



UNIVERSITAS
GADJAH MADA



Algoritma *Data Mining* untuk Optimasi Suhu dan Waktu *Roasting* Nibs Biji Kakao di *Cocoa Teaching Industry* (CTI) UGM

Rizky Alif Ramadhan

19/446785/TK/49890

LOCALLY ROOTED,
GLOBALLY RESPECTED

ugm.ac.id

Daftar Konten

- Pendahuluan
- Tinjauan Pustaka dan Dasar Teori
- Metode Penelitian
- Hasil dan Pembahasan
- Kesimpulan



UNIVERSITAS
GADJAH MADA

Pendahuluan

Latar Belakang



Sumber : WorldAtlas

Nilai Ekspor
US\$1,21 miliar
(IDR 18 Triliun)

Latar Belakang



Luas Kebun Kakao
1.497.467 Ha
(>3x luas Provinsi DIY)

Latar Belakang



UGM CTLI
(*Cocoa Teaching and Learning Industry*)



Cocoa Butter dan Bubuk

Latar Belakang

Suhu/temperatur

Durasi



Roasting/Penyangraian

- Kandar Air /Moist
- Warna
- Aroma

- *Nibs Capacity*
- *Bean Source*
- ...

Latar Belakang

Proses Roasting di UGM CTLI

- Operator **menyetel suhu/temperatur** *roasting*
- Pengecekan kadar air dan pH **secara berkala**
- Jika sudah **sesuai standar** maka proses *roasting* akan **dihentikan**

Optimize!!!

Permasalahan

- Suhu/temperatur yang **terlalu kecil/besar**
- Durasi yang **terlalu singkat/panjang**

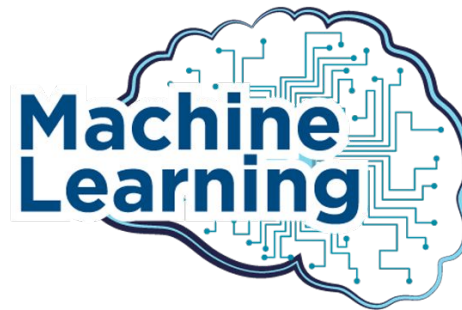


***Under
roasted/over
roasted***

Latar Belakang

Data Mining Algorithm can solve this problem?

- Kadar Air /Moist
- **Durasi**
- *Nibs Capacity*
- *Bean Source*
- ...



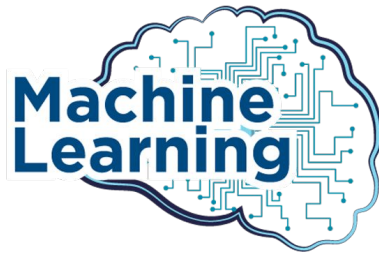
Suhu/temperatur

Algoritma Data Mining/ML

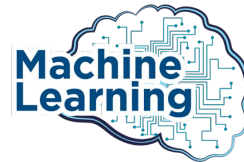
Rumusan Masalah

- Apa saja **variabel** yang mempunyai **tingkat signifikansi** yang **tinggi** terhadap suhu *roasting*?
- Bagaimana **kinerja algoritma *data mining*** dalam merekomendasikan suhu yang optimal berdasarkan durasi roasting tertentu pada proses roasting biji kakao?
- Apa **algoritma *data mining*** yang mempunyai **performa terbaik** dalam merekomendasikan suhu yang optimal berdasarkan durasi roasting tertentu pada proses roasting biji kakao?

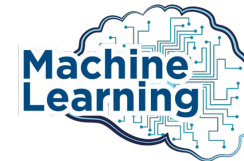
Tujuan Penelitian



Merancang Algoritma



vs



vs



vs



Menentukan Algoritma Terbaik

Manfaat Penelitian



Penelitian dan *Engineering*



Operasional



Industri/Pabrik



UNIVERSITAS
GADJAH MADA

Tinjauan Pustaka dan Dasar Teori

Tinjauan Pustaka

Pelita Perkebunan 2005, 21(3), 169–183

Optimasi Suhu dan Lama Penyangraian Biji Kakao Menggunakan Penyangrai Skala Kecil Tipe Silinder

Optimization of Temperature and Duration of Cocoa Beans Roasting in a Cylindrical Roaster

Misnawi¹⁾, Sri-Mulato¹⁾, Sukrisno Widyotomo¹⁾, Awad Sewet²⁾ and Sugiyono²⁾

Respon : Peroksida lemak, warna, serta sifat organoleptik.

International Food Research Journal 19(4): 1355-1359 (2012)

Journal homepage: <http://www.ifrj.upm.edu.my>



Optimization of cocoa beans roasting process using Response Surface Methodology based on concentration of pyrazine and acrylamide

¹Farah, D.M.H., ¹Zaibunnisa, A.H. and ²Misnawi

¹Faculty of Applied Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

²Indonesia Coffee and Cocoa Research Institute, Jl. PB. Sudirman No. 90 Jember 68100, Indonesia

Respon : Konsentrasi dari pyrazine dan acrylamide

ISSN 0101-2061

Food Science and Technology

DOI: <http://dx.doi.org/10.1590/1678-457X.16416>

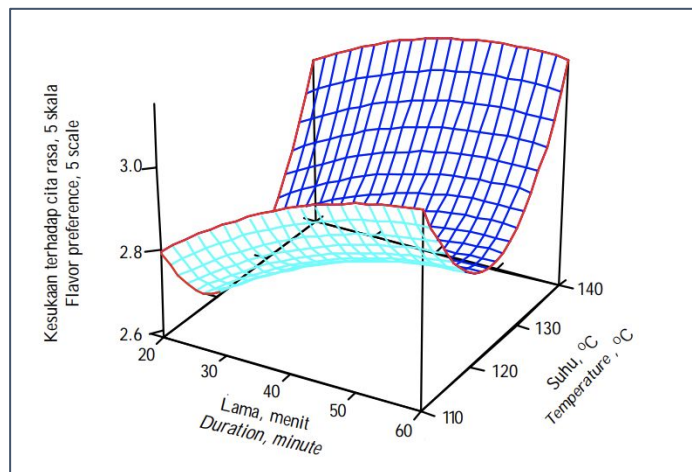
Effect of the roasting temperature and time of cocoa beans on the sensory characteristics and acceptability of chocolate

Ismara Santos ROCHA¹, Ligia Regina Radomille de SANTANA^{2*}, Sérgio Eduardo SOARES³, Eliete da Silva BISPO³

Respon : Penampilan, aroma, rasa, dan tekstur.

Tinjauan Pustaka

Suhu dan durasi optimal menggunakan *Response Surface Methodology*



140°C dan 20 menit

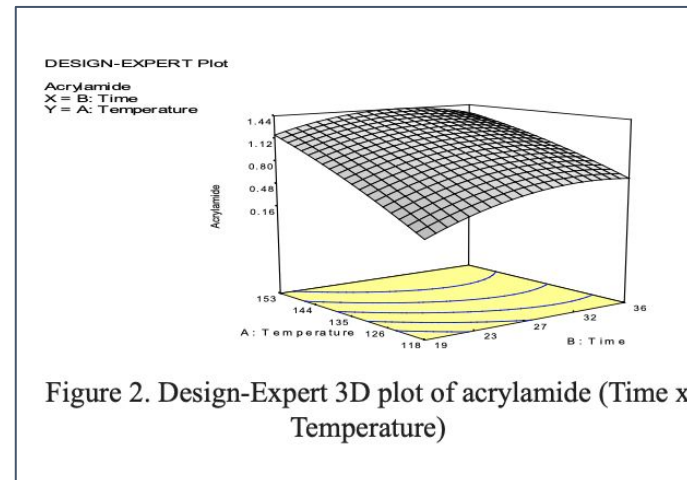
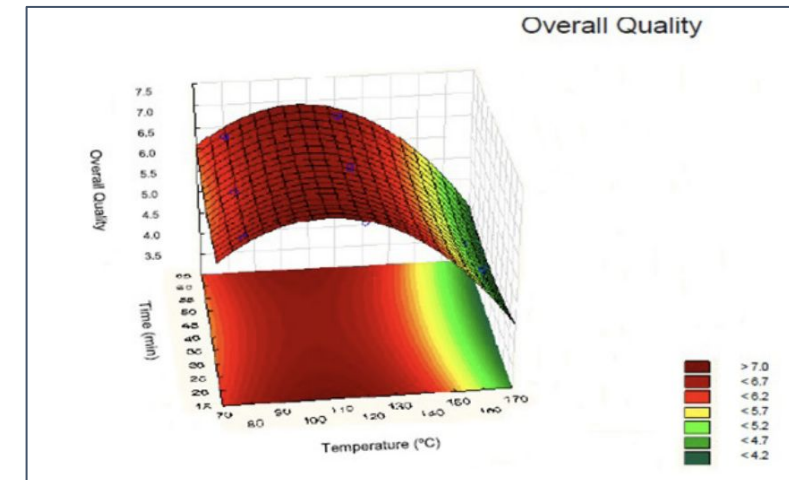


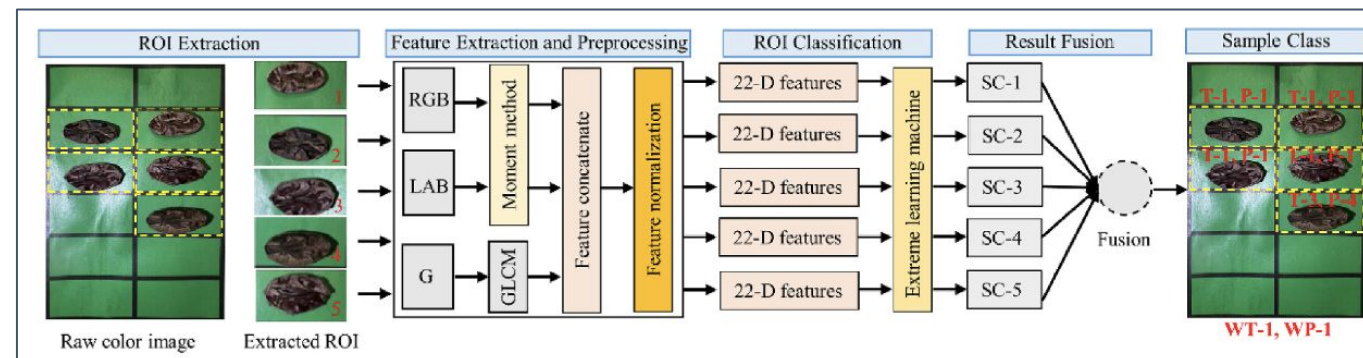
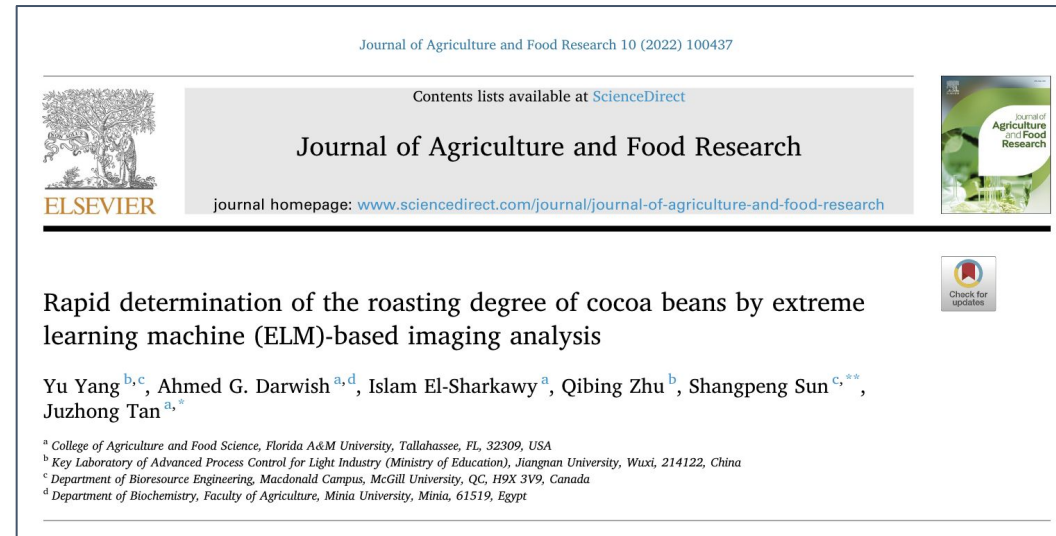
Figure 2. Design-Expert 3D plot of acrylamide (Time x Temperature)

116°C dan 23 menit

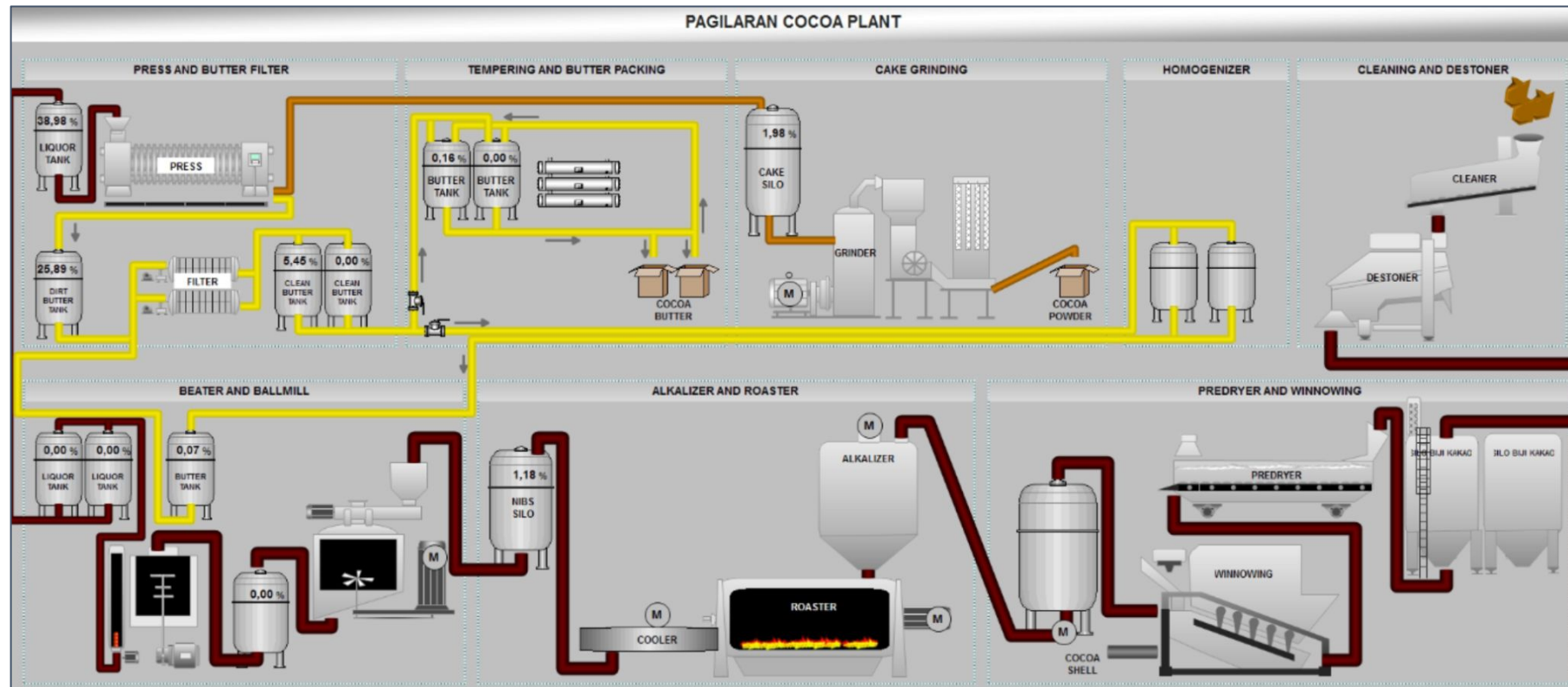


90 sampai 110°C

Tinjauan Pustaka



Dasar Teori



UGM CTLI Process

Dasar Teori



Suhu dan Durasi

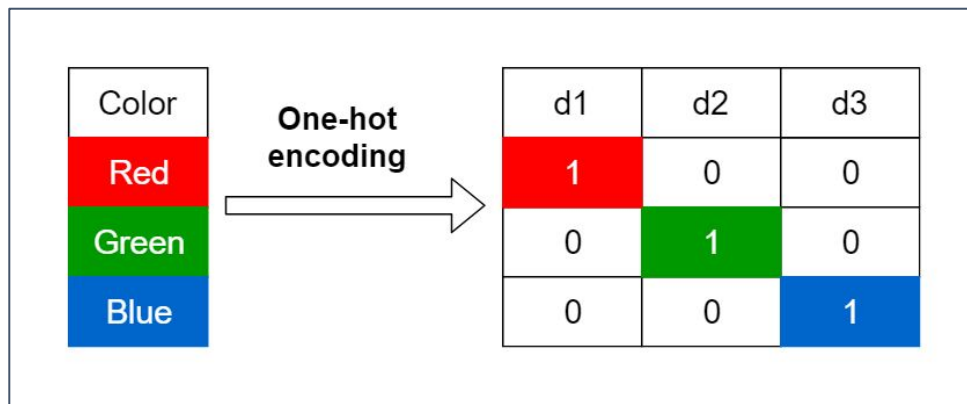


Data Produksi

UGM CTLI Data Collecting

Dasar Teori

Data Preprocessing



One-hot Encoding

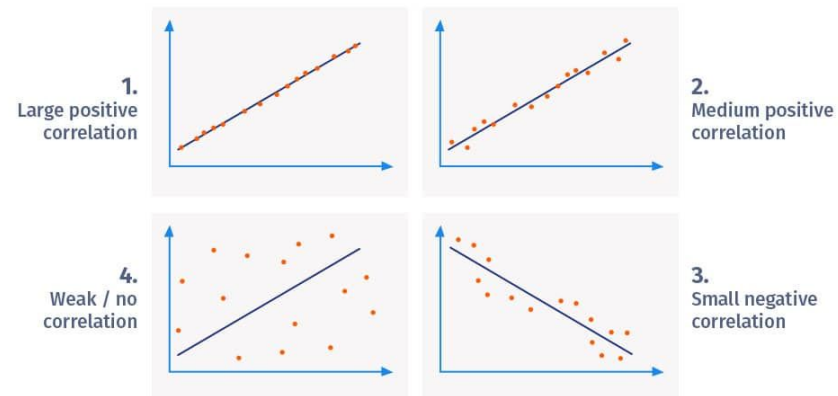
$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Min-max Normalization

Dasar Teori

Seleksi Fitur

Numerik

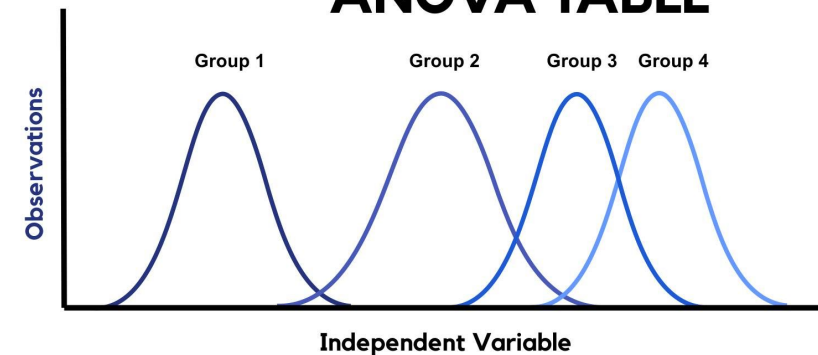


$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Kategorik

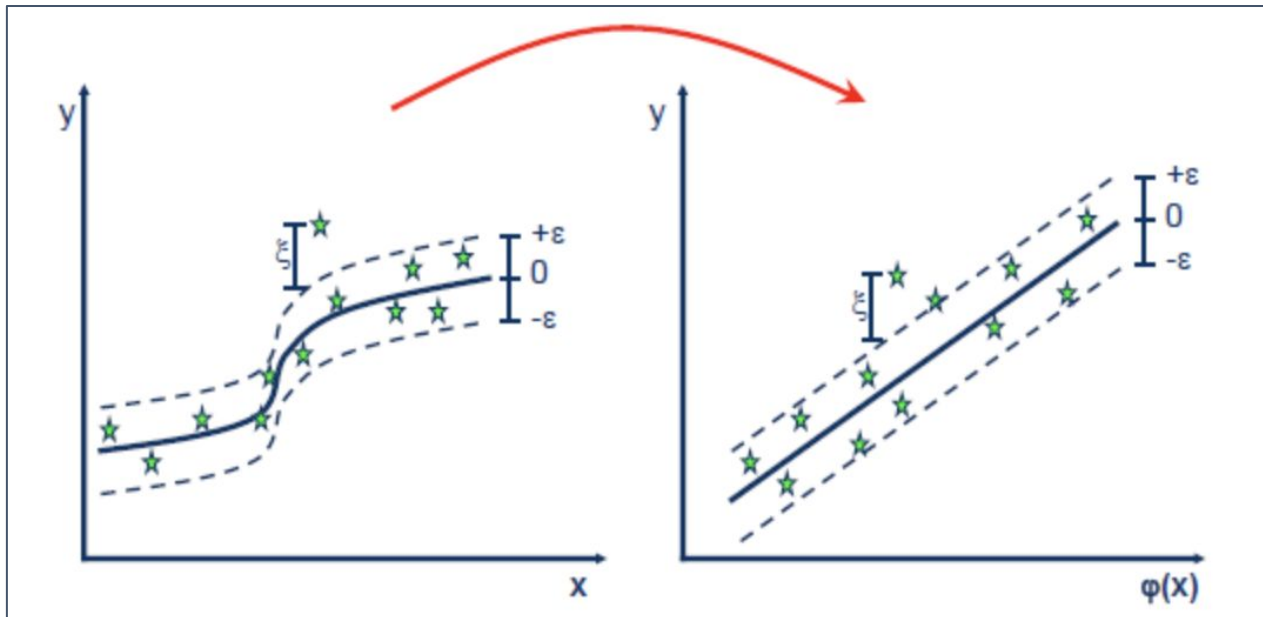
ANOVA TABLE



Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F Value
Between Groups	$SSB = \sum n_j(\bar{X}_j - \bar{X})^2$	$df_1 = k - 1$	$MSB = SSB / (k - 1)$	$f = MSB / MSE$
Error	$SSE = \sum \sum (X - \bar{X}_j)^2$	$df_2 = N - k$	$MSE = SSE / (N - k)$	
Total	$SST = SSB + SSE$	$df_3 = N - 1$		

Dasar Teori

Support Vector Regression (SVR)



$$f(X) = w^T \varphi(X) + B$$

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{N} \sum_{i=1}^N L_E(y_i, f(X_i))$$

$$y_i - w\varphi(X_i) - b \leq E$$

$$w\varphi(X_i) - y_i + b \leq E, \quad i = 1, 2, 3, \dots, N$$

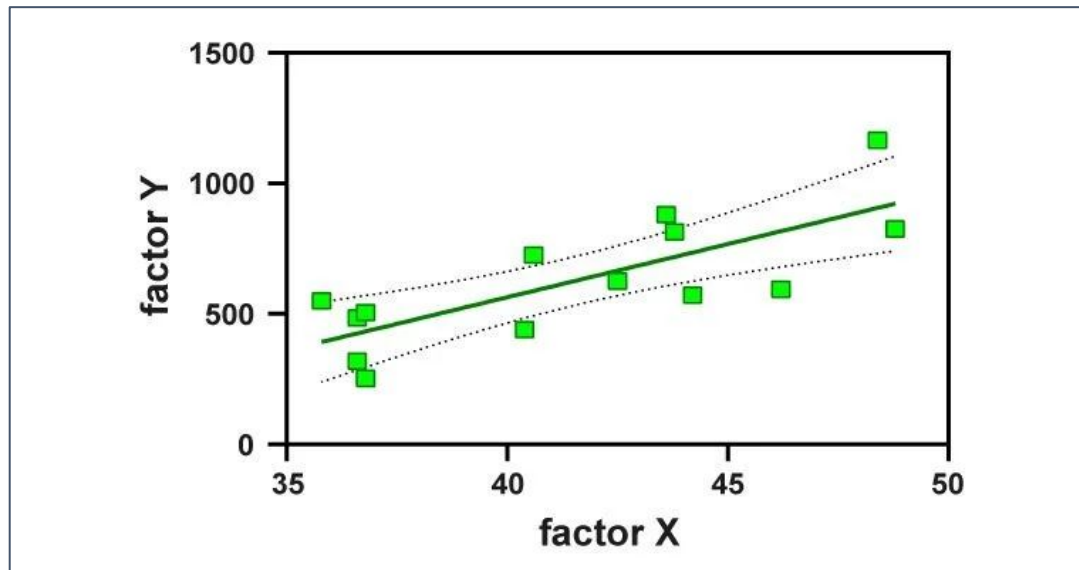
Dimana,

$$L_E(y_i, f(X_i)) = |y_i - f(X_i)| - E|y_i - f(X_i)|$$

- Linear Kernel
 $k(x, y) = x^T y + C$
- Polynomial Kernel
 $k(x, y) = (ax^T y + C)^d$
- Radial Basis Function (RBF) Kernel
 $k(x, y) = \exp(-\gamma \|x - y\|^2)^d$

Dasar Teori

Multiple Linear Regression (MLR)



$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

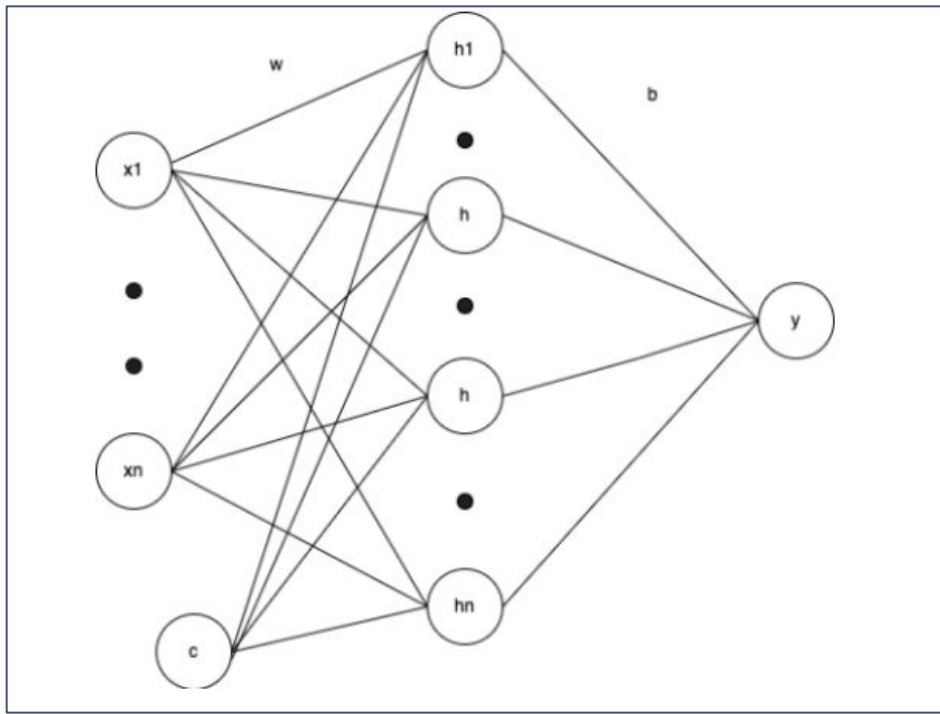
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$Y = X\beta + \varepsilon$

$$b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} = (X'X)^{-1}X'Y$$

Dasar Teori

Extreme Learning Machine (ELM)



$$y_j = \sum_{i=1}^H b_i f(w_i x_j + c_i), \quad j = 1, 2, \dots, N$$

$$Y = FB$$

$$\hat{B} = F^+ Y$$

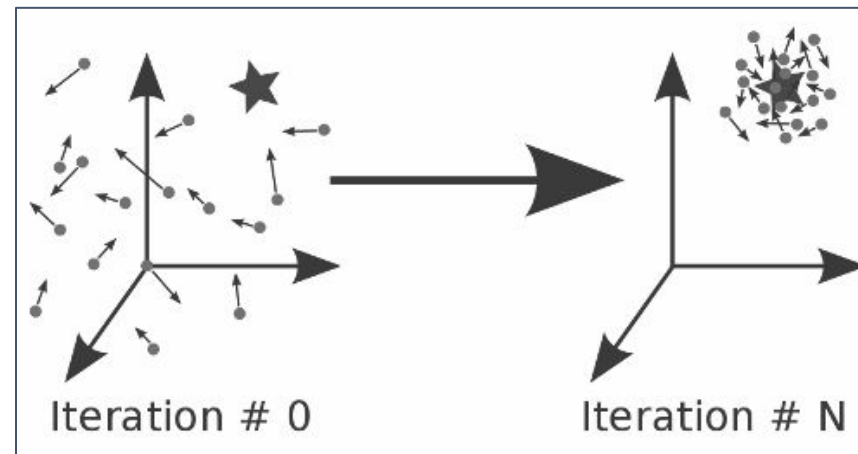
$$F^+ = (F^T F)^{-1} F^T$$

Dasar Teori

Particle Swarm Optimization (PSO)

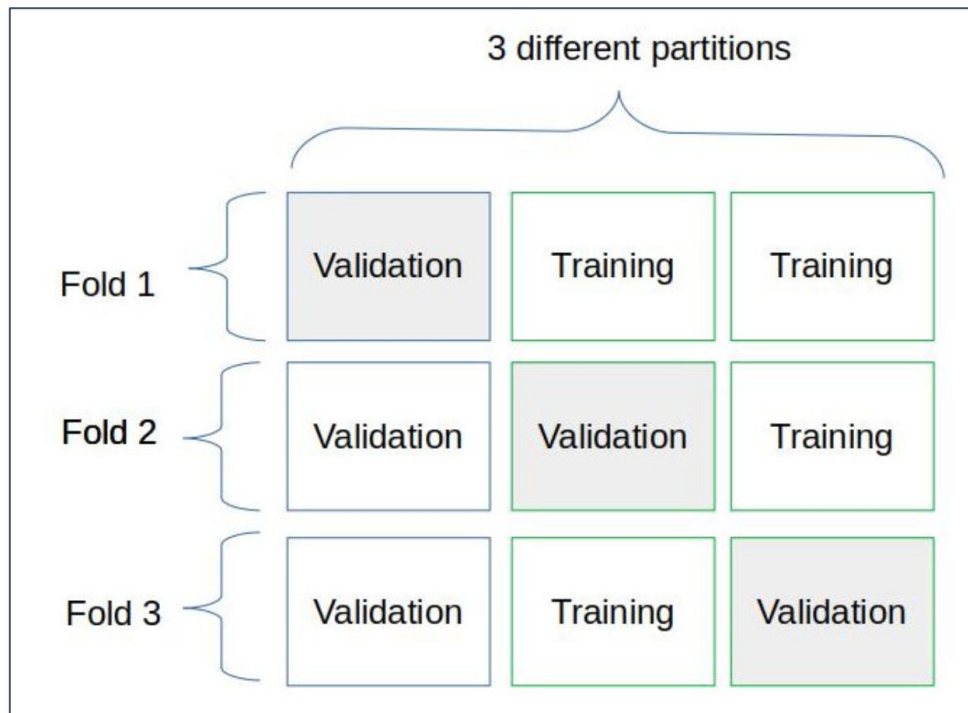
$$v_k^i(t + i) = wv_k^i(t) + c_1rand() \left(p_k^i(t) - x_k^i(t) \right) + c_2rand() \left(g_k^i(t) - x_k^i(t) \right)$$

$$x_k^i(t + i) = x_k^i(t) + v_k^i(t + i)$$



Dasar Teori

Evaluation Metrics



$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{A_t - F_t}{A_t} \right|$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (A_i - F_i)^2}$$

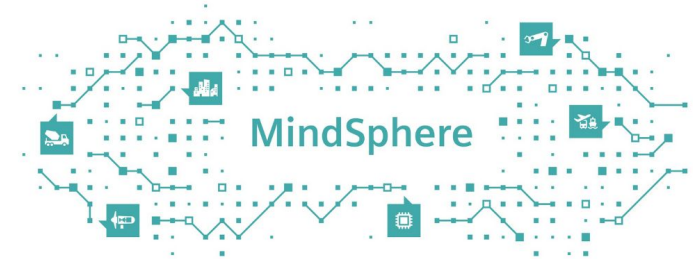


UNIVERSITAS
GADJAH MADA

Metode Penelitian

Dataset

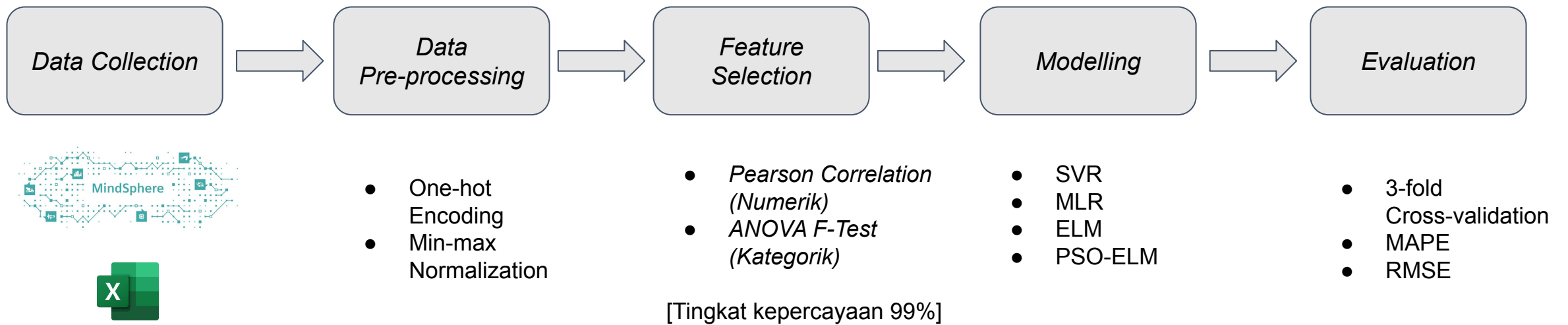
Kolom	Keterangan
nibs_capacity	kapasitas nibs (kg)
solution_load	jumlah air yang dibutuhkan (L)
beans_source	wilayah kebun dari biji tersebut
is_alkalized	apakah proses tersebut mengalami alkilasi?
product_type	tipe produk
durasi_roasting	durasi <i>roasting</i>
suhu	suhu pengaturan
pH_0	pH awal biji kakao
pH_N	pH akhir biji kakao
moist_0	kadar air (%) awal biji kakao
moist_N	kadar air (%) akhir biji kakao



April 2021 sampai Juli 2022

43
Baris Data

Alur Penelitian





UNIVERSITAS
GADJAH MADA

Hasil dan Pembahasan

Seleksi Fitur

Variabel	r_suhu	pvalue_suhu
nibs_capacity	-0.5089	0.0009
solution_load	-0.2962	0.0671
pH_0	0.0392	0.8125
pH_N	0.1251	0.4480
delta_pH	-0.1771	0.2807
moist_0	-0.0248	0.8809
moist_N	-0.5860	0.0001
delta_moist	0.2155	0.1876

Variabel	fscore_suhu	pvalue_suhu
beans_source	29.522289	0.000004
product_type	0.512261	0.478655
is_alkalized	6.490727	0.015131

Uji Hipotesis (Tingkat Signifikansi 99%)

- H0 : variabel bebas x tidak memiliki hubungan dengan variabel target y (suhu)
- H1 : variabel bebas x memiliki hubungan dengan variabel target y (suhu)

Variabel yang mempunyai **tingkat signifikansi tinggi** dan **dipilih dalam pemodelan** pada penelitian ini adalah **nibs_capacity**, **moist_N**, dan **bean_source**.

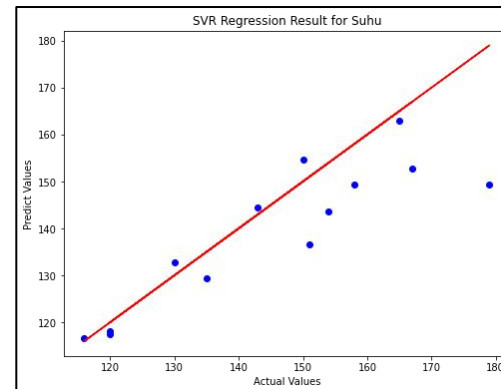
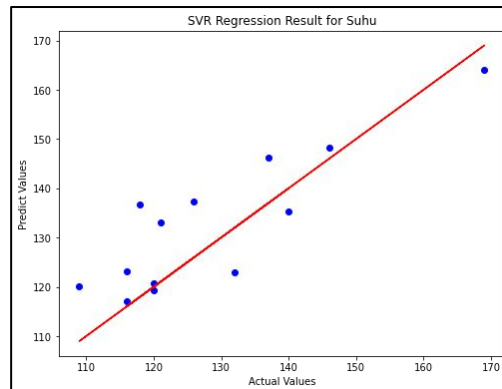
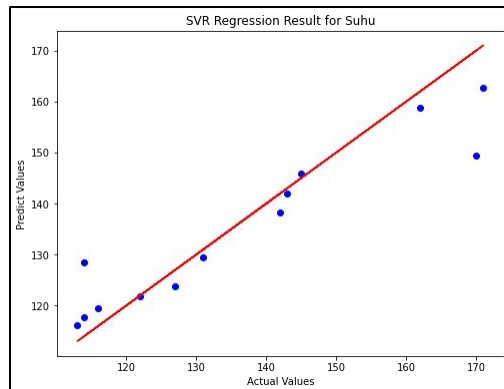
Support Vector Regression (SVR)

Hyperparameter Tuning

```
parameter = {'C': [0.1, 0.001, 1, 10, 12, 14, 16, 18, 20, 22],
             'gamma': [0.001, 0.01, 0.1, 1, 2, 5],
             'epsilon': [0.001, 0.01, 0.1, 1, 2, 4],
             'kernel': ("rbf", "poly", "linear")}
```



- $C = 22$
- $\epsilon = 1$
- $\gamma = 1$
- $\text{kernel} = \text{rbf}$

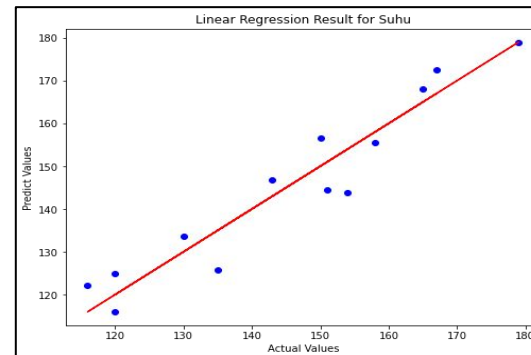
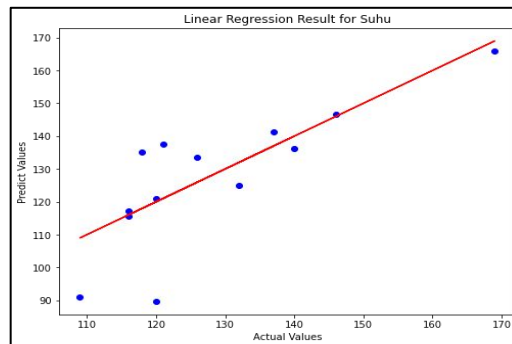
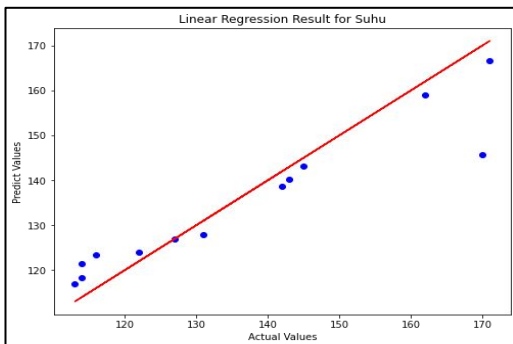


SVR		
Ukuran	MAPE	RMSE
Fold 1	3.71	7.72
Fold 2	5.73	8.85
Fold 3	4.84	10.94
Rata-rata	4.76	9.17
Std Dev	1.01	1.63

Multiple Linear Regression (MLR)

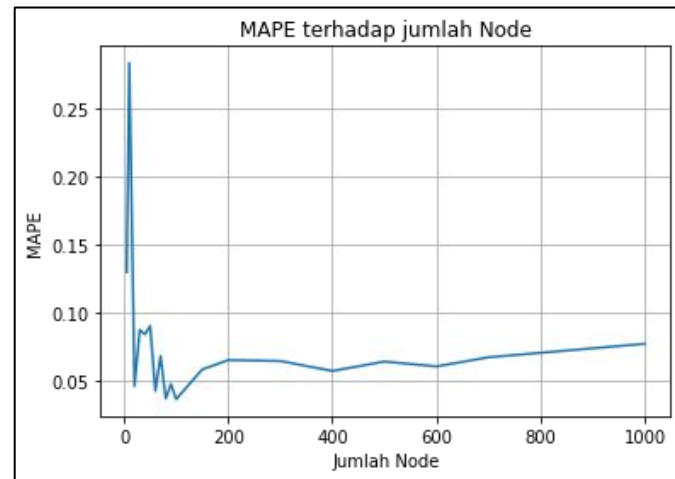
$$suhu = 153.84 - 6.36 \cdot nibs_capacity - 15.99 \cdot moist_N - 16.97 \cdot durasi_roasting + categoric$$

<i>ADACHI</i>	-18.26
<i>BONDO</i>	0.10
<i>GOBEL</i>	-19.46
<i>HMGN</i>	-17.18
<i>PAPUA</i>	13.08
<i>PGL</i>	2.34
<i>SGU</i>	21.89
<i>SIKLON</i>	36.42
<i>UNIDKOREA</i>	7.23

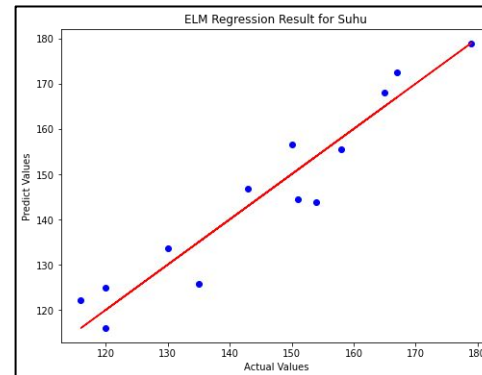
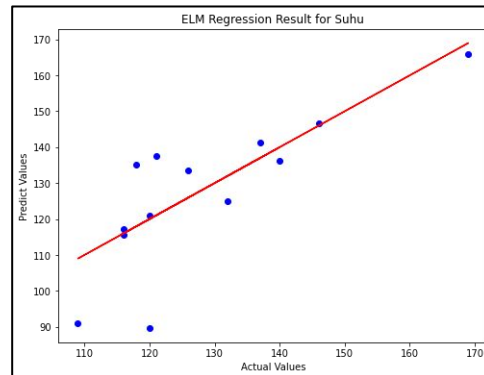
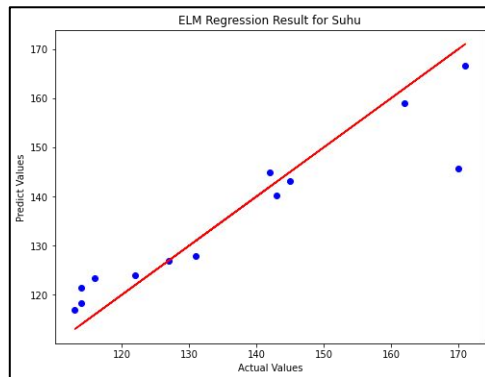


MLR		
Ukuran	MAPE	RMSE
Fold 1	3.72	7.83
Fold 2	7.06	12.32
Fold 3	3.62	5.73
Rata-rata	4.8	8.63
Std Dev	1.96	3.37

Extreme Learning Machine (ELM)

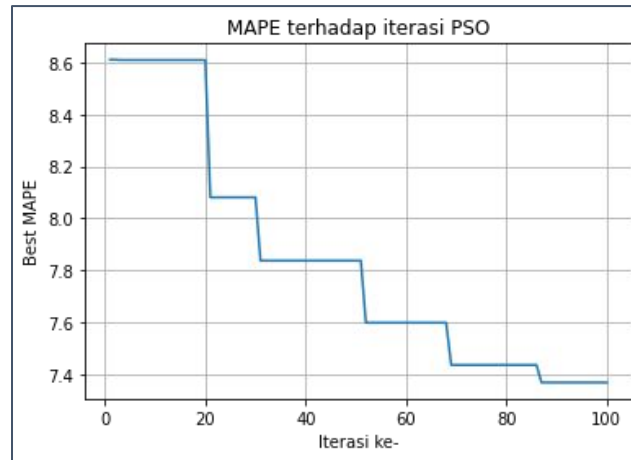


Best N Node
100

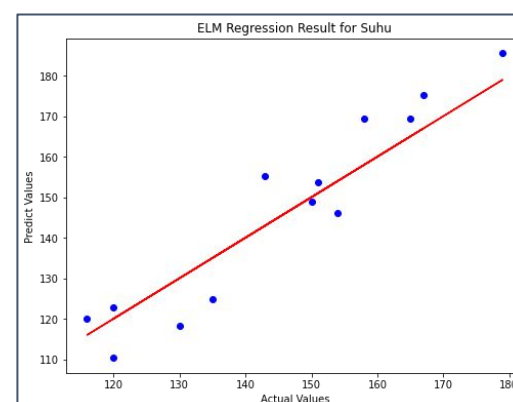
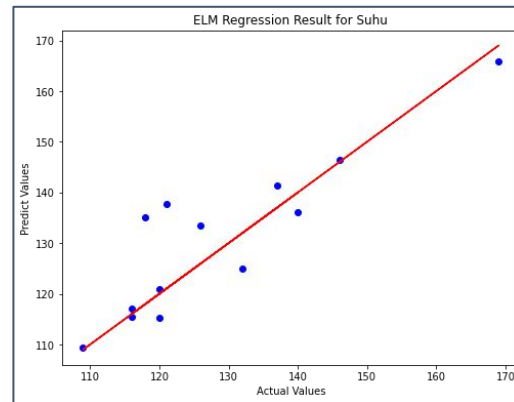
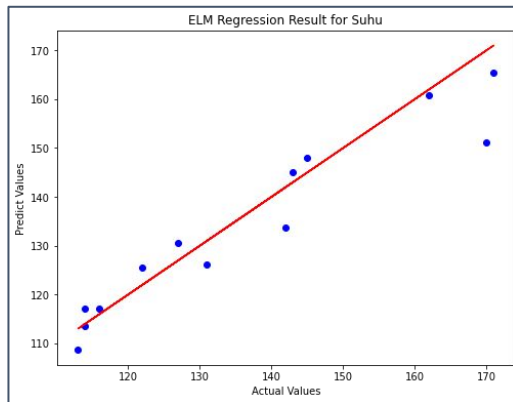


ELM		
Ukuran	MAPE	RMSE
Fold 1	3.69	7.81
Fold 2	7.06	12.32
Fold 3	3.62	5.73
Rata-rata	4.79	8.62
Std Dev	1.97	3.37

PSO-ELM



PSO was
Optimizing
**RMSE to
7.37**



PSO-ELM		
Ukuran	MAPE	RMSE
Fold 1	3.21	6.48
Fold 2	4.19	7.59
Fold 3	5.01	8.03
Rata-rata	4.14	7.37
Std Dev	0.90	0.80

Perbandingan Algoritma

Model	MAPE	RMSE
SVR	4.76	9.17
MLR	4.81	8.63
ELM	4.80	8.62
PSO-ELM	4.14	7.37

PSO-ELM adalah model yang memiliki **performa terbaik** dengan skor **MAPE 4.14%** dan **RMSE 7.37** lebih baik dibandingkan SVR dan MLR. Penambahan **PSO** untuk inisialisasi bobot awal w dan bias dapat **menurunkan nilai RMSE sebesar 14.5%** dari 8.62 menjadi 7.37.



UNIVERSITAS
GADJAH MADA

Kesimpulan dan Saran

Kesimpulan

- Variabel yang mempunyai **tingkat signifikansi** yang **tinggi** dengan variabel suhu adalah **nibs_capacity, moist_N, dan bean_source**.
- **Algoritma data mining terbaik** adalah **PSO-ELM** dengan skor **MAPE** sebesar **4.14%** dan **RMSE** sebesar **7.37**, kemudian **ELM** dengan skor **MAPE** sebesar **4.79%** dan **RMSE** sebesar **8.62**, **MLR** dengan skor **MAPE** sebesar **4.80%** dan **RMSE** sebesar **8.62**, dan **SVR** dengan skor **MAPE** sebesar **4.76%** dan **RMSE** sebesar **9.17**.
- **Algoritma data mining** dikatakan **layak** dengan predikat ***Highly Accurate Forecasting*** karena mempunyai nilai **MAPE < 10%**



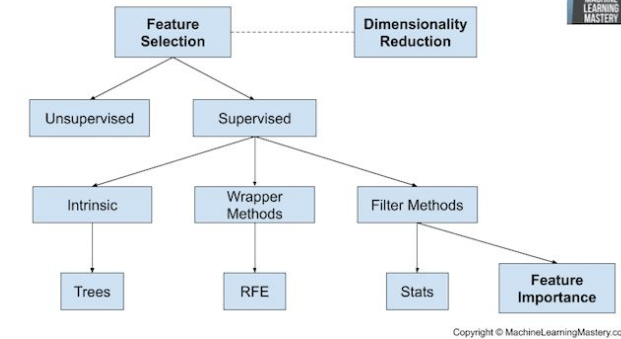
PSO-ELM

Saran

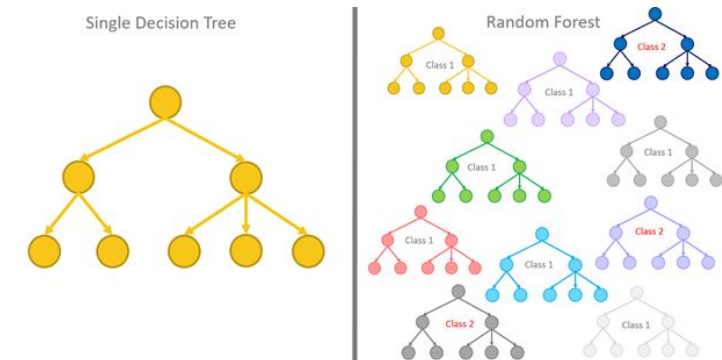


Integrating the Data

Overview of Feature Selection Techniques



Using another Feature Selection Method



Using Tree-based Model



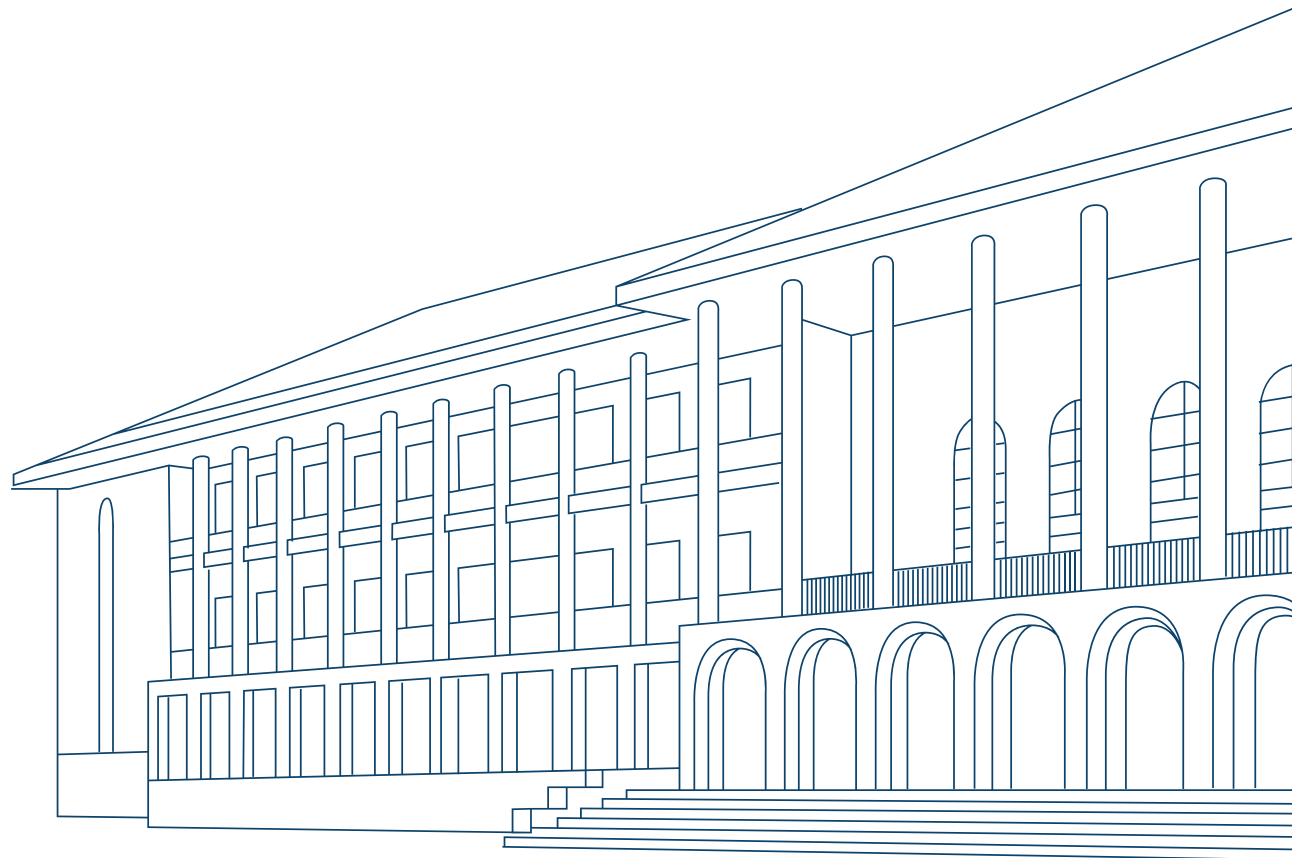
UNIVERSITAS
GADJAH MADA

Terima Kasih



UNIVERSITAS
GADJAH MADA

“Ini adalah text untuk
kutipan
atau qoutes yang
menggunakan
font gama-sans light italic”





UNIVERSITAS
GADJAH MADA



LOCALLY ROOTED,
GLOBALLY RESPECTED

Ini Adalah Judul Presentasi Title Case Dengan Menggunakan Font Gama-sans Bold

Sedangkan ini adalah space untuk
sub-judul
sebagai pelengkap judul utama di atas,
Sentence Case dengan menggunakan
font Gama-Sans regular

ugm.ac.id



UNIVERSITAS
GADJAH MADA



ugm.ac.id

LOCALLY ROOTED, GLOBALLY RESPECTED



UNIVERSITAS
GADJAH MADA

“Ini adalah text untuk
kutipan
atau qoutes yang
menggunakan
font gama-sans light italic”

