diekstrak dapat berupa entitas bernama (misal, nama orang, nama tempat, nama organisasi dll), atau berupa relasi, yang menggambarkan relasi antar entitas, sebagai contoh, CEO dari sebuah perusahaan atau CEO (*orang, perusahaan*). Pada penelitian ini, persoalan ekstraksi informasi dikonversi menjadi persoalan klasifikasi token. Dalam konteks ini, teks yang akan diekstrak dipandang sebagai sekuen kata, dan setiap kata merupakan token yang akan diklasifikasi apakah termasuk dalam kategori informasi yang akan diekstrak atau tidak.

Pada ekstraksi informasi berbasis klasifikasi, setiap token diidentifikasi fitur-fitur dan labelnya (masuk ke kategori tertentu atau tidak) untuk proses pembelajaran. Algoritma klasifikasi selanjutnya dapat diterapkan untuk membangun model klasifikasi dari setiap kategori yang akan diekstrak. Proses ekstraksi dilakukan dengan mengidentifikasi kelas-kelas token dari suatu teks baru, dan hasil ekstraksi akhir diambil dari sekuen token dengan kategori informasi yang sama.

Pada penelitian ini, untuk setiap kategori informasi dipergunakan dua *classifier*: (1) yang pertama untuk mengidentifikasi awal tanda (tag) atau tidak, dan (2) yang kedua untuk mengidentifikasi akhir tanda (tag) atau tidak. Fitur token terdiri dari jenis kata (benda, kerja, sifat dll), ortografi (kapital, huruf besar, huruf kecil dll), jenis token (kata, numerik, simbol dll), *lemma* (bentuk dasar token), ukuran jendela dan *lookup*. Ukuran window menentukan jumlah token sebelum dan sesudahnya yang masih dipertimbangkan. Sedangkan fitur *lookup* berisi nilai yang diasosiasikan dengan token berdasarkan isi kamus yang sudah ditentukan sebelumnya. Sebagai contoh, token “bandung” berdasar isi kamus diasosiasikan sebagai nama kota.

Tabel 3 & 4 memperlihatkan hasil eksperimen penerapan metode Multi-strategi berbasis WTA pada persoalan ekstraksi informasi dengan menggunakan data *Corporate Acquisition* (terdiri dari 600 artikel) dan data *Job Posting* (300 pesan *newsgroup*). Dibandingkan dengan metode ekstraksi informasi yang lain, pendekatan dengan Multi-Strategi berbasis WTA lebih unggul dalam mengekstraksi beragam kategori informasi.

Table 3. Perbandingan kinerja F-measure pada data *Job Posting*.

	RAPIER	LP2	SNOWIE	Mullti-Strategi		Multi-Strategi berbasis WTA
				Voting	Bagging	
<i>application</i>	0.69	0.78	0.61	0.17	0.20	0.74 (PA)
<i>Area</i>	0.42	0.67	0.52	0.06	0.08	0.57 (PA)
<i>City</i>	0.90	0.93	0.89	0.74	0.50	0.95 (SVM)
<i>Company</i>	0.70	0.72	0.75	0.62	0.34	0.82 (PA)
<i>Country</i>	0.93	0.81	0.95	0.64	0.57	0.59 (PA)
<i>desired_degree</i>	0.72	0.65	0.61	0.14	0.06	0.74 (PA)
<i>desired_exp</i>	0.87	0.60	0.79	0.71	0.79	0.86 (SVM)
<i>Id</i>	0.97	1.00	0.99	0.73	0.57	0.99 (SVM)
<i>Language</i>	0.81	0.91	0.82	0.37	0.39	0.88 (PA)
<i>Platform</i>	0.72	0.80	0.74	0.33	0.30	0.82 (PA)
<i>post_date</i>	0.99	0.99	0.99	0.97	0.97	0.99 (SVM)
<i>Recruiter</i>	0.68	0.81	0.85	0.57	0.59	0.87 (PA)
<i>req_degree</i>	0.81	0.85	0.83	0.48	0.37	0.86 (PA)
<i>req_exp</i>	0.67	0.69	0.84	0.63	0.62	0.81 (SVM)
<i>Salary</i>	0.67	0.63	0.73	0.49	0.25	0.84 (PA)
<i>State</i>	0.90	0.85	0.92	0.61	0.40	0.92 (SVM)
<i>Title</i>	0.40	0.44	0.53	0.27	0.16	0.69 (SVM)
Rata-rata	0.75	0.77	0.79	0.50	0.42	82,1

Tabel 4. Kinerja F-measure pada data *Corporate Acquisition*.

Kategori Informasi	RAPIER	SRV	ELIE	Mullti-Strategi		Multi-Strategi berbasis WTA
				Voting	Bagging	
<i>acqabr</i>	0.26	0.38	0.40	0.31	0.24	0.57 (SVM)
<i>acqloc</i>	0.24	0.22	0.34	0.05	0.25	0.47 (SVM)
<i>acquired</i>	0.29	0.39	0.43	0.06	0.22	0.51 (SVM)
<i>dramt</i>	0.39	0.62	0.59	0.09	0.34	0.65 (PA)
<i>purchabr</i>	0.24	0.48	0.29	0.28	0.39	0.49 (SVM)
<i>purchaser</i>	0.28	0.45	0.46	0.22	0.33	0.52 (PA)
<i>seller</i>	0.15	0.23	0.16	0.00	0.23	0.22 (SVM)
<i>sellerabr</i>	0.09	0.25	0.13	0.00	0.24	0.21 (SVM)
<i>status</i>	0.41	0.47	0.50	0.12	0.25	0.53 (PA)
Rata-rata	0.27	0.41	0.39	0.13	0.28	0.46

5.2 Penerapan Multi-Strategi berbasis WTA untuk Klasifikasi Retorik Kalimat

Retorik adalah informasi niat/maksud yang ingin disampaikan oleh seorang penulis kepada pembacanya. Dalam makalah ilmiah, retorik kalimat ini merupakan kalimat-kalimat penting yang sering dicari pembaca, seperti kalimat yang menyatakan tujuan penelitian, kebaharuan dalam makalah, metode yang diusulkan, perbandingan dengan metode lain, hasil-hasil yang diperoleh, kesimpulan dari hasil penelitian ataupun rencana penelitian selanjutnya. Proses identifikasi retorik kalimat ini merupakan tahapan penting dalam proses peringkasan, khususnya untuk peringkasan makalah ilmiah.

Pada persoalan ini, proses klasifikasi dilakukan pada tingkatan kalimat. Untuk proses pelatihan dan klasifikasi, setiap kalimat akan diidentifikasi fitur-fiturnya. Penelitian ini mengadopsi fitur yang diusulkan oleh Teufel yang terdiri dari fitur konten, lokasi absolut kalimat, struktur eksplisit, panjang kalimat, sintaksis kata kerja, sitasi, ekspresi *formulaic* dan fitur *agentivity*. Fitur yang diusulkan oleh Merity juga dipakai seperti penghitung untuk seksi, lokasi dan paragraf. Selain itu penelitian ini juga menambahkan fitur baru seperti konten abstrak dan kemunculan *qualifying adjective*. Fitur konten diambil dari n kata terbaik yang telah diberi bobot dengan menggunakan TF-IDF.

Tabel 5 menampilkan hasil eksperimen penerapan multi-strategi berbasis WTA untuk persoalan klasifikasi retorik kalimat. Hasil eksperimen ini diperoleh dengan menggunakan data yang berjumlah 10.877 kalimat teranotasi yang diambil dari 50 makalah ilmiah. Untuk klasifikasi

retorik kalimat ini, pendekatan multi strategi divariasi dengan menggunakan *classifier* basis yang sama (HO-*Homogeneous*) dan menggunakan basis *classifier* yang beragam (HE-*Heterogeneous*). Seperti yang diperlihatkan pada tabel, pendekatan multi-strategi berbasis WTA lebih unggul dibandingkan pendekatan multi-classifier lainnya.

Tabel 5. Kinerja pada klasifikasi retorik kalimat.

Classifiers Basis	Akurasi (%)			
	Multi-label	Ensemble	Multi-WTA-HO ¹	Multi-WTA-HE ²
SVM	51.0	50.1	80.8	
NB	47.8	48.7	79.5	
C4.5	46.8	47.3	80.6	
Logistic	51.1	52.4	82.4	
MLP	31.1	49.5	78.2	79.6
1-NN	38.2	38.2	76.7	
PART	40.5	42.4	78.6	
RT	35.1	32.8	77.0	
RC	47.5	47.5	80.8	

¹Multi-Homogeneous Classifier

²Multi-Heterogeneous Classifier

6. PERINGKASAN INFORMASI SESUAI KEBUTUHAN PENGGUNA

Peringkasan adalah proses transformasi informasi dari versi aslinya ke bentuk yang lebih pendek/ringkas dengan tetap mempertahankan substansi informasi yang penting. Dalam era ketersediaan informasi yang berlebih saat ini, peringkasan informasi secara otomatis memiliki peran yang penting karena akan mempermudah kita mencerna beragam informasi yang ada. Usaha untuk melakukan peringkasan informasi

sudah dilakukan oleh peneliti pendahulu sejak era 1950-an dan terus berkembang sesuai dengan perkembangan kebutuhan akan peringkasan informasi yang semakin bervariasi.

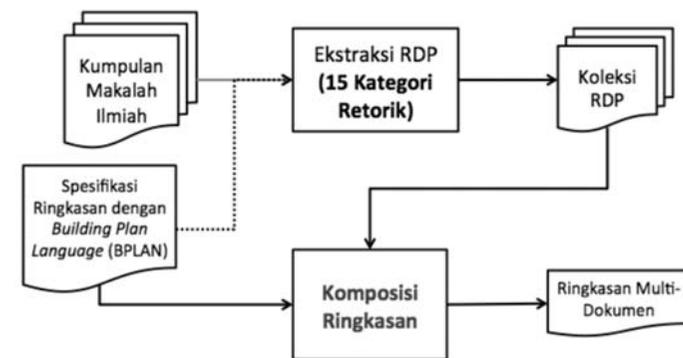
Seperti halnya akan kebutuhan informasi yang berbeda-beda, bentuk peringkasan yang dibutuhkan pun juga berbeda dari satu pengguna ke pengguna lainnya. Dalam dunia penelitian, sebagai contoh, seorang peneliti yang sedang melakukan studi literatur ingin melihat bagaimana perkembangan metoda yang dipakai untuk menyelesaikan suatu persoalan dari waktu ke waktu, yang lainnya ingin mengetahui domain apa saja yang pernah menjadi aplikasi dari sebuah metode, atau ingin mengetahui trend perkembangan terkini untuk suatu topik tertentu. Contoh lainnya adalah dalam dunia *e-commerce* di mana peran opini sangat penting untuk memutuskan apakah seseorang akan membeli produk atau tidak. Dengan banyaknya opini tentang suatu produk yang didapat dari Internet, pembaca akan semakin dipusingkan dengan pendapat yang bervariasi. Sementara itu, pembaca juga memiliki preferensi sendiri terkait kriteria apa saja yang penting dalam memilih suatu produk. Dengan demikian, untuk kedua contoh dunia yang berbeda tersebut, peringkasan yang sifatnya jenerik, walaupun dalam kadar tertentu cukup membantu, bukanlah bentuk peringkasan yang betul-betul dibutuhkan.

Pada bagian ini akan dijabarkan tiga jenis kegiatan dan kontribusi penelitian dalam pengembangan sistem peringkasan otomatis yang dapat disesuaikan dengan kebutuhan pengguna. Yang pertama dalam peringkasan otomatis multi-dokumen untuk makalah ilmiah. Penelitian lainnya adalah peringkasan makalah ilmiah dalam bentuk visualisasi

jejaring sitasi yang dinamik. Sedang penelitian yang ketiga adalah peringkasan opini.

6.1 Peringkasan Kumpulan Makalah Ilmiah berbasis Spesifikasi Ringkasan

Penelitian peringkasan kumpulan makalah ilmiah ini dimotivasi oleh beberapa hal. Pertama, membantu pembaca dalam memahami kumpulan makalah yang jumlahnya banyak serta adanya informasi yang berulang. Kedua, peringkasan makalah ilmiah saat itu masih lebih banyak berfokus pada peringkasan abstrak dengan banyaknya informasi yang terbatas serta menghasilkan struktur ringkasan yang seragam, yang mungkin kurang sesuai dengan kebutuhan pengguna. Selain itu, masih adanya keterbatasan representasi kebutuhan informasi dari sistem peringkasan yang dapat disesuaikan dengan kebutuhan pengguna.



Gambar 5. Sistem peringkasan multi-dokumen berbasis spesifikasi ringkasan.

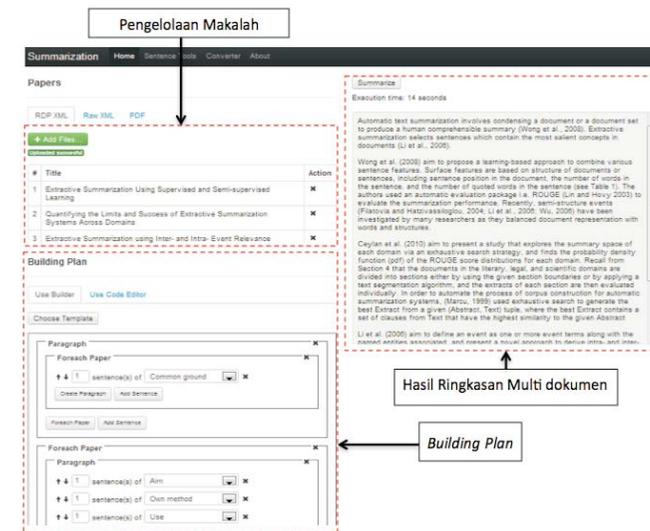
Gambar 5 menyajikan ilustrasi umum dari sistem peringkasan kumpulan makalah ilmiah yang telah dikembangkan. Salah satu inti dari sistem yang membedakannya dari sistem lainnya yang mirip adalah adanya spesifikasi ringkasan yang berbasis BPLAN (*Building Plan Language*), yaitu sebuah bahasa spesifik untuk mendefinisikan spesifikasi ringkasan yang diinginkan oleh penggunanya. Dengan BPLAN, pengguna dapat menyusun komposisi ringkasan yang diinginkan serta mendefinisikan proses perbaikan kalimat (*surface repair*) dari ringkasan yang dihasilkan. Sebagai contoh, dengan BPLAN dapat ditentukan jenis retorik kalimat apa saja yang akan muncul dalam sebuah paragraf dan bagaimana urutan serta pengulangannya. Selain itu, hasil ringkasan multi dokumen dapat ditingkatkan keterbacaannya dengan proses *surface repair*, misal dengan mengganti kata ganti orang (*We*) atau frasa 'In this paper' dengan nama penulisnya.

Komponen utama lainnya yang penting adalah kumpulan *Rhetorical Document Profile* (RDP) dari setiap dokumen dalam kumpulan makalah ilmiah yang akan diringkas. RDP dokumen berisi kumpulan retorik kalimat yang telah diidentifikasi dari suatu dokumen (makalah ilmiah). Pada sistem yang dibuat, RDP terdiri dari 15 jenis retorik kalimat (mulai dari tujuan, sampai kesimpulan serta rencana penelitian selanjutnya). Kumpulan RDP ini berisi bahan yang siap untuk dikomposisikan menjadi ringkasan, dan dihasilkan melalui proses klasifikasi retorik kalimat dari makalah ilmiah lengkapnya. Klasifikasi retorik kalimat menerapkan pembelajaran multi-strategi berbasis WTA seperti yang sudah dijelaskan sebelumnya.

Proses penyajian ringkasan dilakukan dengan mengikuti skenario

pembentukan ringkasan yang telah dispesifikasikan dengan BPLAN. Kalimat-kalimat ringkasan diambil dari kumpulan RDP sehingga menghasilkan draft ringkasan. Proses *surface repair* selanjutnya diterapkan untuk menghaluskan draft ringkasan sehingga lebih mudah dipahami konteks dari kalimat-kalimat ringkasannya.

Gambar 6 memberikan contoh tampilan dari sistem peringkasan multi dokumen yang telah dikembangkan. Pada sisi kiri atas dipakai untuk mengunggah kumpulan makalah ilmiah yang akan diringkas. Spesifikasi ringkasan didefinisikan pada antarmuka yang disediakan di sisi kiri bawah. Sedangkan di sisi kanan menampilkan hasil ringkasannya. Hasil evaluasi yang melibatkan 21 responden (mahasiswa S3) menyimpulkan bahwa *Surface repair* secara keseluruhan dapat membuat ringkasan menjadi lebih mudah dibaca.

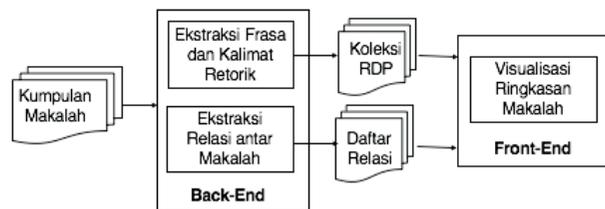


Gambar 6. Contoh tampilan sistem peringkasan multi-dokumen.

6.2 Visualisasi Dinamis Ringkasan Makalah Ilmiah berbasis Jaringan Sitasi

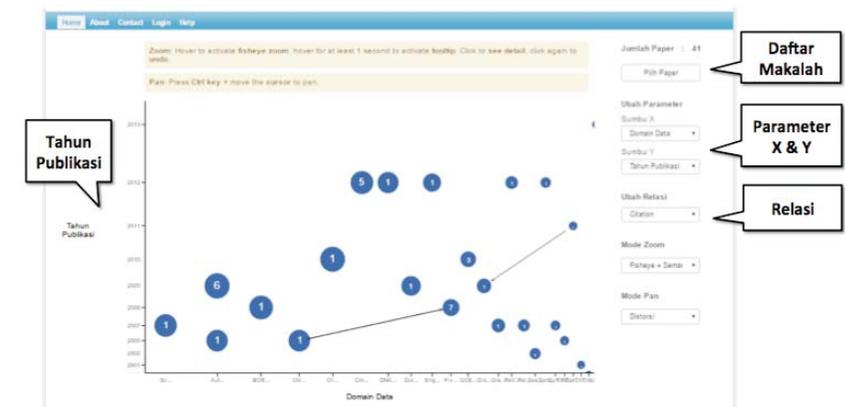
Saat ini banyak sekali inovasi dan topik penelitian baru yang muncul dari literatur. Dalam proses penelitian, untuk mengidentifikasi persoalan riset, peneliti biasanya perlu mencari tahu dengan membaca dan merangkum literatur yang terkait dengan topik penelitiannya. Peneliti selanjutnya dapat melakukan analisis dengan memetakan rangkuman dari setiap referensi, berdasarkan sudut pandang yang menjadi fokusnya saat itu untuk mendapatkan gambaran keterkaitan antar topik riset yang sedang menjadi perhatiannya. Proses ini tentu akan memerlukan waktu jika dilakukan secara manual.

Untuk membantu mengatasi persoalan tersebut, penulis mengembangkan sebuah perangkat bantu untuk melakukan visualisasi ringkasan makalah ilmiah yang hasil visualisasinya dapat disesuaikan dengan kebutuhan analisis dari penggunanya. Sistem yang dikembangkan terdiri dari dua komponen utama (lihat Gambar 7). Pertama adalah bagian *front-end*, yang menampilkan hasil visualisasi dan berinteraksi dengan pengguna. Bagian kedua adalah bagian *back-end*, yang bertugas untuk melakukan ekstraksi meta-data, frasa retorik serta relasi antar makalah dalam jaringan sitasi.



Gambar 7. Sistem Visualisasi Ringkasan Makalah

Gambar 8 memperlihatkan contoh tampilan dari bagian *front-end*. Untuk menggunakan sistem, pengguna menentukan terlebih dahulu kumpulan makalah ilmiah yang akan divisualisasikan. Selanjutnya sistem akan menampilkan visualisasi dari ringkasan makalah dalam sistem koordinat cartesian dua dimensi dengan sumbu X dan Y. Representasi setiap sumbu bersifat parametrik dan ditentukan sesuai dengan kebutuhan pengguna. Parameter tersebut dapat dimodifikasi berdasarkan informasi meta data dan jenis-jenis frasa retorik yang diidentifikasi dari makalah ilmiah. Sebagai contoh, parameter tersebut dapat berupa tahun publikasi, nama penulis, persoalan riset, domain data, metode yang dipakai, hasil yang diperoleh, dll.



Gambar 8. Contoh tampilan aplikasi visualisasi jejaring sitasi.

Makalah ditampilkan sebagai lingkaran pada posisi koordinat yang sesuai, dengan ukuran lingkaran yang berbanding lurus dengan jumlah makalah yang memenuhi kriteria untuk tampil pada koordinat tersebut.

Jika lingkaran berisi satu makalah, maka isi ringkasan makalah dapat ditampilkan dengan menempatkan kursor pada lingkaran tersebut. Jika berisi lebih dari satu makalah, maka makalah di dalamnya dapat di-zoom secara visual untuk melihat daftar makalah di dalamnya. Selain zooming, navigasi jejaring sitasi dapat dilakukan dengan cara panning, baik dengan mode linier maupun distorsi.

Hasil pengujian usability yang melibatkan 20 responden (mahasiswa S3) memperlihatkan efektivitas sistem visualisasi ini dalam membantu mereka melakukan eksplorasi literatur serta memberikan pengalaman pengguna (*user experience*) yang positif.

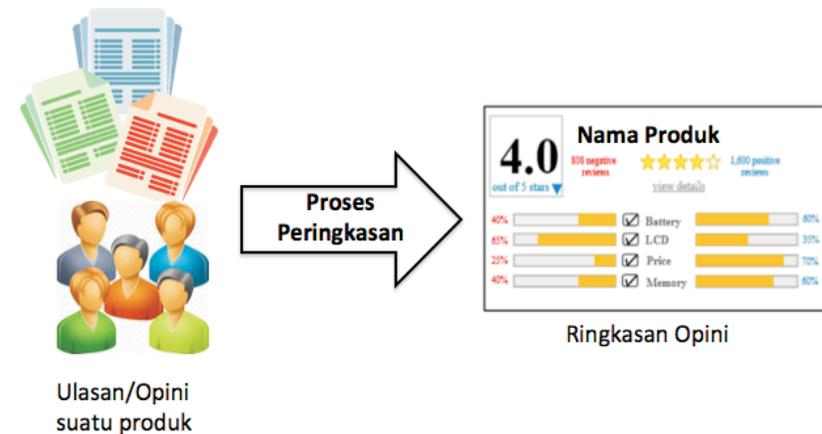
Masukan dari sistem visualisasi ini adalah kumpulan makalah ilmiah (*full text*), yang untuk saat ini fokus pada jenis makalah eksperimental. Setiap makalah akan diekstrak kontennya (metadata, retorik frasa & kalimat, serta relasi antar makalah) pada bagian back-end dari sistem. Metadata di sini termasuk *judul, daftar nama penulis & tahun publikasi*. Retorik frasa dan kalimat adalah frasa dan kalimat penting yang merepresentasikan informasi yang akan banyak dicari oleh pembaca, seperti persoalan penelitian, domain penelitian, metode dan data yang dipakai serta hasil eksperimen. Sedangkan relasi antar makalah mengidentifikasi jenis hubungan antara makalah yang merujuk dan yang dirujuk, apakah merupakan kritik, perbandingan atau ekstensi metode.

Kecuali ekstraksi metadata, proses ekstraksi dilakukan dengan pendekatan klasifikasi, dengan menerapkan algoritma pembelajaran mesin yang standar. Inovasi pada proses ekstraksi ini difokuskan pada pengembangan fitur-fitur baru yang meningkatkan kinerja pada proses

ekstraksi, baik untuk retorik frasa dan kalimat maupun pada klasifikasi jenis relasi antar makalah.

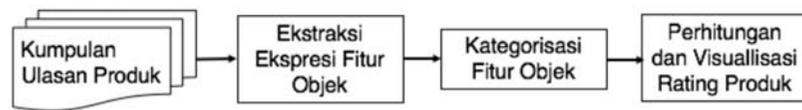
6.3 Peringkasan Opini

Penelitian peringkasan opini didorong oleh melimpahnya ulasan dan opini terkait produk yang tersedia di Web. Ulasan ini perannya sangat penting bagi pengguna pada umumnya karena membantu memutuskan pembelian suatu produk. Namun karena banyaknya dokumen ulasan, membaca seluruh ulasan tersebut menjadi hal yang sulit dilakukan, terlebih lagi apabila ulasan tersebut memiliki sudut pandang yang berbeda-beda dengan isi ulasan yang bisa saling bertentangan. Selain itu, orang yang ingin membeli suatu produk pada umumnya memiliki kriteria tertentu yang belum tentu sama dengan dengan kriteria yang diuraikan dalam ulasan.



Gambar 9. Ilustrasi peringkasan opini suatu produk.

Peringkasan opini pada penelitian ini mencoba melakukan ekstraksi opini suatu produk dari kumpulan teks opini serta menampilkan statistik dari setiap aspek hasil ekstraksi secara visual yang dapat disesuaikan dengan kebutuhan informasi penggunaanya (lihat ilustrasinya pada Gambar 9). Sebagai contoh, dari kumpulan teks yang mengulas telepon seluler, maka akan diekstrak ulasan yang terkait dengan fitur-fiturnya seperti desain, memori, audio, kamera, jaringan, konektivitas, harga, sensor, dlsb. Ulasan yang diberikan pada suatu kategori fitur dapat positif, negatif, obyektif ataupun subyektif. Ulasan-ulasan yang berhasil diidentifikasi selanjutnya ditampilkan secara visual yang menggambarkan statistik dari ulasan setiap kategori. Tampilan ini juga memberikan antarmuka yang memungkinkan pengguna berinteraksi untuk menyesuaikan kebutuhan ringkasan informasi dan tampilan visualisasi.



Gambar 10. Sistem peringkasan opini berbasis kebutuhan pengguna.

Gambar 10 memberikan ikhtisar sistem peringkasan opini berbasis fitur objek yang dikembangkan. Masukan sistem berupa kumpulan ulasan (dalam bentuk teks) dari suatu produk. Tahap pertama yang dilakukan adalah melakukan ekstraksi fitur objek dengan mengembangkan teknik *clue propagation* untuk menjaring ulasan fitur yang lebih banyak. Karena fitur objek dapat direpresentasikan dengan dengan ekspresi yang berbeda-beda, maka fitur-fitur objek yang diperoleh dinormalisasi dengan melakukan kategorisasi fitur. Proses ini dilakukan

dengan mengkombinasikan secara linier pendekatan berbasis kamus dan pendekatan berbasis distribusional. Kumpulan opini yang terkait dengan fitur objek tersebut selanjutnya diringkas dengan menghitung *rating* produk yang berbasis fitur objek. Perhitungan rating dilakukan dengan mempertimbangkan reputasi pengguna dan distribusi ulasan opini. Hasil perhitungan rating selanjutnya divisualisasikan dalam bentuk *5-star* rating. Nilai rating yang dihitung sangat bergantung pada jenis-jenis fitur objek yang dipilih pengguna untuk ditampilkan. Berdasarkan survey yang dilakukan terhadap 200 respondent, 45% lebih menyukai desain visualisasi ringkasan opini yang diusulkan dalam penelitian ini, jauh melebihi vote yang didapatkan oleh empat desain visualisasi rating lainnya yang pernah dikembangkan oleh pihak lain (25%, 11%, 11% dan 8%).

7. PENUTUP

Pembelajaran mesin, dalam bentuk konkritnya, adalah sebuah program yang dapat meningkatkan kinerjanya pada suatu tugas tertentu sejalan dengan bertambahnya pengalaman yang diperoleh. Sesuai fungsinya, kajian dan pengembangan di bidang ini sangat relevan dengan kondisi saat ini dan akan semakin relevan ke depannya mengingat bertambahnya informasi yang semakin tidak terbandung dari waktu ke waktu. Sebagai salah satu perangkat bantu utama dalam bidang lainnya yang sangat terkait seperti *datamining*, *big data*, dan *data science*, kolaborasi bidang pembelajaran mesin dengan disiplin ilmu lainnya baik dalam pengembangan keilmuan maupun dalam penerapannya sangatlah terbuka.

8. UCAPAN TERIMA KASIH

Pada kesempatan ini, saya menyampaikan terima kasih kepada Rektor dan Pimpinan ITB, Pimpinan dan seluruh Anggota Forum Guru Besar ITB serta Senat Akademik ITB atas dukungan yang diberikan sehingga pada akhirnya mendapatkan kesempatan untuk menyampaikan orasi ilmiah di hadapan para hadirin sekalian pada forum yang terhormat ini.

Ucapan terima kasih juga saya sampaikan secara khusus kepada Dr. Jaka Sembiring (selaku Dekan STEI), Dr. Nana Rachmana Syambas (selaku Wakil Dekan bidang Sumber Daya), Ketua dan anggota senat STEI, serta Prof. Yanuarsyah Haroen yang telah memfasilitasi dan mendukung dari proses pengajuan sampai keluarnya SK Guru Besar.

Kepada ayahanda Bapak Moch. Amin dan ibunda Ibu Marfuah (Almarhumah), terima kasih atas usahanya yang tidak kenal lelah membesarkan saya serta doanya yang tulus untuk mendapatkan hasil yang terbaik baik bagi putra-putrinya. Terima kasih juga saya sampaikan kepada Bapak dan Ibu Mertua, Dr. Suparto Siswowitzo dan Ibu. Sri Suharti atas doa dan dukungan moril yang selalu diberikan kepada kami.

Terimakasih kepada Dr. Lina Handayani, istri saya tercinta yang selalu dengan setia mendampingi, mendorong dan menyemangati saya untuk segera mendapatkan gelar akademik yang tertinggi ini, juga kepada putra-putra kami: Risma, Della dan Ardian yang telah menjadi inspirasi bagi bapak/ibunya.

Pencapaian ini tentu tidak lepas dari dukungan para profesor pemberi rekomendasi. Oleh karena itu, saya ucapkan terima kasih kepada Prof. Iping Supriana, Prof. Kuspriyanto, Prof. Andrian Bayu Suksmono serta

Prof. Edy Soewono.

Perolehan jabatan Guru Besar ini juga tidak lepas dari dukungan dan kerjasama dengan para kolaborator penelitian sehingga menghasilkan publikasi. Terimakasih disampaikan kepada Prof. John Yen, Prof. Thomas Ioerger, Prof. Bambang Riyanto, Prof. Kuspriyanto, Prof. Benhard Sitohang, Prof. Tati L. Mengko, Prof. Liliyasi, Prof. Suhono H. Supangkat, Prof. Aminudin Aziz, Prof. Christophoros Nikou, Dr. Marina E. Plisiti, Dr. Adi Pancoro, Dr. Oerip S. Santoso, Dr. Ayu Purwarianti, Dr. Masayu L. Khodra, Dr. Dessi Puji Lestari, Dr. Rinaldi Munir, Dr. Nur Ulfa Maulidevi, Dr. Bayu Hendradjaya, Dr. Kridanto Surendro, Dr. Husni S. Sastramihardja, Dr. Emir Mauludi Husni, Dr. Ahmad Dahlan, Dr. Kurnia Muludi, Dr. Siti Maimunah, Dr. Dwiza Riana, Dr. Afrida Helen, Dr. Yudi Wibisono, Dr. Z.K. Abdurahman Baizal, dan Dr. Rani Megasari, serta para mahasiswa bimbingan pada program Magister dan Doktor di STEI.

Akhir kata, terima kasih yang tak terhingga saya sampaikan kepada pihak-pihak dan individu lainnya yang telah banyak membantu, baik yang secara langsung maupun tidak langsung, yang tidak dapat saya sebutkan satu persatu di sini. Semoga pencapaian ini menjadi anugrah serta barokah bagi kita semua. Aamin Ya Rabbal Aalamiin.

9. DAFTAR PUSTAKA

1. Mitchell, T. *Machine Learning*. McGraw-Hill, 1997.
2. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*. The MIT Press, Cambridge, Massachusetts, 2012.
3. Kelleher, J.D., Namee, B.M., D'Arcy, A. *Machine Learning for Predictive*

Data Analytics. The MIT Press, Cambridge, Massachusetts, 2015.

4. Sibaroni, Y., Widyantoro, D.H., Khodra, M.L. *Information Extraction of Extend Relation in Scientific Papers*. Prosiding International Conference on Data & Software Engineering. Bali, Oktober 2016
5. Hartono, W.S., Widyantoro, D.H. *Fisheye Zoom and Semantic Zoom on Citation Network Visualization*. Prosiding International Conference on Data & Software Engineering. Bali, Oktober 2016.
6. Sibaroni, Y., Widyantoro, D.H., Khodra, M.L. *Extend Relation Identification In Scientific Papers based on Supervised Machine Learning*. Prosiding International conference on Advanced Computer Science and Information Systems. Malang, Oktober 2016.
7. Widyantoro, D.H., Oenang, Y. *System Development for Research Map Visualization*. Prosiding International Conference on Electrical Engineering and Informatics. Bali, Agustus 2015.
8. Riska, Widyantoro, D.H. *Panning Interaction Development for Citation Network Visualization*. Prosiding International Conference on Telecommunications System Services and Applications. Bali, Oktober 2016
9. Maharani, W., Widyantoro, D.H, Khodra, M.L. *Discovering Users' Perceptions on Rating Visualizations*. Prosiding ACM ICPS International Human Computer Interaction and User Experience Conference. Jakarta, Indonesia, April 2016.
10. Helen, A., Purwarianti, A., Widyantoro, D.H. *Document Rhetorical sentences classification based on section class and title of paper for experimental technical papers*. Jurnal ICT Research and Applications,

Vol. 9 No 3, pp. 288-310.

11. Maharani, W., Widyantoro, D.H. & Khodra, M.L. *Aspect Extraction in Customer Reviews using Syntactic Pattern*. Procedia Computer Science 59 (2015), 244-253. ScienceDirect, Elsevier.
12. Maharani, W., Widyantoro, D.H. & Khodra, M.L. *Learning-based aspect identification in customer review products*. Prosiding International Conference on Electrical Engineering and Informatics. Bali, Agustus 2015.
13. Helen, A., Purwarianti, A., Widyantoro, D.H. *Extraction and Classification of Rhetorical Sentences of Experiment Technical Paper based on Section Class*. Prosiding International Conference on Information and Communication Technology. Bandung, 2014.
14. Widyantoro, D.H., Muludi, K., Kuspriyanto *Winner-Takes-All based Multistrategy Learning for Information Extraction*, Telkonnika, Vol. 12 No.11, 2014.
15. Widyantoro, D.H., Masayu L. Khodra, Bambang Riyanto, E. A. Aziz (2013), *A Multiclass-based Classification Strategy for Rhetorical Sentence Categorization from Research Paper*, Journal of ICT Research and Applications, Vol.7, No. 3, 2013, 235-249
16. Helen, A., Widyantoro, D.H., *Pengembangan Framework Pembangunan Peta Penelitian untuk menggambarkan positioning research secara otomatis*. Prosiding Konferensi Nasional Sistem Informasi, Lombok, 2013.
17. Khodra, M.L., Dimas, M. Widyantoro, D.H., Aziz, E.A., Trilaksono, B.R. *Automatic Tailored Multi-Paper Summarization based on Rhetorical Document Profile and Summary Specification*. Jurnal ITB Information and

Communication Technology, Vol. 6, No. 3, 2012, pp 221-240.

18. Widyantoro, D.H., *Exploiting Unlabeled Data in Concept Drift Learning*, Jurnal Teknik Informatika – Petra, Mei, 2007
19. Widyantoro, D.H., *Exploiting Homogeneity of Density in Incremental Hierarchical Clustering*. In Proceedings ITB on Engineering Science, Vol. 38 B, No. 2, 2006.
20. Widyantoro, D.H., and Yen, J. (2005) Relevant Data Expansion for Learning Concept Drift from Sparsely Labeled Data. *IEEE Transactions on Knowledge and Data Engineering*, 17(3):401-412.
21. Widyantoro, D.H., Ioerger, T.R. and Yen, J. (2003) Tracking Changes in User Interests with a Few Relevance Judgments. In *Proceedings of the Twelfth ACM International Conference on Information and Knowledge Management*, pp. 548-551.
22. Widyantoro, D.H., Ioerger, T.R. and Yen, J. (2002) An Incremental Approach to Building a Cluster Hierarchy. In *Proceedings of the Second IEEE International Conference on Data Mining*, pp. 705-708.
23. Widyantoro, D.H., Ioerger, T.R., and Yen, J. (2001). Learning User Interest Dynamics with a Three-Descriptor Representation. *Journal of the American Society for Information Science*, 52(3):212-225.

CURRICULUM VITAE



Nama : **DWI HENDRATMO WIDYANTORO**
Tmpt. & tgl. lhr. : Ambarawa, 7 Desember 1968
Kel. Keahlian : Informatika
Alamat Kantor : Jalan Ganesha 10 Bandung
Nama Istri : Lina Handayani
Nama Anak : 1. Risma Cahyani Widyantoro
2. Della Nurdwiani Widyantoro
3. Ardian Putra Widyantoro

I. RIWAYAT PENDIDIKAN

- Doctor of Philosophy (Ph.D.), bidang Computer Science, Texas A&M University, USA, 2003
- Master of Science (MSc), bidang Computer Science, Texas A&M University, USA, 1999
- Sarjana Teknik Informatika (Ir.), Institut Teknologi Bandung (ITB), 1991.

II. RIWAYAT KERJA di ITB:

- Staf Pengajar Fakultas Teknologi Industri ITB, 1994 - 2005
- Staf Pengajar Sekolah Teknik Elektro & Informatika ITB, 2006 - Sekarang
- Wakil Dekan bidang Akademik, 2011 - Sekarang.

III. RIWAYAT JABATAN FUNGSIONAL

- CPNS, 1 Februari 1994

- Penata Muda, III/A, 1 Juli 1995
- Penata, III/C, 1 April 2006
- Penata TKI/IIID, 1 April 2008
- Pembina, IVA, 1 April 2010
- Pembina TKI, IVB, 1 April 2012
- Pembina Utama Muda, IV/C, 1 April 2017.

IV. RIWAYAT JABATAN FUNGSIONAL

- Asisten Ahli Madya, 1 September 1995
- Asisten Ahli, 1 Januari 2001
- Lektor, 1 Januari 2006
- Lektor Kepala, 1 Desember 2009
- Guru Besar, 1 Oktober 2016.

V. KEGIATAN PENELITIAN (LIMA TAHUN TERAKHIR)

- Pengembangan Sistem Penerjemah Teks Bahasa Sunda ke dalam Bahasa Indonesia dengan menggunakan metode Hybrid (Riset Inovasi ITB, 2017)
- Pengembangan *Chatbot* Berbahasa Indonesia (Program Magister menuju Doktor untuk Sarjana Unggul, 2016-2017)
- Peringkasan Terpandu Otomatis (Program Magister menuju Doktor untuk Sarjana Unggul, 2016-2017)
- Sistem Pengenalan Ucapan untuk Bahasa Indonesia (Program Magister menuju Doktor untuk Sarjana Unggul, 2016-2017)
- *Conversational Recommender System* berbasis Kebutuhan

Fungsional Produk (Hibah Kompetensi – Dikti, 2016-2017)

- Pengembangan *Open Information Extraction* (Open IE) untuk Pembangunan Basis Pengetahuan Skala Besar dari Web (Penelitian Unggulan Perguruan Tinggi, 2016; Program Penelitian, Pengabdian kepada Masyarakat & Inovasi ITB, 2017)
- Pengembangan Sistem untuk Prediksi Kredibilitas Informasi dari Jejaring Sosial Online (Penelitian Unggulan Perguruan Tinggi, 2014-2016)
- Pengembangan Sistem Peringkasan Opini (Program Riset & Inovasi ITB, 2015-2016)
- Model Deteksi Kemacetan berdasarkan Tingkat Kepadatan dan Kecepatan Kendaraan (Penelitian Unggulan Perguruan Tinggi, 2016)
- Pembangkitan Peta Penelitian Otomatis dari Kumpulan Makalah Ilmiah (Program Riset & Inovasi ITB, 2013-2014)
- Analisis Sentimen Realtime pada Microblogging (Program Riset & Inovasi ITB, 2012)
- Pengembangan Sistem Interaktif untuk Peringkasan Multi-Paper (Program Desentralisasi Dikti, 2012)
- Peringkasan Multi-Dokumen untuk Makalah Survey (Program Riset dan Inovasi ITB, 2011).

VI. PUBLIKASI (LIMA TAHUN TERAKHIR)

1. Romadhony, A., **Widyantoro, D.H.**, Purwarianti, A. *Employing semantic knowledge on event trigger clustering*. Jurnal Theoretical and Applied Information Technology. Vol 95, No. 6, Maret 2017.

2. Maharani, W., **Widyantoro, D.H.**, Khodra, M.L. *Aspect-based opinion summarization: A survey*. Jurnal of Theoretical and Applied Information Technology. Vol 95, No. 2, Januari 2017.
3. Saelan, A., Purwarianti, A., **Widyantoro, D.H.** *Question analysis for Indonesian comparative question*. Prosiding International Conference on Computing and Applied Informatics. Medan, Desember 2016
4. Baizal, Z.K.A., **Widyantoro, D.H.**, Maulidevi, N.U. *Factors influencing user's adoption of conversational recommender system based on product functional requirements*. Telkomnika (Telecommunication Computing Electronics and Control). Vol 14, No. 4, pp.1598-1607
5. Rachman, G.H., Khodra, M.L., **Widyantoro, D.H.** *Rhetorical Sentence Categorization for Scientific Paper Using Word2Vec Semantic Representation*. Prosiding International Conference on Computing and Applied Informatics. Medan, Desember 2016
6. Enjat Munajat, M.D., Munir, R., **Widyantoro, D.H.** *Traffic Congestion Models: Challenges and Opportunities*. Prosiding International Conference on Computing and Applied Informatics. Medan, Desember 2016.
7. Setyorini, Kuspriyanto, **Widyantoro, D.H.**, Pancoro, A. *The implementation of bit-parallelism for DNA sequence alignment*. Journal of Physics: Conference Series, 2016.
8. Enjat Munajat, M.D., **Widyantoro, D.H.**, Munir, R. *Vehicle Detection and Tracking Based on Corner and Lines Adjacent Detection Features*. Prosiding International Conference on Science in Information Technology. Balikpapan, Oktober 2016.

9. Jatmiko, A.B., **Widyantoro, D.H.** *Modified Breadth-First Order-based Link Categorization for Finding Financial Statements Documents*. Prosiding International Conference on Cybernetics & Computational Intelligence. Makassar, November 2016
10. Mallafi, H., **Widyantoro, D.H.** *Prediction modelling in career management*. Prosiding International Conference on Cybernetics & Computational Intelligence. Makassar, November 2016
11. Megasari, R., Kuspriyanto, Husni, E.M., **Widyantoro, D.H.** *The strategies for quorum satisfaction in host-to-host meeting scheduling negotiation*. Telkomnika (Telecommunication Computing Electronics and Control), Vol. 14, No. 4, pp. 1575-1585
12. Sibaroni, Y., **Widyantoro, D.H.**, Khodra, M.L. *Information Extraction of Extend Relation in Scientific Papers*. Prosiding International Conference on Data & Software Engineering. Bali, Oktober 2016
13. Baizal, Z.K.A., **Widyantoro, D.H.**, Maulidevi, N.U. *Design of Knowledge for Conversational Recommender System based on Product Functional Requirement*. Prosiding International Conference on Data & Software Engineering. Bali, Oktober 2016.
14. Hartono, W.S., **Widyantoro, D.H.** *Fisheye Zoom and Semantic Zoom on Citation Network Visualization*. Prosiding International Conference on Data & Software Engineering. Bali, Oktober 2016
15. Enjat Munajat, M.D., **Widyantoro, D.H.**, Munir, R. *Vehicle detection and tracking based on corner and lines adjacent detection features*. International Conference on Science in Information Technology. Balikpapan, Oktober 2016

16. Romadhony, A., **Widyantoro, D.H.**, Purwarianti, A. *Using relation similarity on open information extraction-based event template extraction*. Prosiding International conference on Advanced Computer Science and Information Systems. Malang, Oktober 2016.
17. Baizal, Z.K.A., **Widyantoro, D.H.**, Maulidevi, N.U. *Query Refinement in Recommender System based on Product Functional Requirements*. Prosiding International conference on Advanced Computer Science and Information Systems. Malang, Oktober 2016
18. Sibaroni, Y., **Widyantoro, D.H.**, Khodra, M.L. *Extend Relation Identification In Scientific Papers based on Supervised Machine Learning*. Prosiding International conference on Advanced Computer Science and Information Systems. Malang, Oktober 2016
19. Waisakurnia, W., **Widyantoro, D.H.** *Traffic Light Candidate Elimination Based on Position*. Prosiding The Tenth International Conference on Telecommunications System Servicess and Applications. Bali, Oktober 2016
20. Riska, **Widyantoro, D.H.** *Panning Interaction Development for Citation Network Visualization*. Prosiding International Conference on Telecommunications System Servicess and Applications. Bali, Oktober 2016
21. Setiawan, E.B., **Widyantoro, D.H.**, Surendro, K. *Feature Expansion using Word Embedding for Tweet Topic Classification*. Prosiding International Conference on Telecommunications System Servicess and Applications. Bali, Oktober 2016

22. Binangkit, J.L., **Widyantoro, D.H.** *Increasing Accuracy of Traffic Light Color Detection and Recognition using Machine Learning*. Prosiding International Conference on Telecommunications System Servicess and Applications (TSSA). Bali, Oktober 2016
23. Karisma, H., **Widyantoro, D.H.** *Comparison Study of Neural Network and Deep Neural Network on Repricing GAP Prediction in Indonesian Conventional Public Bank*. Prosiding Sixth International Conference on System Engineering & Technology. Bandung, Oktober 2016
24. Shulhan, M., **Widyantoro, D.H.** *Detecting Vandalism on English Wikipedia Using LNSMOTE Resampling and Cascaded Random Forest Classifier*. Prosiding International Conference on Advanced Informatics: Concepts, Theories and Applications. Penang, Malaysia, Agustus 2016.
25. Wibisono, Y., **Widyantoro, D.H.**, Maulidevi, N.U. *Document Comparison of distributional semantic models for recognizing textual entailment*. Jurnal Theoretical and Applied Information Technology, Vol.93, No.2, pp. 270-277.
26. Maharani, W., **Widyantoro, D.H.**, Khodra, M.L. *Discovering Users' Perceptions on Rating Visualizations*. Prosiding ACM ICPS International Human Computer Interaction and User Experience Conference. Jakarta, Indonesia, April 2016.
27. Samuel, Y.T, **Widyantoro, D.H.**, Wuryandari, A.I. *A Framework of Fundamental News Summarization To determine the direction of Foreign Exchange Rate Using Adaptive Indicator Scheme*. Prosiding International Conference on Advanced Informatics: Concepts, Theories and Applicatoins. Chonburi, Thailand, Agustus 2015.

28. Suryani, A.A., **Widyantoro, D.H.**, Purwarianti, A., Sudaryat, Y. *Experiment on a Phrase-Based Statistical Machine Translation Using PoS Tag Information for Sundanese into Indonesian*. Prosiding International Conference on Information Technology System and Innovation. Bandung, November 2015.
29. Sibaroni, Y., **Widyantoro, D.H.**, Khodra, M.L. *Survey on Research Paper Relation*. Prosiding International Conference on Information Technology System and Innovation. Bandung, November 2015.
30. Helen, A., Purwarianti, A., **Widyantoro, D.H.** *Document Rhetorical sentences classification based on section class and title of paper for experimental technical papers*. Jurnal ICT Research and Applications, Vol. 9 No 3, pp. 288-310
31. Munajat, M.D.E., **Widyantoro, D.H.**, Munir, R. *Road Detection System based on RGB Histogram Filterization and Boundary Classifier*. Prosiding International Conference on Advanced Computer Science and Information System. Jakarta, 2015.
32. Romadhony, A., **Widyantoro, D.H.**, Purwarianti, A. *Phrase-based Clause Extraction for Open Information Extraction System*. Prosiding Internatinal Conference on Advanced Computer Science and Information System. Jakarta, 2015.
33. Riana, D., **Widyantoro, D.H.**, Mengko, T.L. *Document Extraction and classification texture of inflammatory cells and nuclei in normal pap smear images*. Prosiding International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering, ICICI-BME 2015, pp. 65-69.
34. Megasari, R., Husni, E.M., Kuspriyanto, **Widyantoro D.H.** (2015),

- Negotiation strategies for meeting scheduling conflict management*. Prosiding International Conference on Science in Information Technology, pp. 276-281.
35. **Widyantoro, D.H.** Saputra, K.I. *Traffic Lights Detection and Recognition based on Color Segmentation and Hough Transform*. Prosiding International Conference on Data and Software Engineering. Yogyakarta, November 2015.
36. **Widyantoro, D.H.**, Oenang, Y. *System Development for Research Map Visualization*. Prosiding International Conference on Electrical Engineering and Informatics. Bali, Agustus 2015.
37. Maharani, W., **Widyantoro, D.H.**, & Khodra, M.L. *SAE: Syntactic-based Aspect and Opinion Extraction from Product Reviews*. Prosiding International Conference on Advanced Informatics: Concepts, Theories and Applicatoins. Chonburi, Thailand, Agustus 2015.
38. Maharani, W., **Widyantoro, D.H.** & Khodra, M.L. *Aspect Extraction in Customer Reviews using Syntactic Pattern*. Procedia Computer Science 59 (2015), 244-253. ScienceDirect, Elsevier.
39. Maharani, W., **Widyantoro, D.H.** & Khodra, M.L. *Learning-based aspect identification in customer review products*. Prosiding International Conference on Electrical Engineering and Informatics. Bali, Agustus 2015.
40. Mariyah, S., & **Widyantoro, D.H.** *A Multi-strategy Approach for Information Extraction of Financial Report Documents*. Prosiding International Conference on Information & Communication Technology and Systems. Surabaya, September 2015.
41. Romadhony, A., **Widyantoro, D.H.**, Purwarianti, A. *Using Second Dependency Level Feature in Open Information Extraction*. Prosiding

- Conference on Pasific Association of Computational Linguistics. Bali, Mei 2015.
42. **Widyantoro, D.H.**, Sambora, T.K. *System Development of Commercial Logo Analysis on Online Social Media*. Prosiding International Conference on Advanced Informatics: Concepts, Theories and Applicatoin. Chonburi, Thailand, Agustus 2015.
 43. Riana, D., Plissiti, M.E., Nikou, C., **Widyantoro, D.H.**, Mengko, T.L.R., Kalsoem, O. *Inflammatory Cell Extraction and Nuclei Detection in Pap Smear Images*. International Journal of E-Health and Medical Communications, Vol. 6 No. 2, 27-43, April-June 2015.
 44. **Widyantoro, D.H.**, Wibisono, W. *Modeling Credibility Assessment and Explanation for Tweets based on Sentiment Analysis*. Jurnal Theoretical & Applied Information Technology. Vol. 70, No. 3, pp 540-54, Desember 2014.
 45. Wibisono, Y., **Widyantoro, D.H.**, Maulidevi, N.U. *Sentence Extraction in Recognition Textual Entailment Task*. Prosiding International Conference on Data & Software Engineering. Bandung, November 2014.
 46. **Widyantoro, D.H.**, Helen, A. *Preposition-based Pattern Sequence for Rhetorical Phrase Extraction in Titles of Scientific Papers*. Prosiding Regional Conference on Computer and Information Engineering. Yogyakarta, 2014.
 47. Helen, A., Purwarianti, A., **Widyantoro, D.H.** *Extraction and Classification of Rhetorical Sentences of Experiment Technical Paper based on Section Class*. Prosiding International Conference on Information and Communication Technology. Bandung, 2014.

48. **Widyantoro, D.H.**, Baizal, Z.K. *A Framework of Conversational Recommender System based on User Functional Requirements*. Prosiding International Conference on Information and Communication Technology. Bandung, 2014.
49. Reinanda, R., **Widyantoro, D.H.** *Performance Comparison of Learning to Rank Algorithms for Information Retrieval*. Prosiding International Conference on Information, Communication Technology and System. Surabaya, 2014.
50. **Widyantoro, D.H.**, Muludi, K., Kuspriyanto *Winner-Takes-All based Multistrategy Learning for Information Extraction*, Telkomnika, Vol. 12 No. 11, 2014.
51. **Widyantoro, D.H.**, Amin, I. *Citation Sentence Identification and Classification for Related Work Summarization*. Prosiding International Conference on Advanced Computer Science and Information System. Jakarta, 2014.
52. Wibisono, Y., **Dwi H. Widyantoro**, Maulidevi, N.U. *Rancangan Sistem Pembangkit Anotasi Sistematis untuk Kredibilitas dan Reliabilitas Informasi dalam Jejaring Sosial Online*. Prosiding Konferensi Nasional Sistem Informasi, 2014.
53. Perkasa, O., **Widyantoro, D.H.**, *Video-based System Development for Automatic Traffic Monitoring*. Prosiding International Conference on Electrical Engineering & Computer Science, pp. 281-285, Denpasar, November 2014.
54. Riana, D., **Widyantoro, D.H.**, Mengko., T.L., *Ekstraksi Fitur Kuantitatif Tekstur dan Klasifikasi Sel Nukleus dan Sel Radang pada Citra Pap Smear*. Konferensi Nasional Ilmu Komputer, Makassar,

- Desember 2014, pp.179-184.
55. **Widyantoro, D.H.**, Ungkawa, U., Hendradjaya, B.. *Case-based Reasoning Approach for Form Interface Design*. Prosiding Internatioal Conference on Data and Software Engineering. Bandung, November 2014.
 56. **Widyantoro, D.H.**, Munajat, M.D.E. *Fuzzy Traffic Congestion Model based on Speed and Density of Vehicle*. Prosiding International Conference on Advanced Informatics: Concepts, Theory and Applications, 332-336, 2014.
 57. Riana, D., Dewi, D. E. O., **Widyantoro, D. H.**, & Mengko, T. L. R. *Color canals modification with canny edge detection and morphological reconstruction for cell nucleus segmentation and area measurement in normal pap smear images*. Prosidng AIP Conference, 414-417, 2014.
 58. **Widyantoro, D.H.**, Khodra, M.L., Trilaksono, B.R., Aziz, E.A. *A Multiclass-based Classification Strategy for Rhetorical Sentence Categorization from Research Paper*. Jurnal ICT Research and Applications, Vol. 7, No. 3, 2013, 235-249, 2013.
 59. Helen, A., **Widyantoro, D.H.**, *Pengembangan Framework Pembangkitan Peta Penelitian untuk menggambarkan positioning research secara otomatis*. Prosiding Konferensi Nasional Sistem Informasi, Lombok, 2013.
 60. Khodra, M.L., **Widyantoro, D.H.**, *BPLAN-SUMM: Sistem Peringkasan Otomatis Kumpulan Makalah sesuai Kebutuhan Informasi Pengguna*. Prosiding Konferensi Nasional Sistem Informasi. Lombok, 2013
 61. Wibisono, Y., **Widyantoro, D.H.**, *Purwarupa Sistem Analisis*

- Sentimen pada Microblob*. Prosiding Konferensi Nasional Sistem Informasi. Lombok, 2013
62. Gunawan, Agus Setiawan, **Dwi H. Widyantoro** (2013). *Model Virtual Laboratory Fisika Modern untuk Meningkatkan Ketrampilan Generik Sains Calon Guru*. Jurnal Pendidikan dan Pembelajaran. Volume 20, Nomor 2, April 2013.
 63. Khodra, M.L., Dimas, M. **Widyantoro, D.H.**, Aziz, E.A., Trilaksono, B.R. *Automatic Tailored Multi-Paper Summarization based on Rhetorical Document Profile and Summary Specification*. Jurnal ITB Information and Communication Technology, Vol. 6, No. 3, 2012, pp 221-240.
 64. Khodra, M.L., Dimas, M., **Widyantoro, D.H.**, Aziz, E.A., Trilaksono, B.R. *Transforming Rhetorical Document Profile Into Tailored Summary Of Scientific Paper*. Jurnal Kursor/Informatika, Trunojoyo, 2012.
 65. Khodra, M.L., Aziz, E.A., **Widyantoro, D.H.**, Trilaksono, B.R. *Peringkasan Makalah Secara Otomatis: Sebuah Contoh Kajian Linguistik Komputasional*. Bagian buku "Dari Menapak Jejak Kata Sampai Menyigi Tata Bahasa", Pusat Kajian Bahasa dan Budaya. Unika Atmajaya, 2012.
 66. Khodra, M.L., **Widyantoro, D.H.**, Aziz, E.A., Trilaksono, B.R., *Sistem Informasi Inteligen Peringkasan Makalah Ilmiah*. Prosiding Konferensi Nasional Sistem Informasi. Denpasar, 2012.
 67. Maimunah, S., Sastramihardja, H.S., **Widyantoro, D.H.**, Kuspriyanto., *CT-FC: more Comprehensive Traversal Focused Crawler*. TELKOMNIKA, Vol 10, No. 1, 2012.

68. Khodra, M.L., **Widyantoro, D.H.**, Aziz, E.A., B Trilaksono, B.R., *Automatic Tailored Multi-Paper Summarization*. Prosiding The Asia-Oceania Top University League on Engineering, Beijing, Oktober 2011.
69. Khodra, M.L., **Widyantoro, D.H.**, E.A. Aziz, B. R. Trilaksono. *Information Extraction for Scientific Paper Using Rhetorical Classifier*. Prosiding International Conference of Electrical Engineering and Informatics,, Bandung, Juli 2011.
70. Muludi,K., **Widyantoro, D.H.**, Kuspriyanto, Santoso, O.S., *Multi-Inductive Learning Approach for Information Extraction*. Prosiding International Conference of Electrical Engineering and Informatics, Bandung, Juli 2011.
71. Maimunah, S., **Widyantoro, D.H.**, Kuspriyanto, Sastramihardja, H.S., *Co-citation & Co-reference Concepts to Control focused Crawler Exploration*. Prosiding International Conference of Electrical Engineering and Informatics, pp. 1211-1217, Bandung, Juli 2011.
72. Purwarianti, A., **Widyantoro, D.H.**, Nazar Iskandar F, Satrya Pratama. *Development of Learning Model in Virtual Class for Primary School in Rural Area*. Prosiding International Conference on Rural Information and Conference Technology 2011, November 2011.
73. Abednego, L., **Widyantoro, D.H.**, *Genetic Programming Hyper-Heuristics for Solving Dynamic Production Scheduling Problem*. Prosiding International Conference of Electrical Engineering and Informatics, Bandung, Juli 2011.
74. Khodra, M.L., **Widyantoro, D.H.**, Aziz, E.A., Trilaksono, B.R. *Konstruksi Koleksi Retorik Kalimat*. Prosiding Konferensi Nasional Sistem Informasi. Medan, Februari 2011.

VII. PENGHARGAAN

- Satya Lencana Karya Satya 10 Tahun, Presiden Republik Indonesia.

