

Rancang Bangun Aplikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa Dengan Algoritma Backpropagation Neural Network

Development Of Data Mining Application For Student Completion Time Using Backpropagation Neural Network

Desi Hanindya Ujjanti¹, Affandy²

^{1,2}Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

^{1,2}Jl. Nakula I, No. 5-11, Semarang, Kode Pos 50131, Telp. (024) 3520165 Fax: 3569684

E-mail : ¹11220104203@mhs.dinus.ac.id, ¹affandy@dsn.dinus.ac.id²

Abstrak

Saat ini sering didapati besarnya jumlah mahasiswa baru pada sebuah perguruan tinggi yang tidak sebanding dengan jumlah lulusan tiap tahunnya. Hal ini merupakan salah satu permasalahan yang harus diselesaikan oleh institusi perguruan tinggi. Agar program studi dapat melakukan langkah antisipasi atau penanganan lebih dini terhadap mahasiswa yang berpotensi tidak lulus tepat waktu, diperlukan suatu mekanisme yang dapat mendeteksi kemungkinan hal ini terjadi. Dewasa ini perkembangan data mining berkembang pesat dalam mengatasi berbagai masalah pengolahan data dengan jumlah besar. Salah satu aplikasi dari data mining adalah untuk keperluan prediksi/klasifikasi, dan salah satu algoritma yang umum digunakan adalah Neural Network. Dalam penelitian ini algoritma Backpropagation Neural Network diterapkan untuk memprediksi masa studi mahasiswa. Dari hasil pengolahan data training dengan algoritma tersebut didapatkan hasil tingkat akurasi yang didapat sebesar 99.49% dengan atribut yang digunakan ialah Indeks Prestasi (IP) dari semester 1 (satu) sampai 4 (empat) dan Indeks Prestasi Kumulatifnya. Pengolahan data tersebut diproses menggunakan RapidMiner dengan menerapkan algoritma Backpropagation Neural Network untuk membentuk pola dari data training. Untuk memperoleh pengetahuan atau knowledge, dari pola tersebut akan diterjemahkan ke dalam bentuk yang dapat dimengerti dan diproses dalam aplikasi Matlab.

Kata Kunci: Data Mining, Prediksi, Algoritma Neural Network (Back Propagation), Prototipe

Abstract

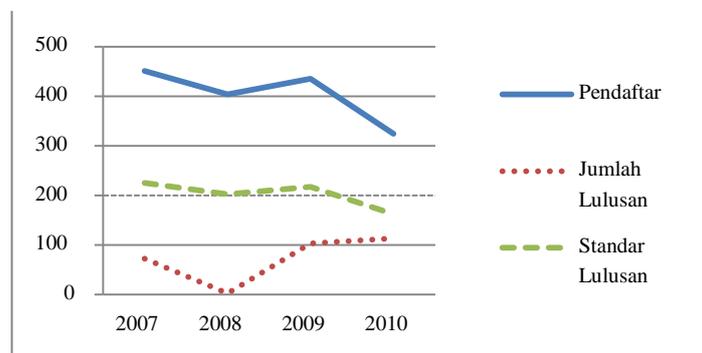
It is found currently the large number of student enrollment was not proportional with the number of graduates every year, where it can be problem that must be resolved by institution of higher education. Head of department needs a mechanism that possible to anticipate or to prepare special treatments for those who potentially could not pass on time. Nowadays, development of data mining grows rapidly to solve the various problems of data processing in a large. One of the data mining application is related with prediction/ classification, and one of the algorithm that generally used is Neural Network. This study applied the Neural Network (Back Propagation) algorithm to predict the student graduation. The results show that the neural network (Back Propagation) algorithm is able to establish effective prediction model from the existing data training and has an accuracy approximately 99.49 % using attributes such as Grade Point Average (GPA) from the first semester to fourth semester and the

Cumulative Grade Point Average (CGPA). The data is processed using RapidMiner by applying the Neural Network (Back Propagation) algorithm to build pattern from the training data. In order to acquire the knowledge of the patterns, the patterns will be transformed into a form that can be understood and it will be processed in Matlab application.

Keywords: Data Mining, Prediction, Neural Network (Back Propagation) Algorithm, Prototype.

1. PENDAHULUAN

Pada institusi pendidikan perguruan tinggi, mahasiswa merupakan aspek penting dalam pencapaian keberhasilan program studi. Hal ini nampak dari borang akreditasi program studi yang menjadikan elemen mahasiswa, sebagai salah satu indikator acuan [1]. Fakta di lapangan menunjukkan rendahnya jumlah mahasiswa program studi xyz yang dapat lulus tepat waktu. Gambar berikut menunjukkan perbandingan antara jumlah mahasiswa pendaftar, jumlah mahasiswa yang lulus tepat waktu, dan target jumlah lulusan sesuai ketentuan BAN-PT [2] yaitu jumlah lulusan harus mencapai minimal 50% dari jumlah pendaftar



Gambar 1. Grafik Pendaftar dan Lulusan

Tampak pada Gambar 1, kesenjangan antara jumlah lulusan mahasiswa progdi xyz dengan jumlah standar minimal kelulusan yang telah ditetapkan oleh BAN-PT. Agar program studi dapat melakukan langkah antisipasi untuk dapat melakukan penanganan lebih dini terkait dengan kondisi mahasiswa yang berpotensi untuk lulus tidak tepat waktu, maka diperlukan suatu mekanisme yang dapat mendeteksi kemungkinan hal ini terjadi.

Dewasa ini *data mining* banyak digunakan untuk proses analisa data dari berbagai perspektif yang kemudian di-*summary* ke dalam bentuk informasi yanga lebih bermanfaat bagi organisasi [3]. Riset ini memanfaatkan *Neural Network* untuk mendapatkan pola prediksi yang akurat [4][5]. Hasil penelitian yang sistematis diharapkan mampu memudahkan program studi untuk memprediksi waktu lulusan mahasiswa (tepat waktu atau tidak tepat waktu). Hal ini sangat diperlukan guna menentukan strategi yang tepat agar mahasiswa dapat terhindar dari waktu lulus yang melebihi 4 tahun. Dengan prediksi yang akurat dan treatment yang tepat diharapkan riset dapat membantu program studi x,y,z untuk meningkatkan jumlah lulusan agar memenuhi standar minimal yang ditentukan oleh BAN-PT.

2. NEURAL NETWORK

2.1 Konsep *Neural Network*

Jaringan syaraf tiruan (*Artificial Neural Network*) adalah sistem pemroses informasi dengan karakteristik dan performa yang mendekati syaraf biologis [16]. Secara umum *Artificial Neural Network* adalah jaringan dari kelompok unit pemrosesan kecil yang dimodelkan berdasarkan jaringan syaraf manusia. Kekuatan komputasi yang luar biasa dari otak manusia ini merupakan sebuah keunggulan di dalam kajian ilmu pengetahuan. *Neuron* adalah bagian dasar dari pemrosesan suatu *Neural Network*.

2.2 *Back Propagation Neural Network* (BPNN)

Dalam pelatihan BPNN terdapat 3 fase, yaitu fase maju (*forward propagation*), fase mundur (*backpropagation*), dan fase modifikasi bobot [18]. Pada fase maju, pola masukan dihitung maju dimulai dari lapisan input hingga lapisan output. Dalam fase mundur, tiap unit output menerima target pola yang berhubungan dengan pola input untuk dihitung nilai kesalahan, dan kesalahan tersebut akan dipropagasikan mundur. Sedangkan fase modifikasi bobot bertujuan untuk menurunkan kesalahan yang terjadi. Ketiga fase tersebut diulang secara terus menerus hingga kondisi penghentian dipenuhi. Berikut ini merupakan algoritma dalam pelatihan BPNN :

Fase I : Propagasi Maju (*forward propagation*)

1. Setiap unit input ($x_i, i = 1, 2, 3, \dots, n$) menerima sinyal x_i dan meneruskan sinyal ke seluruh lapis/unit tersembunyi (*hidden layer*).
2. Setiap unit tersembunyi ($z_j, j = 1, 2, 3, \dots, p$), menjumlahkan bobot sinyal input dengan persamaan berikut. $z_{inj} = v_0 + \sum_{i=1}^n x_i v_{ij}$. Dan menerapkan fungsi aktivasi untuk menghitung sinyal outputnya: $z_j = f(z_{inj})$. Biasanya fungsi aktivasi yang digunakan adalah sigmoid, kemudian mengirimkan sinyal tersebut ke semua output
3. Setiap unit output ($y_k, k = 1, 2, 3, \dots, m$) menjumlahkan bobot sinyal input $y_{ink} = w_0 + \sum_{j=1}^p z_j w_{jk}$ Dan menerapkan fungsi aktivasi untuk menghitung sinyal outputnya:

$$y_k = f(y_{ink})$$

Fase II : Propagasi mundur (*Back propagation*)

1. Setiap unit output ($y_k, k = 1, 2, 3, \dots, m$) menerima pola target yang sesuai dengan pola input pelatihan, kemudian hitung error dengan persamaan berikut:

$$(\delta_k = t_k - y_k) f'(y_{ink})$$

f' adalah turunan dari fungsi aktivasi kemudian hitung koreksi bobot dengan persamaan sebagai berikut:

$$\Delta w_0 k = \alpha * \delta_k * z_j$$

Dan menghitung koreksi bias dengan persamaan berikut, $\Delta w_0 k = \alpha * \delta_k$

Sekaligus mengirimkan δ_k ke unit-unit yang ada di layer paling kanan

2. Setiap unit tersembunyi ($z_j, j = 1, 2, 3, \dots, p$) menjumlahkan delta inputnya (dari unit-unit yang berada pada lapisan di kanannya) : $\delta_{inj} = \sum \delta_{kwjk}^m, k = 1$

Untuk menghitung informasi error, kalikan nilai ini dengan turunan dari fungsi aktivasinya: $\delta_j = \delta_{inj} f'(z_{inj})$, kemudian hitung koreksi bobot dengan persamaan sebagai berikut:

$$\Delta v_{jk} = \alpha * \delta_j * x_k$$

Dan menghitung koreksi bias dengan persamaan berikut: $\Delta v_{0j} = \alpha * \delta_j$

Fase III : Perubahan bobot dan bias

1. Setelah unit output ($y_k, k = 1, 2, 3, \dots, m$) dilakukan perubahan bobot dan bias ($z_j, j = 1, 2, 3, \dots, p$) dengan persamaan berikut :

$$W_{jk}(\text{baru}) = W_{jk}(\text{lama}) + \Delta W_{jk}$$

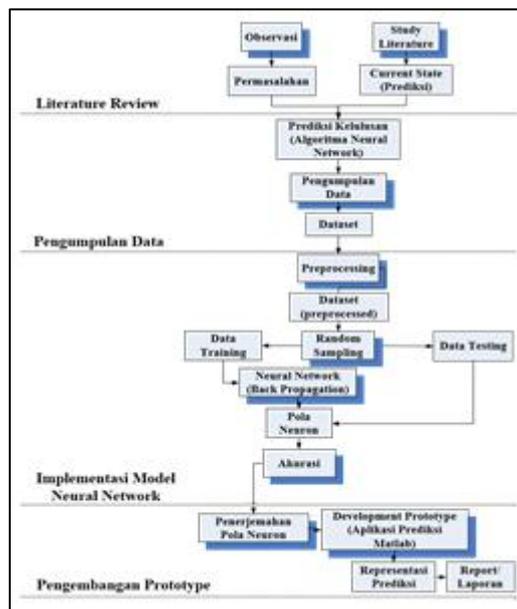
Setiap unit tersembunyi ($z_j, j = 1, 2, 3, \dots, p$) dilakukan perubahan bobot dan bias ($z_i, i = 1, 2, 3, \dots, m$) dengan persamaan berikut:

$$V_{ij}(\text{baru}) = V_{ij}(\text{lama}) + (2.13) V_{ij}$$

2. Pengujian kondisi berhenti

3. METODE PENELITIAN

Proses penelitian secara umum meliputi inisiasi, pengembangan model, dan pengembangan prototype.



Gambar 2. Tahapan Penelitian

3.1 Data Training

Setelah melalui tahap *preprocessing dan random sampling*, data training siap diproses dalam *RapidMiner* dengan jumlah 396 record. Atribut yang digunakan untuk implementasi model Neural Network (Back Propagation) adalah indeks prestasi semester (ips 1, ips 2, ips 3, dan ips 4) dan indeks prestasi kumulatif (ipk) dari sampel data mahasiswa tahun 2007 sampai 2010.

3.2 Hasil Validasi pada RapidMiner

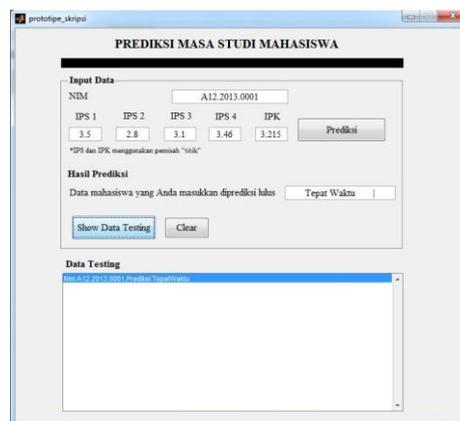
Dari eksperimen yang dilakukan dengan memanfaatkan data training di atas didapatkan pola neuron yang kemudian menjadi usulan model untuk dikembangkan sebagai pola prediksi waktu kelulusan. Setelah pola terbentuk selanjutnya dilakukan pengukuran instrumen-instrumen evaluasi yang meliputi akurasi, presisi, faktor recall, dan AUC. Dari hasil pengukuran didapati bahwa pemanfaatan BPNN memberikan hasil yang positif terkait pengukuran di semua elemen yang telah ditentukan. Rincian hasil evaluasi sebagaimana dipaparkan dalam tabel 1, dimana pada tiga elemen evaluasi pertama didapatkan nilai yang sangat signifikan dengan rata-rata melebihi 98%. Hal ini mengindikasikan pemanfaatan algoritma Neural Network (Back Propagation) sangat sesuai untuk prediksi waktu kelulusan mahasiswa.

Tabel 1 Hasil Validasi

Validasi	Hasil	Keterangan
Akurasi	99.49 %	Good
Precision	98.40 %	Good
Recall	100 %	Excellent
AUC	1	Excellent Classification

3.3 Pengembangan Prototipe

Untuk memudahkan pihak pengelola program studi dalam memprediksi data selanjutnya dengan memanfaatkan pola neuron yang telah terbentuk maka dibangunlah aplikasi sederhana yang memungkinkan pengguna untuk melakukan prediksi tanpa harus membentuk pola baru lagi.



Gambar 3. Prototipe Aplikasi Prediksi Kelulusan

Pada aplikasi prediksi masa studi mahasiswa ini output yang akan ditampilkan adalah label prediksi untuk ketepatan masa kuliah (tepat waktu atau tidak tepat waktu). Output ditampilkan berdasarkan data yang diinputkan oleh user (data testing) yang meliputi Nim, IPS 1, IPS 2, IPS 3, IPS 4, dan IPK. Tombol prediksi digunakan untuk mengaktifasi proses kalkulasi dengan memanfaatkan pola/pattern yang telah dihasilkan dari data training sebelumnya hingga menghasilkan label output. Hasil dari percobaan tersebut adalah pernyataan

4 KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka dapat disimpulkan bahwa prototipe aplikasi data mining dengan algoritma Neural Network (Back Propagation) terbukti dapat memprediksi waktu kelulusan mahasiswa dengan menghasilkan nilai akurasi terbaik yaitu sebesar 99.49% dan nilai AUC (*Area Under Curve*) sebesar 1 (satu) dengan kategori *Excellent classification*. Pengembangan prototipe aplikasi secara umum cukup membantu pihak pengelola program studi dalam memprediksi waktu kelulusannya melalui user interface grafis yang sederhana.

Untuk lebih memutakhirkan pola neuron yang terbentuk seiring dengan penggunaan yang berkelanjutan, perlu dikembangkan pada aplikasi fitur yang memungkinkan system untuk *recalculate* berdasarkan sejumlah data testing yang digunakan. Hal ini perlu dilakukan mengingat *treatment* yang dilakukan program studi besar kemungkinan akan mempengaruhi pola waktu kelulusan yang pada akhirnya juga mempengaruhi pola neuron yang terbentuk. Pengembangan prototipe selanjutnya dapat ditambahkan fitur yang memungkinkan system melakukan import data dalam jumlah yang besar (misalnya, data untuk satu angkatan mahasiswa) hal ini akan memudahkan user dalam mengelola data dari pada harus melakukan entri manual per record.

DAFTAR PUSTAKA

- [1] Kementerian Pendidikan Nasional dan Kebudayaan RI. (2014, maret) Badan Akreditasi Nasional Perguruan Tinggi. [Online]. <http://www.ban-pt.kemdiknas.go.id>
- [2] Kementrian Pendidikan Nasional dan Kebudayaan RI. (2015, Mar.) Kementrian Pendidikan dan Kebudayaan Koordinasi Perguruan Tinggi Swasta Wilayah IV. [Online]. <http://www.kopertis4.or.id/>
- [3] Ian H. Witten, Eibe Frank, and Mark A. Hall, *Data Mining Practical Learning Machine Tools and Techniques*, 3rd ed.: Elsevier, 2011.
- [4] Bogdan Oancea, Raluca Dragoescu, and Stefan Ciucu, "Predicting Students Results in Higher Education Using Neural Networks," *International Conference on Applied Information and Communication Technologies (AICT2013)*, pp. 25-26, April 2013.
- [5] Mujib Ridwan, Hadi Suryono, and M. Sarosa, "Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *EECCIS*, vol. 7, Juni 2013.

-
- [6] Novi Yanti, "Penerapan Metode Neural Network dengan Struktur Back Propagation untuk Prediksi Stok Obat di Apotek (Studi Kasus : Apotek ABC)," in *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, Yogyakarta, Juni 2011.
- [7] Budi Rahmani and Hugo Aprilianto, "Early Model of Student's Graduation Prediction Based on Neural Network," *Telkonnika*, vol. 12, pp. 465-474, June 2014.
- [8] Samy Abu Naser, Ihab Zaqout, Mahmoud Abu Ghosh, Rasha Atallah, and Eman Alajrami, "Predicting Student Performance Using Artificial Neural Network: in the Faculty of Engineering and Information Technology," *International Journal of Hybrid Information Technology*, vol. 8, pp. 221-228, 2015.
- [9] Muhamad Hanief Meinanda, Metri Annisa, Narendri Muhandri, and Kadarsyah Suryadi, "Prediksi Masa Studi Sarjana dengan Artificial Neural Network," *Internet Working Indonesia Journal*, vol. 1, 2009.
- [10] Heri Susanto and Sudiyatno, "Data Mining untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan dan Prestasi Masa Lalu," *Jurnal Pendidikan Vokasi*, vol. 4, Juni 2014.
- [11] Mohammed J. Zaki and Wagner Meira JR, *Data Mining and Analysis : Fundamental Concepts and Algorithms*. New York, United States of America: Cambridge University Press, 2014.
- [12] Kusriani and Emha Taufiq Luthfi, *Algoritma Data Mining*. Yogyakarta, 2009.
- [13] Ogunde A. O and Ajibade D A, "A Data Mining System for Predicting University Students' Graduation Grades Using ID3 Decision Tree Algorithm," *Journal of Computer Science and Information Technology*, vol. 2, pp. 21-46, Maret 2014.
- [14] Hamidah Jantan, Abdul Razak Hamdan, and Zulaiha Ali Othman, "Human Talent Prediction in HRM using C4.5 Classification Algorithm," *IJCSE*, vol. 2, pp. 2526-2534, 2010.
- [15] Muchamad Piko Henry Widiarto, "Pengambilan Pola Kelulusan Tepat Waktu pada Mahasiswa STMIK Amikom Yogyakarta Menggunakan Data Mining Algoritma C4.5," 2011.
- [16] Prabowo Pudjo Widodo and Rahmadya Trias Handayanto, *Penerapan Soft Computing dengan Matlab*. Bandung, Indonesia: Rekayasa Sains, 2012.
- [17] Mark Hudson Beale, Martin T Hagan, and Howard B Demuth, *Neural Network Toolbox*. US: Mathworks, 2015.
- [18] Nazla Nurmila, Aris Sugiharto, and Eko Adi Sarwoko, "Algoritma Back Propagation Neural Network untuk Pengenalan Pola Karakter Huruf Jawa," *Jurnal Masyarakat Informatika*, vol. 1.
- [19] Florin Gorunescu, *Data Mining Concepts, Models, and Techniques.*: Springer, 2011.
- [20] Sebastian Land and Simon Fischer, *RapidMiner 5 : RapidMiner in academic use.*: Rapid-I GmbH, 2012.
- [21] The MathWorks, *Learning MATLAB 7*. Amerika Serikat, 2005.
- [22] Nurhayati, "Studi Perbandingan Metode Sampling Antara Simple Random dengan Stratified Random," *Jurnal Basis Data, CT Research Center UNAS*, vol. 3, no. 1, Mei 2008.

-
- [23] Johan Hendrikson Purba, "Sistem Prediksi Jumlah Penerimaan Anggota Kepolisian dengan Menggunakan Metode Analysis Of Variance ," *Informasi dan Teknologi Ilmiah (INTI)*, vol. 5, Januari 2013.
- [24] Adi Nugroho, *Analisis dan Perancangan Sistem Informasi dengan Metodologi Berorientasi Objek*, Edisi Revisi ed.: Informatika, 2005.
- [25] Ph.D. Roger S. Pressman, *Software Engineering A Practitioner's Approach*, 7th ed., 2010.