

# Regression Model Selection Analysis of Methanol Conversion Based on Temperature, Residence Time, Concentration, Oxygen Ratio, and Reactor System

Andre Marvero, Fahmi Amri, Muhammad Fadhil Irsyad, Yenni Kurniawati\*

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

\*Corresponding author: yennikurniawati@fmipa.unp.ac.id

Submitted : 08 Februari 2025

Revised : 12 Februari 2025

Accepted : 28 Februari 2025

## ABSTRACT

*This study aims to determine the best regression model that explains the effect of temperature, residence time, methanol concentration, oxygen to methanol ratio, and reactor system on methanol conversion in supercritical water. Preliminary analysis showed a violation of the multicollinearity assumption, which affected the validity of the multiple linear regression model. To overcome this and determine the optimal model, variable selection was performed using the stepwise selection method. This method was evaluated based on predictive power, model accuracy and statistical validity. The results showed that the stepwise method produced an optimal model in predicting conversion. Reactor system and temperature were the most significant variables affecting methanol conversion. The conclusion of this study shows that the variable selection approach with stepwise selection can be effectively used to identify the best regression model, when classical assumptions are met. These findings make an important contribution to the optimization of supercritical water-based chemical processes.*

**Keywords:** *Methanol Conversion, Supercritical Water, Multicollinearity, Variable Selection, Stepwise Regression*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

## I. PENDAHULUAN

Konversi metanol dalam air superkritis adalah proses inovatif yang menarik perhatian untuk aplikasi industri, terutama dalam produksi bahan bakar dan bahan kimia berkelanjutan. Pada kondisi superkritis, air berfungsi sebagai pelarut efisien dengan kepadatan tinggi, viskositas rendah, dan kemampuan solvatisasi yang baik, memungkinkan reaksi kimia berlangsung lebih cepat dan selektif. Keunggulan ini menjadikannya pilihan ideal untuk proses industri ramah lingkungan yang mendukung prinsip kimia hijau dengan fokus pada pengurangan limbah dan efisiensi energi (Marcus, 2018). Proses ini dipengaruhi oleh beberapa variabel operasional, termasuk suhu, waktu tinggal, konsentrasi metanol, rasio oksigen, dan jenis sistem reaktor yang digunakan. Pemahaman yang mendalam tentang pengaruh masing-masing variabel terhadap konversi metanol sangat penting untuk mengoptimalkan proses ini.

Namun, dalam analisis regresi yang digunakan untuk memodelkan hubungan antara variabel-variabel ini dan konversi metanol, sering kali ditemukan pelanggaran terhadap asumsi klasik dalam regresi linier, khususnya multikolinearitas. Sebuah model regresi dianggap valid jika memenuhi asumsi klasik, yaitu normalitas, homoskedastisitas, tidak adanya autokorelasi, dan tidak adanya multikolinearitas (Aditiya et al., 2023). Akan tetapi, dalam banyak kasus dengan lebih dari satu variabel independen, asumsi-asumsi ini sulit dipenuhi, dan pelanggaran terhadap multikolinearitas dapat mengganggu hubungan antara variabel independen terhadap variabel dependen (Setya Budi et al., 2024).

Multikolinearitas adalah kondisi di mana terdapat hubungan linear di antara variabel-variabel independen (X) dalam sebuah model regresi (Debataraja et al., 2019). Jika terjadi multikolinieritas, masalah ini dapat diatasi dengan beberapa langkah, seperti menganalisis hubungan antara variabel independen, meningkatkan jumlah data, mengkombinasikan data time series dan cross-section, melakukan transformasi pada variabel, atau menghapus variabel yang diidentifikasi sebagai penyebab multikolinieritas (Nurchaya et al., 2023). Dalam kasus ini, untuk mendapatkan model yang lebih baik dan lebih dapat diandalkan, diperlukan teknik seleksi variabel yang efektif. Oleh karena itu, pendekatan seperti stepwise regression dapat digunakan untuk memilih variabel-variabel yang paling signifikan dan membangun model regresi yang optimal. Teknik ini memungkinkan peneliti untuk mengurangi jumlah variabel yang

tidak relevan dan memfokuskan model pada variabel-variabel yang memiliki pengaruh paling besar terhadap konversi metanol.

Penggunaan seleksi variabel dalam regresi tidak hanya bertujuan untuk meningkatkan akurasi model, tetapi juga untuk meningkatkan interpretabilitas dan efisiensi analisis. Beberapa penelitian yang berkaitan dengan pemilihan model terbaik dengan masalah multikolinieritas telah banyak berkembang seperti Pemilihan model regresi terbaik dalam kasus pengaruh premi, klaim, hasil investasi dan hasil underwriting terhadap laba asuransi jiwa (Fariha et al., 2018), Penanganan masalah multikolinieritas pada pemodelan pertumbuhan ekonomi Indonesia berdasarkan teori pertumbuhan ekonomi endogenous (Yanke et al., 2022). Dengan mengaplikasikan teknik seleksi variabel ini, diharapkan dapat ditemukan model regresi yang paling efisien dan akurat untuk menggambarkan konversi metanol dalam kondisi superkritis, yang dapat memberikan kontribusi signifikan dalam pengembangan proses kimia yang lebih berkelanjutan dan efisien.

Penelitian ini akan menganalisis pengaruh variabel-variabel tersebut terhadap konversi metanol, mengidentifikasi pelanggaran asumsi multikolinieritas, dan menentukan hasil model regresi yang diperoleh melalui teknik stepwise selection untuk menemukan model yang paling sesuai. Diharapkan, temuan ini akan memberikan wawasan yang berguna dalam merancang sistem konversi metanol yang lebih optimal dan efisien.

## II. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis regresi linier berganda. Penelitian dilakukan dalam beberapa tahap, meliputi pengumpulan data, mengatasi masalah multikolinieritas, uji asumsi klasik, dan pemodelan regresi menggunakan teknik seleksi variabel (*stepwise regression*). Langkah ini bertujuan untuk mendapatkan model regresi terbaik yang dapat menjelaskan hubungan antara variabel-variabel independen terhadap variabel dependen.

Dalam analisis regresi, salah satu indikator utama yang digunakan untuk menilai kualitas model adalah koefisien determinasi berganda ( $R^2$ ). Koefisien ini menunjukkan proporsi varians dalam variabel dependen yang dapat dijelaskan oleh variabel independen dalam model. Semakin tinggi nilai  $R^2$ , semakin baik model dalam menjelaskan perubahan pada variabel dependen. Nilai  $R^2$  dihitung menggunakan rumus: (Amalita & Kurniawati, 2013)

$$R^2 = \frac{JKR}{JKT}$$

Penelitian ini menggunakan data sekunder dari penelitian Tester et.al (1993), mengenai kinetika oksidasi metanol dalam air superkritis, yang menjadi dasar analisis lebih lanjut dalam penelitian ini. Variabel independen yang dianalisis meliputi suhu, waktu tinggal, konsentrasi metanol, rasio oksigen, dan jenis sistem reaktor, dengan konversi metanol sebagai variabel dependennya.

Dalam penelitian ini, langkah-langkah yang diambil meliputi pencarian model dugaan dengan menggunakan analisis regresi berganda melalui Program R dan Minitab, mengatasi masalah multikolinieritas yang muncul, melakukan uji asumsi klasik termasuk normalitas, homoskedastisitas, autokorelasi, linieritas, dan multikolinieritas, dan mengatasi masalah multikolinieritas yang muncul. Pengujian normalitas dilakukan untuk memastikan bahwa distribusi data mendekati distribusi normal, yang diuji menggunakan metode statistik seperti Uji *Shapiro-Wilk*. Homoskedastisitas diuji untuk melihat apakah varians kesalahan konstan pada seluruh rentang data, yang dapat diuji menggunakan Uji *Breusch-Pagan*. Selanjutnya, uji autokorelasi dilakukan untuk mendeteksi adanya korelasi antar residual dalam model menggunakan Uji *Durbin-Watson*. Linearitas hubungan antara variabel independen dan dependen diuji dengan *RESET test*, sedangkan multikolinieritas yang menunjukkan adanya hubungan kuat antar variabel independen dapat diidentifikasi dengan melihat nilai *Variance Inflation Factor* (VIF).

## III. HASIL DAN PEMBAHASAN

Dalam penelitian ini, langkah pertama yang dilakukan adalah menemukan persamaan regresi dugaan pengaruh suhu, waktu tinggal, konsentrasi metanol, rasio oksigen, dan system reaktor terhadap konversi metanol yang didapatkan dari perhitungan melalui Program R adalah

$$\hat{Y} = -2669.172 + 22.261X_1 + 3.881X_2 + 101.668X_3 + 0.809X_4 - 1.633X_5$$

dengan  $X_1$  = sistem reaktor,  $X_2$  = suhu,  $X_3$  = waktu tinggal,  $X_4$  = konsentrasi metanol, dan  $X_5$  = rasio oksigen.

Setelah didapatkan hasil linier dugaannya, selanjutnya melakukan uji asumsi klasik.

### 1. Uji Normalitas

Model regresi dianggap baik jika data yang digunakan memiliki distribusi normal atau mendekati normal. Untuk mendeteksi hal ini, dapat dilakukan Uji *Shapiro Wilk*.

**Tabel 1.** Hasil Uji Shapiro Wilk

Nilai uji W	P-value
0.97074	0.8114

Berdasarkan hasil perhitungan dengan tingkat signifikansi 5%, diperoleh nilai  $p$ -value sebesar 0.8114, yang menunjukkan bahwa data tersebut berdistribusi normal.

2. Uji Homoskedastisitas

Salah satu kriteria untuk memastikan bahwa model regresi berkualitas baik adalah bebas dari masalah heteroskedastisitas. Untuk mendeteksi adanya heteroskedastisitas, Uji *Breusch-Pagan* dapat digunakan.

**Tabel 2.** Hasil Uji Breush-Pagan

Nilai uji BP	P-value
0.72832	0.9814

Hasil perhitungan menunjukkan bahwa pada tingkat signifikansi 5%, nilai  $p$ -value yang diperoleh adalah 0.9814, yang mengindikasikan bahwa data tidak mengalami masalah heteroskedastisitas.

3. Uji Autokorelasi

Pendeteksian autokorelasi dilakukan melalui Uji *Durbin-Watson*. Jika tidak ada korelasi antara residu, maka residu tersebut dianggap acak dan tidak berkorelasi. Model regresi yang baik seharusnya bebas dari autokorelasi.

**Tabel 3.** Hasil Uji Durbin-Watson

Nilai uji DW	P-value
2.377	0.5048

Hasil perhitungan menunjukkan bahwa pada tingkat signifikansi 5%, nilai  $p$ -value yang diperoleh adalah 0.5048, yang mengindikasikan bahwa data tidak mengalami autokorelasi.

4. Uji Linieritas

Pendeteksian linieritas dilakukan dengan menggunakan Uji *RESET test*. Jika hubungan antara variabel independen dan dependen bersifat linier, maka model regresi dapat dianggap valid.

**Tabel 4.** Hasil Uji RESET

Nilai uji RESET	P-value
4.0142	0.05251

Berdasarkan hasil perhitungan, diperoleh nilai  $p$ -value sebesar 0.05251. Hal ini menunjukkan bahwa terdapat indikasi linieritas dalam model yang diuji.

5. Uji Multikolinieritas

Multikolinieritas dideteksi menggunakan nilai VIF yang dapat di lihat pada Tabel 1.

**Tabel 5.** Nilai VIF masing-masing variabel

X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>
1.519064	26.283999	26.447032	2.202201	1.922689

Terlihat pada Tabel 5, nilai VIF X<sub>2</sub> dan X<sub>3</sub> lebih dari 10 yang menunjukkan ada masalah multikolinieritas. Jika terjadi multikolinieritas dapat ditangani dengan beberapa cara, antara lain bisa dengan tranformasi data atau mengeluarkan variabel yang memiliki korelasi yang tinggi (Nurchaya et al., 2023). Nilai VIF setelah dilakukan tranformasi data terhadap variabel X<sub>3</sub> ditunjukkan dalam Tabel 6.

**Tabel 6.** Nilai VIF setelah transformasi data

X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>
1.904192	40.279306	42.460819	2.047210	1.642512

Dari Tabel 6. terlihat bahwa masalah multikolinieritas masih belum teratasi diantara X2 dan X3 bahkan membuat korelasinya meningkat. Selanjutnya, dengan menghapus X3 karena memiliki korelasi yang sangat tinggi.

**Tabel 7.** Nilai VIF setelah menghapus X3

X <sub>1</sub>	X <sub>2</sub>	X <sub>4</sub>	X <sub>5</sub>
1.439308	1.173061	2.047208	1.617418

Hasil nilai VIF dari menghapus variabel X3, menunjukkan tidak ada pelanggaran multikolinieritas yang terjadi. Sehingga diperoleh persamaan regresi baru dugaan pengaruh suhu, konsentrasi metanol, rasio oksigen, dan system reaktor terhadap konversi metanol.

$$\hat{Y} = -639 + 28.7X_1 + 1.281X_2 + 4.89X_4 + 10.2X_5$$

dengan X<sub>1</sub> = sintem reaktor, X<sub>2</sub> = suhu, X<sub>4</sub> = konsentrasi metanol, dan X<sub>5</sub> = rasio oksigen.

Selanjutnya dilakukan pemilihan model terbaik menggunakan *stepwise regression*. Metode *stepwise regression* adalah metode seleksi variabel yang secara sistematis menambah atau menghapus variabel independen berdasarkan signifikansi statistiknya terhadap model. Proses dimulai dengan memasukkan semua variabel independen, atau tanpa memasukkan variabel apa pun, kemudian secara bertahap menambah atau menghapus variabel yang tidak relevan hingga diperoleh model yang memenuhi kriteria statistik yang optimal. Keunggulan metode ini adalah kemampuannya untuk mengidentifikasi kombinasi variabel terbaik secara efisien dan menjaga validitas model yang dihasilkan.

**Tabel 8.** Pemilihan model terbaik menggunakan *Stepwise Regression*

	----Step 1----		----Step 2----	
	Coef	P	Coef	P
Constatn	-651		-595.4	
X <sub>2</sub>	1.376	0.000	1.248	0.000
X <sub>1</sub>			23.44	0.014
S		19.5618		16.4380
R-s		70.47%		80.45%
R-sq(adj)		68.63%		77.85%
Mallows' Cp		7.37		2.15
AICc		163.72		159.66
BIC		164.68		160.15

Metode *stepwise regression* didapatkan hasil model regresi linier yang baru adalah:

$$\hat{Y} = -595.4 + 23.44X_1 + 1.248X_2$$

Setiap kenaikan satu unit pada sistem reaktor berkorelasi dengan peningkatan konversi metanol sebesar 23,44 unit, sedangkan kenaikan satu unit suhu menghasilkan peningkatan konversi metanol sebesar 1,248 unit, asalkan faktor-faktor lain dijaga konstan.

**Tabel 9.** Analisis terhadap model  $\hat{Y} = -595.4 + 23.44X_1 + 1.248X_2$

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-595.4	94.9	-6.27	0.000	
X <sub>1</sub>	23.44	8.47	2.77	0.014	1.06
X <sub>2</sub>	1.248	0.193	6.47	0.000	1.06
S = 16.4380	R-sq = 80.45%	R-sq(adj) = 77.85%		R-sq(pred) = 67.71%	

Pada model ini memiliki R – sq sebesar 80.45% yang cukup tinggi dalam menjelaskan pengaruh variabel independen terhadap variabel variabel bebas.

#### IV. KESIMPULAN

Berdasarkan hasil penelitian dalam pemilihan model terbaik dalam kasus pengaruh suhu, waktu tinggal, konsentrasi metanol, dan rasio oksigen terhadap konversi metanol menggunakan *stepwise regression*, diperoleh persamaan regresi terbaik yang sama yaitu:

$$\hat{Y} = -595.4 + 23.44X_1 + 1.248X_2$$

dengan  $X_1$  = sistem reaktor dan  $X_2$  = suhu. Hasil analisis menunjukkan bahwa setiap peningkatan satu unit pada sistem reaktor akan meningkatkan konversi metanol sebesar 23.44, sementara setiap kenaikan satu unit pada suhu meningkatkan konversi metanol sebesar 1.248, dengan asumsi variabel lainnya tetap.

Model terbaik ini diperoleh ketika masalah multikolinieritas yang terdeteksi telah diatasi dan sudah memenuhi asumsi klasik lainnya. Hasil penelitian menunjukkan bahwa metode *stepwise regression* secara efektif mampu memilih model terbaik dengan nilai R-sq sebesar 80.45%, yang cukup tinggi dalam menjelaskan pengaruh variabel independen terhadap variabel dependen. Selain itu, model ini bebas dari masalah multikolinieritas, sehingga dapat diandalkan untuk menggambarkan hubungan antara variabel-variabel yang diteliti.

#### DAFTAR PUSTAKA

- Aditiya, N. Y., Evani, E. S., & Maghfiroh, S. (2023). Konsep Uji Asumsi Klasik Pada Regresi Linier Berganda. *Jurnal Riset Akuntansi Soedirman*, 2(2), 102–110. <https://doi.org/10.32424/1.jras.2023.2.2.10792>
- Amalita, N., & Kurniawati, Y. (2013). Model Regresi Dummy dalam Memprediksi Performansi Akademik Mahasiswa Jurusan Matematika FMIPA UNP. *Prosiding SEMIRATA 2013*, 19(19), 19.
- Debataraja, N. N., Anggraini, N., & Kusnandar, D. (2019). Metode Generalized Ridge Regression Dalam Mengatasi Multikolinieritas. *Bimaster : Buletin Ilmiah Matematika, Statistika Dan Terapannya*, 8(4), 679–686. <https://doi.org/10.26418/bbimst.v8i4.35879>
- Fariha, N. F., Subekti, R., Matematika, P., Matematika, F., Ilmu, D., & Alam, P. (2018). Pemilihan Model Regresi Terbaik Dalam Kasus Pengaruh Premi, Klaim, Hasil Investasi Dan Hasil Underwriting Terhadap Laba Asuransi Jiwa (Studi Kasus Pt. Asuransi Jiwasraya (Persero)). *Prodi Matematika Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta*, 674–684.
- Marcus, Y. (2018). Extraction by subcritical and supercritical water, methanol, ethanol and their mixtures. *Separations*, 5(1). <https://doi.org/10.3390/separations5010004>
- Nurchaya, W. A., Arisanti, N. P., & Hanandhika, A. N. (2023). Penerapan Uji Asumsi Klasik untuk Mendeteksi Kesalahan Pada Data Sebagai Upaya Menghindari Pelanggaran Pada Asumsi Klasik. *Madani: Jurnal Ilmiah Multidisiplin*, 1(12), 472–481.
- Setya Budi, A. D. A., Septiana, L., & Panji Mahendra, B. E. (2024). Memahami Asumsi Klasik dalam Analisis Statistik: Sebuah Kajian Mendalam tentang Multikolinieritas, Heterokedastisitas, dan Autokorelasi dalam Penelitian. *Jurnal Multidisiplin West Science*, 3(01), 01–11. <https://doi.org/10.58812/jmws.v3i01.878>
- Tester, J. W., Webley, P. A., & Holgate, H. R. (1993). Revised Global Kinetic Measurements of Methanol Oxidation in Supercritical Water. *Industrial and Engineering Chemistry Research*, 32(1), 236–239. <https://doi.org/10.1021/ie00013a032>
- Yanke, A., Zentrato, N. E., & Soleh, A. M. (2022). Handling Multicollinearity Problems in Indonesia's Economic Growth Regression Modeling Based on Endogenous Economic Growth Theory. *Indonesian Journal of Statistics and Its Applications*, 6(2), 228–244. <https://doi.org/10.29244/ijjsa.v6i2p214-230>