

Prediksi Tingkat Kelulusan Mahasiswa Menggunakan *Machine Learning* dengan Teknik *Deep Learning*

Martanto ^{1*}, Irfan Ali², Mulyawan³

¹ Program Studi Manajemen Informatika, STMIK IKMI Cirebon

^{2,3} Program Studi Teknik Informatika, STMIK IKMI Cirebon

^{12,3} Jln. Perjuangan no 10 B Majasem kota Cirebon

email: ¹martantomusijo@gmail.com, ²irfanaali0.0@gmail.com, ³kui.stmikikmi@gmail.com

Special Issue on Seminar Nasional - Inovasi Dalam Teknologi Informasi & Teknologi Pembelajaran 2019

Abstract – The graduation rate of students on time at the Informatics Engineering study program STMIK IKMI Cirebon greatly affects the accreditation assessment. Graduation prediction is difficult to do, but many have done predictions using a variety of methods. Graduation prediction is needed in order to determine preventive policies for students who graduate not on time. The method used in this research is Machine learning with deep learning techniques. The data set used as many as 405 data of students who graduated on time or who were not on time. The research attributes used are the Nim attribute, the GPA value of students who have graduated and the status of graduating or not graduating. The results of this study are the level of accuracy using Machine Learning by 72.84%.

Keywords - prediction, graduation, machine learning, deep learning

Abstrak – Tingkat kelulusan mahasiswa tepat waktu pada program studi Teknik Informatika STMIK IKMI Cirebon sangat mempengaruhi penilaian akreditasi. Prediksi tingkat kelulusan sulit untuk dilakukan, tetapi sudah banyak yang melakukan prediksi menggunakan berbagai metode. Perlu adanya prediksi tingkat kelulusan agar dapat menentukan kebijakan pencegahan terhadap mahasiswa yang lulus tidak tepat waktu. Metode yang digunakan dalam penelitian ini adalah Machine learning dengan teknik *deep Learning*. Data set yang digunakan sebanyak 405 data mahasiswa yang lulus tepat waktu maupun yang tidak tepat waktu. Atribut penelitian yang digunakan adalah atribut Nim, nilai IPK mahasiswa yang telah lulus dan Status lulus atau tidak lulus. Hasil penelitian ini adalah tingkat akurasi menggunakan *Machine Learning* sebesar 72.84%.

Kata Kunci – prediksi, kelulusan, machine learning, deep learning

I. PENDAHULUAN

Saat ini terdapat 3.276 [1] perguruan tinggi di Indonesia, yang terdiri dari berbagai jenis yaitu Universitas, institute, sekolah tinggi, akademi, akademi komunitas dan politeknik. Jumlah mahasiswa yang terdaftar berjumlah 6.924.511 [1], semua mahasiswa tersebut belum tentu lulus tepat waktu.

*) **penulis korespondensi:** Martanto
Email: martantomusijo@gmail.com

Salah satu indikator keberhasilan sebuah perguruan tinggi adalah jumlah lulusan tidak berbeda jauh dengan jumlah saat mahasiswa baru masuk. Sehingga setiap perguruan tinggi berusaha agar semua mahasiswanya dapat lulus tepat waktu. Masing-masing perguruan tinggi memiliki jumlah lulusan yang berbeda dari tahun ke tahun tergantung dari keadaan mahasiswanya. Lulusan agar dapat lulus dengan tepat waktu, maka perlu dilakukan pengukuran perkiraan jumlah lulusan setiap angkutannya agar dapat lebih di kontrol lulusannya agar dapat tepat waktu.

STMIK IKMI Cirebon merupakan salah satu perguruan tinggi yang ada di Wilayah Jawa barat yang memiliki 1098 [2] mahasiswa. Jumlah tersebut sebagian besar berasal dari program studi Teknik informatika sebesar 79,5 %. Jika jumlah mahasiswa pada program studi teknik informatika yang tidak lulus tepat waktu banyak, maka akan mempengaruhi penilaian akreditasi[3] terhadap STMIK IKMI Cirebon.

Dari pemaparan yang telah diuraikan, dilakukan prediksi tingkat kelulusan menggunakan metode deep learning. *Metode deep learning* adalah teknik dalam NN yang menggunakan teknik tertentu seperti *Restricted Boltzmann Machine* (RBM) untuk mempercepat proses pembelajaran dalam NN yang menggunakan lapis yang banyak atau lebih dari 7 lapis [4],[5].

II. PENELITIAN YANG TERKAIT

Penelitian yang dilakukan oleh Indah Puji Astuti dengan topik prediksi ketepatan waktu kelulusan dengan algoritma *data mining* C4.5. Penelitian ini membahas tentang banyaknya mahasiswa yang lulus tidak tepat waktu. Metode yang digunakan untuk menyelesaikan masalah menggunakan algoritma C4.5 dengan memanfaatkan data jenis sekolah asal, asal daerah, pekerjaan orang tua, dan kelas. Hasil dari penelitian ini menunjukkan bahwa algoritma C4.5 dapat memprediksi tingkat kelulusan dengan akurasi 82% [6]. Perbedaan dengan penelitian yang dilakukan adalah terletak pada penggunaan atribut dan metode penyelesaiannya.

Penelitian lainnya dilakukan oleh Ike Fitriyaningsih, dkk mengangkat topik Aplikasi prediksi curah hujan, debit air, dan kejadian banjir berbasis web dengan *machine learning* di Deli Serdang. Masalah yang diangkat dalam penelitian ini adalah Perlu adanya prediksi banjir pada daerah yang sering terkena banjir agar masyarakat lebih dini dapat

mempersiapkan diri. Metode yang digunakan untuk memprediksi banjir menggunakan *Support Vector Machine*. Hasil dari penelitian ini menjelaskan bahwa akurasi terjadi atau tidaknya banjir sebesar 94,4% namun tidak dapat memprediksi hari terjadinya banjir [7]. Perbedaan dengan penelitian ini adalah kasus permasalahannya namun metode yang digunakan sama yaitu menggunakan *machine learning*.

Agus Romadhona, dkk pada penelitiannya mengangkat topik prediksi kelulusan mahasiswa tepat waktu berdasarkan usia, jenis kelamin, dan indeks prestasi menggunakan algoritma *decision tree*. Dasar masalah penelitian ini adalah penentuan keakuratan masa studi siswa untuk menentukan kebijakan pencegahan masa studi yang tidak tepat waktu dapat dilakukan. Metode yang dilakukan dengan membandingkan tingkat keakuratan menggunakan algoritma C4.5, algoritma ID3 dan algoritma CHAID. Hasil dari penelitian ini adalah akurasi algoritma C4.5 lebih baik dari pada algoritma ID3 dan algoritma CHAID [8]. Perbedaan dengan penelitian ini terletak pada atribut dan metode yang digunakan.

Syarli, dkk pada penelitiannya yang berjudul metode *Naive Bayes* untuk prediksi kelulusan, mengangkat masalah tentang prediksi kelulusan mahasiswa yang baru lulus menggunakan data klasifikasi. Atribut yang terpilih adalah prodi, pilihan pertama, pilihan kedua dan Nilai rata-rata. Metode yang digunakan menggunakan *Naive Bayes*. *Naive Bayes* adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk machine learning dan data mining. Hasil dari penelitian ini adalah naive bayes dapat mengklasifikasikan menggunakan metode probabilitas statistik [9]. Perbedaan dengan penelitian ini adalah penggunaan atribut dan metode yang digunakan.

Banjarsari, dkk pada penelitiannya dengan judul penerapan k-optimal pada algoritma KNN untuk prediksi kelulusan tepat waktu mahasiswa program studi ilmu komputer FMIPA unlam berdasarkan IP sampai dengan semester 4. Masalah yang diangkat dalam penelitian ini adalah belum optimalnya pemanfaatan data pada database system informasi akademik Program studi Ilmu Komputer FMIPA Unlam. Metode yang digunakan adalah K-Nearest Neighbor. Metode ini mengklasifikasikan data terhadap jarak yang paling dekat dengan objek. Hasil yang didapat adalah tingkat keakuratan sebesar 80% [10].

III. METODE PENELITIAN

A. Data Collection

Data kelulusan diperoleh dari bagian akademik program studi teknik informatika pada tahun 2017 yang berjumlah 405. Data tersebut akan dilakukan proses *pre-processing* terlebih dahulu untuk memastikan semua data telah terisi atau tidak ada data yang missing. Pre-processing juga dimaksudkan untuk menentukan atribut-atribut yang akan dipergunakan dalam dataset, yaitu data yang bersifat numerik (3). Berikut adalah *dataset* yang menjadi fokus penelitian

- Nim, yaitu nomor induk mahasiswa yang telah lulus
- Status, yaitu status dari lulusan yang terdiri atas “lulus tepat waktu”, “lulus tidak tepat waktu”
- atribut yang diolah, yaitu IPK1, IPK2, IPK3, IPK4, IPK5, IPK6, IPK7, IPK8

Proses konversi data, dilakukan terhadap atribut nim menjadi id dan atribut status menjadi label. Pada Gambar 1 menunjukkan hasil konversi atribut data.

column index	attribute name	data information
0	NIM	integer * id *
1	IPK1	real * attribute *
2	IPK2	real * attribute *
3	IPK3	real * attribute *
4	IPK4	real * attribute *
5	IPK5	real * attribute *
6	IPK6	real * attribute *
7	IPK7	real * attribute *
8	IPK8	real * attribute *
9	STS	polynomial * label *

Gbr. 1 Atribut data mahasiswa

TABEL I
CROSS VALIDATION

No	Parameter	Value	Ket
1	Number Of Fold	10	
2	Sampling type	Stratified sampling	

TABEL II.
PARAMETER DEEP LEARNING

No	Parameter	Value	Ket
1	activation	Rectifier	
2	Epoch	10	
3	Hidden layer	50	
3	Train sample periteration	-2	
4	Epsilon	1.0e-8	
5	Rho	0.99	
6	L1	1.0e-5	
7	L2	0	
8	Max W2	10	
9	Loss function	Cross entropy	
10	Distribution function	multinomial	
11	Max runtime second	0	

Selanjutnya dilakukan proses Normalisasi data menggunakan Z-Normalization. Untuk parameter dari cross validation seperti pada Tabel I. Pada Tabel I menunjukkan bahwa parameter yang digunakan ada dua yaitu Number OFFold sebesar 10 dan *sampling type* adalah *Stratified sampling*.

B. Menentukan parameter deep learning

Parameter yang ditetapkan dalam *deep learning* seperti diperlihatkan pada Tabel II. Sedangkan untuk parameter yang ditetapkan dalam *deep learning* ada 11. Sesuai dengan Tabel 2 menunjukkan bahwa nilai parameter *activation* adalah *rectifier*, nilai *Epoch* sebesar 10, nilai *hidden layer* adalah 50, *Train sample periteration* bernilai -2, *Epsilon* adalah 1.0e-8, *Rho* sebesar 0.99, L1 sebesar 1.0e-5, L2 sebesar 0, *Max W2* sebesar 10, *Loss function* menggunakan *Cross entropy*, *Distribution function* menggunakan *multinomial* dan *Max runtime second* sebesar 0.

C. Experimen Dan Pengujian Algoritma

Algoritma yang digunakan untuk klasifikasi adalah algoritma Deep learning. Deep learning merupakan salah satu metode classification yang dapat mengklasifikasikan data lulusan ke dalam beberapa klasifikasi, sehingga data lulusan dapat diprediksi menurut data yang telah dimuat modelnya. Implementasi menggunakan Rapidminer digunakan untuk membantu menemukan nilai yang akurat. Atribut yang digunakan untuk dijadikan target penelitian adalah status kelulusan. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan strategi status kelulusan dari mahasiswa di program studi teknik informatika STMIK IKMI Cirebon.

Berikut ini adalah langkah – langkah dalam melakukan teknik klasifikasi dengan deep learning (4):

- Menentukan data kelulusan mahasiswa yang akan dijadikan dataset
- Melakukan preprocessing data dan normalisasi data
- Membuat model proses dan evaluasi untuk algoritma deep learning
- Menghitung akurasi dari model deep learning

IV. HASIL DAN PEMBAHASAN

Berdasarkan hasil dari *confusion matrix* untuk metode *Machine Learning* dengan teknik *Deep learning* dapat dilihat dari Gbr.2 menunjukkan bahwa *class precision* prediksi tidak tepat waktu adalah 80,22 %, sedangkan *class precision* prediksi tepat waktu sebesar 70,70%. *Class recall true* tidak tepat waktu sebesar 44,24% dan *class recall true* tepat waktu sebesar 92,50%.

accuracy: 72.84% +/- 5.71% (micro average: 72.84%)

	true Total Tepat waktu	true Total Akut	class precision
pred Total Tepat waktu	73	18	80.22%
pred Total Akut	92	222	70.70%
class recall	44.24%	92.50%	

Gbr .2 Akurasi prediksi tingkat kelulusan

Tingkat akurasi yang dihasilkan menggunakan metode *Machine Learning* dengan teknik *Deep learning* adalah sebesar 72.84% dengan perhitungan manualnya sebagai berikut:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} = \frac{222+73}{222+73+18+92} = 72.84\%$$

Presisi dari perhitungan menggunakan *deep learning* seperti terlihat pada gambar 3 dapat dilihat bahwa *class precision* untuk prediksi tidak tepat waktu sebesar 80,22% dan class precision prediksi tepat waktu sebesar 70,70%. Class recall true tidak tepat waktu sebesar 44,24% dan class

recall true tepat waktu sebesar 92,50% sehingga didapat nilai presisi sebesar 71.32%. Pada Gbr.3 merupakan hasil presisi prediksi tingkat kelulusan.

precision: 71.32% +/- 5.71% (macro average: 70.70%) (positive class: Tepat Waktu)

	true Total Tepat waktu	true Total Akut	class precision
pred Total Tepat waktu	73	18	80.22%
pred Total Akut	92	222	70.70%
class recall	44.24%	92.50%	

Gbr. 3 presisi prediksi tingkat kelulusan



Gbr. 4 Grafik AUC Deep learning tingkat kelulusan

Status of Neuron Layer (predicting STS, 3-class classification, multinomial distribution, CrossEntropy loss, 1,122 weights/biases)

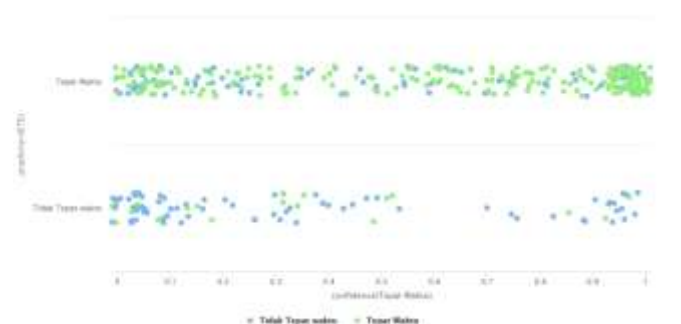
Layer Index	Type	Dropout	LL	LR	Rate	Rate	RMS	Norm	Wgt	Wgt	RMS	Norm	LR	RMS
1	0	Input	0.01	0	0.00000	0.00000	0.002749	0.002075	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2	50	Rectifier	0.01	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
3	50	Rectifier	0.01	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
4	2	softmax	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Gbr. 5 Status neuron layer

Scoring History:

Timestamp	Duration	Training Speed	Epochs	Iterations	Simple Training RMS	Training RMS	R² Training	LogLoss Training	AUC
2019-11-08 13:42:15	0.000 sec	0.00000	0	0.00000	NaN	NaN	NaN	NaN	NaN
2019-11-08 13:42:15	0.000 sec	0.00000	1	400.00000	0.00049	0.13643	0.00000	0.72470	0.72470
2019-11-08 13:42:15	0.062 sec	45322 cond/sec	10.00000	10.000.00000	0.17942	0.26685	0.54518	0.80223	0.80223

Gbr.6 Scoring history



Gbr. 7 Grafik prediksi kelulusan

Untuk hasil dari Grafik AUC yang diperoleh dapat dilihat pada Gbr. 4. Pada Gbr. 4 menunjukkan bahwa posisi status mahasiswa “tepat waktu” memiliki nilai AUC sebesar 0.771 +/- 0.080. Sedangkan hasil pengolahan menggunakan rapidminer diperoleh status dari neuron layer dapat dilihat pada Gbr.5. Pada Gbr. 5 menunjukkan bahwa status neuron layer dengan *hidden Layer sizes* sebesar 50 menggunakan CrossEntropy sebagai loss function menghasilkan RMS sebesar 0.002071 dan 0.025465.

Pada Gbr. 6 menunjukkan bahwa pengulangan dilakukan sebanyak 10 kali dengan waktu selama 0.062 detik dengan

jumlah sampel adalah sebanyak 405 dengan hasil *Training AOC* lebih besar yaitu sebesar 0.80225. Hasil grafik scatter prediksi kelulusan ditunjukkan pada Gbr.7. Pada Gbr. 7 menunjukkan bahwa warna biru mewakili status mahasiswa yang lulus tepat waktu, sedangkan warna biru mewakili mahasiswa yang lulus tidak tepat waktu. Sesuai gambar 7 dapat disimpulkan bahwa jumlah warna hijau lebih dominan dari warna biru

V. KESIMPULAN

Kesimpulan dari penelitian ini adalah prediksi tingkat kelulusan mahasiswa menggunakan metode *Machine Learning* dengan teknik *deep learning* mendapatkan tingkat akurasi sebesar 72.84%. Hasil tersebut menunjukkan bahwa teknik *deep learning* tingkat akurasinya masih kurang bagus ditinjau dari hasil penelitian dengan topik yang sama metode yang berbeda. Kurangnya keakurasian disebabkan data yang digunakan masih terlalu sedikit untuk prediksi dengan teknik *deep Learning*.

UCAPAN TERIMA KASIH

Terima kasih kami sampaikan kepada STMIK IKMI Cirebon yang telah membantu baik dalam hal sarana dan prasarana dalam penelitian ini.

DAFTAR PUSTAKA

- [1] Pdd. Kemenristekdikti, "Statistik Pendidikan Tinggi," 2017.
- [2] "SRV5 PDDIKTI: Pangkalan Data Pendidikan Tinggi." [Online]. Available: <https://forlap.ristekdikti.go.id/perguruan tinggi/detail/Q0QxNjhDRDctOTQ4Qy00N0Y5LTBNjctQUYxQjBGOTNFMMDA4>. [Accessed: 08-Nov-2019].
- [3] R. P. S. Putri and I. Waspada, "Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, p. 1, 2018.
- [4] Y. Kali and M. Linn, "Science," *Int. Encycl. Educ.*, vol. 313, no. July, pp. 468–474, 2010.
- [5] Yoshua Bengio, Patrice Simard, and Paolo Frasconi, "Learning Long-term Dependencies with Gradient Descent is Difficult," *IEEE Transactions on Neural Network*, vol. 5, no. 2, p. 157, 2014.
- [6] I. P. Astuti, "Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5," *Fountain Informatics J.*, vol. 2, no. 2, p. 5, 2017.
- [7] I. Fitriyaningsih *et al.*, "Aplikasi Prediksi Curah Hujan, Debit Air, Dan Kejadian Banjir Berbasis Web Dengan Machine Learning Di Deli Serdang," *J. Penelit. Komun. dan Opini Publik*, vol. 22, no. 2, 2018.
- [8] H. H. Agus Romadhona1, Suprapedi2, "Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma Decision Tree," *J. Teknol. Inf.*, vol. 13, pp. 69–83, 2017.
- [9] S. Syarli and A. A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *J. Ilmu Komput.*, vol. 2, no. 1, pp. 22–26, 2018.
- [10] M. A. Banjarsari, I. Budiman, and A. Farmadi, "Penerapan K-Optimal Pada Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4," *Klik - K J. Ilmu Komput.*, vol. 2, no. 2, pp. 159–173, 2016.