VOLUME 2, No 1, Juni 2023
e-ISSN: 2987-206X
https://ejournal.unib.ac.id/diophantine,

Performa Teknik Regularisasi Dalam Penanganan Masalah Multikolinieritas

Alin Febianti Fikri[1], Winalia Agwil[2*], Dian Agustina[3]

[1,2,3] Program Studi Statistika, Universitas Bengkulu

* Corresponding Author: winaliaagwil@unib.ac.id

Article Information

Article History:
Submitted: 06 23 2023
Accepted: 06 28 2023
Published: 06 30 2023

Key Words:
Elastic Net
LASSO
Multicollinearity
Ridge

DOI:
https://doi.org/10.33369/diophantine.v2i01.28480

Abstract

Multicollinearity is a condition in which there is a linear relationship between independent variables that are correlated with each other. Multicollinearity can be handled using regularization techniques that adjust or reduce the estimated coefficient to zero. There are several regularization techniques, such as ridge regression, LASSO, and elastic net. This study discusses the handling of multicollinearity using these three methods in several dataset conditions, namely when $p > n$, $p < n$, and $p \gg n$, or high dimensional data. Based on some of the dataset conditions, it was observed that the best model was based on the MSE value. The data used are simulation data and case studies from the official BPS website for $p < n$ and the UCI machine learning repository for $p \gg n$. The results of the simulation data and case studies show that the ridge model is good at modeling datasets with p = 20, 40, and 80 or $p < n$, and the elastic net is good at modeling datasets with p = 100, 160, and 320 or $p \gg n$.

1. PENDAHULUAN

Analisis regresi merupakan salah satu analisis statistik yang digunakan untuk menganalisis hubungan antar variabel yaitu variabel independen dan variabel dependen. Jika dipandang dari jumlah variabel independent yang terdapat dalam model, analisis regresi terbagi menjadi dua macam yaitu analisis regresi sederhana dan analisis regresi berganda. Jika hanya melibatkan satu variabel independen (X) maka disebut analisis regresi linier sederhana sedangkan, jika terdapat lebih dari satu variabel independen maka disebut analisis regresi linier berganda. Metode Kuadrat Terkecil (MKT) adalah salah satu metode yang digunakan untuk menduga parameter regresi dalam analisis linier berganda. Koefisien regresi yang ditaksir menggunakan metode MKT mempunyai sifat Best Linear Unbiased Estimator (BLUE) dimana salah satu asumsi yang harus dipenuhi adalah tidak terjadinya masalah multikolinieritas [1]. Multikolinieritas adalah kondisi terdapat hubungan linier antar variabel independen, dimana diantara variabel independen tersebut saling berkorelasi. Akibatnya akan sulit untuk melihat pengaruh variabel independen terhadap variabel dependen.

Dampak lain dari adanya multikolinieritas pada model regresi antara lain; 1) satu atau lebih variabel independen adalah *redundant* artinya satu variabel independen menjelaskan tentang variabel dependen persis sama dengan yang dijelaskan oleh variabel independen lainnya, 2) mempengaruhi kemampuan model untuk mengestimasi koefisien regresi, dan 3) varians dari estimasi parameter dengan MKT menjadi tinggi [2]. [3] menyatakan bahwa multikolinieritas dapat mengakibatkan parameter regresi yang dihasilkan oleh analisis regresi berganda menjadi sangat lemah atau tidak dapat memberikan hasil analisis yang mewakili sifat dan pengaruh dari variabel independen yang bersangkutan.

Penelitian untuk mengatasi multikolinieritas telah dilakukan sebelumnya menggunakan metode regresi ridge oleh [4]. Masalah multikolinieritas yang terjadi pada metode kuadrat terkecil dapat diatasi dengan menggunakan metode regresi ridge dengan besar koefisien determinasi yaitu 47,43%. Penelitian lain dilakukan oleh [5]. Model regresi LASSO dengan algoritma LARS mampu mengatasi masalah multikolinieritas dalam data dan menghasilkan 6 variabel yang berpengaruh secara signifikan dimana nilai adjusted $R2$ sebesar 73,25%. Penelitian lainnya dilakukan oleh [6] yang mengkaji mengenai performa metode elastic net dalam kasus multikolinieritas pada analisis linier berganda, disimpulkan bahwa metode elastic net mampu menutupi kekurangan LASSO dan ridge. Metode elastic net dapat menyusutkan variabel yang saling berkorelasi dengan bantuan penalti penyusutan gabungan antara regresi ridge dan LASSO.

Regresi LASSO dapat menyusutkan koefisien dan menetapkan koefisien ke angka 0. Oleh karena itu, LASSO dapat menghasilkan model dengan variabel terbaik. Namun, LASSO memiliki beberapa kelemahan. Ketika $p < n$, kinerja LASSO lebih didominasi oleh ridge, p adalah jumlah variabel independent sedangkan n adalah jumlah amatan. Ketika $p > n$, maka LASSO hanya memilih n variabel yang diikutkan dalam model. Sehingga, pada kondisi $p \gg n$ LASSO bukanlah metode yang ideal. Hal ini dikarenakan model LASSO hanya mampu menyeleksi paling banyak n variabel dari p kandidat [7]. Berdasarkan latar belakang tersebut, pada penelitian ini akan dibahas penanggulangan multikolinieritas pada beberapa kondisi dataset yaitu dimensi $p < n$, $p > n$, dan $p \gg n$ menggunakan regresi ridge, LASSO dan elastic net. Teknik regulasi terbaik dipilih berdasarkan penerapan metode pada data simulasi dan studi kasus

2. METODE

2.1 Data

Data yang digunakan dalam penelitian ini adalah data kuantitatif yang diperoleh dari website resmi Badan Pusat Statistik (BPS) [8], UCI machine learning repository [9], serta membangkitkan variabel independen (X) dan variabel dependen (Y) yang berdistribusi normal dengan rataan berada pada interval 0-50 dan simpangan baku berada pada interval 0-20.

2.2 Metode Analisis

Analisis data menggunakan bantuan program aplikasi R. Berikut adalah langkah-langkah analisis yang dilakukan dalam penelitian ini, sebagai berikut.

1. Membangkitkan data simulasi
2. Melakukan pemodelan regresi dengan MKT (metode kuadrat terkecil), yaitu meregresikan variabel X dengan variabel Y.
3. Mendeteksi adanya multikolinieritas dari masing-masing variabel independen (X) dengan melihat nilai *Variance Inflation Factor* (VIF) masing-masing variabel independen. Jika nilai VIF lebih besar dari 10 maka terdapat pelanggaran asumsi multikolinieritas. Semakin tinggi nilai VIF-nya maka semakin serius permasalahan multikolinieritasnya [10]
4. Melakukan analisis regresi dengan regresi ridge.
   a. Memilih nilai λ optimal yang menghasilkan cross validation error minimum untuk δ = 0.
   b. Pemodelan menggunakan metode regresi ridge dengan δ = 0 dan λ optimal.

      Regresi ridge meminimumkan *sum of squares error* (SSE) atau jumlah kuadrat sisaan pada pendugaan koefisien regresi. Namun, dalam metode ridge menambahkan penalti penyusutan dalam meminimumkan SSE. Sehingga persamaan yang terbentuk adalah sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_{ridge} = \underset{\boldsymbol{\beta}}{\arg min} \sum_{i=1}^{n} \left(\mathrm{y}_i - \beta_0 - \sum_{j=1}^{p} X_{ij}\beta_j \right)^2 + \lambda \left(\sum_{j=1}^{p} \beta_j^2 \right)$$

   c. Melihat evaluasi model dengan menghitung MSE pada data testing.
5. Melakukan analisis regresi dengan regresi LASSO.
   a. Memilih nilai λ optimal yang menghasilkan cross validation error minimum untuk δ = 1.
   b. Pemodelan menggunakan metode regresi LASSO dengan δ = 1 dan λ optimal.

      Pada regresi ridge saat menduga parameter terjadi proses kontinu yang mengecilkan koefisien. Namun, tidak menetapkan koefisien apapun menjadi 0 dan karenanya tidak memberikan model yang mudah diinterpretasikan. Sedangkan LASSO dapat menyusutkan beberapa koefisien dan menetapkan yang lain ke 0. Oleh karena itu, LASSO dapat mempertahankan fitur-fitur yang baik dari pemilihan subset dan regresi ridge [11]. LASSO dan regresi ridge memiliki persamaan yang hampir sama, hanya penaltinya saja yang berbeda. Penalti penyusutan pada regresi ridge $\sum_{j=1}^{p} \beta_j^2$ diganti dengan $\sum_{j=1}^{p} |\beta_j|$ sebagai penalti penyusutan LASSO. Sehingga persamaan yang terbentuk adalah sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_{LASSO} = \underset{\boldsymbol{\beta}}{\arg min} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} X_{ij}\beta_j \right)^2 + \lambda \left(\sum_{j=1}^{p} |\beta_j| \right)$$

c. Melihat evaluasi model dengan menghitung MSE pada data testing.

6. Melakukan analisis regresi dengan regresi elastic net.
   a. Menentukan rentang nilai δ yaitu $0 \leq \delta \leq 1$ dengan validasi silang.
   b. Memilih nilai pasangan nilai δ dan λ optimal yang menghasilkan cross validation error minimum.
   c. Analisis regresi menggunakan metode regresi elastic net dengan pasangan δ dan λ optimal.

   Misalkan kumpulan data memiliki n pengamatan dengan p variabel independen. Jika $\boldsymbol{y} = (y_1, \dots, y_n)^T$ adalah variabel dependen dan $\boldsymbol{X} = (\boldsymbol{x}_1 | \dots | \boldsymbol{x}_p)$ adalah matriks model, dimana $\boldsymbol{x}_j = (x_{1j}, \dots, x_{nj})^T, j = 1, \dots, p$, adalah variabel independen. Diketahui kriteria naïve elastis net sebagai berikut:

$$L(\lambda_1, \lambda_2, \boldsymbol{\beta}) = |\boldsymbol{y} - \boldsymbol{X\beta}|^2 + \lambda_2 |\boldsymbol{\beta}|^2 + \lambda_1 |\boldsymbol{\beta}|$$

   Dengan

$$|\boldsymbol{\beta}|^2 = \textstyle\sum_{j=1}^{p} \beta_j^2$$
$$|\boldsymbol{\beta}| = \textstyle\sum_{j=1}^{p} |\beta_j|$$

   Penduga *naïve elastic net* $\widehat{\boldsymbol{\beta}}$ diperoleh dengan meminimalkan $L(\lambda_1, \lambda_2, \boldsymbol{\beta})$

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\arg min} \{L(\lambda_1, \lambda_2, \boldsymbol{\beta})\}$$

   d. Melihat evaluasi model dengan menghitung MSE pada data testing.
7. Melihat kebaikan model ridge, LASSO, dan elastic net pada masing-masing dataset dengan dimensi yang berbeda.
8. Langkah 2-7 juga dilakukan pada data studi kasus yaitu dataset kemiskinan dan dataset Gas Sensor Array Under Flow Modulation

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data Simulasi

Penelitian dilakukan dengan membangkitan dataset yang mengandung multikolinieritas. Bagian pertama yang dirancang adalah cara membuat data yang sesuai dengan kriteria yang diinginkan sehingga dapat dianalisis. Variabel dependen dan independen dibangkitkan menurut sebaran normal dengan rataan berada pada interval 0-50 dan simpangan baku berada pada interval 0-20.

Asumsi multikolinieritas pada model regresi berganda mensyaratkan masing-masing variabel independen memiliki nilai VIF < 10 agar tidak terdapat pelanggaran asumsi multikolinieritas. Jika ingin memperoleh dataset yang mengandung multikolinieritas dapat dilakukan dengan menentukan matriks ragam-peragam variabel independen agar nilai korelasi antar variabel > 0,5. Namun, cara tersebut kurang efektif apabila ingin membangkitkan high dimensional data. Multikolinieritas menyebabkan $(X'X)$ singular sehingga $(X'X)^{-1}$ tidak ada. Hal ini berarti satu atau lebih kolom merupakan kombinasi linier dari kolom-kolom lainnya sehingga $|X'X| = 0$ (Draper dan Smith, 1992). Oleh karena itu, untuk memperoleh model regresi yang mengandung multikolinieritas dapat dilakukan dengan menggunakan kombinasi linier antar variabel independen.

Dataset pada penelitian ini mengandung multikolinieritas pada 2 variabel pertama. Keberhasilan proses ini dapat diamati dari nilai VIF pada uji multikolinieritas. Jika VIF ≥ 10 maka diperoleh dataset yang mengandung multikolinieritas sehingga sesuai untuk digunakan dalam mengkaji teknik regularisasi. Pemilihan jumlah observasi dan banyak variabel independen yang akan dibangkitkan mengandung subjektifitas dan dapat berbeda-beda di setiap percobaan dan pengguna (user). Akan dibangkitkan dataset dengan jumlah amatan 100 dan banyak parameter p = 20, 40, 80, 100, 160 dan 320. Kondisi multikolinieritas dalam data diperoleh dengan menggunakan kombinasi linier. Masing-masing variabel mengikuti sebaran normal. Sebagai kontrol dilakukan 30 kali pengacakan dataset yang mengandung multikolinieritas. Karakterisik dataset dideskripsikan pada Tabel 1

Tabel 1 Karakteristik pembangkitan dataset

Variabel	Karakteristik
X_3	1,2*X2 +rnorm(n = 100, $mean$ = [0 – 50], sd = [0 – 20])
X_6	0,2*X4+rnorm(n = 100, $mean$ = [0 – 50], sd = [0 – 20])

Variabel X_3 dibangkitkan dari perkalian variabel X2 dengan 1,2 untuk selanjutnya ditambah dengan bilangan pengacakan sebaran normal. Sedangkan, variabel X_6 dibangkitkan dari perkalian variabel X4 dengan 0,2 untuk selanjutnya ditambah dengan bilangan pengacakan sebaran normal. Selain itu, variabel dependen dan independen lainnya (X_p) dibangkitkan dengan pengacakan yang berasal dari sebaran normal dengan rataan berada pada interval 0-50 dan simpangan baku berada pada interval 0-20.

Selanjutnya akan dilakukan pengujian multikolinieritas pada dataset dengan 20 variabel independen dengan memperhatikan nilai VIF, nilai VIF setiap variabel independen terlampir pada Tabel 2.

Tabel 2. Hasil uji multikolinieritas dataset simulasi p=20

Variabel	VIF	Variabel	VIF
X_1	1,2508	X_{11}	1,2805
X_2	12,0712*	X_{12}	1,2646
X_3	12,0441*	X_{13}	1,2083
X_4	1,2031	X_{14}	1,3690
X_5	1,1472	X_{15}	1,1611
X_6	1,2861	X_{16}	1,1127
X_7	1,1086	X_{17}	1,1559
X8	1,2799	X_{18}	1,2369
X9	1,1457	X_{19}	1,1878
X_{10}	1,3365	X_{20}	1,1826

Berdasarkan Tabel 2 terdeteksi bahwa 2 variabel dari 20 variabel independen yang digunakan dalam penelitian memiliki korelasi yang tinggi atau dengan kata lain terjadi multikolinieritas didalam dataset. Sehingga dataset yang dibangkitkan telah memenuhi untuk dilakukan analisis dengan menggunakan regresi ridge, regresi LASSO dan *elastic net*. Tabel 2 hanya salah satu contoh dari banyak skenario dataset yang dibangkitkan.

Dilakukan 30 pembangkitan pada masing-masing dataset yaitu skenarion dengan jumlah amatan 100 dan banyak parameter p = 20, 40, 80, 100, 160 dan 320. Pemilihan model terbaik untuk setiap dataset yang dibangun dengan regresi ridge, LASSO, dan elastic net dilihat dari nilai MSE terkecil. Model terbaik antara regresi ridge, LASSO, dan elastic net adalah model yang mempunyai nilai rataan MSE terkecil dari 30 simulasi dataset.

Tabel 3 Perbandingan nilai MSE regresi ridge, LASSO dan elastic net

Model	p=20	p=40	p=80	p=100	p=160	p=320
Ridge	22,014	20,047	26,163	29,401	31,123	25,794
LASSO	22,301	20,339	27,018	28,966	31,083	26,127
Elastic Net	22,077	20,082	26,345	28,395	30,512	25,640

Hasil dari pemodelan dataset simulasi dengan banyak amatan 100 yang dibangkitkan dengan p = 20, 40, dan 80 atau $p < n$ menghasilkan rataan MSE terkecil pada model ridge yaitu 22,014 untuk p = 20, 20,047 untuk p = 40 dan 26,163 untuk p = 80. Dataset dengan p = 100, 160, dan 320 menghasilkan rataan MSE terkecil pada model elastic net yaitu 28,395 untuk p = 100, 30,512 untuk p = 160 dan 25,640 untuk p = 320. Hal ini sejalan dengan pernyataan [7] bahwa, ketika $p < n$, kinerja LASSO lebih didominasi oleh ridge dan jika $p \gg n$ maka model elastic net akan bekerja dengan efektif.

3.2 Analisis Data Studi Kasus (dataset kemiskinan)

Dataset kemiskinan dengan kondisi data jumlah variabel (p) lebih kecil dibandingkan jumlah amatan (n). dataset terdiri dari 6 variabel independen dengan tingkat kemiskinan sebagai variabel dependen. Data diperoleh dari website resmi BPS pada tingkat kabupaten/kota se-Indonesia tahun 2021.

Tabel 4 Variabel Dataset Kemiskinan

Variabel	Keterangan	Satuan
Y	Tingkat kemiskinan	(%)
X1	Persentase rumah tangga sumber air minumnya berasal dari sumur tidak terlindungi	(%)
X2	Persentase rumah tangga yang menggunakan bahan bakar memasak sehari-hari adalah kayu bakar	(%)
X3	Persentase rumah tangga dengan jenis lantai terluas adalah tanah/lainnya	(%)
X4	Persentase angkatan kerja terhadap usia kerja	(%)
X5	Jumlah murid SD	-
X6	Jumlah murid SMP	-

Dataset kemiskinan dimodelkan dengan menggunakan ketiga metode yaitu regresi ridge, LASSO, dan elastic net, kemudian Pemilihan model terbaik dari model dilakukan dengan melihat nilai MSE. Berdasarkan Tabel 5, nilai MSE paling kecil pada dataset studi kasus terdapat pada model ridge.

Tabel 5 Perbandingan nilai MSE regresi ridge, LASSO, dan elastic net pada kondisi p<n

Model	δ	λ	MSE
Ridge	0	0,07396	0,4982
	0,1	0,03056	0,5038
	0,2	0,02020	0,5053
	0,3	0,01622	0,5059
	0,4	0,01216	0,5064
Elastic net	0,5	0,00973	0,5068
	0,6	0,00739	0,5070
	0,7	0,00633	0,5071
	0,8	0,00317	0,5073
	0,9	0,00448	0,5074
LASSO	1	0,00443	0,5076

Hasil dari pemodelan dataset studi kasus menghasilkan MSE 0,4982 dengan nilai λ = 0,07396 dan δ = 0 untuk model ridge. Model LASSO dengan δ = 1 dan λ = 0,00443 menghasilkan MSE 0,5076. Sedangkan, model elastic net optimum memiliki δ = 0,1 dan λ = 0,03056 menghasilkan MSE yaitu 0,5038. Berdasarkan hasil pemodelan Tabel 4, dapat disimpulkan bahwa model ridge mampu memodelkan dataset yang mengandung multikolinieritas secara optimal untuk dataset dengan jumlah $p < n$.

3.3 Analisis Data Studi Kasus (Gas Sensor Array Under Flow Modulation Dataset)

Dataset tersebut memiliki kondisi dimana jumlah variabel lebih banyak dibandingkan jumlah amatan. Dataset dianalisis menggunakan ketiga model sebelumnya kemudian model terbaik dipilih berdasarkan nilai MSE. Berdasarkan Tabel 6, nilai MSE paling kecil pada dataset studi kasus terdapat pada model elastic net

Tabel 6. Perbandingan MSE regresi ridge, LASSO, dan elastic net

Model	δ	λ	MSE
Ridge	0	2,99023	0,01942
	0,1	0,07942	0,01262
	0,2	0,04783	0,01232
	0,3	0,02773	0,01185
	0,4	0,01809	0,01150
Elastic net	0,5	0,01381	0,01135
	0,6	0,01098	0,01119
	0,7	0,00986	0,01111
	0,8	0,00904	0,01124
	0,9	0,01062	0,01139
LASSO	1	0,00956	0,01145

Hasil dari pemodelan dataset studi kasus menghasilkan MSE 0,01942 dengan nilai λ = 2,99023 dan δ = 0 untuk model ridge. Model LASSO dengan δ = 1 dan λ = 0,00956 menghasilkan MSE 0,01145, sedangkan model elastic net optimum memiliki δ = 0,7 dan λ = 0,00986 menghasilkan MSE yaitu 0,01111. Berdasarkan hasil pemodelan Tabel 6, dapat disimpulkan bahwa model elastic net mampu memodelkan dataset yang mengandung multikolinieritas secara optimal untuk dataset dengan jumlah $p \gg n$.

4. SIMPULAN

Berdasarkan hasil dan pembahasan mengenai regresi ridge, LASSO, dan elastic net, maka dapat diambil kesimpulan sebagai berikut :

1. Dari 30 pengacakan diperoleh rataan MSE hasil dari pemodelan dataset simulasi dengan banyak amatan 100 yang dibangkitkan dengan p = 20, 40, dan 80 atau $p < n$ menghasilkan rataan MSE terkecil pada model ridge. Dataset dengan p = 100, 160, dan 320 menghasilkan rataan MSE terkecil pada model elastic net. Hal ini sejalan dengan pernyataan Zhou dan Hastie (2005) bahwa, ketika $p < n$, kinerja LASSO lebih didominasi oleh ridge dan jika $p \gg n$ maka model elastic net akan bekerja dengan efektif.
2. Hasil perbandingan metode ridge, LASSO, dan elastic net pada dataset kemiskin dapat disimpulkan model ridge mampu memodelkan dataset yang mengandung multikolinieritas secara optimal untuk dataset dengan jumlah $p < n$.

Hasil dari perbandingan metode ridge, LASSO, dan elastic net pada model gas sensor array under flow modulation menunjukkan nilai MSE yang dimiliki model ridge sebesar 0,01942. Model LASSO menghasilkan MSE 0,01145 sedangkan, model elastic net optimum memiliki MSE terkecil sebesar 0,01111. Sehingga, model terbaik untuk pemodelan gas sensor array under flow modulation adalah regresi elastic net. Hal ini menunjukkan bahwa regresi elastic net mampu memodelkan dataset yang mengandung multikolinieritas

REFERENSI

[1] H. Pakinde and A. Setiawan, "Studi Simulasi dan Studi Kasus Masalah Multikolinearitas dalam Model Regresi Linier," in Seminar Nasional Sains dan Pendidikan Sains VI, Salatiga, 2009.

[2] S. Sunaryo, S. Setiawan and T. H. Siagian, "Mengatasi Masalah Multikolinearitas dan Outlier dengan Pendekatan ROBPCA (Studi Kasus Analisis Regresi Angka Kematian Bayi Di Jawa Timur)," Jurnal Matematika Sains Dan Teknologi, vol. 12, no. 1, 2011.

[3] W. W. Hines and D. C. Montgomery , Probabilitas dan Statistik dalam Ilmu Rekayasa dan Manajemen, Jakarta: Universitas Indonesia, 1990.

[4] W. R. Anggraeni, N. N. Debataraja and S. W. Rizki, "Estimasi Parameter Regresi Ridge untuk Mengatasi Multikolinearitas," Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster), vol. 7, no. 4, 2018.

[5] M. Robbani, F. Agustiani and N. Herrhyanto, "Regresi Least Absolute Shrinkage and Selection Operator (Lasso) pada Kasus Inflasi di Indonesia Tahun 2014-2017," Jurnal Eurekamatika, vol. 7, no. 2, 2019.

[6] H. E. Sinaga and D. Saputro, "Performa Metode Elastic-Net dalam Kasus Multikolinearitas pada Analisis Linear Berganda," in Seminar Pendidikan Matematika dan Matematika, Yogyakarta, 2021.

[7] H. Zou and T. Hastie, "Regularization and Variable Selection Via the Elastic-Net," Journal of the Royal Statistical Society Series B, vol. 67, pp. 301-320, 2005.

[8] "Badan Pusat Statistik," 2022. [Online]. Available: https://www.bps.go.id/.

[9] "UCI Machine Learning Repository," 2022. [Online]. Available: https://archive.ics.uci.edu/dataset/308/gas+sensor+array+under+flow+modulation.

[10] T. Naes, T. Isaksson, T. Fearn and T. Davies, Multivariate Calibration and Classification, West Sussex: Nir Publication, 2002.

[11] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," Journal of the Royal Statistical Society Series B , vol. 58, pp. 267-288, 1996.