

STUDI REPLIKASI *SPACED REPETITION MODEL* BERBASIS PEMBELAJARAN MESIN UNTUK PEMBELAJARAN BAHASA PADA APLIKASI DUOLINGO

Aditya Bariq Ikhsan (1606889742)
Supervisor: Sabda PS

Outline

Pendahuluan

- Tujuan Penelitian

Ruang Lingkup Penelitian

- Sistematika Penelitian
- Tinjauan Pustaka

Metode Penelitian

- Desain dan Metode Penelitian
- Waktu dan Tempat

Hasil dan Pembahasan

- *Data Preparation*
- *Data Preprocessing*
- *Evaluasi Model*

Kesimpulan

Pendahulaun

Tujuan Penelitian

Membuat *spaced repetition model* menggunakan metode *half-life regression*

Membuat *spaced repetition model* menggunakan metode *linear regression*

Membuat *spaced repetition model* menggunakan metode *tree-based method*

Membuat *spaced repetition model* menggunakan metode *neural network*

RUANG LINGKUP PENELITIAN TINJAUAN PUSTAKA

Spaced Repetition adalah metode belajar sedikit-sedikit (*short study*) dengan jeda dalam suatu periode

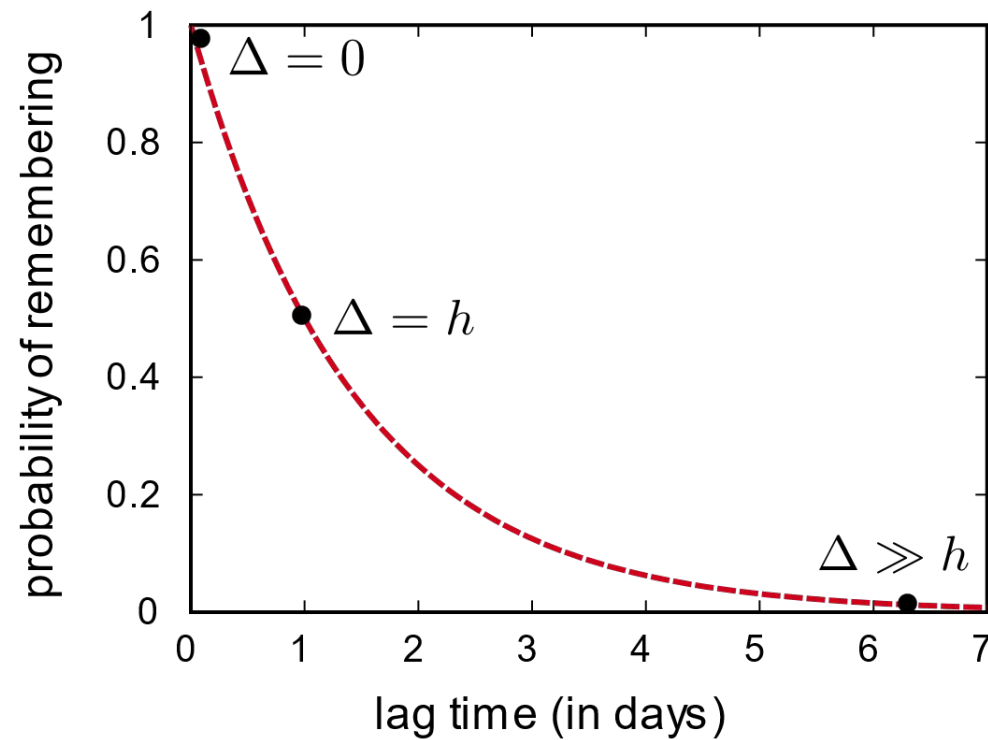
Spacing effect menunjukkan pengamatan orang-orang yang menggunakan metode *spaced repetition* dapat mengingat lebih efektif

lag effect menunjukkan pengamatan orang-orang belajar lebih baik ketika terjadi peningkatan jangka waktu dalam pengulangan belajar

RUANG
LINGKUP
PENELITIAN
TINJAUAN PUSTAKA

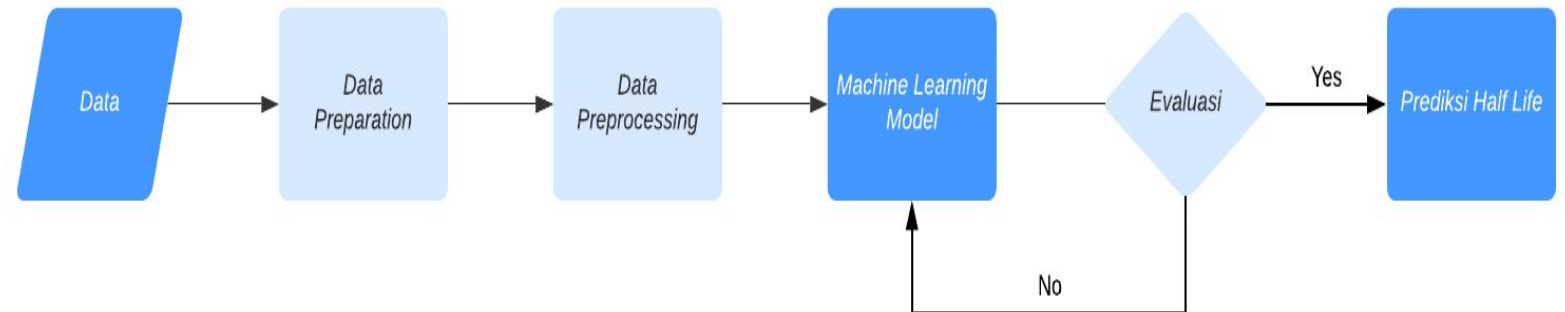
Forgetting Curve

$$P = 2^{-\frac{\Delta}{h}}$$

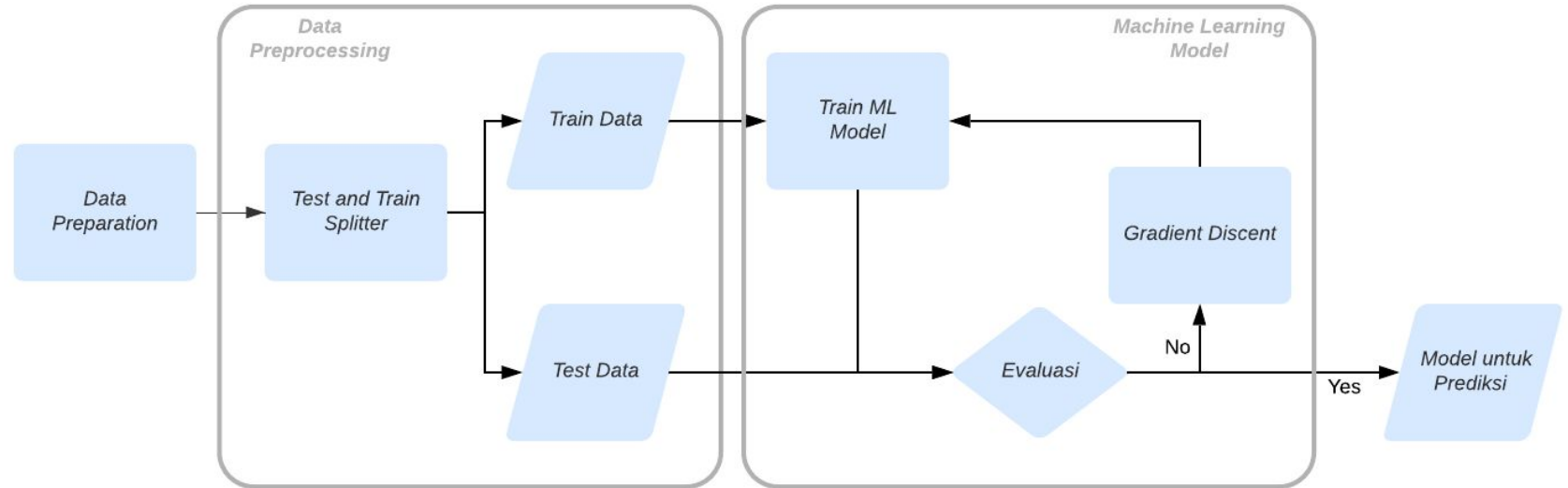


Ruang Lingkup Penelitian Tinjauan Pustaka

- Bagaimana cara mengukur half-life tiap orang?
- Paul Pimsleur
- Sebastian Leitner (*flashcard*)
- Pembelajaran Mesin



Ruang Lingkup Penelitian Tinjauan Pustaka



Ruang Lingkup Penelitian Tinjauan Pustaka

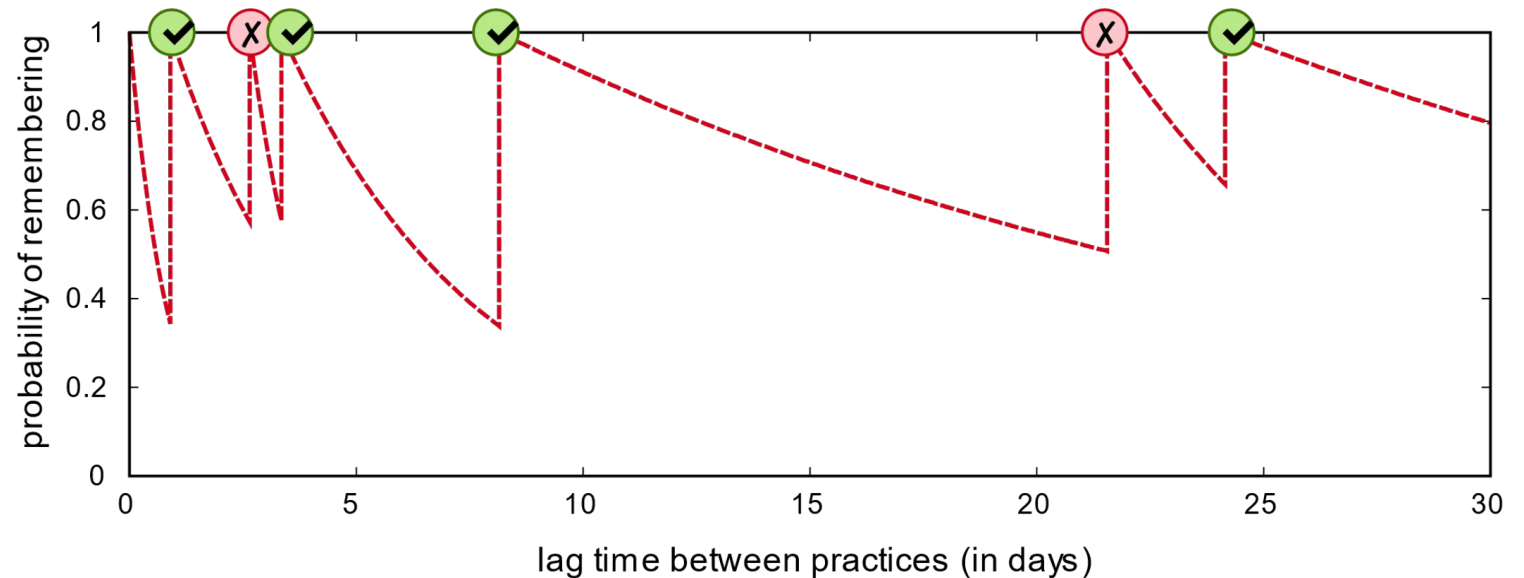


Ruang Lingkup Penelitian Tinjauan Pustaka

Half-life Regression

• $\hat{h}\theta = 2^{\theta \cdot x}$

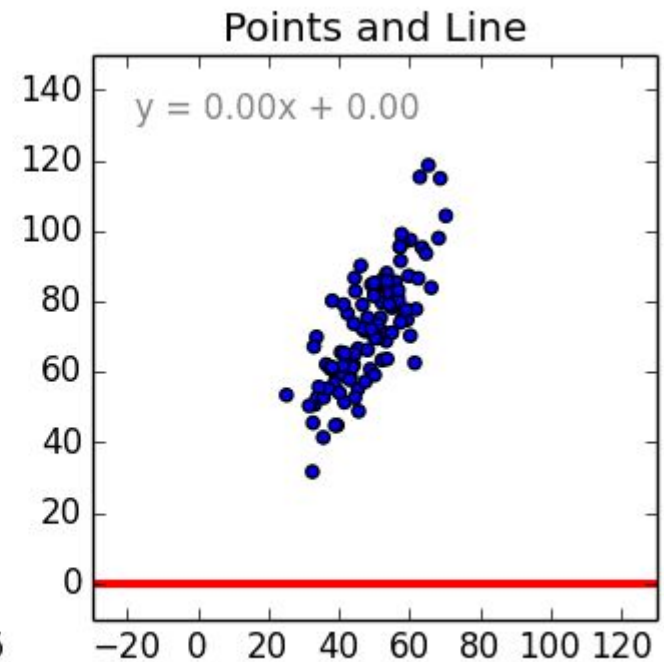
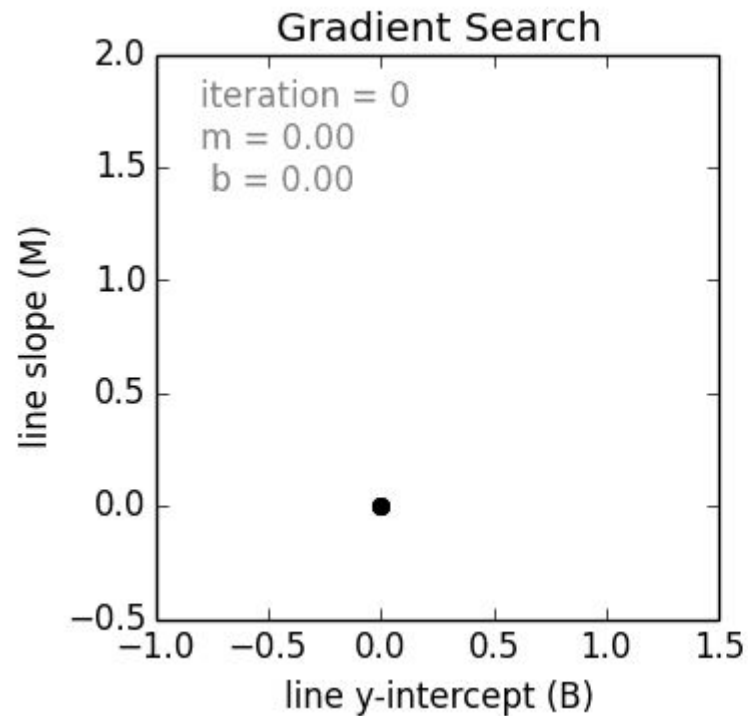
• $\ell = \left(p - 2^{-\frac{\Delta}{2^{\theta \cdot x}}}\right)^2 + \alpha \left(-\frac{\Delta}{\ln(p)} - 2^{\theta \cdot x}\right)^2 + \lambda |\theta|^2$



Ruang Lingkup Penelitian Tinjauan Pustaka

Multi Linear Regression

- $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p + e$

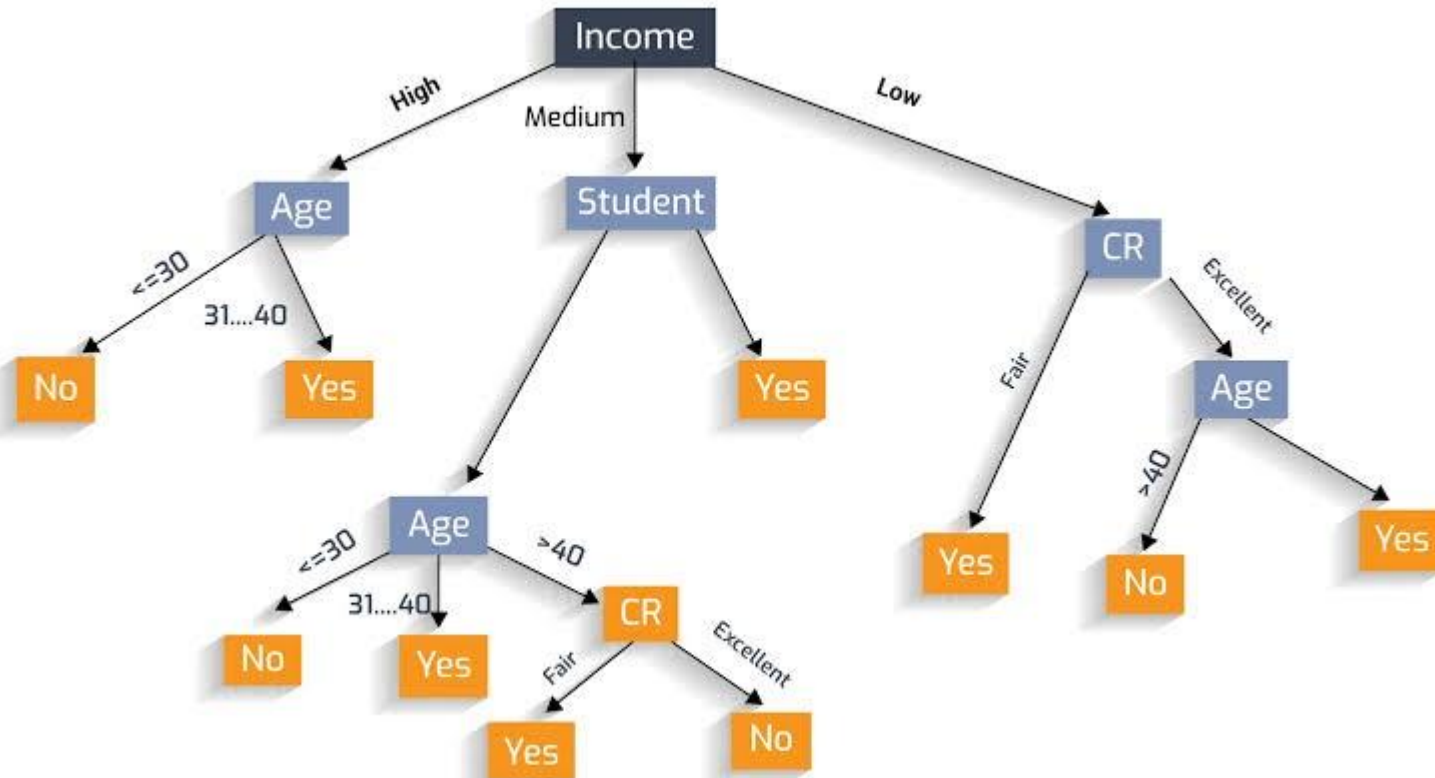


Ruang Lingkup Penelitian Tinjauan Pustaka

Decision Tree

- *LightGBM*

- $Entropy(S) = \sum_i^c - P_i \log_2 P_i$
- $Split\ Information(S, A) = \sum_{i=1}^c - \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$
- $Gain\ Ration(S, A) = \frac{Entropy(S, A)}{Split\ Information(S, A)}$



Ruang Lingkup Penelitian Tinjauan Pustaka

Neural Network

- MultiLayer Perceptron

- Forward pass

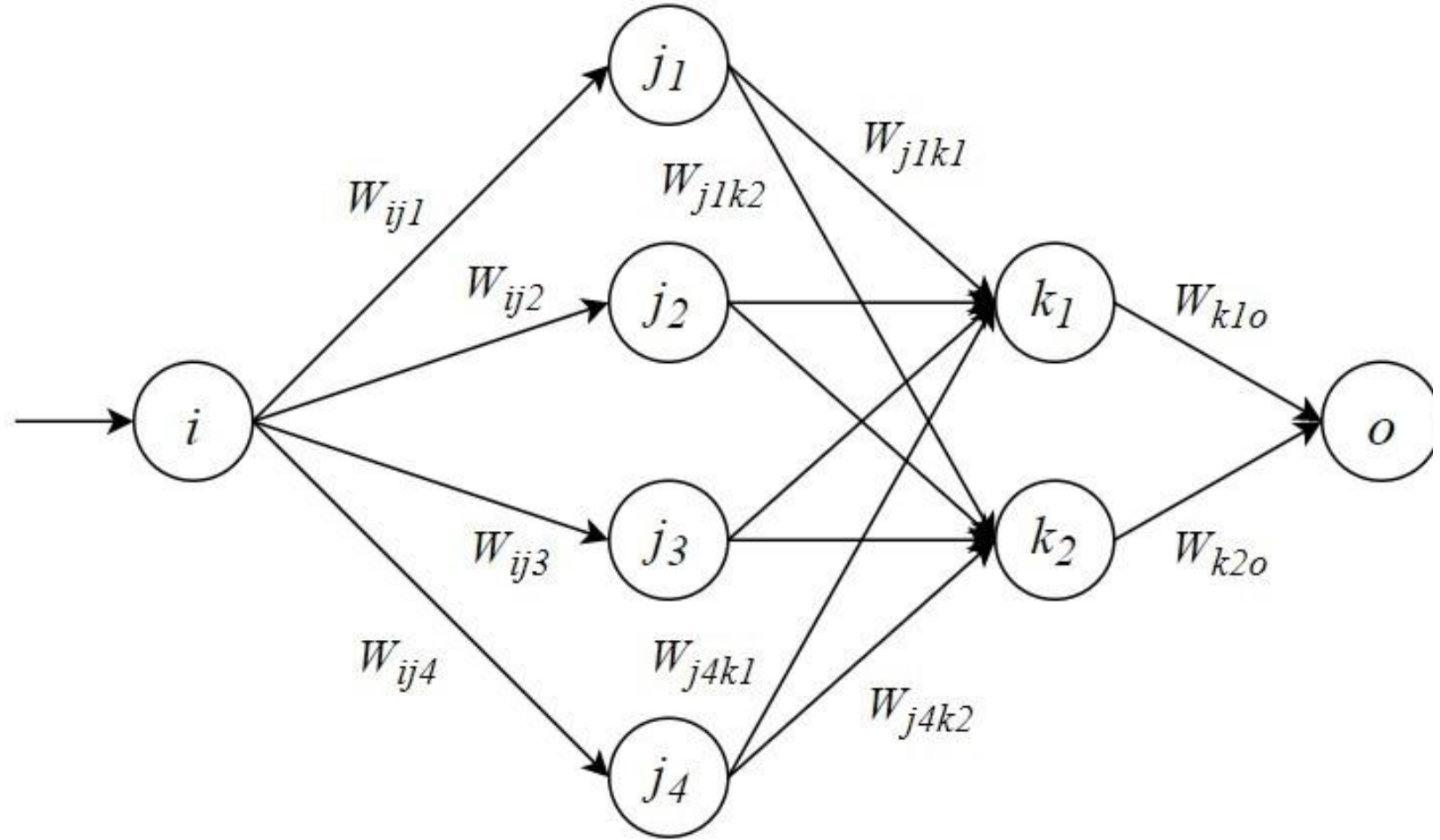
$$dot_j = \sum_i w_{ji} x_i + b_j$$

$$h_j = \sigma(dot_j) = \max(0, dot_j)$$

- Backward pass

- $w'_{ji} = w_{ji} - \alpha \left(\frac{\partial loss}{\partial w_{ji}} \right)$

- $b'_i = b_i - \alpha \left(\frac{\partial loss}{\partial b_i} \right)$



Metode Penelitian

Waktu dan Tempat

Tahapan	Juni	Juli				Agustus
	IV	I	II	III	IV	I
Studi Literatur						
Persiapan Data						
Pembuatan Model						
Analisis Data						
Pembuatan Laporan						

- Kantor Zenius Education Jl. Tebet Barat I No.19, RT.6/RW.2, Tebet Bar., Kec. Tebet, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12810

Metode Penelitian

Desain dan Metode Penelitian

• Data Preparation

	p_recall	timestamp	delta	user_id	learning_language	ui_language	lexeme_id	lexeme_string	history_seen	history_correct	session_seen	session_correct
0	1	1.36E+09	27649635	u:FO	de	en	76390c1350lernt/lerne a8dac3118 n<vblex><p 6187e2fe1e ri><p3><sg 178 >		6	4	2	2
1	0.5	1.36E+09	27649635	u:FO	de	en	7dfd7086f3 671685e2cf die/die<det 1c1da72796><def><f>< d7 sg><nom>		4	4	2	1
2	1	1.36E+09	27649635	u:FO	de	en	35a54c25a 2cda81273 mann/man 43f6a82e6f n<n><m>< 6b7d sg><nom>		5	4	1	1
3	0.5	1.36E+09	27649635	u:FO	de	en	ocf63ffe3d da158bc3d frau/frau<n bd55682b3 ><f><sg>< 55ae nom>		6	5	2	1
4	1	1.36E+09	27649635	u:FO	de	en	84920990d das/das<de 78044db53 t><def><nt c1b012f5bf ><sg><no 9ab5 m>		4	4	1	1

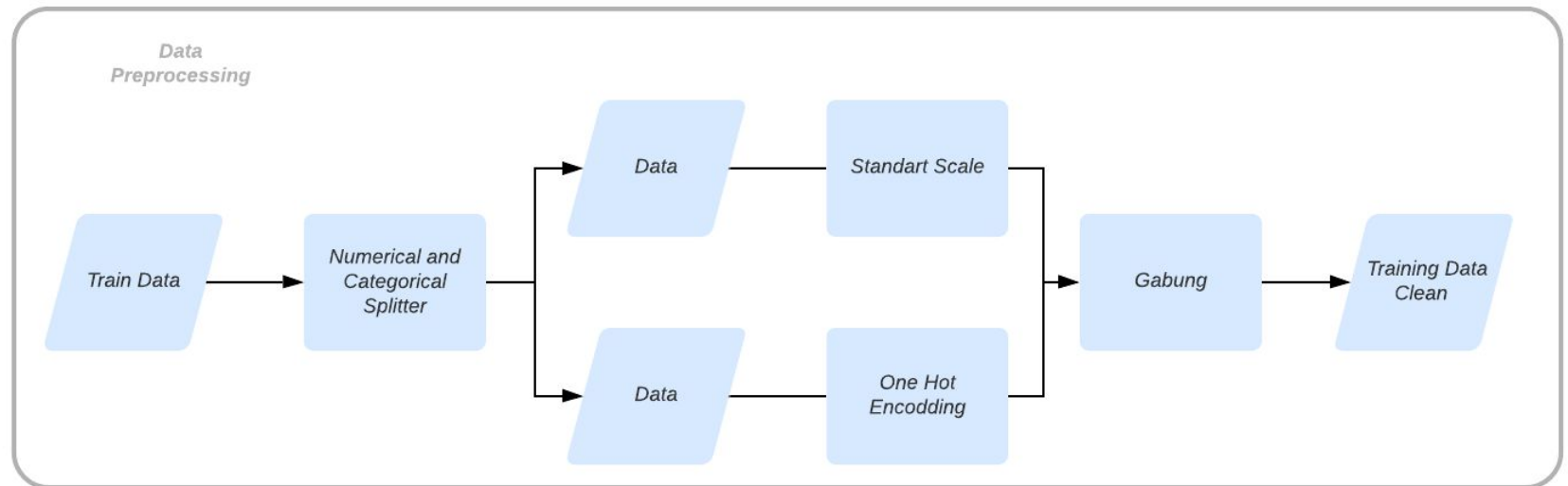
Metode Penelitian

Desain dan Metode Penelitian

- *Data preprocessing*

• *Standar scale*

$$x' = \frac{x - \bar{x}}{\sigma}$$

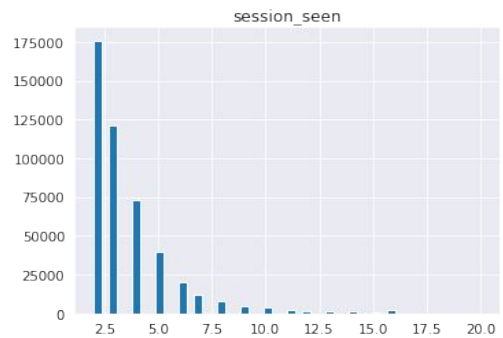
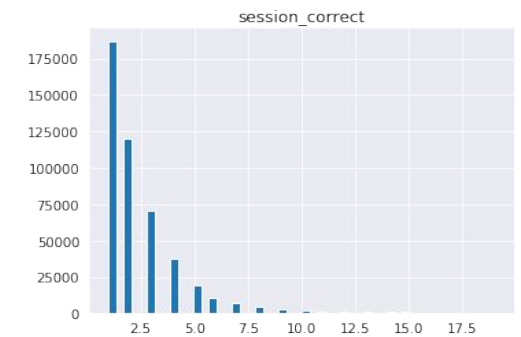
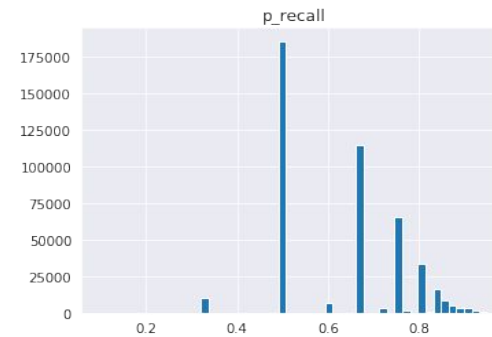
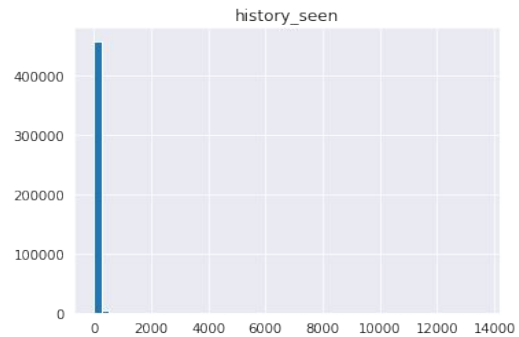
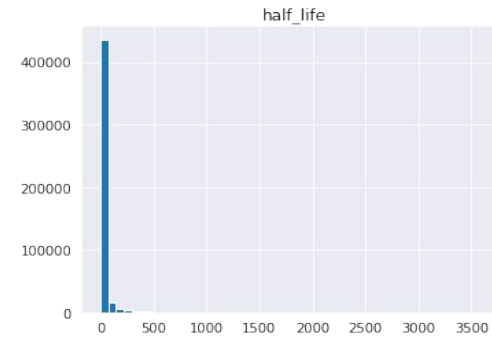
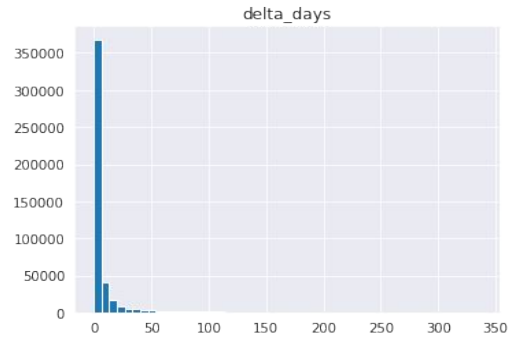


Metode Penelitian Desain dan Metode Penelitian

- Evaluasi
- $MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$
- $RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$
- $RPD = \frac{\sigma}{RMSE}$
- $R^2 = 1 - \frac{\sum_i^n (y_i - \hat{y}_j)^2}{\sum_i^n (y_i - \bar{y}_j)^2}$

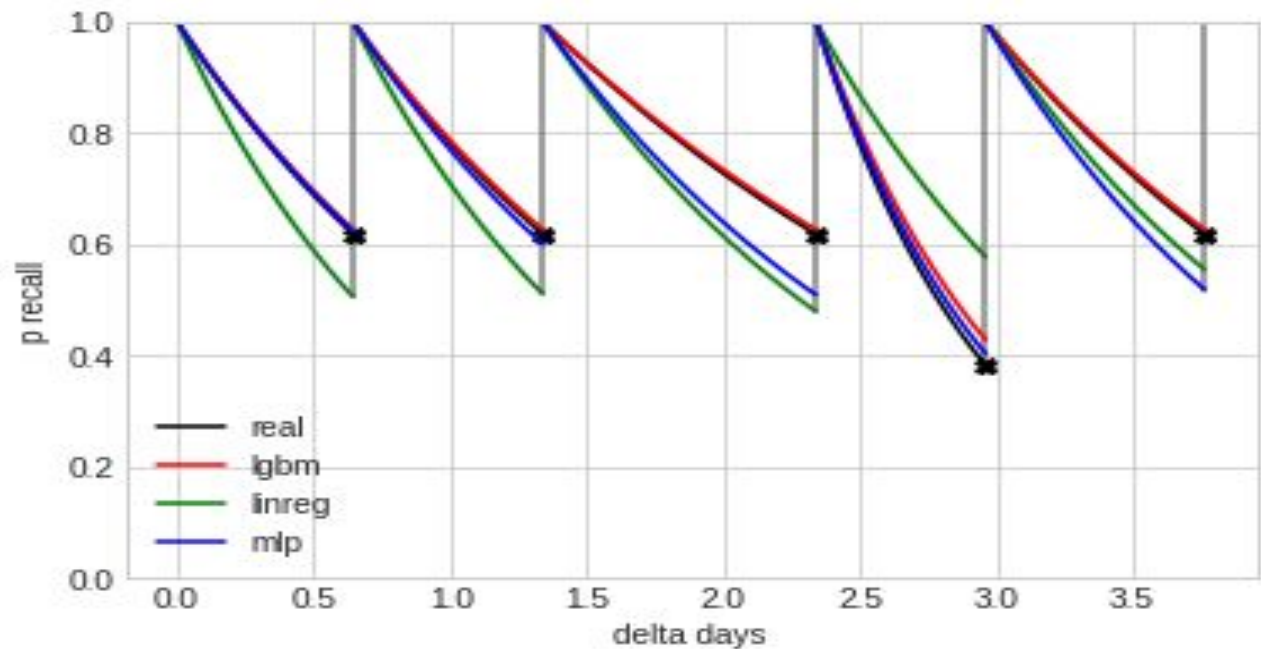
Hasil dan Pembahasan

Data Preparation



Hasil dan Pembahasan Evaluasi

	MAE	RMSE	RPD	R ²
Half-life Regression	17.221797	48.3933	0.011035	-0.054704
Linear Regression	24.215673	94.844316	0.000923	17.221797
LGBM	0.277689	3.705722	24.253665	0.998375
MLP	1.338323	3.662804	24.207490	0.996582



Kesimpulan

Sistem prediksi half-life berbasis *Half-Life Regression*, *Linear Regression*, *Decision Tree*, dan *Neural Network* berhasil dibuat.

Model *Half-Life Regression* memiliki performa sebesar 17.221797 pada MAE, 48.3933 pada RMSE, 0.011035 pada RPD, dan -0.054704 pada R^2 .

Model *Linear Regression* memiliki performa sebesar 24.215673 pada MAE, 94.844316, pada RMSE, 0.000923 pada RPD, dan -0.064572 pada R^2 .

Model LightGBM memiliki performa sebesar 0.277689 pada MAE, 3.705722 pada RMSE, 24.253665 pada RPD, dan 0.998375 pada R^2 .

Model *Multi-Layer Perceptron* memiliki performa sebesar 1.338823 pada MAE, 3.662804 pada RMSE, 24.207490 pada RPD, 0.996582 pada R^2 .

Referensi

B. Settles and B. Meeder. 2016. *A Trainable Spaced Repetition Model for Language Learning*. In *Proceedings of the Association for Computational Linguistics (ACL)*

M. Streeter. 2015. *Mixture Modeling of Individual Learning Curves*

R. Pelanek, J. Rihak and J. Papousek. 2016. *Impact of Data Collection on Interpretation and Evaluation of Student Models*.

R. Palanek. 2015. *Metrics for Evaluation of Student Models*

G. Aurelien. 2018. *Hand-On Machine Learning with Scikit-Learn*

G. Ke, Q. Meng, T. Finley, T. Wang. 2017. *Highly Efficient Gradient Boosting Decision Tree*

Pertanyaan

1. Berapa banyak data dan banyak user yang digunakan pada penelitian kali ini?
 - Ukuran datanya sebesar 12854226 x 12 data
 - Dengan 115222 User
2. Bagaimana cara pengambilan data
 - Data didapatkan dengan menggunakan aplikasi yang sudah dibuat, hasil data dari aplikasi akan disimpan dalam cloud, menyimpan segala bentuk data pengguna aplikasi kapan dia membuka, berapa lama, apa saja yang dibuka dan sebagainya, dari database tersebut perlu dilakukan *query* untuk mengambil data yang diperlukan