

DATA MINING DENGAN ALGORITMA NEURAL NETWORK DAN VISUALISASI DATA UNTUK PREDIKSI KELULUSAN MAHASISWA

Neni Purwati¹, Rini Nurlistiani², Oscar Devinsen³

¹²³Fakultas Ilmu Komputer, Institut Informatika dan Bisnis Darmajaya
Jl. Z.A. Pagar Alam No. 93, Bandar Lampung - Indonesia 35142
Telp. (0721) 787214 Fax. (0721) 700261
e-mail : nenipurwati87@darmajaya.ac.id, rininurlistiani@darmajaya.ac.id,
oscar.1811050052@mail.darmajaya.ac.id

ABSTRACT

Graduating on time is the desire of all students, and is one of the indicators in accreditation, namely the more number of students who graduate on time, the higher / better the score will be, so that graduation on time is an advantage for educational institutions. However, if the status of graduating students is not predicted early, it will result in many students graduating on time, and this will greatly harm students and educational institutions. This research was conducted using data from students who graduated for 4 years from 2016-2019. The classification method is an approach to grouping data in data mining, namely classifying data. The classification data mining method that will be used is the neural network algorithm. Neural network algorithms are used to predict student graduation which is difficult to do manually, while visualization is used to visually display recapitulation data so that it is more interesting and easy to understand. The attributes used in the training data consisted of Gender, Origin, Class, Department, Age, GPA, Judicium Date, Year of Judiciary and Result Class. There are 9 attributes that are parameterized, where 8 attributes are predictor and 1 attribute is result. Training and testing data by changing parameters, namely: Hidden Layer Size: 3, Training Time: 500, Learning Rate: 0.3, Momentum: 0.2 resulting in classification with 87.80% Precision, 86.90% Recall and Accuracy (accuracy level) of 92.83%. In addition, data visualization displays several reporting recapitulations in the form of a very complete dashboard, so that the prediction and visualization of the data can help in graduating students to be able to graduate on time, and provide recommendations for actions that must be taken appropriately by management or the authorities in making decisions.

Keywords— *Business Intelligence, Data Mining, Neural Network, Visualization*

ABSTRAK

Lulus tepat waktu merupakan keinginan seluruh mahasiswa, dan merupakan salah satu indikator dalam akreditasi yaitu semakin banyak jumlah mahasiswa lulus tepat waktu maka nilai yang dihasilkan semakin tinggi/baik, sehingga lulus tepat waktu merupakan keuntungan bagi instansi pendidikan. Namun, status kelulusan mahasiswa jika tidak diprediksi secara dini, akan mengakibatkan banyak mahasiswa lulus tidak tepat waktu, dan hal tersebut akan sangat merugikan mahasiswa dan instansi pendidikan. Penelitian ini dilakukan dengan menggunakan data mahasiswa yang lulus selama 4 tahun dari tahun

2016-2019. Metode klasifikasi adalah pendekatan untuk melakukan pengelompokan data dalam data mining yaitu menggolongkan data. Metode klasifikasi ini juga dapat digunakan untuk melakukan prediksi atas informasi yang belum diketahui sebelumnya. Metode data mining klasifikasi yang akan digunakan adalah algoritma *neural network*. Algoritma *neural network* digunakan untuk memprediksi kelulusan mahasiswa yang sulit dilakukan secara manual, sedangkan visualisasi digunakan untuk menampilkan data rekapitulasi secara visual sehingga lebih menarik dan mudah dipahami. Adapun atribut yang digunakan dalam data *training* terdiri dari Jenis Kelamin, Asal, Kelas, Jurusan, Umur, Kelompok umur, Kelompok IPK, dan Class Hasil. Atribut yang menjadi parameter sebanyak 8 atribut, dimana 7 atribut *predictor* dan 1 atribut hasil. *Training* dan *testing* data dengan mengubah parameter yaitu: *Hidden Layer Size*: 3, *Training Time*: 500, *Learning Rate*: 0.3, *Momentum*: 0.2 menghasilkan klasifikasi dengan *Precision* 87.80%, *Recall* 86.90% dan *Accuracy*(tingkat akurasi) sebesar 92,83%. Selain itu visualisasi data menampilkan beberapa rekapitulasi pelaporan berupa dashboard yang sangat komplit, sehingga dengan prediksi dan visualisasi data tersebut dapat membantu dalam kelulusan mahasiswa untuk dapat lulus tepat waktu, dan memberikan rekomendasi tindakan yang harus dilakukan dengan tepat oleh manajemen atau pihak yang berwenang dalam mengambil keputusan.

Kata Kunci—Business Intelligence, Data Mining, Neural Network, Visualisasi

I. PENDAHULUAN

Lulus tepat waktu adalah keinginan seluruh mahasiswa, tidak hanya itu lulus tepat waktu adalah keuntungan bagi kedua pihak, yaitu mahasiswa dan instansi pendidikan. Dalam mengukur status kelulusan dapat diperoleh dari indek prestasi mahasiswa setiap semester. Jika indek prestasi mahasiswa lebih rendah maka akan mengakibatkan mahasiswa lulus tidak tepat waktu [1]. Namun status kelulusan mahasiswa ketika diprediksi tidak selalu menghasilkan prediksi secara dini, sehingga mampu mengakibatkan lulus tidak tepat waktu. Adapun pendekatan yang sering digunakan ialah menggunakan teknik *data mining*. Metode yang sering digunakan adalah klasifikasi [2].

Manajemen Perguruan Tinggi perlu mengerahkan seluruh sumber daya yang dimiliki untuk membantu mahasiswa menyelesaikan pendidikan mereka dengan prestasi akademik yang baik dan meminimalisir tingkat ketidakkelulusan mahasiswa, Tentunya dalam perguruan tinggi kelulusan mahasiswa sangat penting terutama dalam penilaian akreditasi perguruan tinggi, sedangkan dalam perguruan tinggi memiliki data mahasiswa yang bisa digunakan untuk keperluan pengambilan keputusan untuk pengembangan sistem belajar dalam perguruan tinggi [3].

Dalam sebuah perusahaan atau organisasi, keputusan dibuat pada tingkatan atau level yang berbeda-beda, keputusan juga harus dilakukan secepat mungkin untuk mempertahankan daya saing, sehingga

dapat membuat keputusan yang benar dan menjadi dasar yang solid dalam hal data, informasi dan ketersediaan pengetahuan. Dasar ini dapat bersumber dari informasi atau data berita bisnis, konferensi, pelanggan, tenaga penjualan, dan sebagainya. Sehingga bersama-sama membentuk sebuah konsep mengenai pasar dan dunia bisnis tertentu [4].

Data mining lebih khusus dan lebih spesifik dibandingkan OLAP mengingat *database* bukan satu-satunya bidang ilmu yang mempengaruhi *data mining*, seperti: *information science* (ilmu informasi), *high performance computing*, visualisasi, *machine learning*, statistik, *neural network* (jaringan syaraf tiruan), pemodelan matematika, *information retrieval* dan *information extraction* serta pengenalan pola. Bahkan pengolahan citra (*image processing*) juga digunakan dalam rangka melakukan *data mining* terhadap data *image/spatial*.

Operasi *data mining* menurut sifatnya dibedakan menjadi dua, yaitu:

1. Prediksi (*prediction driven*). Operasi prediksi digunakan untuk validasi hipotesis, *querying* dan pelaporan (misalnya: spreadsheet dan pivot tabel), analisis multidimensi (*dimensional summary*), OLAP

(*Online Analytic Processing*) serta analisis statistik.

2. Penemuan (*discovery driven*) bersifat transparan. Operasi penemuan digunakan untuk analisis data eksplorasi, pemodelan prediktif, segmentasi *database*, analisis keterkaitan (*link analysis*) dan deteksi deviasi [5].

Neural network (jaringan saraf tiruan) adalah model non-linear yang rumit dibangun dari komponen yang secara individu berperilaku mirip seperti model regresi. *Neural network* dapat direpresentasikan sebuah grafik, dan beberapa sub-grafik tampaknya ada integritas yang sama dengan gerbang logika. Struktur dari jaringan *neuron* atau saraf secara terperinci dirancang terlebih dahulu [6]. *Neural network* adalah model mesin pembelajaran yang meniru aspek pembelajaran dari pengalaman masa lalu untuk memprediksi masa depan [7].

Visualisasi adalah sebuah metode penggunaan komputer dalam menjelaskan metode terbaik dalam menampilkan data untuk mengingat informasi dengan nalar alami manusia serta dapat memberikan hasil untuk melihat data yang sulit dilihat dengan pemikiran sehingga peneliti dapat mengamati sebuah simulasi dan komputasi

proses penemuan akademis dan ilmiah serta mengembangkan pemahaman yang lebih baik[8]. Visualisasi adalah pengelolaan dalam pembuatan sebuah gambar, diagram atau bentuk animasi untuk menampilkan suatu informasi. Visualisasi yang disajikan dalam bentuk gambar baik bersifat abstrak maupun nyata sudah ada sejak awal dari peradaban manusia[9].

II. METODE PENELITIAN

Dalam penelitian ini tools yang digunakan pada algoritma *Neural Network* adalah Weka 3.8.3 dan untuk visualisasinya menggunakan Tableau Profesional Edition @2018. Sedangkan tahapan metode penelitian yang dilakukan adalah sebagai berikut [10]:



Gambar 3.1 Metode Penelitian

Dalam pelaksanaannya, proses yang ada data mining dapat dibagi menjadi tiga(3) yaitu: Preprocessing, Data Mining, dan Post Processing[11]. *Preprocessing* merupakan proses awal yang akan mentransformasikan data masukan menjadi data dengan format yang sesuai dan siap untuk diproses. Beberapa contoh hal yang dilakukan dalam *preprocessing* meliputi berbagai proses yang diperlukan

antara lain: penggabungan, perubahan bentuk, ataupun pentransformasian data sebagai cara untuk membersihkan, mengintegrasikan, mereduksi dan mendiskritisasi. Lebih lanjut proses yang ada dalam *preprocessing* dapat terdiri dari salah satu kegiatan proses ataupun gabungan dari beberapa proses diatas[12]. Salah satu masalah *preprocessing* yang krusial adalah pada proses pembersihan data[13]. Pada penelitian ini *preprocessing* berada ditahap *Collecting Data* dan *Seleksi Data*. Berikut ini tahapan metode penelitian yang dilakukan:

1) Collecting data

Ditahap ini dicari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data kedalam dataset.

2) Seleksi data

Pada tahap ini setelah data diperoleh, data akan diseleksi untuk membuat sebuah target data, fokus dalam bagian dari variabel atau sampel data yang akan dilakukan penelitian.

3) Metode yang diusulkan

Pada tahap ini data dianalisis dan dikumpulkan untuk menjadi data yang saling berhubungan satu sama lainnya. Setelah analisis data diterapkan, selanjutnya menerapkan model-model

yang sesuai dengan jenis datanya. Pembagian data dalam data *testing* (data uji) dan data *training* (data latihan) yang digunakan untuk pembuatan model.

4) Penelitian dan Pengujian Metode

Pada tahap ini model yang diusulkan akan diuji untuk melihat hasil berupa aturan yang akan dimanfaatkan dalam pengambilan keputusan.

5) Evaluasi

Pada tahap ini dilakukan evaluasi data terhadap model yang akan ditetapkan untuk mengetahui tingkat akurasi model. Model yang diusulkan tentang prediksi kelulusan mahasiswa dengan parameter *neural network* dan akan memberikan data nilai akurasi dari model tersebut.

III. HASIL DAN PEMBAHASAN

3.1 Neural Network

Nilai learning rate ditentukan dengan melakukan uji coba memasukan Learning Rate. *Performance evaluation* berguna untuk menguji performa dari klasifikasi(*classifier*) terdiri dari *Recall*, *precision*, dan *accuracy*. *Recall* adalah kumpulan data positif yang diklasifikasikan dengan benar sebagai data positif. *Precision* adalah kumpulan data yang diklasifikasikan sebagai positif yang benar – benar positif. *Accuracy* adalah

ketepatan klasifikasi data[14][15] ditentukan dari *training data*.

Adapun rumusnya adalah:

$$Precision = \frac{TP}{TP + FP} \dots\dots\dots(1)$$

$$Recall = \frac{TP}{TP + FN} \dots\dots\dots(2)$$

$$Akurasi = \frac{(TN + TP)}{(TN + FN + TP + FP)} \dots\dots\dots(3)$$

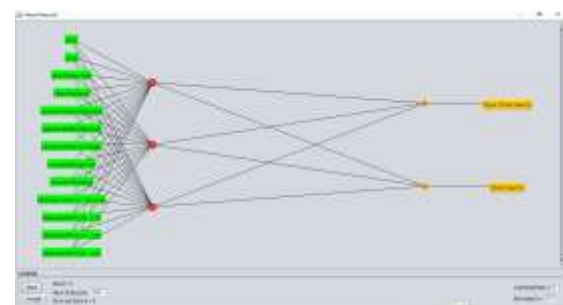
Berikut adalah hasil yang diperoleh dari percobaan yang dilakukan untuk penentu parameter *Learning Rate*(LR) dengan mengubah nilai *Learning Rate* dalam tabel 1 berikut:

LR	Accuracy	Precision	Recall
0.1	92.55%	86.90%	86.90%
0.2	92.26%	87.50%	84.80%
0.3	92.26%	87.50%	84.80%
0.4	92.26%	84.60%	88.90%

Penerapan model *neural network* berdasarkan nilai akurasi tertinggi dari 4 kali *training* dan *testing* data dengan mengubah parameter yaitu:

1. *Hidden Layer* : 3
2. *Training Time* : 500
3. *Learning Rate* : 0.1
4. *Momentum* : 0.2
5. *Num Of Epochs* : 500

terbentuk arsitektur *neural network* seperti pada gambar berikut:



Gambar 1 Arsitektur *Neural Network*.

Pada gambar 1 menjelaskan arsitektur yang terbentuk memiliki 13 lapisan input(JK, Umur, Asal, Kelas, 5 Jurusan, Kelompok Umur, 3 Kelompok IPK), 3 lapisan tersembunyi(*hidden*) dan 2 lapisan output(Tidak Tepat Waktu dan Tepat Waktu).

Pengujian model setelah melakukan penerapan *neural network* dalam menentukan tingkat keakurasian maka *confusion matrix* yang didapat dari model *neural network* itu sendiri berdasarkan 349 record data yang diolah seperti pada tabel 2 berikut:

Classified	a	b
a	237	13
b	13	86

Keterangan:

a: Tidak Tepat Waktu

b: Tepat Waktu

Berdasarkan *confusion matrix* di atas dapat diklasifikasikan sebagai berikut :

True Positive (TP) = 86 record

False Negative (FN) = 13 record

True Negative (TN) = 237 record

False Positive (FP) = 13 record

dengan perhitungan *Precision*, *Recall* dan tingkat akurasi(*Accuracy*) sebagai berikut:

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{86}{86 + 13}$$

$$Precision = \frac{86}{99}$$

$$Precision = 0.8686 \text{ atau } 86.90\%$$

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{86}{86 + 13}$$

$$Recall = \frac{86}{99}$$

$$Recall = 0.8690 \text{ atau } 86.90\%$$

$$Akurasi = \frac{(TN + TP)}{(TN + FN + TP + FP)}$$

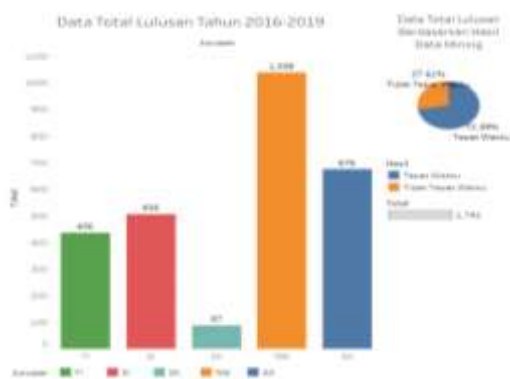
$$Akurasi = \frac{(237 + 86)}{(237 + 13 + 86 + 13)}$$

$$Akurasi = \frac{323}{349} = 0.9255 \text{ atau } 92.55\%$$

Dari hasil klasifikasi dengan menggunakan algoritma *neural network* menunjukkan bahwa *Precision* sebesar 87.80%, *Recall* sebesar 86.90%, dan tingkat akurasi(*Accuracy*) adalah sebesar 92.83%.

3.2 Visualisasi Data

Total data kelulusan dalam 4(empat) tahun terakhir dimulai tahun 2016 sampai 2019 berjumlah 2742 record, dapat divisualisasikan berdasarkan Jurusan dengan jenjang Strata Satu(S1) dan berdasarkan Class “Hasil” data mining dapat ditampilkan pada dashboard berikut:



Gambar 2 Dashboard Total Lulusan Per Jurusan dan Persentase Lulusan “Hasil”.

Pada gambar 2 diatas menjelaskan bahwa:

1. Data Total Lulusan Tahun 2016-2019 dengan rincian: Jurusan TI(Teknik Informatika) 436 orang, Jurusan SI(Sistem Informasi) 505 orang, Jurusan SK(Sistem Komputer) 87 orang, Jurusan MA(Manajemen) 1038 orang, dan Jurusan AK(Akuntansi) 676 orang.
2. Data Total Lulusan Berdasarkan Hasil Data Mining yaitu: Tepat Waktu 72.39% dan Tidak Tepat Waktu 27.62%.

IV. SIMPULAN

Hasil penelitian yang dilakukan dari tahap awal sampai dengan tahap pengujian menggunakan Data Mining dengan algoritma *Neural Network* untuk kelulusan mahasiswa menghasilkan prediksi dengan *Precision* 87.80%, *Recall* 86.90% dan nilai akurasi sebesar 92.83%. Sedangkan dari visualisasi data dari *dataset* yang berjumlah 2742 record menampilkan

beberapa rekapitulasi pelaporan berupa dashboard yang sangat komplit, sehingga dengan prediksi dan visualisasi data tersebut dapat membantu dalam kelulusan mahasiswa dan memberikan rekomendasi tindakan yang tepat dan harus dilakukan oleh manajemen atau pihak yang berwenang untuk mengambil keputusan.

DAFTAR PUSTAKA

- [1] S. Salmu and A. Solichin, “Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes : Studi Kasus UIN Syarif Hidayatullah Jakarta,” 2017.
- [2] E. Rohmawan, “Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Desicion Tree Dan Artificial Neural Network,” *J. Ilm. Matrik*, 2018.
- [3] M. A. Azis and A. Fazriansyah, “PREDIKSI TINGKAT KELULUSAN NILAI MAHASISWA TERHADAP MATAKULIAH WEB PROGRAMMING MENGGUNAKAN METODE NEURAL NETWORK,” *J. Pilar Nusa Mandiri*, 2019, doi: 10.33480/pilar.v15i2.660.
- [4] A. Yumalia and R. E. Indrajit, “Penerapan Konsep Business

- Intelligence untuk Percepatan Penyelesaian Perkara pada Panmud Perdata Khusus Mahkamah Agung RI,” *Ikraith-Informatika*, 2017.
- [5] F. A. Hermawati, *Data Mining*. Yogyakarta: CV Andi Offset, 2013.
- [6] Rudy Ansari, “PREDIKSI KELULUSAN MAHASISWA DENGAN JARINGAN SYARAF TIRUAN,” *J. Teknol. Inf. Univ. Lambung Mangkurat*, 2016, doi: 10.20527/jtiulm.v1i1.4.
- [7] M. Inthachot, V. Boonjing, and S. Intakosum, “Artificial Neural Network and Genetic Algorithm Hybrid Intelligence for Predicting Thai Stock Price Index Trend,” *Computational Intelligence and Neuroscience*. 2016, doi: 10.1155/2016/3045254.
- [8] T. Limbong, “Pemanfaatan Visualisasi dan Animasi untuk Kegiatan Akademik sebagai Sarana Pengumuman pada STMIK Budi Darma Medan,” *Inf. dan Teknol. Ilm.*, 2013, doi: 10.17605/OSF.IO/KQUW3.
- [9] S. A. Hendrawan, R. R. Isnanto, and I. P. Windasari, “Aplikasi Visualisasi 3D Pada Struktur Sistem Rangka Manusia Berbasis Android,” *J. Teknol. dan Sist. Komput.*, 2015, doi: 10.14710/jtsiskom.3.4.2015.426-435.
- [10] I. Ali and L. Sularto, “Optimasi Parameter Artificial Neural Network Menggunakan Algoritma Genetika Untuk Prediksi Kelulusan Mahasiswa,” *J. ICT Inf. Commun. Technol.*, 2019, doi: 10.36054/jict-ikmi.v18i1.52.
- [11] D. B. Setyohadi, F. A. Kristiawan, and E. Ernawati, “Perbaikan Performansi Klasifikasi Dengan Preprocessing Iterative Partitioning Filter Algorithm,” *Telematika*, 2017, doi: 10.31315/telematika.v14i01.1960.
- [12] R. K. D. A. Malathi, “Preprocessing of Various Data Sets Using Different Classification Algorithms for Evolutionary Programming ,” *Int. J. Sci. Res.*, 2015.
- [13] D. B. Setyohadi, A. A. Bakar, and Z. A. Othman, “Optimization overlap clustering based on the hybrid rough discernibility concept and rough K-Means,” *Intell. Data Anal.*, 2015, doi: 10.3233/IDA-150746.
- [14] *Principles of Data Mining*. 2007.
- [15] G. A. Sandag, J. Leopold, and V. F. Ong, “Klasifikasi Malicious Websites Menggunakan Algoritma K-NN Berdasarkan Application

- Layers dan Network Characteristics,” *CogITO Smart J.*, 2018, doi: 10.31154/cogito.v4i1.100.37-45. *Series* (Vol. 1529, No. 2, p. 022063). IOP Publishing. [20]
- [16] Fitria, Y. A. (2019). Visualization of data on earthquake prone areas from the analysis of earthquake data vibrations. *Test Engineering and Management*, 81(11-12).
- [17] Kurniawan, H. (2018, April). Strategy Development Of Human Source Competitiveness Strengthening With Learning Media System Analisis Model. In *Prosiding International conference on Information Technology and Business (ICITB)* (pp. 9-12).
- [18] Purwati, N. (2019, November). Sistem Informasi E-Document Pada Biro Administrasi Akademik dan Kemahasiswaan (BAAK) Institut Informatika dan Bisnis Darmajaya. In *Prosiding Seminar Nasional Darmajaya* (Vol. 1, pp. 293-301).
- [19] Aziz, R. A., Maria, D., Laila, S. N., & Azima, M. F. (2020, April). Development of Knowledge Management System for Determining Organizational Performances, Total Quality Management, And Culture. In *Journal of Physics: Conference*