

DATA MINING UNTUK REKOMENDASI KERJA BAGI ALUMI DENGAN ALGORITMA GARC(GAIN BASED ASSOCIATION RULE CLASSIFICATION)

Steffi Budi Fauziah¹, Shaufiah², Mahmud Dwi Suliyo³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Menjelang kelulusan mahasiswa perlu diarahkan mengenai karirnya, di IT Telkom salah satunya. Untuk itu di IT Telkom ada fasilitas CDC (Career Development Center). CDC berfungsi sebagai penyedia lowongan kerja untuk mahasiswa IT Telkom yang telah lulus. Banyaknya lowongan kerja membuat pelamar perlu diarahkan karirnya supaya sesuai dengan bidangnya. Tujuan dari penelitian ini adalah menemukan pola untuk prediksi pelamar dalam merekomendasikan alumni untuk menentukan pekerjaannya. Penelitian ini menggunakan algoritma GARC (Gain based Association Rule Classification) untuk menentukan rule yang akan digunakan dalam memprediksi pelamar menentukan pekerjaannya. Dari hasil yang diujikan dengan information gain mendapatkan akurasi yang lebih baik dibandingkan dengan yang tidak menggunakan information gain. Sedangkan pada pengujian lainnya menggunakan pengaturan confidence dan support. Dengan hasil confidence 0.51 dan support 0.01 serta confidence 0.51 dan support 0.001 yang terbaik akurasinya yaitu 60%. Hasil dari penelitian ini adalah prediksi pelamar dalam menentukan pekerjaan yang sesuai dengan bidangnya.

Kata Kunci : GARC, job vacancy, confidence, support, information gain

Abstract

Before graduate, students need to be directed to choose their career, for example in IT Telkom. Therefore, IT Telkom provide CDC (Career Development Center). CDC provides job vacancy informations for IT Telkom students whom have been graduated. There are so many job vacancy in which make the applicant need to be directed so that their career match their scope of study. The aim of this study is to fine patterns to predict and recomend applicant determine their job. This study uses GARC (Gain based Association Rule Classification) algorithm to establish the rules which will be used as parameter to predict the job. From the testing result, the used of information gain makes the result more accurate than not using information gain. While for the other testing is using confidence and support. With the result confidence 0.51 and support 0.01 and confidence 0.51 and support 0.001, the best accuration is 60%. The result from this study is to predict applicant job based on their scope of study.

Keywords : GARC, job vacancy, confidence, support, information gain

Telkom
University

1. Pendahuluan

1.1 Latar belakang

Menjelang kelulusan, alumni disetiap perguruan tinggi akan mencari pekerjaan. Begitu juga pada alumni di Yayasan Perguruan Tinggi (YPT) IT Telkom. Tidak jarang pekerjaan yang didapat tidak sesuai dengan bidang yang ditekuninya ataupun rekomendasi tentang informasi pekerjaan kurang. Padahal ini sangat penting untuk membantu alumni mendapatkan pekerjaan yang tepat dan sesuai dengan bidangnya. Selain itu, juga dapat membantu institusi dalam mencari mahasiswa yang tepat untuk suatu pekerjaan tertentu. Untuk itu alumni perlu diarahkan mengenai karir yang akan dijalani. Supaya alumni tidak terjebak dalam memilih karirnya, diperlukan sebuah sistem untuk dapat merekomendasikan alumni mengikuti seleksi pekerjaan sesuai dengan bidang yang diinginkannya. Sehingga alumni tidak akan lagi terjebak pada karirnya yang tidak sesuai dengan bidangnya.

Permasalahan tersebut dapat menggunakan berbagai metode data mining misalnya menggunakan metode *Decision Tree*, *Rule Association*, atau *Classification*. Untuk memilih metode mana yang tepat terdapat beberapa faktor, seperti ukuran dari *dataset*, tipe dari pola yang sudah ada dalam data, apakah data cocok dengan asumsi dari metode yang diterapkan, seberapa banyak *noise* dalam data, dan apa tujuan khusus dari analisa. Teknik klasifikasi dan asosiasi memungkinkan untuk menyelesaikan masalah tersebut. Pada teknik klasifikasi terdapat beberapa metode yang dapat digunakan yaitu ID3, C4.5, C.5, SVM, NN, Naive Bayes dll [2]. Pada asosiasi juga terdapat metode yang dapat digunakan yaitu salah satunya adalah apriori. Dapat juga menggabungkan kedua metode klasifikasi dan asosiasi untuk menyelesaikan permasalahan tersebut, yaitu dengan menggunakan algoritma GARC (*Gain based Association Rule Classification*).

Pada penelitian sebelumnya telah disebutkan bahwa metode GARC adalah metode yang terbaik untuk mendapatkan prediksi dalam pencarian rule. Salah satu pengujian yang dilakukan pada referensi adalah membandingkan akurasi dengan menggunakan *information gain* dan tanpa *information gain* [2]. Dari pengujian tersebut didapatkan bahwa algoritma GARC (dengan menggunakan *information gain*) yang mendapatkan akurasi terbesar [2]. Hal ini terjadi disebabkan oleh adanya rule yang redundan pada pengujian tanpa *information gain* saat pencarian rule. Sedangkan pada algoritma GARC rule yang dihasilkan tidak ada yang redundan. Sesuai dengan kelebihan dari algoritma GARC ini adalah mencari rule yang efisien dan dapat menghindari redundansi dari rule yang telah didapat. Rule yang redundan ini dihilangkan pada metode klasifikasi yaitu menggunakan *information gain*. Pada *information gain* ini mencari nilai *gain* terbesar untuk dijadikan kandidat itemset pada metode asosiasi apriori. Maka, pada tahapan ini tidak akan ada rule yang redundan. Setelah itu, dilakukan pencarian nilai *confidence* dan *support* untuk pencarian rule yang efisien. Rule yang telah didapat tersebut yang akan menjadi *knowledge* untuk memprediksi pelamar dalam mendapatkan pekerjaan yang sesuai dengan bidangnya.

1.2 Perumusan masalah

Adapun pokok perumusan masalah yang akan diselesaikan dalam Tugas Akhir ini adalah perlunya alumni untuk diarahkan mengenai karirnya menjelang kelulusan, maka dari itu pada tugas akhir ini dibuat sebuah sistem untuk merekomendasikan alumni untuk mengetahui karirnya yang sesuai dengan bidangnya dengan metode GARC (*Gain based Association Rule Classification*).

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini antara lain:

1. Penelitian menggunakan data dari CDC IT Telkom dan sebagian dari CV alumni.
2. Atribut yang digunakan pada data Job_Seeker ada beberapa atribut secara keseluruhan yaitu: jenjang, jurusan, IPK, usia, keahlian, pengalaman kerja, dan jenis kelamin.
3. Data CV (*Curriculum Vitae*) yang digunakan berbahasa Indonesia dan file input (data testing) menggunakan beberapa dokumen Microsoft Word (.doc), sebagian data CV menggunakan data CDC yang ada di Microsoft Excel (.xls).
4. Data yang di Microsoft Word hanya berbentuk text (tidak ada gambar).
5. Format CV sudah ditentukan dan tidak bebas, atribut di CV tidak boleh ada yang null, dan menggunakan Microsoft Word 1997- 2003.

1.4 Tujuan

Tujuan yang ingin dicapai dari penelitian ini sebagai berikut :

1. Menemukan pola untuk prediksi dalam merekomendasikan alumni untuk menentukan pekerjaannya sesuai dengan bidangnya.
2. Mengetahui performansi dari algoritma GARC.

1.5 Metodologi Penyelesaian Masalah

Metodologi penyelesaian masalah dalam tugas akhir ini terbagi kedalam beberapa tahapan, yaitu:

1. Studi Literatur
Mencari dan mempelajari literatur tentang teknik data mining menggunakan algoritma GARC dan mempelajari secara mandiri algoritma tersebut.
2. Pengumpulan data
Data yang digunakan untuk merekomendasi pekerjaan alumni menggunakan CV (*Curriculum Vitae*) karena dengan CV dapat mengetahui identitas, latar belakang pendidikan, pengalaman alumni dan dapat terlihat juga kompetensi mahasiswa.
3. Perancangan
Dilakukan perancangan untuk mempermudah melakukan implementasi pada tahap selanjutnya dengan membuat *flowchart* untuk mempermudah melihat perancangan sistemnya.

4. Implementasi
 - a. Dilakukan implementasi pada algoritma GARC dengan dan tanpa menggunakan *information gain*.
 - b. Dilakukan implementasi pada algoritma GARC dengan asosiasi apriori
5. Analisa
 - a. Analisis hasil pengujian testing dan data training
 - b. Kesimpulan hasil pengujian
6. Pembuatan laporan Tugas Akhir

1.6 Sistematika Penulisan

Adapun sistematika penulisan Tugas Akhir ini adalah sebagai berikut:

1. BAB 1 Pendahuluan
Pada Bab 1 diuraikan isi dan rencana pengerjaan Tugas Akhir secara keseluruhan yang meliputi latar belakang, masalah, rumusan masalah, tujuan, batasan masalah, dan metode penyelesaian masalah yang diterapkan.
2. BAB 2 Dasar Teori
Bab 2 memaparkan dasar-dasar teori yang berkaitan dengan Data Mining, algoritma GARC, *preprocessing* dan *information gain*.
3. BAB 3 Perancangan dan Implementasi
Perancangan sistem, analisis kebutuhan sistem, dan contoh implementasi dari sistem yang akan dibangun akan dipaparkan pada bab ini. Selanjutnya akan dilakukan proses implementasi.
4. BAB 4 Pengujian dan Analisis
Pada bab ini akan dibahas scenario dan hasil pengujian yang dilakukan pada hasil implementasi sistem.
5. BAB 5 Kesimpulan dan Saran
Bab ini berisi kesimpulan dan saran yang didapatkan dari hasil implementasi sistem secara keseluruhan.

5. Penutup

5.1 Kesimpulan

Kesimpulan yang dapat diberikan untuk penelitian ini terkait dengan algoritma GARC yang pada studi kasus seleksi pelamar job vacancy, sebagai berikut :

1. Pada pengujian seleksi pelamar pekerjaan menggunakan algoritma GARC dengan *information gain* dengan pengurangan atribut usia mendapatkan akurasi yang tinggi yaitu 60%. Hasil yang didapat nilai training 500 mendapatkan akurasi yang paling tinggi yaitu 60%. Maka semakin banyak nilai training yang digunakan makin banyak juga *knowledge* yang didapat untuk testing.
2. Pada pengujian seleksi pelamar pekerjaan tanpa *information gain* mendapatkan akurasi rendah yaitu pada training 500 yaitu 44%, pada training 400 yaitu 15% dan pada training 300 yaitu 3%. Hanya menggunakan asosiasi apriori saja tanpa menggunakan *information gain* akan menambah jumlah rule yang sama (redundan) dan rule yang dihasilkan lebih sedikit dibandingkan dengan menggunakan *information gain*.
3. Pada pengujian seleksi pelamar pekerjaan menggunakan algoritma GARC dengan pengaturan minimal *confidence* dan minimal *support*. Dilihat dari pengujiannya dengan *confidence* 0.51 dan *support* 0.01 dan 0.001 yang terbaik karena nilai akurasinya 60%. Semakin minimal menggunakan nilai *support* semakin banyak rule yang akan dikeluarkan.

5.2 Saran

Adapun saran perbaikan yang dapat dilakukan untuk penelitian selanjutnya mengenai data mining dengan studi kasus seleksi pelamar job vancy sebagai berikut :

1. Menggunakan semua data CV pelamar yang berformat Microsoft Word (.doc).
2. Pengujian tanpa *information gain* dilakukan dengan menggunakan program, tidak menggunakan aplikasi weka.

6. Referensi

- [1] Al-Radaideh. Qasem A, A. N. (2012). *Using Data Mining Techniques to Build a Classification Model for Predictiong Employees Performance*. Vol.3 N0.2.
- [2] Chen. Guoqing. Et. al, “A new approach to classification based on association rule mining”, Department of Management Science and Engineering, School of Economics and Management, Tsinghua University, Beijing 100084, China.
- [3] Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). “From Data Mining to Knowledge Discovery in Databases”. *AI Magazine*, 38-54.
- [4] Fayyad, U., dan Stolorz, P., (1997). *Data Mining and KDD: Promise and Challenges*. *Future Generation Computer Systems*, 13, 99-115.
- [5] Febriani, Atik. “Perancangan Model Pemilihan Laptop Berdasarkan Customer Behaviour Menggunakan Garc (Gain Based Association Rule Classification)”. 2011. Jurusan Teknik Industri. Fakultas Teknologi Industri. Universitas Islam Indonesia. Yogyakarta.
- [6] Han, Jiawie, and Kamber, Micheline. *Data Mining : Concept and Techniques*. Morgan Kaufmann Publishers, San Francisco, CA. 2001.
- [7] L. Breiman, *Classification and Regression trees*, Wadsworth, Belmont, 1984.
- [8] Larose, Daniel, T. *Discovering Knowledge in Data: An Introduction to Data Mining*. 2005. John Wiley & Sons, Inc., Hoboken, New Jersey.
- [9] M.J. Zaki, S. Parthasarathy, M. Ogihara, W. Li, “New algorithms for fast discovery of association rules”, *American Association for Artificial Intelligence* (1999).
- [10] Olafsson, S., Xiaonan Li, dan Shuning Wu, (2008). *Operations Research and Data Mining*. *European Journal of Operational Research*, **187**, 1429-1448.