

**PENYELEKSIAN INDEKS SPEKTRAL MELALUI MODEL
REGRESI LOGISTIK MULTINOMIAL DENGAN METODE
LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR
(LASSO)**

SKRIPSI

*Diajukan sebagai salah satu persyaratan guna memperoleh gelar
Sarjana Statistika*



**Oleh
SABINA CHAIRUN NAJWA
NIM 20337009**

**PROGRAM STUDI SARJANA STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG
2024**

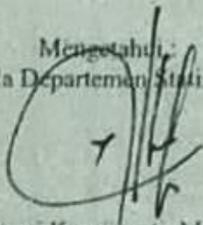
PERSETUJUAN SKRIPSI

**PENYELEKSIAN INDEKS SPEKTRAL MELALUI MODEL
REGRESI LOGISTIK MULTINOMIAL DENGAN METODE
LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR
(LASSO)**

Nama : Sabina Chairun Najwa
NIM : 20337009
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 30 Mei 2024

Mengetahui :
Kepala Departemen Statistika



Dr. Yenni Kurniawati, M.Si.
NIP. 198402232010122005

Disetujui Oleh :
Pembimbing



Dr. Yenni Kurniawati, M.Si.
NIP. 198402232010122005

PENGESAHAN LULUS UJIAN SKRIPSI

Nama : Sabina Chairun Najwa
NIM : 20337009
Program Studi : SI Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

PENYELEKSIAN INDEKS SPEKTRAL MELALUI MODEL REGRESI LOGISTIK MULTINOMIAL DENGAN METODE *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* (LASSO)

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi Departemen
Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Padang

Padang, 30 Mei 2024

Tim Penguji

Nama
Ketua : Dr. Yenni Kurniawati, M.Si.
Anggota : Dr. Syatrindi, M.Si.
Anggota : Fadhilah Fitri, S.Si., M.Stat.

Tanda Tangan



The image shows three handwritten signatures in blue ink, each written over a horizontal line. The signatures are somewhat stylized and difficult to read, but they correspond to the names of the examiners listed in the adjacent text.

SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertanda tangan dibawah ini :

Nama : Sabina Chairun Najwa
NIM : 20337009
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan, bahwa skripsi saya dengan judul “ Penyeleksian Indeks Spektral Melalui Model Regresi Logistik Multinomial Dengan Metode *Least Absolute Shrinkage and Selection Operator (LASSO)*” adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dan etika yang berlaku dalam tradisi keilmuan. Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun masyarakat dan negara.

Demikian pernyataan ini saya buat dengan penuh kesadaran dan rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Padang, 3 Juni 2024

Diketahui Oleh :
Kepala Departemen Statistika



Dr. Yenni Kurniawati, M.Si.
NIP. 198402232010122005

Saya yang menyatakan,



Sabina Chairun Najwa
NIM. 20337009

Penyeleksian Indeks Spektral Melalui Model Regresi Logistik Multinomial Dengan Metode *Least Absolute Shrinkage and Selection Operator* (LASSO)

Sabina Chairun Najwa

ABSTRAK

Perhitungan luas panen padi nasional di Indonesia menggunakan pendekatan Kerangka Sampel Area (KSA). Survei KSA cukup akurat untuk melakukan mengestimasi luas panen, namun biaya survei yang diperlukan cukup tinggi. Salah satu solusi meminimalisir biaya survei KSA adalah dengan memanfaatkan indeks spektral pada citra satelit salah satunya citra satelit Landsat-8. Pemanfaatan citra satelit membantu pemodelan amatan *multiclass* dalam survei KSA menjadi lebih efisien. Salah satu model yang dapat digunakan untuk mengklasifikasikan amatan KSA adalah Regresi Logistik Multinomial (RLM). Cara untuk menyederhanakan model kompleks dari model RLM, yaitu dengan metode *Least Absolute Shrinkage and Selection Operator* (LASSO). Penelitian ini bertujuan untuk menyederhanakan dan mengetahui performa model klasifikasi amatan KSA.

Penelitian ini merupakan penelitian terapan, yaitu penerapan metode *Least Absolute Shrinkage and Selection Operator* (LASSO) pada Regresi Logistik Multinomial (RLM) yang memodelkan amatan KSA. Data yang digunakan adalah indeks-indeks spektral Landsat-8 selama tiga periode waktu pada Provinsi Jawa Barat tahun 2018. Adapun variabel prediktor yang adalah indeks-indeks spektral, yaitu EVI_t , $NDBI_t$, $MNDWI_t$, $NDBI_{t-1}$, $MNDWI_{t-1}$, $NDBI_{t-2}$, $MNDWI_{t-2}$, $dEVI_1$, $dEVI_2$, dan variabel respon yang digunakan adalah kategori amatan KSA.

Penelitian ini bertujuan untuk menyederhanakan model klasifikasi pengamatan KSA pada data citra satelit Landsat-8. Hasil penelitian menunjukkan bahwa LASSO dapat mengurangi kompleksitas model dengan memilih indeks spektral yang berkontribusi. LASSO dapat memodelkan data dan mengestimasi parameter sekaligus memilih parameter secara bersamaan, sehingga lebih efisien. Model RLM dan RLM-LASSO memiliki kinerja yang serupa. Namun, kriteria *confusion matrix* pada model RLM-LASSO meningkat, yang membuktikan bahwa model yang sederhana lebih baik daripada model yang kompleks dalam mengklasifikasikan pengamatan KSA.

Kata Kunci : *Least Absolute Shrinkage and Selection Operator*, Regresi Logistik Multinomial, Kerangka Sampel Area, Citra Satelit Landsat-8, Indeks Spektral.

Spectral Indices Selection Through Multinomial Logistic Regression Model Using Least Absolute Shrinkage and Selection Operator (LASSO) Method

Sabina Chairun Najwa

ABSTRACT

The calculation of the national rice harvest area in Indonesia uses the Area Sampling Frame (ASF) approach. ASF surveys are accurate enough to estimate, but the required survey costs are high. One solution to minimize the cost of ASF surveys is to utilize indices in satellite imagery, one of which is Landsat-8 satellite imagery. Using satellite imagery helps make multiclass observation modeling in ASF surveys more efficient. One model that can be used to classify ASF observations is the Multinomial Logistic Regression Model (MLM). The way to simplify the complex model of the MLM model is with the Least Absolute Shrinkage and Selection Operator (LASSO) method. This study aims to streamline and determine the ASF observation classification model's performance.

This research applies the Least Absolute Shrinkage and Selection Operator (LASSO) method to the Multinomial Logistic Regression Model (MLM) modeling KSA observations. The data used are Landsat-8 spectral indices for three time periods in West Java Province in 2018. The predictor variables are spectral indices, namely EVI_t , $NDBI_t$, $MNDWI_t$, $NDBI_{t-1}$, $MNDWI_{t-1}$, $NDBI_{t-2}$, $MNDWI_{t-2}$, $dEVI_1$, $dEVI_2$, and the response variable used is the ASF observation category.

The results show that LASSO can reduce model complexity by selecting contributing spectral indices. LASSO can model data and estimate parameters while selecting parameters simultaneously, making it more efficient. The MLM and MLM-LASSO models have similar performance. However, the confusion matrix criterion in the MLM-LASSO model increases, proving that simple models are better than complex models in classifying ASF observations.

Keywords : Least Absolute Shrinkage and Selection Operator, Multinomial Logistic Regression, Area Sample Frame, Landsat-8 Satellite Imagery, Spectral Indices.

KATA PENGANTAR

Alhamdulillah puji syukur penulis ucapkan kepada Allah SWT yang telah memberikan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Penyeleksian Indeks Spektral Melalui Model Regresi Logistik Multinomial Dengan Metode *Least Absolute Shrinkage and Selection Operator (LASSO)*”. Shalawat dan salam juga tercurah kepada Nabi Muhammad SAW beserta keluarga, sahabat, dan umatnya.

Skripsi ini disusun sebagai salah satu prasyarat untuk menyelesaikan pendidikan pada program studi Sarjana Statistika, Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang. Pada penulisan skripsi ini, penulis banyak mendapatkan dukungan, bimbingan dan bantuan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terimakasih kepada:

1. Ibu Dr. Yenni Kurniawati, M.Si., sebagai Dosen Pembimbing Skripsi dan Kepala Departemen Statistika yang telah berkenan meluangkan waktu, ilmu dan tenaga dalam memberikan arahan selama penyusunan skripsi.
2. Bapak Dr. Syafriandi, M.Si., sebagai Dosen Pembahas Skripsi yang telah memberikan arahan dan masukan selama penyusunan skripsi.
3. Ibu Fadhilah Fitri S.Si., M.Stat., sebagai Dosen Pembahas Skripsi dan Dosen Pembimbing Akademik yang telah memberikan arahan, bimbingan dan nasehat selama perkuliahan serta selama penyelesaian skripsi.
4. Bapak dan Ibu Dosen, Staf Pengajar Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.

5. Terisitimewa kepada kedua orang tua tercinta, Bapak Hasan Basri dan Ibu Nurbaiti yang selalu memberikan dukungan, kasih sayang, serta doa dalam setiap kesempatan.
6. Teman-teman seperjuangan Angkatan 2020 atas kerjasama dan kebersamaannya selama ini.
7. Semua pihak yang telah ikut berkontribusi dalam penulisan skripsi ini

Semoga semua bimbingan, arahan, masukan, bantuan dan do'a dibalas oleh Allah SWT sebagai amal ibadah. Penulis menyadari bahwa penulisan skripsi ini jauh dari kesempurnaan. Oleh karena itu, saran dan masukan dari berbagai pihak sangat diharapkan demi kesempurnaan penulisan yang akan datang. Semoga skripsi ini bermanfaat untuk berbagai pihak yang membutuhkan.

Padang, Maret 2024

Penulis

Sabina Chairun Najwa

DAFTAR ISI

| | |
|---|-------------|
| ABSTRAK | i |
| KATA PENGANTAR..... | iii |
| DAFTAR ISI..... | v |
| DAFTAR TABEL | vii |
| DAFTAR GAMBAR..... | viii |
| DAFTAR LAMPIRAN | ix |
| BAB I PENDAHULUAN | 1 |
| A. Latar Belakang Masalah | 1 |
| B. Rumusan Masalah | 8 |
| C. Batasan Masalah | 9 |
| D. Tujuan Penelitian..... | 9 |
| E. Manfaat Penelitian..... | 9 |
| BAB II KERANGKA TEORITIS | 11 |
| A. Asumsi Regresi Logistik | 11 |
| B. Regresi Logistik Multinomial..... | 13 |
| C. Estimasi Parameter Regresi Logistik Multinomial..... | 15 |
| D. <i>Least Absolute Shrinkage and Operator Selection (LASSO)</i> | 18 |
| E. <i>Confusion Matrix</i> | 22 |
| F. Kerangka Sampel Area (KSA) | 26 |
| G. Citra Satelit Landsat-8..... | 29 |
| H. Penelitian Relavan | 32 |
| BAB III METODOLOGI PENELITIAN..... | 35 |
| A. Jenis Penelitian | 35 |
| B. Jenis dan Sumber Data | 35 |
| C. Variabel Penelitian | 35 |

| | |
|---|-----------|
| D. Teknik Analisis Data | 37 |
| BAB IV HASIL DAN PEMBAHASAN..... | 40 |
| A. <i>Preprocessing Data</i> | 40 |
| B. Eksplorasi Data..... | 40 |
| C. Analisis Data | 43 |
| D. Pembahasan | 59 |
| BAB V PENUTUP | 63 |
| A. Kesimpulan..... | 63 |
| B. Saran | 65 |
| DAFTAR PUSTAKA | 66 |
| LAMPIRAN..... | 71 |

DAFTAR TABEL

| Tabel | Halaman |
|---|----------------|
| 1. Interpretasi Koefisien Korelasi | 12 |
| 2. <i>Confusion Matrix</i> $N \times N$ | 23 |
| 3. Standar Evaluasi..... | 26 |
| 4. Fase Amatan KSA..... | 28 |
| 5. <i>Band</i> , Panjang Gelombang dan Kegunaannya Citra Satelit Landsat-8..... | 29 |
| 6. Nilai Kriteria <i>Confusion Matrix</i> RLM | 33 |
| 7. Variabel Penelitian..... | 36 |
| 8. Struktur Data | 37 |
| 9. <i>Pre-Processing</i> Label KSA..... | 40 |
| 10. Penduga Parameter LASSO | 48 |
| 11. <i>Confusion Matrix</i> RLM-LASSO..... | 53 |
| 12. Nilai Kriteria <i>Confusion Matrix</i> RLM-LASSO | 55 |

DAFTAR GAMBAR

| Gambar | Halaman |
|---|----------------|
| 1. Gambaran Pencitraan Permukaan Bumi dengan Landsat-8..... | 3 |
| 2. Tahap Penyusunan Kerangka Sampel..... | 26 |
| 3. Ilustrasi Pembagian Wilayah dalam Blok dan Segmen | 27 |
| 4. Diagram Alir Penelitian | 39 |
| 5. (a) <i>Boxplot</i> EVI_t terhadap Amatan KSA, (b) <i>Boxplot</i> $NDBI_t$ terhadap Amatan KSA, (c) <i>Boxplot</i> $MNDWI_t$ terhadap Amatan KSA, (d) <i>Boxplot</i> $NDBI_{t-1}$ terhadap Amatan KSA, (e) <i>Boxplot</i> $MNDWI_{t-1}$ terhadap Amatan KSA, (f) <i>Boxplot</i> $NDBI_{t-2}$ terhadap Amatan KSA, (g) <i>Boxplot</i> $MNDWI_{t-2}$ terhadap Amatan KSA, (h) <i>Boxplot</i> $dEVI_1$ terhadap Amatan KSA, (i) <i>Boxplot</i> $dEVI_2$ terhadap Amatan KSA..... | 41 |
| 6. Frekuensi Kategori Amatan KSA | 43 |
| 7. Matriks Korelasi Indeks-Indeks Spektral Landsat-8..... | 44 |
| 8. Nilai <i>Cross Validation</i> menggunakan $\log(\lambda)$ | 45 |
| 9. (a) Pendugaan Koefisien LASSO Logit ₁ , (b) Pendugaan Koefisien LASSO Logit ₂ , (c) Pendugaan Koefisien LASSO Logit ₃ , (d) Pendugaan Koefisien LASSO Logit ₄ , (e) Pendugaan Koefisien LASSO Logit ₆ , (f) Pendugaan Koefisien LASSO Logit ₇ , (g) Pendugaan Koefisien LASSO Logit ₈ | 47 |
| 10. (a) Nilai <i>Sensitivity</i> RLM dan RLM-LASSO, (b) Nilai <i>Specificity</i> RLM dan RLM-LASSO, (c) Nilai <i>Balanced Accuracy</i> RLM dan RLM-LASSO... | 57 |
| 11. Nilai Akurasi Model RLM dan RLM-LASSO..... | 58 |

DAFTAR LAMPIRAN

| Lampiran | Halaman |
|---|----------------|
| 1. Data Indeks Spektral Landsat-8 Provinsi Jawa Barat Tahun 2018 | 71 |
| 2. Nilai <i>Lambda</i> Optimal dan Log <i>Lambda</i> Optimal | 72 |
| 3. Estimasi Parameter RLM-LASSO | 73 |
| 4. <i>Confusion Matrix</i> dan Kriteria <i>Confusion Matrix</i> RLM-LASSO | 76 |
| 5. <i>Syntax</i> Metode LASSO Pada Model Regresi Logistik Multinomial | 77 |

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Pangan menjadi kebutuhan primer bagi manusia sehingga pemenuhan sumber pangan, khususnya tanaman pangan memerlukan perhatian yang serius (Suprpto, 2022). Menurut Hakim dan Irawan (2021) membangun ketahanan pangan secara strategis diperlukan untuk mendorong kemajuan ekonomi dan sosial. Ketahanan pangan sangat penting bagi pertumbuhan bidang perekonomian lainnya karena tidak ada negara yang dapat maju secara ekonomi kecuali krisis pangannya teratasi (Ali dkk, 2022). Oleh karena itu, ketahanan pangan merupakan komponen kunci dari kerangka pembangunan nasional dan sarana untuk mencapai tujuan *Sustainable Development Programs* (SDGs) atau tujuan pembangunan berkelanjutan yang kedua, yaitu memberantas kelaparan (*zero hunger*) (Darmawan, 2023).

Perencanaan kebijakan pangan nasional tidak terlepas dari kehadiran data pangan yang akurat. Data tanaman pangan khususnya data produksi padi di Indonesia sejak tahun 1997 diduga terjadi overestimasi sebesar 17.07 persen (Marsuhandi dkk, 2019). Overestimasi merupakan perkiraan mengenai sesuatu yang terlalu tinggi dibandingkan dengan nilai sebenarnya. Data overestimasi pada produksi padi dapat dihasilkan dari pengambilan data lapangan yang tidak berdasarkan hasil pengukuran, melainkan melalui Survei Pertanian (SP) lahan yang menggunakan pengamatan mata (*eye estimate*) (Kurniawati dkk, 2023). Hal ini dapat menimbulkan bias atau kesenjangan dalam data lapangan, sehingga menyebabkan beberapa kebijakan tidak efektif dan menyia-nyiakan sumber daya. Oleh karena itu, pemerintah melalui Badan Pusat Statistik (BPS) dan Badan

Pengkajian dan Penerapan Teknologi (BPPT) bekerja sama untuk memperbaiki metodologi pengumpulan data luas panen nasional dalam rangka meningkatkan keakuratan data pangan menggunakan metode Kerangka Sampel Area (KSA) pada tahun 2018 (BPS, 2018b).

Penerapan survei KSA di Indonesia tahun 2018 hanya mencakup komoditas padi. Survei ini dilaksanakan di seluruh provinsi di Indonesia dengan jumlah sampel segmen sebanyak 22.088 dan titik pengamatan sebanyak 198.792 (BPS, 2018a). Segmen merupakan wilayah yang akan disurvei dan di setiap segmen terdapat 9 subsegmen yang menjadi titik pengamatan fase tumbuh tanaman padi yang diamati setiap bulannya oleh petugas. Karena diamati setiap bulan, survei KSA mampu memberikan kualitas pendugaan yang tinggi. Namun, terdapat biaya yang besar untuk dapat mengamati setiap area sampel segmen di lapangan.

Pemerintah Indonesia telah mengalokasikan anggaran sekitar 64 miliar rupiah pertahun untuk melaksanakan survei KSA secara nasional (Kurniawati dkk, 2023). Oleh karena itu beberapa penelitian mencoba menggabungkan metode KSA dengan pengamatan citra satelit untuk meminimalisir biaya survei lapangan. Beberapa penelitian sebelumnya yang telah menggunakan pengamatan citra satelit untuk klasifikasi amatan KSA adalah penelitian Triscowati dkk (2019), Marsuhandi dkk, (2019) dan Kurniawati dkk (2023) dengan penggunaan citra satelit Landsat-8 terhadap hasil amatan KSA.

Citra satelit Landsat-8 adalah suatu satelit yang diluncurkan oleh *United States Geological Survey* (USGS) yang dapat diakses secara gratis melalui *platform Google Earth Engine* (GEE). Landsat-8 memiliki dua sensor, yaitu

sensor *Operasional Land Imager* (OLI) yang memiliki sembilan *band* dan *Thermal Infra Red Sensor* (TIRS) yang memiliki dua *band* (Pradipta dkk, 2019). Pencitraan oleh sensor Landsat-8 dapat dilihat pada Gambar 1.



Sumber : Sitanggang (2010:50)

Gambar 1. Gambaran Pencitraan Permukaan Bumi dengan Landsat-8

Gambar 1 menggambarkan pencitraan permukaan bumi oleh citra satelit Landsat-8 dengan menggunakan dua sensor, yaitu OLI dan TIRS. Sensor tersebut akan merekam frekuensi gelombang elektromagnetik yang dipantulkan oleh benda-benda dipermukaan bumi dengan resolusi tertentu (Sitanggang, 2010). Pencitraan satelit memiliki empat resolusi: spasial, spektral, temporal, dan radiometrik (Yusuf dan Rijal, 2017). Masing-masing berfungsi untuk menampilkan jumlah informasi yang dikumpulkan sensor, lebar kisaran dari masing-masing *band* spektral, interval waktu antar pengukuran, dan perbedaan dari energi yang dipantulkan objek. Kerincian informasi yang direkam dari data penginderaan jauh sangat bergantung pada resolusi tersebut (Yusuf dan Rijal, 2017).

Landsat-8 memiliki resolusi spasial 30 m dan resolusi temporal dengan waktu liput ulang 16 hari (Sitanggang, 2010). Resolusi radiometrik Landsat-8

pada sensor OLI memiliki resolusi sebesar 16 bit yang mampu membedakan objek-objek di permukaan bumi dengan sangat baik (Oktaviani dan Johan, 2016). Sedangkan resolusi spektral akan menghasilkan panjang gelombang pada setiap *band* spektral. Di mana *band* merupakan saluran yang merekam frekuensi gelombang satelit. Sebelas *band* yang terdapat pada Landsat-8 dapat disusun menjadi beberapa indeks spektral dengan berbagai fungsi.

Indeks spektral merupakan kombinasi dari dua atau lebih panjang gelombang yang digunakan untuk mengekstrak informasi terkait sifat spesifik permukaan bumi (Rouse dkk, 1974). Indeks spektral dapat digunakan untuk memantau keadaan vegetasi suatu wilayah, seperti pemantauan fase pertumbuhan padi. Beberapa indeks spektral yang bisa digunakan untuk memantau fase pertumbuhan padi adalah *Enhanced Vegetation Index* (EVI), *Normalized Difference Bare Index* (NDBI), dan *Modified Normalized Difference Water Index* (MNDWI) (Kurniawati dkk, 2023). Indeks EVI, NDBI, dan MNDWI masing-masing digunakan untuk mengamati kehijauan tanaman padi, mengidentifikasi fase panen dan menunjukkan kadar air yang ada untuk mengidentifikasi fase padi saat pertama kali ditanam. Saat ingin memodelkan klasifikasi fase pertumbuhan tanaman padi, ketiga indeks spektral tersebut dapat digunakan sebagai variabel prediktor.

Selain penggunaan indeks spektral, fase pertumbuhan padi tidak dapat diidentifikasi hanya dalam satu periode saja. Hal tersebut dikarenakan pertumbuhan padi yang berubah-ubah setiap waktunya. Sehingga penambahan periode waktu amatan dapat membantu klasifikasi kelas dengan tepat. Pada penelitian Kurniawati (2023) dan Marsuhandi dkk (2019) digunakan waktu

amatan tiga periode, yaitu waktu survei (t), satu periode sebelum survei ($t-1$), dan dua periode sebelum survei ($t-2$) yang cukup untuk membantu klasifikasi fase pertumbuhan padi. Penggunaan tiga periode waktu pengamatan temporal tersebut juga dapat meningkatkan keakuratan model klasifikasi (Kurniawati, 2023).

Pemodelan klasifikasi fase tanaman padi dalam amatan KSA yang menggunakan citra satelit banyak dilakukan menggunakan metode *Machine Learning*, seperti *Random Forest* dan *Boosting*. Pada penelitian Triscowati dkk (2019) menggunakan pendekatan *Random Forest* untuk memilih variabel terbaik pada klasifikasi fase tanaman padi dan diperoleh akurasi tertinggi sebesar 72.73%. Penelitian Marsuhandi dkk (2019) mencoba membandingkan algoritma klasifikasi *Random Forest* dan *Boosting* untuk mengklasifikasikan amatan KSA. Penelitian tersebut menemukan bahwa *Random Forest* lebih baik daripada *Bosting* dalam pengklasifikasian amatan KSA dengan akurasi sebesar 76,52%.

Penggunaan kedua metode *Machine Learning* tersebut mampu meningkatkan ketepatan klasifikasi, namun terbatas untuk melihat pendugaan parameter dan pengaruh signifikansi parameter dikarenakan kedua metode tersebut merupakan model *opaque*. Model *opaque* menghasilkan banyak parameter model yang tidak dapat diinterpretasikan karena tidak menampilkan nilai taksiran dan *p-value* dari parameter (Kurniawati dkk, 2024). Pengaruh dari parameter seringkali dibutuhkan, oleh karena itu Kurniawati (2023) mengembangkan metode klasifikasi lain yang digunakan pada klasifikasi KSA, yaitu model Regresi Logistik Multinomial (RLM). Pada penelitian tersebut model yang diperoleh sangat kompleks dengan sembilan parameter ditambah kategori

variabel respons yang bersifat *multiclass*. Sehingga, untuk mereduksi kompleksitas model diperlukan suatu metode pemilihan variabel.

Beberapa metode yang populer digunakan untuk pemilihan variabel dalam model linier adalah *best subset selection* dan *stepwise selection* (Hastie dkk, 2020). Metode seleksi menggunakan *best subset selection* dan *stepwise selection* dapat mengurangi ragam prediksi dengan mengorbankan sedikit bias. Kekurangan dari dua metode tersebut menurut adalah penduga model yang tidak stabil khususnya dalam Regresi Linier, di mana perubahan kecil pada data dapat menghasilkan model dan *subset* variabel yang berbeda (Soleh dan Aunuddin, 2013).

Alternatif lain metode pemilihan variabel adalah dengan menggunakan metode *shrinkage* yang merupakan suatu pendekatan untuk menyesuaikan model dengan mengatur estimasi parameter mendekati atau menjadi nol sehingga yang dapat mengurangi varians secara signifikan (James dkk, 2013). Adapun metode *shrinkage* yang paling terkenal menurut James dkk (2013) adalah Regresi *Ridge* dan *Least Absolute Shrinkage and Selection Operator* (LASSO). Regresi *Ridge* menghasilkan penduga model yang lebih stabil, namun relatif sulit diinterpretasi karena tidak mampu menyusutkan koefisien menjadi nol. Oleh sebab itu, dikembangkan metode *shrinkage* lainnya, yaitu LASSO untuk mengatasi masalah keakuratan interpretasi dan pendugaan model regresi dengan tetap mempertahankan keuntungan dari Regresi *Ridge*.

LASSO melakukan estimasi dengan meminimalkan jumlah kesalahan kuadrat dengan menyusutkan sejumlah parameter menjadi nol (James dkk, 2013). Hal tersebut membuat model lebih mudah diinterpretasi (*interpretable*) dan

variabel-variabel yang terdapat pada model hanyalah yang memiliki kontribusi (Khamidah dkk, 2024). Regresi LASSO memiliki model yang lebih stabil dibandingkan Regresi Linier dan Regresi *Ridge*, serta mampu mengatasi masalah multikolinearitas pada data dengan menyusutkan koefisien regresi yang berkorelasi tepat menjadi nol (Abriantoro dan Khana, 2023; Robbani dkk, 2019).

Selain mengatasi masalah multikolinearitas, LASSO juga biasa digunakan untuk seleksi variabel pada data berdimensi tinggi. Data dianggap berdimensi tinggi jika memiliki banyak variabel. Landsat-8 sendiri memiliki beberapa *band* kanal yang bisa disusun menjadi beberapa indeks spektral, sehingga data memiliki kompleksitas yang besar. Triscowati dkk (2019) yang menggunakan 7 *band* dan 4 indeks spektral dengan waktu amatan multitemporal 4 periode waktu menjadikan variabel yang digunakan berjumlah 44 variabel. Selanjutnya, pada penelitian Marsuhandi dkk (2019) menggunakan 11 *band* dan 4 indeks spektral, serta statistik dari *band* dengan 3 periode waktu temporal yang membuat lebih dari 50 variabel digunakan pada penelitian tersebut. Oleh karena itu, pemodelan yang menggunakan dengan data citra satelit Landsat-8 tergolong ke dalam *high dimensional* data, karena ada banyak band ataupun indeks spektral yang digunakan dalam pemodelan. Penelitian ini menggunakan sembilan variabel yang terdiri dari EVI, NBDI, dan MNDWI dalam tiga periode pengamatan temporal.

Penerapan LASSO tidak hanya digunakan pada model Regresi Linier dengan variabel respon yang bertipe numerik, namun juga pada Regresi Logistik dengan tipe data kategorik. Penelitian Kurniawati dkk (2020) menerapkan LASSO pada Regresi Logistik Biner yang menyederhanakan model dengan 13 parameter

menjadi hanya 8 parameter. Kemudian penelitian Efendi dan Ramadhan (2018) menerapkan LASSO pada Regresi Logistik Multinomial (RLM). Hasil dari penelitian tersebut menyederhanakan model dengan 12 variabel prediktor yang menjadikan model logit pertama hanya memiliki 7 variabel, model logit kedua memiliki 3 variabel, dan model logit ketiga memiliki 2 variabel yang berkontribusi. Penerapan LASSO pada Regresi Logistik mampu memberikan kualitas model yang baik terutama pada masalah multi klasifikasi dimana terdapat beberapa efek interaksi antara kovariat dan kumpulan peubah (Asenso dkk, 2022).

Penelitian ini mencoba menggunakan metode LASSO untuk menyeleksi variabel indeks spektral pada model klasifikasi amatan KSA yang bersifat *multiclass* dengan model data yang dikembangkan oleh Kurniawati (2023) menggunakan Regresi Logistik Multinomial (RLM). Penerapan metode LASSO digunakan untuk mereduksi kompleksitas model amatan KSA. Sehingga berdasarkan penelitian sebelumnya, dilakukan penelitian yang berjudul **“Penyeleksian Indeks Spektral Melalui Model Regresi Logistik Multinomial Dengan Metode *Least Absolute Shrinkage and Selection Operator* (LASSO)”**.

B. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Apa saja variabel indeks spektral Landsat-8 yang berkontribusi pada model Regresi Logistik Multinomial dengan metode LASSO (RLM-LASSO)?
2. Bagaimana kompleksitas pada model Regresi Logistik Multinomial dengan metode LASSO (RLM-LASSO) untuk memodelkan fase amatan KSA?

3. Bagaimana kinerja model Regresi Logistik Multinomial LASSO (RLM-LASSO) dalam mengklasifikasi fase amatan KSA dibandingkan model Regresi Logistik Multinomial (RLM) pada penelitian sebelumnya?

C. Batasan Masalah

Berdasarkan latar belakang masalah yang telah dikemukakan, maka batasan masalah dalam penelitian ini adalah seleksi variabel indeks spektral Landsat-8 selama tiga periode waktu.

D. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dirumuskan, maka tujuan dalam penelitian ini adalah sebagai berikut.

1. Mengetahui variabel indeks spektral apa saja yang berkontribusi pada model Regresi Logistik Multinomial dengan metode LASSO (RLM-LASSO).
2. Mengetahui kompleksitas pada model Regresi Logistik Multinomial dengan metode LASSO (RLM-LASSO) untuk memodelkan fase amatan KSA.
3. Mengetahui kinerja model Regresi Logistik Multinomial LASSO (RLM-LASSO) dalam mengklasifikasi fase amatan KSA dibandingkan model Regresi Logistik Multinomial (RLM) pada penelitian sebelumnya.

E. Manfaat Penelitian

Adapun manfaat yang diharapkan dari hasil penelitian ini adalah.

1. Bagi Peneliti, sebagai tambahan ilmu pengetahuan khususnya di bidang pertanian dan penginderaan jauh dengan melakukan seleksi variabel indeks spektral menggunakan model Regresi Logistik Multinomial dengan

metode LASSO (RLM-LASSO) untuk memodelkan fase Amatan KSA melalui data citra satelit Landsat-8.

2. Bagi Instansi Pertanian, sebagai acuan untuk mengambil tindakan strategis terhadap pemantauan fase pertumbuhan padi.
3. Bagi Pembaca, sebagai pengetahuan mengenai seleksi variabel menggunakan model Regresi Logistik Multinomial dengan metode LASSO (RLM-LASSO) yang dapat digunakan sebagai referensi untuk studi kasus lebih lanjut.