P-ISSN:2527-5321
E-ISSN:2527-5941

PENENTUAN KLASIFIKASI DENGAN CRISP-DM DALAM MEM*PREDIKSI* KELULUSAN MAHASISWA PADA SUATU MATA KULIAH

Shedriko[1], Muhammad Firdaus[2]

[1,2]Program Studi Teknik Informatika, Universitas Indraprasta PGRI
Jl. Nangka No 58C, Tanjung Barat, Jagakarsa, Jakarta Selatan – 12530
[1]shedriko@gmail.com, [2]dasurichi@gmail.com

ABSTRAK

Universitas XYZ merupakan suatu perguruan tinggi yang memiliki mahasiswa yang relatif banyak. Keberadaan suatu *tool prediksi* kelulusan dalam suatu mata kuliah sangat diperlukan dalam menunjang proses belajar mengajar di universitas tersebut, untuk mendorong tingkat kelulusan yang diinginkan. Kelulusan mahasiswa dalam suatu mata kuliah dapat di*prediksi*kan berdasarkan data masukkan yang menjadi parameternya. Banyak metode klasifikasi yang dapat digunakan dengan keunggulan dan kekurangannya masing-masing yang dapat digunakan untuk melakukan *prediksi* tersebut. Dengan menggunakan metodologi data *mining* CRISP-DM, yaitu mengkomparasi beberapa klasifikasi dalam *supervised learning*, dapat diperoleh nilai terbaik yang berkaitan dengan akurasi dan *error*. Klasifikasi yang diperbandingkan dengan menggunakan *software Orange Data Mining* tersebut adalah *Naive Bayes* (NB), *Neural Network* (NN), *Logistic Regression* (LR) dan *Support Vector Machine* (SVM). Objek penelitian dilakukan terhadap mahasiswa yang mengikuti mata kuliah PTI (Pengantar Teknologi Informasi). Tujuan dari penelitian ini adalah mendapatkan metode klasifikasi yang efektif dan efisien dalam menentukan kelulusan mahasiswa dalam suatu mata kuliah. Hasil penelitian menunjukkan bahwa SVM merupakan salah satu klasifikasi yang sangat *robust*.
Kata Kunci: crisp-dm, data *mining*, klasifikasi, kelulusan, *prediksi*.

ABSTRACT

The university of XYZ is one univerisity with a bulk number of student. A prediction tool is needed to support the teaching process for the pass rate achievement. The pass of college student from one subject can be predicted on the base of input parameters. Many used classification methods have their own positive and negative impacts in prediction. By using CRISP-DM data mining methodology, which is comparing several supervised learning classification, it will be obtaining the best result regarding accuracy and error value. This classification which is supported by Orange Data Mining software, are Naive Bayes (NB), Neural Network (NN), Logistic Regression (LR) and Support Vector Machine (SVM). The objects of the research are students in Introduction to Information Technology subject. The research purpose is determining the best classification method which is effective and efficient in the student pass of a subject.. The result denotes that SVM is one of robust methods in classification.
Key Word: *crisp-dm, data mining, classification, the pass, prediction.*

PENDAHULUAN

Penelitian ini dilakukan di Universitas XYZ yang merupakan salah satu universitas yang berlokasi di Jakarta, yang memiliki 5 fakultas yang salah satunya merupakan fakultas pasca sarjana. Universitas ini memiliki visi, di samping sebagai penyelenggara pendidikan yang berguna untuk mencerdaskan kehidupan bangsa sehingga kelak dapat meningkatkan taraf kehidupan para peserta didik, juga memiliki visi unik yaitu menjadi perguruan tinggi yang membantu mengentaskan kemiskinan yaitu dengan memberikan biaya pendidikan yang relatif murah dan terjangkau kepada para peserta didik. Biaya yang relatif terjangkau ini menarik minat banyak calon peserta didik untuk mendaftar menjadi mahasiswanya. Meskipun memberikan biaya pendidikan yang relatif murah, namun universitas tetap berfokus pada kualitas. Hal tersebut dibuktikan dengan peningkatan kualitas pengajar secara berkesinambungan serta kualitas peserta didik dan fasilitas belajar mengajar yang digunakan.

Dengan jumlah mahasiswa yang relatif banyak, kehadiran suatu *tool* mengenai perkiraan kelulusan mahasiswa dalam suatu mata kuliah sangat mendukung proses belajar mengajar. Sehingga dilakukan penelitian terhadap mahasiswa yang mengikuti mata kuliah

P-ISSN:2527-5321
E-ISSN:2527-5941

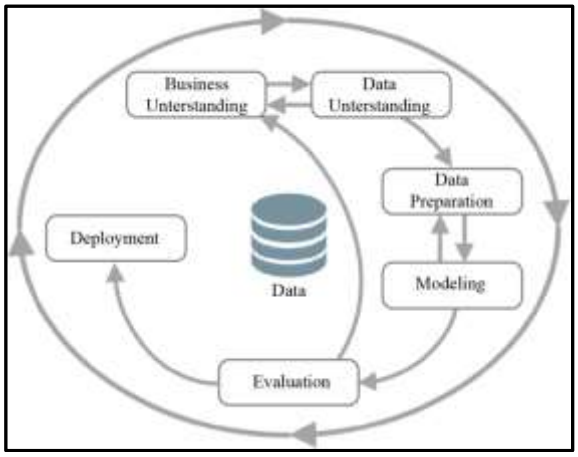
Pengenalan Teknologi Informasi (PTI) dengan beberapa kelas yang diambil sebagai *sample*nya. Dalam data homogen tersebut diambil sejumlah 175 mahasiswa yang akan diamati parameter penilaiannya yang kemudian dianggap sebagai data *training*. Selanjutnya data tersebut akan diproses menggunakan application, yaitu model yang diperoleh dari data training yang digunakan untuk menentukan hasil dari suatu set data *test* baru (Han, Kamber, & Pei, 2012). Menurut Han dkk terdapat 2 kategori penilaian tersebut, yaitu:

1. *Descriptive*, bertugas mencirikan atau mengklasifikasikan sifat-sifat data tertentu ke dalam suatu set data tujuan
2. *Predictive*, bertugas melakukan induksi atau rangsangan terhadap suatu set data untuk dapat melakukan proses *prediksi* atau ramalan atau perkiraan

Dari kategori penilaian tersebut, dapat kita pahami bahwa penelitian ini merupakan suatu *prediksi* dari parameter inputnya.

Dalam bukunya, Dean menyatakan bahwa dalam kumpulan banyak data terdapat informasi yang dapat merubah tidak hanya kondisi seorang pasien, tapi bahkan tatanan dunia (Dean, 2014). Ungkapan dari Dean tersebut merupakan gambaran dari data *mining*. Dengan melakukan ekstraksi data, maka akan diperoleh pola-pola tertentu. Dari pola-pola tersebut dapat dilanjutkan dengan menggunakan algoritma yang dipilih untuk menghasilkan *prediksi-prediksi*, dimana penggunaan algoritma merupakan bagian dari *machine learning* (Manrai, 2020). CRISP-DM atau *Cross-industry Standard Process for Data Mining* merupakan metode yang menyediakan standar baku dalam data *mning* yang dapat diterapkan ke dalam strategi pemecahan masalah umum pada bisnis atau pada unit penelitian yang terdiri dari beberapa fase (Huber, 2018).

Sedangkan *machine learning* terdiri dari 3 jenis, yaitu *Supervised Learning*, *Unsupervised Learning* dan *Reinforcement Learning* (Hertzmann & Fleet, 2012).



Gambar 1. Proses CRISP-DM (Huber, 2018)

Supervised Learning merupakan tipe pembelajaran dimana data *training* yang berisi jawaban yang benar diberikan terlebih dulu sebagai acuan. Banyak contoh algoritma yang masuk ke dalam kategori *supervised*, yang mana untuk penelitian ini akan diperbandingkan beberapa algoritma untuk mendapatkan yang terbaik berdasarkan nilai akurasi dan *error*nya, yaitu *Niave Bayes* (NB), *Neural Network* (NN) atau yang dikenal dengan *Perceptron*, *Logistic Regression* (LR) dan *Support Vector Machine* (SVM). *Naive Bayes Classifier* adalah kerangka kerja probabilitas untuk penyelesaian masalah klasifikasi, dimana naive dimaksudkan bahwa suatu fitur diasumsikan independent terhadap fitur lainnya yang tidak mungkin ada dalam dunia nyata, sedangkan bayes diambil dari nama seorang pembuat teorema statistik (Sawla, 2018). *Perceptron* atau *Neural Network* adalah model biologi neuron yang terdiri dari bobot dan bias yang *adjustable* (dapat diubah-ubah nilainya), yang dapat mengklasifikasi dengan hanya dua kelas, dimana algoritma ini akan melakukan konvergensi dan *positioning* posisi hasil pada bidang yang membagi dua kelas dengan memperbaiki nilai bobot terus menerus hingga posisi klasifikasi diperoleh (Bennamoun, 2003). Sedangkan Logistic Regression adalah metoda analisa statistika untuk mendeskripsikan hubungan antara dua atau lebih variabel terikat dengan variabel bebas (Kleinbaum & Klein, 2010). Dan Support Vector Machine merupakan metode dalam *Supervised Learning* yang digunakan untuk mencari *hyperplane* terbaik dengan memaksimalkan jarak antar kelas (Deng, Tian, & Zang, 2013).

P-ISSN:2527-5321
E-ISSN:2527-5941

Tabel 1. Perbandingan Metode Klasifikasi

No	Metode	Kelebihan	Kekurangan
1	Naive Bayes (NB)	1. Bekerja baik pada data training yang sedikit 2. Mengkonver-gensi secara sangat cepat dibanding model lain 3. Dapat mengatasi fitur yang tidak relevan serta mendukung klasifikasi biner dan multi-class	1. Fitur-fitur independen, tidak ada dalam dunia nyata 2. Kesalahan akan terjadi bila populasi tidak terwakili 3. Mengekstrak data diskrit ke continue berpotensi terjadi data loss
2	Neural Network (NN)	1. Dapat bekerja baik dengan data training yang kurang 2. Toleran terhadap fault 3. Robust terhadap noise dari data training	1. Membutuhkan processor yang bekerja secara paralel 2. Durasi penyelesaian tidak dapat dipastikan 3. Permasalahan tidak terlihat dari iterasi
3	Logistic Regression (LR)	1. Mudah, cepat dan simple 2. Klasifikasi untuk multi-class capable 3. Variable independen terhadap dependen nampak jelas	1. Tidak capable pada klasifikasi non-linier 2. Dibutuhkan pemilihan fitur yang tepat 3. Kurang baik bila terdapat data noise
4	Support Vector Machine (SVM)	1. Dapat memberikan solusi bagi masalah yang kompleks 2. Dapat menahan loss sehingga memberikan akurasi yang lebih baik 3. Nilai optimal minimum selalu dapat dicapai	1. Menahan loss dapat berakibat kurangnya kerapatan informasi 2. Perlakuan khusus untuk parameter hiperaktif 3. Semakin besar dataset akan semakin besar pula waktu trainingnya

(Shrestha, 2019; Varghese, 2018a, 2018b)

METODE PENELITIAN

Penelitian dilakukan dengan menggunakan data yang diperoleh melalui pengamatan parameter nilai input dan output terhadap mahasiswa yang mengikuti mata kuliah PTI (Pengantar Teknologi Informasi) yang berjumlah 175 orang, secara quantitatif (Kothari, 2004), untuk kemudian memprosesnya menggunakan metodology beberapa klasifikasi yang dipilih untuk diperbandingkan. Analisis dilakukan terhadap informasi yang sudah ada dengan melakukan perhitungan dari parameter input hingga menjadi output, dilanjutkan dengan menghitung nilai akurasi dan errornya, untuk membuat evaluasi kritis berkaitan dengan masalah tertentu (Kothari, 2004). Evaluasi tersebut dilakukan terhadap 4 jenis klasifikasi dengan melakukan langkah penghitungan yang sama dengan yang telah dijelaskan di atas, namun disesuaikan dengan rumusan masing-masing.

HASIL DAN PEMBAHASAN

Beberapa kriteria penilaian yang umum dilakukan pada kelulusan mahasiswa dalam suatu mata kuliah mencakup keseluruhan nilai dari kehadiran, tugas, ujian tengah semester, ujian akhir semester serta nilai-nilai penunjang lainnya seperti keaktifan di kelas dan kuis. Sehingga uraian dalam proses CRISP-DM nya adalah sebagai berikut (Hananto, 2017).

1. Pemahaman Bisnis

Tujuan bisnis dari penelitian ini adalah menentukan model klasifikasi yang tepat untuk dapat melakukan *prediksi* kelulusan yang baik dari beberapa model. Kondisi saat ini, belum adanya satu *tools* yang dapat digunakan untuk melakukan klasifikasi tersebut. Sedangkan tujuan dari data *mining*nya adalah model klasifikasi yang telah dipilih dapat memberikan nilai akurasi yang baik berdasarkan pola kelulusan yang ada berkaitan dengan parameter nilai inputnya.

2. Pemahaman Data

Pada fase ini dilakukan pengumpulan data awal sebagai syarat kelulusan dengan rincian sebagai berikut:

a. Nilai akhir harus mencapai nilai lebih besar atau sama dengan 56
b. Absensi harus memenuhi setidaknya 80 persen
c. Nilai tugas minimal 60
d. Menghadiri kuis yang diadakan
e. Aktif di kelas dengan bertanya atau menjawab pertanyaan-pertanyaan yang terlontar di kelas baik yang dikemukakan oleh dosen maupun mahasiswa pada sesi diskusi

Namun demikian, beberapa penilaian subyektif dapat saja terjadi dengan anggapan bahwa seorang mahasiswa mungkin lengah pada saat ujian, padahal dalam kegiatan belajar mengajar sehari-hari di kelas dosen melihat bahwa mahasiswa tersebut memiliki potensi yang cukup besar di perkuliahan tersebut. Data dideskripsikan sebagai berikut: nilai tugas diambil berdasarkan pengumpulan dari tugas yang diberikan termasuk kuis, nilai ujian tengah semester diambil dari ujian yang dilaksanakan pada tengah semester dan nilai akhir semester yang diambil pada akhir semester.

P-ISSN:2527-5321
E-ISSN:2527-5941

3. Persiapan Data

Pada fase ini dilakukan pemilihan dan pembersihan data. Beberapa *noise* tetap dipertahankan untuk mewakili kondisi sesungguhnya di lapangan. Adapun atribut yang digunakan sebagai *evidence* dalam algoritma yang diperbandingkan pada penelitian ini terdiri dari 3 *item*, yaitu.

a. Nilai tugas
b. Nilai ujian tengah semester (UTS)
c. Nilai akhir semester (UAS)

Data yang menjadi nilai inputan tersebut nantinya akan dimasukkan ke dalam s*oftware* pengolah data, yaitu *Orange Data Mining*. *Software* tersebut merupakan pengolah data *mining open source* yang cukup *powerful* disamping penggunaannya yang relatif mudah dan *simple*. Beberapa penampakan data *training* dimana mahasiswa dinyatakan tidak lulus dengan nilai target = 0 dan lulus dengan nilai target = 1 dapat dilihat pada tabel di bawah berikut ini.

Tabel 2. Beberapa data *training* yang menyatakan mahasiswa tidak lulus

1	Nilai Tugas	Nilai UTS	Nilai UAS	Kelulusan	Target
2	9,0	5,0	2,8	TIDAK LULUS	0
3	9,0	4,5	1,7	TIDAK LULUS	0
4	0,0	3,0	0,0	TIDAK LULUS	0
5	10,0	3,0	1,0	TIDAK LULUS	0
6	0,0	0,0	0,0	TIDAK LULUS	0
7	10,0	5,0	1,8	TIDAK LULUS	0
8	8,0	5,0	1,8	TIDAK LULUS	0
9	9,0	3,0	4,2	TIDAK LULUS	0
10	0,0	0,0	0,0	TIDAK LULUS	0
11	0,0	6,8	0,0	TIDAK LULUS	0
12	9,0	6,2	2,8	TIDAK LULUS	0
13	8,0	6,2	0,9	TIDAK LULUS	0
14	9,0	5,0	2,1	TIDAK LULUS	0
15	1,5	4,5	8,1	TIDAK LULUS	0
16	1,5	6,8	6,8	TIDAK LULUS	0
17	1,5	7,2	8,2	TIDAK LULUS	0
18	1,5	9,0	3,5	TIDAK LULUS	0
19	1,5	8,5	5,6	TIDAK LULUS	0
20	0,0	0,0	0,0	TIDAK LULUS	0
21	8,0	8,0	6,0	TIDAK LULUS	0

Tabel 3. Beberapa data *training* yang menyatakan mahasiswa lulus

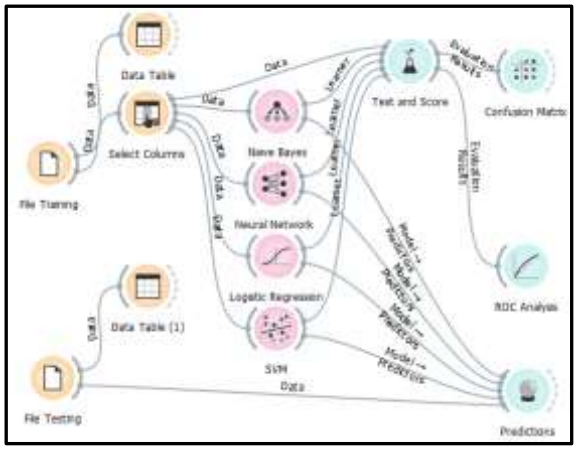
1	Nilai Tugas	Nilai UTS	Nilai UAS	Kelulusan	Target
157	10,0	7,2	7,1	LULUS	1
158	10,0	6,2	4,7	LULUS	1
159	9,0	6,8	4,2	LULUS	1
160	10,0	8,0	6,2	LULUS	1
161	9,0	5,0	6,3	LULUS	1
162	9,0	8,5	7,2	LULUS	1
163	9,0	5,0	6,1	LULUS	1
164	10,0	7,2	6,3	LULUS	1
165	9,0	7,2	6,2	LULUS	1
166	10,0	7,2	7,5	LULUS	1
167	10,0	6,8	4,9	LULUS	1
168	10,0	7,2	6,3	LULUS	1
169	9,0	6,2	5,8	LULUS	1
170	10,0	5,0	4,2	LULUS	1
171	8,0	5,5	5,0	LULUS	1
172	8,0	7,2	6,3	LULUS	1
173	10,0	4,0	5,8	LULUS	1
174	8,0	7,2	5,9	LULUS	1
175	9,0	6,8	5,9	LULUS	1
176	10,0	6,2	4,8	LULUS	1

4. Permodelan

Pada pemilihan teknik permodelan diambil beberapa algoritma klasifikasi yang akan diperbandingkan untuk mendapatkan klasifikasi terbaik untuk memprediksi kelulusan mahasiswa dalam suatu mata kuliah. Algoritma yang diperbandingkan tersebut adalah *Naive Bayes* (NB), *Neural Network* (NN), *Logistic Regression* (LR) dan *Support Vector Machine* (SVM). Dalam membangun model, digunakan *software Orange Data Mining*, yang direkonstruksi sedemikian rupa sehingga suatu rangkaian model yang berkaitan dengan keempat algoritma tersebut di atas dibuat. Beberapa fitur yang disediakan oleh software ini adalah sebagai berikut:

a. Banyak menyediakan algoritma yang berbeda untuk *machine learning* dan *data mining*
b. Merupakan aplikasi *open source* dan tersedia secara bebas atau gratis
c. Platform yang dapat berdiri sendiri atau *independent*
d. Mudah digunakan oleh siapa saja, tidak hanya spesialis data *mining*
e. Fasilitas yang disediakan fleksibel untuk eksperimen
f. Dilakukan *up-to-date* secara berkala dengan penambahan fitur dan algoritma baru

P-ISSN:2527-5321
E-ISSN:2527-5941



Gambar 2. Model CRISP-DM Penelitian Dengan Orange Data Mining

Penilaian model yang merupakan perbandingan akurasi model dari keempat model data *mining* dan algoritma *machine learning* yang diperbandingkan dapat dilihat pada tabel 4. Akurasi merupakan nilai yang diperoleh dari hasil perhitungan jumlah nilai true positive dengan true negative yang dibagi dengan total populasi. Sedangkan error rate diperoleh dari hasil perhitungan dari jumlah nilai false positive dan false negative dibagi dengan total populasi. Dalam penelitian ini, total populasi yang diambil adalah 175 mahasiswa.

Tabel 4. Perbandingan akurasi dan error

Classification Algorithms	Accuracy Rate	Error Rate
Naive Bayes (NB)	89,7%	10,3%
Neural Network (NN)	85,7%	14,3%
Logistic Regression (LR)	88,6%	11,4%
Support Vector Machine (SVM)	95,4%	4,6%

Sedangkan nilai-nilai statistik lainnya dapat dilihat pada table 5. Dalam table tersebut diperlihatkan nilai yang merupakan *Precision*, *Recall* dan AUC (*Area Under Curve*), yang kesemuanya diklasifikasikan ke dalam 2 kelompok, yaitu lulus (*passed*) dan tidak lulus (*failed*), dengan menggunakan *Orange Data Mining*. *Precision* merupakan rasio *prediksi* benar positif dibandingkan dengan keseluruhan hasil yang di*prediksi* positif, sedangkan *Recall* merupakan rasio benar positif dibandingkan dengan keseluruhan data yang benar positif (Arthana, 2019). Dan AUC merupakan daerah di bawah ROC yang digunakan untuk menentukan model yang terbaik dalam *prediksi* (Abhigyan, 2020).

Tabel 5. Perbandingan analisis statistik

Algoritma	Class	Precision	Recall	AUC
NB	Passed	0,935	0,935	0,947
	Failed	0,750	0,750	0,947
NN	Passed	0,848	1,000	0,902
	Failed	1,000	0,306	0,902
LR	Passed	0,89	0,993	0,971
	Failed	0,944	0,472	0,971
SVM	Passed	0,958	0,986	0,969
	Failed	0,938	0,833	0,969

Kemudian, dengan menggunakan beberapa data testing, dilakukan pengujian terhadap keempat algoritma. Hasil pengujian dapat dilihat pada tabel 6.

Tabel 6. Hasil Pengujian

	NB	NN	LR	SVM	Homework	MidTest	FinalTest
1	Failed	Failed	Failed	Failed	42	42	42
2	Failed	Passed	Failed	Failed	43	43	43
3	Failed	Passed	Failed	Failed	45	45	45
4	Failed	Passed	Passed	Failed	46	46	46
5	Failed	Passed	Passed	Failed	58	58	58
6	Passed	Passed	Passed	Passed	59	59	59

Dari tabel hasil pengujian, terlihat bahwa nilai ambang kelulusan dari *Naive Bayes* (NB) terdapat pada nilai 58 dan 59, *Neural Network* (NN) pada nilai 42 dan 43, *Logistic Regression* (LR) pada nilai 45 dan 46, serta *Support Vector Machine* (SVM) pada nilai 58 dan 59.

SIMPULAN DAN SARAN

Dari beberapa parameter yang telah dijelaskan di atas, diperoleh hasil penilaian sebagai berikut:

1. Nilai *precision* kelas lulus tertinggi diraih oleh algoritma SVM
2. Nilai *recall* kelas lulus tertinggi diraih oleh algoritma LR
3. Nilai AUC tertinggi diraih oleh algoritma LR
4. Nilai akurasi tertinggi diraih oleh algoritma SVM
5. Nilai uji pada ambang kelulusan terhadap nilai yang diharapkan, poin tertingginya diraih oleh algoritma NB dan SVM

Kelima poin di atas menunjukkan, bahwa algoritma SVM memberikan 3 kali kemunculan dalam hal peraihan nilai tertinggi. Sementara

P-ISSN:2527-5321
E-ISSN:2527-5941

algoritma LR muncul sebanyak 2 kali. Sedangkan NB hanya muncul sebanyak 1 kali. Sehingga dapat ditarik kesimpulan bahwa SVM merupakan algoritma yang terbaik yang dapat diambil sebagai tools dalam memprediksi kelulusan mahasiswa pada suatu mata kuliah. Dengan tetap mempertahankan *noise* yang terdapat pada data yang diolah, dapat disimpulkan bahwa SVM cukup *robust* dalam mengatasi anomali data pada penelitian ini. Penelitian ini dapat dijadikan acuan sebagai penilai dalam memprediksi kelulusan mahasiswa, tidak hanya pada mata kuliah Pengantar Tenologi Informasi, tapi juga untuk mata kuliah lainnya.

DAFTAR PUSTAKA

Abhigyan. (2020). Understanding The AUC-ROC Curve. Retrieved July 31, 2021, from medium.com website: https://medium.com/analytics-vidhya/understanding-the-auc-roc-curve-cdc754d7b58a

Arthana, R. (2019). Mengenal Accuracy, Precision, Recall dan Specificity serta yang diprioritaskan dalam Machine Learning. Retrieved July 31, 2021, from medium.com website: https://rey1024.medium.com/mengenal-accuracy-precission-recall-dan-specificity-serta-yang-diprioritaskan-b79ff4d77de8

Bennamoun, M. (2003). *Single Layer Perceptron (SLP) Classifiers*. Lecture Slide of Neural Computation, University of Western Australia.

Dean, J. (2014). *Big data, data mining, and machine learning [internet resource]: value creation for business leaders and practitioners*. John Wiley & Sons Inc.

Deng, N., Tian, Y., & Zang, C. (2013). *Support vector machines (SVM): Optimization Based Theory, Algorithms and Extension*. CRC Press,Taylor & Francis Group.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (Third). https://doi.org/10.1016/C2009-0-61819-5

Hananto, V. R. (2017). Analisis Penentuan Metode Data Mining Untuk Prediksi Kelulusan Mahasiswa Sebagai Penunjang Angka Efisiensi Edukasi. *Jurnal Ilmiah SCROLL*, *5*(1), 1–11.

Hertzmann, A., & Fleet, D. (2012). *Machine Learning and Data Mining*. Computer Science Department, University of Toronto.

Huber, S. (2018). A holistic extension to the CRISP-DM model. *Science Direct*, (12th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 18-20 July 2018, Gulf of Naples, Italy).

Kleinbaum, D. G., & Klein, M. (2010). *Statistics for biology and health: Logistical regression* (pp. 1–709). pp. 1–709.

Kothari, C. R. (2004). *Research Methodology, Methods & Techniques* (Second Rev). https://doi.org/10.1016/0022-460X(82)90541-7

Manrai, T. (2020). How Data Mining is Different Than Machine Learning. Retrieved July 26, 2021, from medium.com website: https://manrai-tarun.medium.com/how-data-mining-is-different-than-machine-learning-cdcaf559d2a7

Sawla, S. (2018). Introduction to Naive Bayes for Classification. Retrieved July 28, 2021, from medium.com website: https://medium.com/@srishtisawla/introduction-to-naive-bayes-for-classification-baefefb43a2d

Shrestha, P. (2019). The Advantages and Disadvantages of Neural Networks. Retrieved July 28, 2021, from gkstuffs.com website: https://gkstuffs.com/future-tech/advantages-and-disadvantages-of-neural-networks/

Varghese, D. (2018a). Comparative Study on Classic Machine Learning Algorithms. Retrieved July 28, 2021, from towardsdatascience.com website: https://towardsdatascience.com/comparative-study-on-classic-machine-learning-algorithms-24f9ff6ab222

Varghese, D. (2018b). Comparative Study on Classic Machine Learning Algorithms Part-2. Retrieved July 28, 2021, from medium.com website: https://medium.com/@dannymvarghese/comparative-study-on-classic-machine-learning-algorithms-part-2-5ab58b683ec0