

ANALISIS DAN IMPLEMENTASI CHURN PREDICTION MENGGUNAKAN ALGORITMA GENETIKA (STUDI KASUS MOBILE TELECOMMUNICATION)

Quvin Nola Prihandani¹, Moch Arif Bijaksana², Bambang Gunadi A Ir³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Industri telekomunikasi merupakan bisnis yang sedang berkembang saat ini. Banyak perusahaan baru yang muncul dengan menawarkan berbagai layanan dan fasilitas yang sanggup menarik pelanggan. Sehingga banyak pelanggan yang dapat dengan mudah melakukan churn. Keterbatasan jumlah staff untuk menghubungi semua pelanggan juga menjadi kendala untuk mengurangi fenomena ini. Oleh karena itu diperlukan suatu sistem data mining yang mencakup berbagai macam teknik untuk memprediksi apakah seorang pelanggan berpotensi churn atau tidak. Dalam tugas akhir ini digunakan suatu Algoritma Genetika dengan salah satu variant-nya yaitu Data Mining Evolutionary Learning (DMEL) [8] untuk melakukan klasifikasi dalam menentukan pelanggan tersebut. Metode ini digunakan karena DMEL dapat dengan efektif menemukan rule dan dapat menangani missing value. Studi kasus yang digunakan dalam tugas akhir ini adalah churn prediction untuk pelanggan salah satu pelanggan perusahaan telekomunikasi di Indonesia. Dalam tugas akhir ini dilakukan perbandingan keakuratan hasil klasifikasi yang dihasilkan oleh perangkat lunak Churn Prediction dengan hasil prediksi menggunakan tools Clementine 9.0 terhadap algoritma C.5, Neural Network dan CHAID. Setelah dilakukan percobaan, DMEL terbukti lebih baik dari model-model yang dibangun menggunakan tools Clementine 9.0.

Kata Kunci : churn, data mining, Data Mining Evolutionary Learning, tools

Abstract

Telecommunications industry is expanding business in this time. Many new company which emerge by offering various service and facility which ready to draw customers. So many customers easily to churn. Limitedness sum of staff to contact all customers is also become the problem to lessen this phenomenon. Therefore needed a system of data mining including assorted technique for prediction that customers have potency to churn or not churn. This final project used an Genetic Algorithm, it have one variant that is Data Mining Evolutionary Learning (DMEL) [8] to build classification in determining the customers. This method is used because DMEL can effectively find the rule and can handle the missing value. The accuracy of classification yielded by software of Churn Prediction in this final project was comparison with with result using Clementine 9.0 with C.5 algorithm, Neural Network and CHAID. After conducted with the attempt, DMEL proven is better than models which constructed by Clementine 9.0

Keywords : churn, data mining, Data Mining Evolutionary Learning, tools

1. Pendahuluan

1.1 Latar belakang

Perkembangan telekomunikasi yang sedemikian pesat mendorong terbentuknya perusahaan-perusahaan baru yang bergerak dalam bidang telekomunikasi. Perkembangan jaman, gaya hidup dan kebutuhan akan fasilitas komunikasi menjadi beberapa latar belakang orang banyak menggunakan telepon selular baik GSM (*Global System Mobile*) maupun CDMA (*Code Division Multiple Access*) sebagai media komunikasi. Alat komunikasi ini semakin mempermudah komunikasi yang dapat digunakan kapan saja dan dimana saja. Harga yang terjangkau oleh masyarakat juga menjadi salah satu faktor pertimbangan orang banyak menggunakannya.

Untuk dapat menarik banyak pelanggan, banyak layanan yang ditawarkan untuk memenuhi kebutuhan pelanggan. Persaingan harga pun menjadi *trend* perusahaan untuk menarik pelanggan sebanyak-banyaknya. Sehingga jika perusahaan telekomunikasi tidak melakukan analisis bisnis dan identifikasi karakteristik pelanggan yang tepat, maka tidak mustahil bila perusahaan tersebut dapat kehilangan pelanggannya.

Saat ini sangat mudah bagi seorang pelanggan untuk *churn* dan berpindah dari satu perusahaan telekomunikasi ke perusahaan telekomunikasi yang lain. Pengertian *churn* disini adalah ketika semua jasa yang digunakan diputus oleh pelanggan ataupun oleh perusahaan. Ada banyak faktor atau alasan yang melatarbelakangi seorang pelanggan dapat memutuskan untuk *churn*. Beberapa faktor itu antara lain perusahaan telekomunikasi lain yang menawarkan harga yang lebih rendah, *traffic* yang tinggi sehingga membuat pelanggan sulit untuk melakukan *call*, pelayanan yang tidak memuaskan terhadap pelanggan dan masih banyak lagi faktor lain.

Dengan keterbatasan jumlah staff pada suatu perusahaan untuk menghubungi sebagian kecil dari pelanggan, maka diperlukan usaha untuk mencirikan pelanggan dengan tingkat kemungkinan *churn* yang berbeda-beda. Agar dengan sumber daya yang terbatas tersebut, pelanggan yang mempunyai kemungkinan untuk *churn* tinggi dihubungi terlebih dahulu. Oleh karena itu diperlukan suatu sistem *data mining* yang mencakup berbagai macam teknik untuk memprediksi apakah seorang pelanggan berpotensi *churn* atau tidak. Hal ini dilakukan agar suatu perusahaan dapat memiliki analisis bisnis dan identifikasi karakteristik pelanggan yang tepat dalam meningkatkan kompetisi demi keuntungan perusahaan, menguasai *market share* pada bidang marketing dan perusahaan dapat melakukan penanggulangan yang tepat kepada karakteristik pelanggan tersebut.

Metode yang dapat digunakan dalam hal ini adalah algoritma Genetika dengan salah satu *variant*-nya yaitu *Data Mining Evolutionary Learning* (DMEL). Metode ini digunakan karena DMEL dapat dengan efektif menemukan *rule*, terutama dapat meramalkan *churn* dengan teliti dibawah tingkat *churn* yang berbeda serta dapat menangani *missing value* [8]. DMEL ini menggunakan *lift curve* untuk memperkirakan fraksi dari *churner* yang dapat ditangani oleh jumlah staff yang terbatas, sehingga staff dapat menghubungi fraksi pelanggan yang benar-benar *churn*. *Lift curve* ini dapat mengevaluasi performansi DMEL yang

dapat menggambarkan perbedaan *churner* dengan *non churner* dengan sempurna [8]. DMEL juga merupakan pengganti dari *random generation* pada algoritma Genetika yang mencari rule secara non random. Sehingga pada saat *initial population* yang terdiri dari *first-order rule* akan di-generate menggunakan teknik *probabilistic induction*. Berdasarkan rule yang telah diperoleh maka rule yang mempunyai *higher order* (mempunyai 2 atau lebih kondisi) akan diperoleh secara iteratif menggunakan *initial population* yang telah diperoleh pada awal dari setiap iterasi. Selain efektif untuk melakukan pencarian rule, DMEL ini juga dapat menangani *missing value*[8].

Sehingga dalam tugas akhir ini akan dibuat suatu implementasi DMEL untuk melakukan klasifikasi terhadap pelanggan salah satu perusahaan telekomunikasi tersebut *churn* atau tidak. Data yang digunakan telah dilakukan proses *preprocessing* tetapi masih terdapat *missing value*. Hasil akhir dari tugas akhir ini akan dilakukan perbandingan keakuratan hasil klasifikasi yang dihasilkan oleh perangkat lunak *Churn Prediction* dengan hasil prediksi menggunakan tools *Clementine 9.0* dengan algoritma C5.0, Neural Network dan CHAID. Perangkat lunak tersebut diharapkan dapat mengatasi masalah *churn prediction*. Dan pada akhirnya dapat membantu perusahaan untuk meningkatkan kompetisi demi keuntungan dan menguasai *market share* pada bidang *marketing* dan perusahaan dapat melakukan penanggulangan yang tepat kepada karakteristik pelanggan .

1.2 Perumusan masalah

Dengan mengacu pada latar belakang masalah diatas, maka permasalahan yang akan dibahas dan diteliti adalah :

1. Bagaimana menerapkan algoritma Genetika dengan salah satu *variant*-nya yaitu *Data Mining Evolutionary Learning* (DMEL) dalam kasus *churn prediction*.
2. Bagaimana menerapkan *Data Mining Evolutionary Learning* sebagai suatu metode melakukan pencarian rule dengan efektif dan dapat menggambarkan perbedaan *churner* dengan *non churner* menggunakan *lift curve*.

1.3 Tujuan

Berdasarkan rumusan masalah di atas, maka tujuan dari tugas akhir ini adalah:

- 1 Menghasilkan suatu perangkat lunak *Churn Prediction* untuk kasus pelanggan salah satu perusahaan telekomunikasi yang banyak terdapat *missing value*
- 2 Menganalisis keakuratan hasil klasifikasi yang dihasilkan oleh perangkat lunak *Churn Prediction* menggunakan variant algoritma Genetika yaitu DMEL
- 3 Membandingkan keakuratan hasil klasifikasi dari *variant* algoritma Genetika yaitu DMEL dengan dengan model yang dibangun menggunakan tool *Clementine 9.0*

1.4 Batasan Masalah

Dalam Tugas Akhir ini, yang akan dibahas adalah suatu implementasi *variant* dari algoritma Genetika yaitu *Data Mining Evolutionary Learning* dalam masalah *churn prediction* dengan batasan masalahnya sebagai berikut :

1. Perangkat lunak yang akan dihasilkan untuk menangani studi kasus pelanggan salah satu perusahaan telekomunikasi.
2. Data pelangganyang digunakan masih terdapat *missing value*.
3. Tidak membahas penggunaan *tool Clementine 9.0* dalam pengujian terhadap algoritma-algoritma yang digunakan.
4. Input data yang digunakan berupa *artefact*
5. Volume data tidak terlalu besar yaitu menggunakan data < 50.000 pelanggan.
6. Aplikasi ini merupakan aplikasi yang berdiri sendiri, tidak menempel pada *tool Clementine 9.0*.

1.5 Metodologi penyelesaian masalah

Metode yang digunakan dalam penyelesaian tugas akhir ini adalah menggunakan metode studi pustaka atau studi literatur dan analisis dengan langkah kerja sebagai berikut :

1. Studi Literatur :
 - a. Pencarian referensi, mencari referensi dan sumber-sumber lain yang layak yang berhubungan dengan *data mining*, *Data Mining Evolutionary Learning*, dan *Genetic Algorithm*.
 - b. Pendalaman materi, mempelajari dan memahami materi yang berhubungan dengan tugas akhir.
2. Mempelajari konsep dari *Data Mining Evolutionary Learning* yang akan digunakan dalam implementasi perangkat lunak
3. Melakukan analisis *Data Mining Evolutionary Learning* dalam perancangan perangkat lunak
4. Melakukan implementasi perancangan perangkat lunak
5. Melakukan pengujian perangkat lunak dengan memasukkan data yang sudah dilakukan proses *preprocessing* serta mencatat hasil keluaran program.
6. Pengambilan kesimpulan dan penyusunan laporan tugas akhir

5. Penutup

5.1 Kesimpulan

1. Semakin kecil parameter ukuran populasi maka performansi DMEL dapat lebih meningkat. Sebaliknya, parameter ukuran probabilitas mutasi yang semakin kecil belum tentu meningkatkan performansi DMEL. DMEL dengan parameter ukuran populasi 30 dan probabilitas mutasi 0.001 menghasilkan performansi paling baik dengan mendapatkan *recall minor* tertinggi.
2. Semakin besar ukuran populasi belum tentu meningkatkan prosentase *actual churn* yang didapatkan. Probabilitas mutasi juga belum tentu berpengaruh terhadap meningkatnya prosentase *actual churn* yang didapatkan untuk membentuk *lift curve*.
3. DMEL dengan parameter ukuran populasi 50 dan probabilitas mutasi 0.001 mampu mendeteksi jumlah *actual churn* yang lebih banyak dari pada yang lain. Untuk pengujian dengan prosentase 2.5% mampu mendeteksi *actual churn* 14.28%, pada prosentase 5% mampu mendeteksi *actual churn* 22.32%, pada prosentase 7.5% mampu mendeteksi *actual churn* 32.14%. Pada prosentase 10% mampu mendeteksi *actual churn* 36.60%.
4. DMEL terbukti lebih baik dari model-model yang dibangun menggunakan Clementine 9.0 dimana DMEL dapat mendeteksi *actual churn* lebih banyak/menghasilkan *lift curve* yang lebih bagus dari pada model-model yang dibangun menggunakan Clementine 9.0.
5. Bila dibandingkan dengan model yang dibangun menggunakan Clementine 9.0, DMEL membutuhkan waktu yang lama untuk melakukan *testing* dan *learning*

5.2 Saran

1. Sistem *churn prediction* ini dapat juga ditambah dengan penanganan khusus terhadap data *imbalance* misalnya menggunakan *boosting*.

Telkom
University

Daftar Pustaka

- [1] Alex Alves Freitas. *Evolutionary Computation*, Curitiba. Brazil.
- [2] Jiawei Han and Micheline Kamber. *Data Mining : Concepts and Techniques*. Intelligent Database Systems Research Lab, School of Computing Science, Simon Fraser University
- [3] DeJong, K.A; Spears, W.M; Gordon, D.F. *Learning Concept Classification Rules Using Genetic Algorithm*. 1993
- [4] John Hadden, Ashutosh Tiwari, Rajkumar Roy, and Dymtr Ruta,. *Churn Prediction using Complaints Data*, ISSN 2006
- [5] Jyothsna R. Nayak and Diane J. Cook. *Approximate Association Rule Mining*, Departement Of Computer Science and engineering Arlington.
- [6] Kusumadewi, Sri. *Artificial Intelligence (Teknik dan Aplikasi. edisi pertama)*. Graha Ilmu. Jogjakarta. 2003. [hal. 279-331]
- [7] Mattison, Rob. *Data Warehousing and Data mining for Telecommunications*. Artech house, INC. 1997
- [8] Ngurah Putu Sumantrika. *Data Mining Task Klasifikasi Menggunakan Algoritma Genetika*. Jurnal TA, Jurusan Teknik Informatika STTTELKOM 2005
- [9] Pang-Ning Tan, Michael Steinbach and Vipin Kumar. *Introduction to Data Mining*. University of Minnesota and Army High Performance Computing Research Center
- [10] Suyanto, *Algoritma Genetika dalam Matlab*, ANDI. Yogyakarta. 2005
- [11] Wai-Ho Au, Keith C. C. Chan, and Xin Yao,. *A Novel Evolutionary Data Mining Algorithm With Application to Churn Prediction*, Fellow, IEEE
- [12] Wai-Ho Au, Keith C. C. Chan, and Xin Yao,. *An Evolutionary Approach for Discovering Changing Patterns in Historical data*, Departement of Computing, The Hong Kong Polytechnic University
- [13] Wai-Ho Au, Keith C. C. Chan, and Xin Yao,. *An Effective Algorithm For Mining Interesting Quantitative Association Rules*, Departement of Computing, The Hong Kong Polytechnic University

Telkom
University