

## KOMPARASI MODEL KLASIFIKASI ALGORITMA KETERLAMBATAN SISWA MASUK SEKOLAH

Imam Amirulloh<sup>1\*</sup>, Taufiqurrochman<sup>2</sup>

<sup>\*12</sup>Ilmu Komputer, SMTIK Nusa Mandiri, Jakarta

Jl. Kramat Raya, Senen, Kota Jakarta Pusat

\*Email: imamamirulloh@gmail.com

### ABSTRAK

Kedisiplinan bagian dari pendidikan yang diselenggarakan di sekolah, terutama disiplinnya siswa untuk datang ke sekolah tepat waktu atau datang lebih awal, beberapa faktor yang perlu diketahui sehingga terjadinya keterlambatan siswa datang ke sekolah serta bisa menjadi evaluasi siswa. Algoritma Model Klasifikasi menjadi metode untuk memprediksi keterlambatan siswa, pada penelitian ini akan menguji serta membandingkan lima algoritma model klasifikasi pada dataset dari hasil penelitian di SMK YPC Tasikmalaya, adapun lima algoritma tersebut : naïve bayes, k-nn, decision tree, logistic regression, Deep Learning dengan membandingkan algoritma tersebut akan di dapat metode mana yang lebih bagus dalam menyelesaikan permasalahan penelitian ini. Hasil dari penelitian ini Decision Tree dan K-NN menghasilkan performa yang lebih bagus dibandingkan metode yang lain.

**Kata kunci:** Komparasi, Algoritma, siswa, sekolah, Klasifikasi

### ABSTRACT

*Discipline part of education held in school, especially discipline students to come to school on time or coming early, some factors that need to be known so that the occurrence of student delay in coming to school and can become student evaluation. The classification algorithm becomes a method to predict student delay, in this research will test and compare five algorithms of classification model in dataset from research result in SMK YPC Tasikmalaya, as for the five algorithms: naïve bayes, k-nn, decision tree, logistic regression, Deep Learning by comparing the algorithm will be able to which method is better in solving this research problem. The results of this research Decision Tree and K-NN produce better performance than other methods.*

*Keywords: Comparison, Algorithm, student, school, Classification*

### PENDAHULUAN

Sekolah bukan hanya tempat pendidikan pencarian ilmu, tapi juga tempat melatih kedisiplinan untuk semua yang terlibat di sekolah tersebut salah satunya siswa. kedisiplinan siswa datang ke sekolah menjadi yang pertama yang dilihat, beberapa faktor yang terjadi keterlambatan bisa terjadi antara lain: jarak ke sekolah, jam bangun, jam berangkat, kondisi perjalanan, dan kendaraan yang dipakai. Faktor-faktor tersebut alasan yang sering disampaikan dan dikeluhkan siswa.

Algoritma Model Klasifikasi menjadi metode untuk memprediksi keterlambatan siswa, dengan beberapa metode algoritma klasifikasi model klasifikasi yang ada, pada penelitian ini akan dilakukan pengujian pada 5 metode algoritma yang bertujuan untuk mencari

metode mana yang lebih bagus dalam memprediksi keterlambatan siswa.

Lima algoritma yang dimaksud adalah K-NN, Naïve Bayes, Decision Tree, Logistic Regression, dan Deep Learning.

### LANDASAN TEORI

#### 1. Data Mining

Data mining dalam istilah sederhana adalah penemuan pola yang berguna dalam pengolahan data, data mining juga disebut sebagai ilmu pengetahuan, *machine learning*, dan *analisis prediksi*

#### 2. Metode Algoritma

##### 2.1. Decision Tree

Mempelajari pohon keputusan dari record pada kelas yang diberi label. Decision tree adalah sebuah flowchart yang seperti struktur pohon, dimana setiap node internal (node tidak berdaun) menandakan sebuah tes pada atribut, setiap branch merepresentasikan hasil dari tes tersebut, dan setiap leaf node (atau node terminal) memegang label kelas. Node yang paling atas di pohon disebut node akar.

### 2.2. Naive Bayes

Klasifikasi bayes adalah dasar dari teori bayes. Ilmu membandingkan klasifikasi algoritma telah menemukan Bayesian classifier sederhana yang dikenal naïve Bayesian classifier yang bisa jadi pembanding kinerja dengan decision tree dan neural network. Bayesian classifier juga bisa menunjukkan akurasi dan kecepatan tinggi apabila diterapkan pada database besar.

### 2.3. K-NN

Klasifikasi Nearest-Neighbor adalah dasar pembelajaran dengan analogi, yaitu dengan membandingkan tupel tes dengan tupel training. Tupel training digambarkan oleh atribut. Setiap tupel merupakan titik dalam ruang dimensi, dengan cara ini semua tupel training diruang dimensi, ketika diberi tupel yang tidak diketahui. K-NN mencari ruang pola tupel training k yang paling dekat dengan tupel yang tidak diketahui. Tupel training k adalah k-nn dari tupelo yang tidak diketahui.

### 2.4. Logistic Regression

Logistic regression memperluas gagasan pada beberapa linear regression untuk situasi dimana variabel saling ketergantungan, y adalah diskrit. Pada logistic regression (Homer & Lemeshow 2000) tidak ada asumsi membuat tentang distribusi pada variabel yang independen. Pemberian set pada sampel N (xi, yi) dengan xi ∈ Rd, dimana d adalah nomor dimensi dan label kelas yang sesuai yi ∈ {1, 2, ..., K}. kemudian, logistic regression mencoba untuk memperkirakan probabilitas posterior pada sampel x baru seperti:

$$p(y = k | x) = \frac{\exp(-(w_{k0} + w_k^T x))}{1 + \sum_{l=1}^{K-1} \exp(-(w_{l0} + w_l^T x))}, k = 1, \dots, K - 1,$$

### 3. Cross Validation

K-Fold cross validation adalah sebuah teknik intensif komputer yang menggunakan keseluruhan data yang ada sebagai training set dan test set [BEN04]. Seluruh data secara acak dibagi menjadi K buah subset Bk dengan ukuran yang sama dimana Bk merupakan himpunan bagian dari {1, ..., n} sedemikian sehingga dan . Setelah itu dilakukan iterasi sebanyak K kali. Pada iterasi k, subset Bk menjadi test set, sedangkan subset yang lain menjadi training set. Setelah itu dihitung nilai rata-rata error dengan menggunakan hasil dari K buah iterasi. [SCH97].

### 4. Rapidminer

Salah satu software untuk pengolahan data mining. Pekerjaan yang dilakukan oleh RapidMiner text mining adalah berkisar dengan analisis teks, mengekstrak pola-pola dari data set yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan database. Tujuan dari analisis teks ini adalah untuk mendapatkan informasi bermutu tertinggi dari teks yang diolah.

### METODE

Metode yang di usulkan pada penelitian ini bertujuan untuk melakukan klasifikasi terhadap data survei dan melakukan komparasi lima algoritma klasifikas dengan metode evaluasi algoritma cross validation, sedangkan metode komparasi menggunakan algoritma T-Test. Metodol kerangka pemikiran yang ditujukan pada penelitian ini.

Tahapan pertama untuk melakukan komparasi algoritma adalah menentukan objek data yang akan di olah, tahap kedua dilakukan pemisahan data otomatis training dan testing melalui cross validation. Tahap ketiga dilakukan proses ekstraksi data mining terhadap data set yang telah di siapkan sebelumnya dengan lima algoritma Decision tree, K-nn, Naïve Bayes, Logistic Regression, Deep Learning. Tahap ke empat melakukan komparasi hasil klasifikasi menggunakan algoritma T-Test.

Data set yang digunakan pada penelitian ini adalah data set hasil kuesioner 64 siswa di SMK YPC Tasikmalaya

### HASIL DAN PEMBAHASAN

dataset dari hasil survei dari 64 siswa yang ada di SMK YPC Tasikmalaya, ada beberapa variabel dalam survei ini yaitu jarak, jam bangun siswa, jam berangkat siswa, jam masuk, kemacetan dan kendaraan yang dipakai siswa serta hasil secara fakta kedatangan siswa terlambat atau tidak terlambat berdasarkan variabel-variabel yang diketahui.

Tabel 1. Dataset

NAMA	JARAK (KM)	JAM BANGUN	JAM BERANGKAT	JAM MASUK	KEMACETAN	KENDARAAN	DATANG
Ai widya sari	8,00	05:00:00	06:00:00	07:00:00	tidak	motor	tidak terlambat
Angga Rizal F	10,00	04:30:00	06:15:00	07:00:00	tidak	umum	tidak terlambat
Anis Yustika	11,70	05:00:00	06:10:00	07:00:00	tidak	motor	tidak terlambat
Anwar Nuralam	1,00	05:00:00	06:00:00	07:00:00	tidak	jalan kaki	tidak terlambat
Ari khoerul anwar	2,00	05:00:00	06:30:00	07:00:00	tidak	motor	tidak terlambat
Ari Rizqi M	3,80	04:00:00	06:15:00	07:00:00	tidak	motor	tidak terlambat
Dede Nurul N	9,00	04:00:00	05:45:00	07:00:00	tidak	motor	tidak terlambat
Dalis Oktaviani	12,00	05:00:00	06:00:00	07:00:00	tidak	umum	tidak terlambat
Derry CA	0,35	03:00:00	07:05:00	07:00:00	tidak	jalan kaki	terlambat
Desi Susanti	10,00	04:00:00	06:00:00	07:00:00	tidak	umum	tidak terlambat
Erna Novianti	0,35	03:00:00	07:00:00	07:00:00	tidak	jalan kaki	terlambat
Friska Hdayani	0,35	03:00:00	07:00:00	07:00:00	tidak	jalan kaki	terlambat
Gita Milenia	10,00	04:00:00	05:00:00	07:00:00	tidak	umum	tidak terlambat
Haru Rukmana	3,70	05:00:00	06:45:00	07:00:00	sedang	motor	tidak terlambat
Ihsan Nurzaman	5,30	05:00:00	06:20:00	07:00:00	tidak	umum	tidak terlambat
Intan	7,00	05:00:00	06:15:00	07:00:00	tidak	umum	tidak terlambat
Ira Wiranti	2,70	04:00:00	06:45:00	07:00:00	tidak	umum	tidak terlambat
Jejen Jaelani	6,00	05:00:00	06:45:00	07:00:00	tidak	motor	tidak terlambat
Lia Iskandar	4,00	05:30:00	06:45:00	07:00:00	tidak	motor	tidak terlambat
Muhamad Fauzi	8,80	05:00:00	06:15:00	07:00:00	tidak	umum	tidak terlambat
Muhammad Alif	8,80	05:30:00	06:30:00	07:00:00	sedang	motor	tidak terlambat
Nadia Nurmeira	2,70	05:00:00	06:45:00	07:00:00	tidak	umum	tidak terlambat
Nur Azizah	4,00	05:00:00	06:50:00	07:00:00	tidak	umum	tidak terlambat
Rani Nurani	4,00	04:30:00	06:40:00	07:00:00	sedang	motor	tidak terlambat
Ragiana H	0,35	03:00:00	07:10:00	07:00:00	tidak	jalan kaki	terlambat
Rina Marlina	5,00	05:00:00	06:30:00	07:00:00	tidak	umum	tidak terlambat
Riza Abdul M	5,00	04:45:00	06:40:00	07:00:00	tidak	umum	tidak terlambat
Siffa Dwi N	6,90	05:00:00	06:35:00	07:00:00	tidak	motor	tidak terlambat
Siti Asizah H	11,00	04:30:00	06:05:00	07:00:00	tidak	umum	tidak terlambat
Siti Eka Febrianti	0,60	03:20:00	07:00:00	07:00:00	tidak	jalan kaki	terlambat
Siti Hanipah	6,90	05:00:00	06:45:00	07:00:00	sedang	motor	terlambat

Dari dataset ini dapat langsung dilakukan pengolahan menggunakan Rapidminer, berikut hasil dari perbandingan dari 5 metode algoritma:

1. *Desicion Tree*

accuracy: 78.81% +/- 18.48% (mikro: 79.37%)

	true tidak terlambat	true terlambat	class precision
pred. tidak terlambat	38	0	80.30%
pred. terlambat	7	12	63.16%
class recall	84.44%	66.67%	

Gambar 1. Hasil Accuracy Decision Tree

Pada Gambar 1 menunjukkan Jumlah data yang diprediksi tidak terlambat dan kenyataannya terlambat 38, jumlah data yang diprediksi tidak terlambat dan kenyataannya terlambat 6, jumlah data yang diprediksi terlambat dan kenyataan tidak terlambat 7, jumlah data yang diprediksi terlambat dan kenyataannya

terlambat 12. Hasil dari komparasi AUC – nya bernilai 0,693

2. *Naive Bayes*

accuracy: 46.19% +/- 23.24% (mikro: 46.03%)

	true tidak terlambat	true terlambat	class precision
pred. tidak terlambat	13	2	86.67%
pred. terlambat	32	16	33.33%
class recall	28.89%	68.89%	

Gambar 2. Hasil Accuracy Decision Tree

Pada Gambar 2 menunjukkan jumlah data yang diprediksi tidak terlambat dan kenyataannya terlambat 13, jumlah data yang diprediksi tidak terlambat dan kenyataannya terlambat 2, jumlah data yang diprediksi terlambat dan kenyataan tidak terlambat 32, jumlah data yang diprediksi terlambat dan kenyataannya terlambat 16. Hasil dari komparasi AUC – nya bernilai 0,668

3. *K-NN*

accuracy: 87.38% +/- 9.04% (mikro: 87.30%)

	true tidak terlambat	true terlambat	class precision
pred. tidak terlambat	38	1	97.44%
pred. terlambat	7	17	70.83%
class recall	84.44%	94.44%	

Gambar 3. Hasil Accuracy K-NN

Pada Gambar 3 menunjukkan jumlah data yang diprediksi tidak terlambat dan kenyataannya terlambat 38, jumlah data yang diprediksi tidak terlambat dan kenyataannya terlambat 1, jumlah data yang diprediksi terlambat dan kenyataan tidak terlambat 7, jumlah data yang diprediksi terlambat dan kenyataannya terlambat 17. Hasil dari AUC bernilai 0,500

4. *Logistic Regression*

accuracy: 71.19% +/- 6.94% (mikro: 71.43%)

	true tidak terlambat	true terlambat	class precision
pred. tidak terlambat	45	18	71.43%
pred. terlambat	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 4. Hasil Accuracy Logistic Regression

Pada gambar 4 menunjukkan jumlah data yang diprediksi tidak terlambat dan kenyataannya

terlambat 45, jumlah data yang diprediksi tidak terlambat dan kenyataannya terlambat 18, jumlah data yang diprediksi terlambat dan kenyataan tidak terlambat 0, jumlah data yang diprediksi terlambat dan kenyataannya terlambat 0. Hasil dari AUC – nya bernilai 0,500

5. Deep Learning

accuracy: 67.86% +/- 18.48% (mikro: 68.25%)

	true tidak terlambat	true terlambat	class precision
pred. tidak terlambat	34	9	79.07%
pred. terlambat	11	9	45.00%
class recall	75.56%	50.00%	

Gambar 5. Hasil Accuracy Deep Learning

Pada gambar 5 menunjukkan jumlah data yang diprediksi tidak terlambat dan kenyataannya terlambat 34, jumlah data yang diprediksi tidak terlambat dan kenyataannya terlambat 9, jumlah data yang diprediksi terlambat dan kenyataan tidak terlambat 11, jumlah data yang diprediksi terlambat dan kenyataannya terlambat 9. Hasil dari AUC – nya bernilai 0,657

6. T-Test

Tabel 2. T-Test

	Decision Tree	Naive Bayes	K-NN	Logistic Reg	Deep L
Decision Tree		0,003	0,204	0,238	0,178
Naive Bayes			0	0,004	0,031
K-NN				0	0,005
Logistic Reg					0,55
Deep L					

KESIMPULAN

Dari pengujian dataset oleh masing – masing algoritma tersebut dapat disajikan tabel T-Test.

Tabel 3. Komparasi Accuracy dan AUC

	DT	KNN	NB	LogR	DL
Accuracy	78,81	87,38	46,19	71,19	67,86
AUC	0,693	0,5	0,668	0,5	0,657

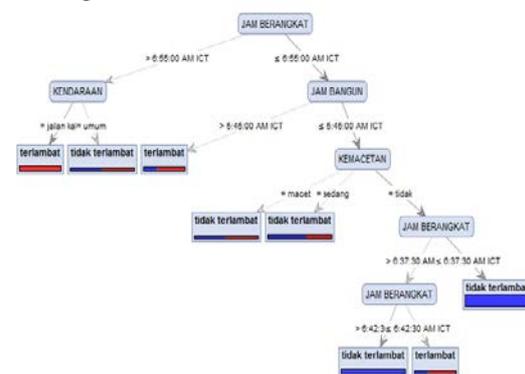
Berdasarkan nilai akurasi algoritma yang lebih akurat K-NN lebih baik disusul oleh DT, LogR, dan DL.

Berdasarkan nilai AUC decision tree, naïve bayes, dan deep Learning menunjukkan nilai yang bagus.

Pada tabel pengujian T-Test dapat disimpulkan bahwa algoritma Decision Tree, K-NN, Log R, dan Deep L menampilkan nilai yang dominan.

Sehingga dari berbagai komparasi tersebut Decision tree dan K-NN lebih baik di banding algoritma yang lain.

Serta faktor yang paling berpengaruh keterlambatan siswa adalah waktu atau jam berangkat



DAFTAR PUSTAKA

Wahono, RS. A Comparison Framework of Classification Models for Software Defect Prediction. Adv. Sci. Lett. 20, 1945–1950, 2014

Lan Yu. “Application and Comparison of Classification Techniques in Controlling Credit Risk”

Vijay Kotu. “Predictive Analytics and Data Mining”. 225 Wyman Street, Waltham, MA 02451, USA. 2015

Ary, D., Jacobs, L.C. & Razavieh, A. 1976. *Pengantar Penelitian Pendidikan*. Terjemahan oleh Arief Furchan. 1982. Surabaya: Usaha nasional