



ANALISIS KESENJANGAN ANTARA TOPIK SKRIPSI MAHASISWA DAN TOPIK RISET DI PUBLIKASI ILMIAH

Fardanuddin Sufyan¹⁾, Tubagus Mohammad Akhriza²⁾, Evy Sophia³⁾

¹⁾ Sistem Informasi, STMIK Pradnya Paramita (STIMATA) Malang
Email : adanrezept@gmail.com

²⁾ Teknik Informatika, STMIK Pradnya Paramita (STIMATA) Malang
Email : akhriza@stimata.ac.id

³⁾ Sistem Informasi, STMIK Pradnya Paramita (STIMATA) Malang
Email : evysophia@yahoo.co.id

Abstrak

Tingkat penyerapan mahasiswa terhadap perkembangan teknologi dapat diamati salah satunya dari topik skripsi yang diambil oleh mahasiswa dimaksud. Permasalahan yang sering dihadapi oleh perguruan tinggi adalah bagaimana metode yang secara efektif dan efisien mampu mengukur kesenjangan antara topik skripsi mahasiswa dengan perkembangan teknologi terkini. Penelitian ini mengusulkan metode untuk mengukur kesenjangan dimaksud dengan cara membandingkan topik skripsi dengan topik yang diangkat di publikasi ilmiah baik nasional maupun internasional. Metode yang digunakan adalah association rule mining, dimana kata-kata yang paling sering disebut (*frequent termset*) di dalam abstrak skripsi dan abstrak riset yang di publikasi ilmiah dieksploitasi. Dari *frequent termset* yang dikumpulkan, hubungan antara satu kata dengan yang lain ditemukan, untuk kemudian, perbedaan antara topik yang dibahas di dalam skripsi dengan yang diangkat oleh publikasi ilmiah dianalisis dan disimpulkan. Hasil analisis akan menunjukkan topik yang sering diangkat pada skripsi namun topik tersebut tidak pernah diangkat di riset, dan sebaliknya.

Kata kunci: *association rule, data mining, topik skripsi, topik publikasi ilmiah*

Abstract

The level of student absorption on technological developments can be observed through their choice on their thesis topics. The problems commonly faced by universities are how to effectively and efficiently measure the gap between students' thesis topics and the latest technological development. This research proposes a method to measure the gap by comparing the students' research topics with the scientific publication topics both nationally and internationally. The method used is association rule mining, where the most frequent words (frequent termset) in abstracts of the thesis and abstracts of scientific publications are exploited. From the frequent termset collected, the relationship between one word and the others is found, to then, the difference between the topics discussed in the thesis with those raised by the scientific publications is analyzed and summarized. The result of the analysis will show the topics usually found in thesis but rarely found in published researches and vice versa.

Keyword: *association rule, data mining, thesis topics, scientific publication topics.*

PENDAHULUAN

Secara global, tugas perguruan tinggi ada 3 yang tertuang dalam tri dharma, yaitu

1) pendidikan dan pengajaran 2) penelitian dan pengembangan 3) pengabdian masyarakat (Djojodibroto, 2004). Salah satu



tugas perguruan tinggi adalah penelitian dan pengembangan yang dilakukan oleh dosen dan mahasiswa. Penelitian mahasiswa dapat dilihat dari hasil skripsi mereka. Selain itu penelitian mahasiswa melalui skripsi juga merupakan tolak ukur tingkat penyerapan kampus terhadap perkembangan teknologi yang terjadi saat ini. Namun demikian perkembangan teknologi informasi dan komunikasi (TIK) begitu cepat dan massif, dan masih sulit bagi mahasiswa khususnya program Sarjana untuk mengikuti perkembangan TIK dimaksud. Sebagai akibatnya, terjadi kesenjangan antara penelitian skripsi mahasiswa dengan perkembangan TIK.

Permasalahan yang sering dihadapi oleh perguruan tinggi adalah bagaimana metode yang secara efektif dan efisien mampu mengukur kesenjangan antara penelitian skripsi mahasiswa dengan perkembangan TIK terkini.

Di sisi lain, informasi mengenai perkembangan TIK dapat diperoleh dari berbagai sumber. Penelitian Akhriza dkk (2017) dan Litecky dkk (2010) menggunakan iklan lowongan kerja sebagai sumber informasi mengenai *skillset* yang dibutuhkan oleh industri TIK. Di penelitian lain, Son dkk (2015) menggunakan social media sebagai sumber informasi perkembangan

TIK yang mutakhir. Para peneliti tersebut menggunakan teknik *clustering* dan visualisasi informasi untuk menggambarkan kesenjangan antara topik penelitian dengan topik yang dibicarakan di jejaring *social media*.

Artikel ini mengusulkan metode untuk mengukur kesenjangan antara topik penelitian skripsi mahasiswa dengan perkembangan TIK mutakhir dengan cara membandingkan abstrak skripsi dengan abstrak riset yang di publikasi ilmiah baik jurnal maupun prosiding konferensi. Dataset skripsi mahasiswa program Sarjana diambil dari STMIK Pradnya Paramita Malang (STIMATA), sedangkan dataset abstrak riset dikumpulkan dari IEEE Explore. Pendekatan eksploitasi *Association Rule (AR mining)* digunakan untuk menemukan hubungan kata-kata yang sering muncul (*frequent termset*) dari kedua dataset tersebut.

Hasil implementasi metode yang diusulkan berhasil menemukan topik-topik penelitian yang sering dibahas di jurnal dan prosiding konferensi namun belum pernah dibahas di penelitian skripsi mahasiswa; dan sebaliknya.



KAJIAN LITERATUR

Association rule

Association rules digunakan untuk menemukan hubungan di antara data atau bagaimana suatu kelompok data mempengaruhi suatu keberadaan data yang lain. Metode ini dapat membantu mengenali pola-pola tertentu di dalam kumpulan data yang besar. Dalam association rules, suatu kelompok item dinamakan itemset. Support dari itemset X adalah persentase transaksi di D yang mengandung X, biasa ditulis dengan $supp(X)$.

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap:

1. Analisis pola frekuensi tinggi

Tahapan ini mencari kombinasi item yang memenuhi syarat minimum dari nilai support dalam database. Nilai support sebuah item diperoleh dengan rumus berikut.

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}}$$

Sementara itu, nilai support dari 2 item diperoleh dari rumus 2 berikut.

$$\text{Support (A,B)} = \frac{\sum \text{Transaksi mengandung A dan B}}{\sum \text{Transaksi}}$$

2. Pembentukan aturan asosiasi Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk confidence dengan

menghitung confidence aturan asosiatif $A \rightarrow B$.

Nilai confidence dari aturan $A \rightarrow B$ diperoleh dari rumus berikut.

$$\text{Confidence} = P(B|A) = \frac{\sum \text{Transaksi mengandung A dan B}}{\sum \text{Transaksi mengandung A}}$$

Apriori

Apriori adalah algoritma yang digunakan dalam melakukan pencarian *frequent itemset* untuk mendapatkan aturan asosiasi. Sesuai dengan namanya, algoritma ini menggunakan prior knowledge mengenai frequent itemset properties yang telah diketahui sebelumnya untuk memproses informasi selanjutnya. Apriori menggunakan pendekatan secara iterative yang disebut juga sebagai level-wish search dimana k-itemset digunakan untuk mencari (k+1)- itemset. Pertama-tama dicari set dari frequent 1-itemset, set ini dinotasikan sebagai L1. L1 yaitu large itemset pertama yang digunakan untuk menemukan L2, kemudian set dari frequent 2-itemset digunakan untuk menemukan L3, dan seterusnya sampai tidak ada lagi frequent k-itemset yang dapat ditemukan. Large itemset adalah itemset yang sering terjadi atau itemset-itemset yang sudah melewati batas minimum support yang telah ditentukan.

Adapun dua proses utama yang dilakukan dalam algoritma apriori, yaitu:



1. Join (penggabungan). Pada proses ini setiap item dikombinasikan dengan item yang lainnya sampai tidak terbentuk kombinasi lagi.
2. Prune (pemangkasan). Pada proses ini, hasil dari item yang telah dikombinasikan tadi lalu dipangkas dengan menggunakan minimum support yang telah ditentukan oleh user.

Metode Penemuan Kesenjangan

Akhriza dkk, (2017) pada jurnalnya *Revealing The Gap Between Skillset Of Student And Evolving Skills Required By The Industry Of Information And Communication Technology* dan juga Litecky dkk, (2010) pada jurnalnya *Mining Computing Jobs* kedua penelitian ini memiliki kesamaan yaitu mengeksploitasi lowongan kerja yang dibutuhkan industri sebagai acuan perkembangan TIK, namun yang membedakan, pada jurnal Akhriza dkk, memiliki tujuan lain yaitu menemukan kesenjangan antara keterampilan siswa dengan keterampilan yang dibutuhkan pihak industri, sehingga pihak sekolah atau kampus bisa melakukan evaluasi terhadap kurikulum yang ada. Namun penelitian ini menggunakan frequent itemset saja, sehingga hubungan keterkaitan topik satu dengan yang lain belum ditemukan.

Son dkk, (2015) pada jurnal yang berjudul *Visualization of e-Health Research Topics and Current Trends Using Social Network Analysis* [6]. penelitian ini bertujuan untuk secara kritis meninjau topik penelitian utama dan tren internasional e-kesehatan melalui analisis jaringan sosial. Dalam penelitian ini menggunakan pendekatan *clustering* atau pengelompokan.

Penelitian tentang kesenjangan yang dilakukan oleh Akhriza dkk (2017), Litecky dkk, (2010) dan Son dkk (2015) belum mengukur *frequent termset* atau himpunan kata yang ada di dalam abstrak. Mereka hanya fokus mengukur kesenjangan metode dengan menggunakan *frequent itemset* dan belum menggunakan *Association Rule* (AR Mining), padahal association rule dapat mengungkapkan hubungan topik yang satu dengan yang lain.

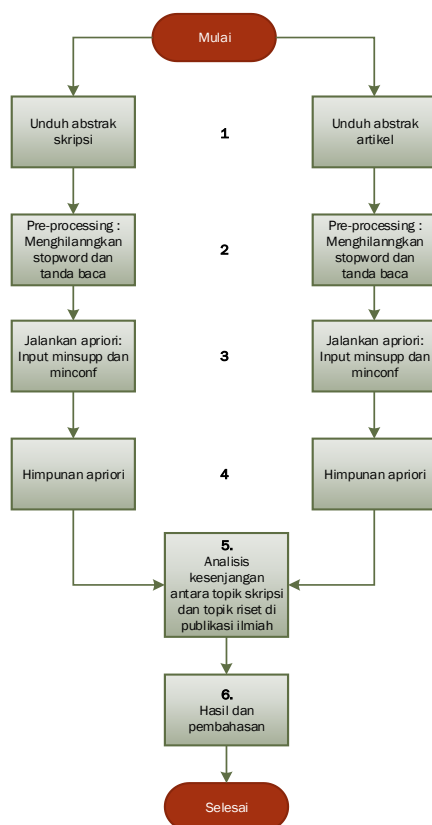
Maka diusulkan penelitian ini bertujuan untuk mengungkapkan kesenjangan tersebut dengan menggunakan *Association Rule* (AR Mining), dimana kata-kata yang paling sering muncul (*frequent termset*) di dalam abstrak skripsi dan abstrak riset dieksploitasi. Dari *frequent termset* yang dikumpulkan, hubungan antara satu kata dengan yang lain ditemukan, untuk kemudian, perbedaan antara topik yang dibahas di dalam skripsi dengan yang

diangkat oleh publikasi ilmiah dianalisis dan disimpulkan. Dari hasil simpulan tersebut, maka dibandingkan dan dihasilkan topik yang sering diangkat di skripsi, namun tidak pernah diangkat di riset yang sudah dipublikasi dan sebaliknya, topik yang tidak pernah diangkat di skripsi, namun sudah sering diangkat di riset.

METODE PENELITIAN

Framework Penelitian

kerangka kerja penelitian ini seperti *framework* pada **Gambar 1**. Angka yang tertera pada gambar adalah urutan alur proses penelitian.



Gambar 1. Kerangka kerja penelitian

- Proses ke-1 unduh abstrak skripsi dan abstrak riset. Pada proses **ke-1** file abstrak di simpan dalam bentuk **.txt**.
- Pada proses **ke-2** setelah abstrak sudah dalam bentuk **.txt** dilanjutkan dengan proses **pre-processing** yaitu pembersihan *stopword* dan tanda baca pada abstrak. Sehingga abstrak berisi inti dari pembahasan skripsi dan riset jurnal.
- Dari proses pre-processing berlanjut ke proses **ke-3** yaitu menjalankan aplikasi **apriori**, dengan menentukan ambang batasnya itu *minsupp* dan *minconf* yang diinginkan.
- Setelah proses aplikasi **Apriori**, maka proses **ke-4** menghasilkan *frequent termset* dari aplikasi **Apriori**.
- Setelah itu, hasil *frequent termset* dari apriori berlanjut ke tahap selanjutnya proses **ke-5** yaitu analisis kesenjangan antara topik skripsi dan topik riset. fokus disini adalah untuk menemukan yang kata kunci sebelah kirinya adalah sistem informasi, yang disebut kesenjangan atau *gap* adalah seberapa banyak kata yang tidak sama antara topik skripsi dan topik riset.
- Selanjutnya proses **ke-6**, maka diperoleh hasil analisis kesenjangan sehingga terlihat topik keterkaitan



topik yang dibahas. Dan akan terlihat topik yang sering digunakan di skripsi, namun tidak pernah digunakan di riset ataupun topik yang jarang diangkat tetapi menarik untuk dibahas.

Rancangan Eksperimen

Data abstrak yang digunakan pada penelitian ini ada dua macam, yaitu data abstrak skripsi dan data abstrak riset yang di publikasi. Data abstrak skripsi diperoleh abstrak skripsi program sarjana dari STMIK PPKIA Pradnya Paramita Malang (STIMATA) prodi sistem informasi dengan 5 periode terakhir dan data abstrak riset diperoleh dari situs www.ieee.com dengan kata kunci "Information System" *filter* tahun 2012-2017, dan artikel yang diambil hanya yang dipublikasi di Jurnal dan Konferensi yang diindeks oleh IEEE Explore.

Pada penelitian ini, ambang batas data abstrak skripsi diatur *minsupp* 3 dan *minconf* 30% dan pada data abstrak riset dengan *minsupp* 10 dan *minconf* 50%, setelah itu pada data abstrak skripsi mengambil fokus pencarian kata kunci "sistem informasi", "sistem penunjang", "sistem berbasis" dan "database", dan pada abstrak publikasi focus pencarian pada kata kunci "system information", "decision

support", "database" dan "web". *Filtering* digunakan sebagai acuan untuk menentukan pola tertentu.

Jika *filtering* sudah dilakukan dengan menggunakan kata kunci dan diperoleh hasil, maka selanjutnya dilakukan perbandingan antara hasil *filtering* data skripsi dan *filtering* abstrak publikasi. Dalam penelitian ini perbandingan menggunakan operasi matematika, yaitu irisan, sehingga dapat mengungkap topik yang sering digunakan di riset, namun tidak pernah digunakan pada skripsi dan topik yang sering digunakan di skripsi, namun belum pernah digunakan di riset. hal tersebut dapat mengungkap kesenjangan dengan membandingkan topik yang dibahas.

HASIL PENELITIAN DAN PEMBAHASAN

Hasil dari penelitian ini dapat di lihat dari tabel berikut. Pada **Tabel 1** adalah hasil *filtering* dari *frequent termset* AR *mining* data abstrak skripsi. Pada tabel tersebut ada pembahasan scaffolding \Leftarrow android informasi sistem (3, 33.3333) pada abstrak skripsi, artinya jika mahasiswa membahas android, system dan informasi, maka mahasiswa juga pasti membahas scaffolding. Angka 3 didalam kurung adalah menunjukkan *support*, yaitu menunjukkan bahwa



keseluruhan dari total pembahasan mahasiswa tentang scaffolding dengan android, system dan informasi muncul secara bersamaan yaitu sebanyak 3 kali. Sedangkan angka 33.3% adalah *confidence*, artinya, ada *confidence* (peluang) sebesar 33.3% bahwa orang yang membahas android informasi juga membahas scaffolding. Pembahasan yang sejenis dg ini bisa menyesuaikan.. *scaffolding* sendiri adalah suatu struktur sementara yang digunakan untuk menyangga manusia dan material dalam konstruksi atau perbaikan gedung dan bangunan-bangunan besar lainnya. Pembahasan tersebut notabene tidak ada hubungannya dengan TIK, namun penelitian mahasiswa melalui skripsi sudah membahas hal tersebut. Terlihat juga ada pembahasan seo \Leftarrow saw pemilihan additive simple pendukung keputusan sistem (3, 33.3333), penggabungan seo atau *search engine optimization* dengan metode saw atau *simple additive weighting*. Penggabungan tersebut bisa dibilang jarang terjadi.

Pada **Tabel 2** adalah hasil *filtering* dari frequent termset AR mining data abstrak riset. Pada pembahasan tersebut terdapat agile \Leftarrow software development systems system information (10, 50). Pembahasan Agile dan *system information* dapat dikatakan banyak dibahas dan *confidence* nya terbilang tinggi 50%. Agile adalah sebuah

metode dari beberapa kumpulan prinsip untuk pengembangan software di mana persyaratan dan solusi melalui upaya kolaboratif dari antar tim fungsional dan klien. Dan ada juga pembahasan health \Leftarrow decision performance information (10, 50), hal ini memberi penilaian baru, bahwa sistem keputusan bukan hanya berhubungan dengan bidang industri saja, namun juga bias masuk di bidang kesehatan.

Tabel 1. Hasil *filtering* data abstrak skripsi

No.	Hasil
1	scaffolding \Leftarrow android informasi sistem (3, 33.3333)
2	bigbluebutton \Leftarrow website pendukung sistem (3, 33.3333)
3	seo \Leftarrow saw pemilihan additive simple pendukung keputusan sistem (3, 33.3333)
4	saw \Leftarrow gerabah pendukung sistem (3, 66.6667)
5	ahp \Leftarrow pengambilan penunjang keputusan sistem (3, 66.6667)
6	akademik \Leftarrow ahp penunjang keputusan sistem (3, 33.3333)
7	sms \Leftarrow android informasi sistem (3, 33.3333)
8	geografis \Leftarrow website mysql informasi sistem (3, 66.6667)



Tabel 2. Hasil *filtering* data abstrak riset

No	Hasil
1	agile \Leftarrow software development systems system information (10, 50)
2	communication \Leftarrow realize system information (10, 50)
3	health \Leftarrow decision performance information (10, 50)
4	mapping \Leftarrow web software system information (11, 54.5455)
5	hierarchy \Leftarrow assessment risk process system information (11, 63.6364)
6	analysis \Leftarrow decisionmaking system information (13, 69.2308)

Tabel selanjutnya menjelaskan tentang kesenjangan antara topik skripsi dan topik riset yang di publikasi. seperti Pada **Tabel 3** menjelaskan tentang topik yang dibahas publikasi, namun tidak pernah di angkat pada skripsi, seperti metodologi penelitian TIK semacam *Agile software development* yang ternyata sudah banyak dipakai di publikasi.

Pada **Tabel 4** menjelaskan tentang topik yang dibahas pada skripsi, namun tidak ada pada topik publikasi, seperti Bigbluebutton tidak pernah dibahas pada publikasi.

Dari hal ini kita dapat mengambil simpulan kesenjangan antara kedua topik, yaitu topik skripsi dan topik publikasi.

Tabel 3. hasil kesenjangan abstrstrak skripsi dan abstrak riset

No	Hasil
1	Agile
2	Hierarchy
3	Evaluate
4	Communication
5	Resources
6	Resources
7	Study

Tabel 4. hasil kesenjangan riset dan skripsi

No	Hasil
1	Scaffolding
2	Bigbluebutton
3	Administrasi
4	Fuzzy
5	Seo
6	Mysql
7	Ahp

KESIMPULAN DAN SARAN

Kesimpulan

Dari hasil penelitian terhadap analisis kesenjangan antara topik skripsi dan topik riset yang di pulikasi menggunakan



eksploitasi *Association rule* (AR Mining) dapat di simpulkan bahwa:

1. Penelitian ini berhasil mengungkap kesenjangan antara topik skripsi dan topik riset yang di publikasi dengan efektif dan efisien, sehingga hasil analisis dapat dipergunakan untuk mahasiswa atau perguruan tinggi ke depan.
2. Pada penelitian ini pula, dapat ditemukan keterkaitan topik satu dengan yang lain, sehingga dapat menambah wawasan dan pengembangan ranah penelitian untuk selanjutnya.

Saran

Untuk meningkatkan keakuratan hasil, data abstrak skripsi dan abstrak riset ilmiah dapat ditambahkan lebih banyak lagi.

REFERENSI

- Akhriza, T. M. Ma, Y. H. Dan Li. J. H. 2017. *Revealing The Gap Between Skillset Of Student And Evolving Skills Required By The Industry Of Information And Communication Technology*. International Journal Of Software Engineering And Knowledge Engineering. Vol 27, No. 5 2017: 675-698.
- Abhiesa, Alif. 2016, “Cara Instal Bigbluebutton Web Conferencing”, <http://www.noobsgnu.com/2015/06/cara-instal-bigbluebutton-web.html>. diakses 3 Agustus 2017
- Anggraeni, H. D. Saputra, Ragil. Dan Noranita, Beta. 2013. *Aplikasi Data Mining Analisis Data Transaksi Penjualan Obat Menggunakan Algoritma Apriori (Studi Kasus di*

Apotek Setya Sehat Semarang) Journal Of Informatics And Technology, vol. 2, no. 2, pp. 22 - 28, 2013.

- Djojodibroto, R. D.,. 2004. *Tradisi Kehidupan Akademik*. Galang Press. Yogyakarta.
- Hernanda, Fikhri. 2016, “Agile Software Development, salah satu metode pengembangan perangkat lunak”, <https://www.dictio.id/t/agile-software-development-salah-satu-metode-pengembangan-perangkat-lunak/3242> , diakses 3 agustus 2017.
- Kusrini dan Luthfi, E. T. 2009. *Algoritma Data Mining*. CV. Andi Offset. Yogyakarta.
- Litecky, C. Aken, A. A. dan J. Nelson. *Mining For Computing Jobs*, IEEE Software 27(1) (2010) 78–85.
- Son, Youn-jun., Jeon. Senator, Kang. Byeon-Gwon, Kim. Sun-Hyung, dan Lee, Soo-Kyoung. 2015. *Visualization of e-Health Research Topics and Current Trends Using Social Network Analysis*. Mary Ann Liebert, Inc. Vol. 21 No. 5.