

**KLASIFIKASI PENERIMA PROGRAM KELUARGA HARAPAN
DI KOTA PADANG MENGGUNAKAN *K-NEAREST NEIGHBORS*
DENGAN *RECURSIVE FEATURE ELIMINATION CROSS
VALIDATION***



**YURIVO RIANDA SAPUTRA
NIM 20337063**

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

**KLASIFIKASI PENERIMA PROGRAM KELUARGA HARAPAN
DI KOTA PADANG MENGGUNAKAN *K-NEAREST NEIGHBORS*
DENGAN *RECURSIVE FEATURE ELIMINATION CROSS
VALIDATION***

SKRIPSI

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



**Oleh
YURIVO RIANDA SAPUTRA
NIM 20337063**

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

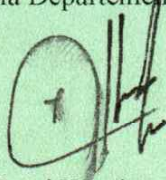
PERSETUJUAN SKRIPSI

KLASIFIKASI PENERIMA PROGRAM KELUARGA HARAPAN DI KOTA PADANG MENGGUNAKAN *K-NEAREST NEIGHBORS* DENGAN *RECURSIVE FEATURE ELIMINATION CROSS VALIDATION*

Nama : Yurivo Rianda Saputra
NIM : 20337063
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

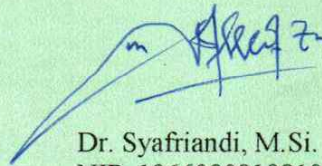
Padang, 30 Mei 2024

Mengetahui:
Kepala Departemen Statistika



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

Disetujui Oleh:
Pembimbing



Dr. Syafriandi, M.Si.
NIP. 196609081991031003

PENGESAHAN LULUS UJIAN SKRIPSI

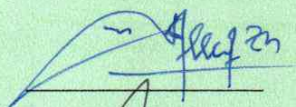
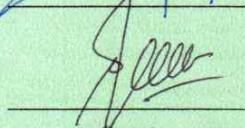
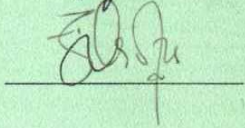
Nama : Yurivo Rianda Saputra
NIM : 20337063
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

KLASIFIKASI PENERIMA PROGRAM KELUARGA HARAPAN DI KOTA PADANG MENGGUNAKAN *K-NEAREST NEIGHBORS* DENGAN *RECURSIVE FEATURE ELIMINATION CROSS VALIDATION*

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	Tanda Tangan
Ketua	: Dr. Syafriandi, M.Si.	
Anggota	: Dr. Dony Permana, S.Si., M.Si.	
Anggota	: Zilrahmi, S.Pd., M.Si.	

SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertanda tangan dibawah ini :

Nama : Yurivo Rianda Saputra
NIM : 20337063
Program Studi : S1.Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan, bahwa skripsi saya dengan judul “Klasifikasi Penerima Program Keluarga Harapan di Kota Padang Menggunakan *K-Nearest Neighbors* dengan *Recursive Feature Elimination Cross Validation*” 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, 30 Mei 2024

Diketahui Oleh:
Kepala Departemen Statistika



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

Saya yang menyatakan,



Yurivo Rianda Saputra
NIM. 20337063

**Klasifikasi Penerima Program Keluarga Harapan di Kota Padang
Menggunakan *K-Nearest Neighbors* dengan *Recursive Feature Elimination
Cross Validation***

Yurivo Rianda Saputra

ABSTRAK

Program Keluarga Harapan (PKH) merupakan program bantuan sosial dari pemerintah yang bertujuan untuk memberikan perlindungan sosial dalam upaya pemerintah memajukan kesejahteraan sosial. PKH memberikan manfaat kepada keluarga miskin, khususnya ibu hamil dan anak, dengan memanfaatkan berbagai layanan kesehatan dan pendidikan yang tersedia. Manfaat PKH juga mencakup penyandang disabilitas dan lanjut usia dengan menjaga tingkat kesejahteraan sosialnya sesuai dengan Konstitusi dan Nawacita Republik Indonesia. Dalam penyaluran PKH terdapat kesalahan penyaluran yang menyebabkan kerugian negara hingga Rp 6,93 triliun. Untuk mengantisipasi agar PKH tersalurkan kepada yang benar membutuhkan, maka diperlukan suatu sistem untuk mengambil keputusan agar penyaluran PKH tepat sasaran.

Sistem yang dapat dimanfaatkan adalah algoritma *machine learning* dalam klasifikasi. Algoritma klasifikasi yang sederhana dan mudah dipelajari serta efektif jika data pelatihan berukuran besar adalah *K-Nearest Neighbors* (KNN). Klasifikasi dilakukan dengan membandingkan jumlah tetangga (k) pada KNN serta jumlah atribut hasil seleksi fitur dengan metode *Recursive Feature Elimination Cross Validation* (RFECV). Sebelum dilakukan klasifikasi perlu dilakukan penyeimbangan data dengan metode *Synthetic Minority Oversampling Technique Edited Nearest Neighbors* (SMOTEENN). Sumber data diperoleh dari data SUSENAS 2023 Kota Padang.

Hasil penelitian menunjukkan bahwa KNN dengan sepuluh atribut hasil RFECV dan nilai $k = 3$ merupakan algoritma dengan nilai *accuracy* dan *precision* terbesar. Sedangkan KNN dengan semua atribut dan nilai $k = 3$ merupakan algoritma dengan nilai *recall* terbesar. Perbandingan nilai evaluasi antara dua situasi dataset menghasilkan algoritma KNN dengan sepuluh atribut hasil RFECV lebih baik dibandingkan algoritma KNN dengan seluruh atribut dalam klasifikasi penerima PKH di Kota Padang.

Kata Kunci: PKH, Klasifikasi, KNN, RFECV, SMOTEENN

Classification of Program Keluarga Harapan Recipients in City of Padang Using K-Nearest Neighbors with Recursive Feature Elimination Cross Validation

Yurivo Rianda Saputra

ABSTRACT

Program Keluarga Harapan (PKH) is a social assistance program from the government aimed at providing social protection in the government's efforts to promote social welfare. PKH provides benefits to poor families, especially pregnant women and children, by utilizing various health and education services available. PKH benefits also include people with disabilities and the elderly by maintaining their level of social welfare in accordance with the Constitution and the Nawacita of the Republic of Indonesia. In the distribution of PKH there were distribution errors which caused state losses of up to IDR 6.93 trillion. To anticipate that PKH will be distributed to those who really need it, a system is needed to make decisions so that PKH distribution is right on target.

The system that can be utilized is by utilizing algorithms machine learning in classification. A simple, easy-to-learn, and effective classification algorithm for large training data is K-Nearest Neighbors (KNN). Classification is performed by comparing the number of neighbors (k) in KNN and the number of attributes resulting from feature selection using the Recursive Feature Elimination Cross Validation (RFECV) method. Before classification, data balancing is required using the Synthetic Minority Oversampling Technique Edited Nearest Neighbors (SMOTEENN) method. The data source is obtained from the 2023 SUSENAS data of Padang.

The research results show that KNN with ten RFECV attribute results and a value of $k = 3$ is an algorithm with value accuracy and precision biggest. Meanwhile, KNN with all attributes and value $k = 3$ is an algorithm with values recall biggest. A comparison of the evaluation values between the two dataset situations produces a KNN algorithm with ten RFECV attributes that is better than the KNN algorithm with all attributes in the classification of PKH recipients in Padang.

Keywords: PKH, Classification, KNN, RFECV, SMOTEENN

KATA PENGANTAR

Bismillahirrahmanirrahim

Alhamdulillah rabbil 'aalamiin, puji dan syukur penulis panjatkan kepada Allah SWT, yang telah memberikan hidayah, kemudahan dan nikmat kepada penulis dalam menjalankan amanah yang menjadi tanggung jawab penulis. Shalawat serta salam penulis panjatkan kepada baginda Rasulullah Nabi Muhammad SAW, karena dengan syafa'atnya telah membawa umatnya dari zaman jahiliyah ke zaman dengan ilmu pengetahuan saat ini.

Atas izin yang maha kuasa penulis dapat menyelesaikan skripsi dengan judul **“Klasifikasi Penerima Program Keluarga Harapan di Kota Padang Menggunakan *K-Nearest Neighbors* dengan *Recursive Feature Elimination Cross Validation*”**. Skripsi ini disusun untuk memenuhi persyaratan dalam memperoleh gelar Sarjana pada Program Studi Sarjana Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang. Penyusunan skripsi ini tidak terlepas dari bantuan dan kerjasama berbagai pihak. Terimakasih penulis sampaikan kepada:

1. Bapak Dr. Syafriandi, M.Si. sebagai Dosen Pembimbing Akademik dan sekaligus sebagai Dosen Pembimbing Skripsi.
2. Bapak Dr. Dony Permana, S.Si., M.Si. dan Ibu Zilrahmi, S.Pd., .M.Si. sebagai Dosen Penguji Skripsi.
3. Ibu Dr. Yenni Kurniawati, S.Si., M.Si. sebagai Kepala Departemen Statistika, serta Kepala Prodi Sarjana Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang.

4. Bapak dan Ibu Dosen serta Staf Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang yang telah membantu penulis selama menimba ilmu di Program Studi Statistika.
 5. Teristimewa kepada kedua orang tua tercinta, Bapak Djailani dan Ibu Delliyetti, dan kakak-kakak beserta keluarga yang telah memberikan doa, semangat dan dukungan secara moril dan materiil selama proses perkuliahan sampai dengan penyelesaian skripsi hingga saat ini.
 6. Teman-teman seperjuangan di program studi Sarjana Statistika Angkatan 2020 yang telah memberikan masukan, semangat dan dukungan kepada penulis dalam pembuatan skripsi ini.
 7. Semua pihak terkait yang telah ikut berkontribusi dalam penulisan skripsi ini.
- Semoga skripsi ini kedepannya dapat bermanfaat bagi penulis dan semua pihak. Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, penulis mengharapkan masukan dan saran dari pembaca. Akhir kata penulis ucapkan terima kasih

Padang, 30 Mei 2024

Yurivo Rianda Saputra

DAFTAR ISI

ABSTRAK	i
ABSTRACT	ii
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	9
C. Batasan Masalah.....	10
D. Tujuan Penelitian.....	10
E. Manfaat Penelitian.....	10
BAB II KERANGKA TEORITIS	12
A. Program Keluarga Harapan.....	12
B. <i>Machine Learning</i>	16
C. Klasifikasi.....	19
D. Pembagian Data.....	20
E. <i>Synthetic Minority Oversampling Technique Edited Nearest Neighbors</i>	20
F. <i>Recursive Feature Elimination Cross Validation</i>	24
G. <i>K-Nearest Neighbors</i>	31
H. <i>Confusion Matrix</i>	34
I. Penelitian Relevan.....	36
BAB III METODOLOGI PENELITIAN	39
A. Jenis Penelitian.....	39
B. Sumber Data.....	39
C. Variabel Penelitian	39
D. Struktur Data	41
E. Teknik Analisis Data.....	42

F. Diagram Alir	45
BAB IV HASIL DAN PEMBAHASAN	46
A. Deskripsi Data	46
B. Hasil Analisis	58
C. Pembahasan	67
BAB V KESIMPULAN DAN SARAN	71
A. Kesimpulan.....	71
B. Saran.....	71
DAFTAR PUSTAKA	73
LAMPIRAN.....	79

DAFTAR TABEL

Tabel	Halaman
1. <i>Confusion Matrix</i>	34
2. Variabel Penelitian	39
3. Struktur Data	41
4. Nilai <i>Cross Validation</i> Dari Setiap Jumlah Variabel	60
5. Nilai Koefisien Setiap Variabel	61
6. Pembagian Dataset	62
7. Nilai <i>Confusion Matrix</i> Dari Klasifikasi Menggunakan Semua Variabel	63
8. Nilai Evaluasi Dari KNN Menggunakan Semua Variabel.....	63
9. Nilai <i>Confusion Matrix</i> Dari Klasifikasi Menggunakan Variabel Hasil Seleksi Variabel	64
10. Nilai Evaluasi Dari KNN Menggunakan Variabel Hasil Seleksi Variabel...	65
11. Nilai Evaluasi Dari KNN Menggunakan Nilai $K = 3$	66

DAFTAR GAMBAR

Gambar	Halaman
1. Jumlah Penduduk Miskin (Ribuan Jiwa) Kabupaten/Kota di Sumatera Barat Tahun 2022 dan 2023.....	2
2. Ilustrasi SMOTEENN.....	23
3. Ilustrasi K-fold Cross Validation dengan $K = 10$	27
4. Ilustrasi SVM Linear.....	28
5. Ilustrasi KNN.....	33
6. Flowchart Penelitian.....	45
7. Rumah Tangga Penerima PKH.....	46
8. Rumah Tangga Penerima PKH Berdasarkan Status Kepemilikan Rumah.....	47
9. Rumah Tangga Penerima PKH Berdasarkan Banyak Anggota Rumah Tangga.....	48
10. Rumah Tangga Penerima PKH Berdasarkan Luas Lantai Rumah.....	48
11. Rumah Tangga Penerima PKH Berdasarkan Kepemilikan Rumah Lain.....	49
12. Rumah Tangga Penerima PKH Berdasarkan Bahan Atap Rumah.....	50
13. Rumah Tangga Penerima PKH Berdasarkan Bahan Dinding Rumah.....	51
14. Rumah Tangga Penerima PKH Berdasarkan Bahan Lantai Rumah.....	52
15. Rumah Tangga Penerima PKH Berdasarkan Daya Listrik.....	53
16. Rumah Tangga Penerima PKH Berdasarkan Sumber Air Minum.....	53
17. Rumah Tangga Penerima PKH Berdasarkan Bahan Bakar Masak.....	54
18. Rumah Tangga Penerima PKH Berdasarkan Kepemilikan Kulkas.....	55
19. Rumah Tangga Penerima PKH Berdasarkan Kepemilikan AC.....	55
20. Rumah Tangga Penerima PKH Berdasarkan Kepemilikan Telepon.....	56
21. Rumah Tangga Penerima PKH Berdasarkan Kepemilikan Motor.....	57
22. Rumah Tangga Penerima PKH Berdasarkan Kepemilikan Mobil.....	57
23. Rumah Tangga Penerima PKH Berdasarkan Kepemilikan TV.....	58
24. Proporsi Data Setelah Diseimbangkan.....	59
25. Nilai Evaluasi Dari KNN Menggunakan Semua Variabel.....	64
26. Nilai Evaluasi Dari KNN Menggunakan Variabel Hasil Seleksi Variabel.....	65
27. Nilai Evaluasi Dari KNN Menggunakan Nilai $K = 3$	66

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Data Penelitian	79
2. Syntax Penelitian dengan Google Colab.....	80
3. Link Google Colab Syntax Penelitian.....	85

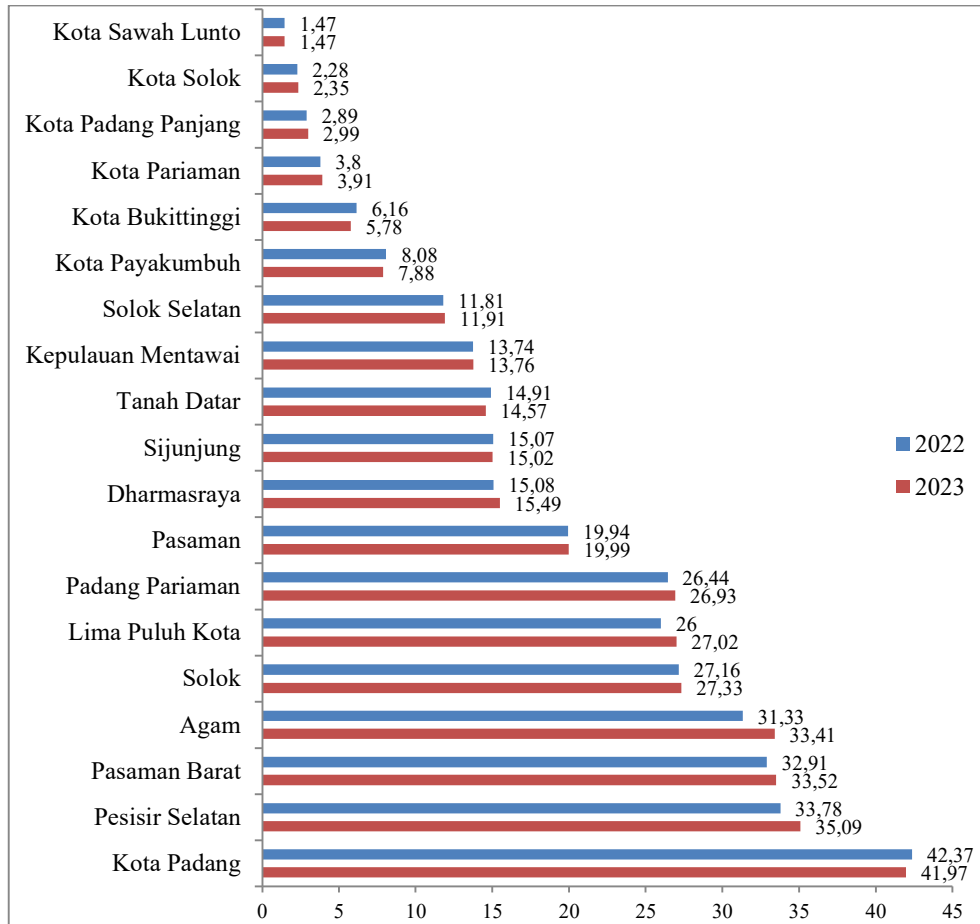
BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Badan Pusat Statistik (BPS), memproyeksikan jumlah penduduk Indonesia tahun 2023 sebanyak 278,7 juta jiwa (Badan Pusat Statistik, 2023). Jumlah tersebut naik sebesar 1,1% dibandingkan tahun 2022. Penduduk Indonesia sangat beragam dengan berbagai latar belakang budaya, etnis, sosial dan ekonomi yang menimbulkan banyak masalah. Permasalahan yang banyak dihadapi pemerintah dan masyarakat Indonesia terutama dalam aspek ekonomi seperti kemiskinan. Kemiskinan merupakan permasalahan global. Menurut Quraisy & Arifin (2016), kemiskinan merupakan keadaan dimana terjadinya kekurangan dalam memenuhi kebutuhan pokok seperti makanan, pakaian, tempat berlindung, dan hal-hal yang berkaitan dengan kualitas hidup. BPS mencatat jumlah penduduk miskin di Indonesia pada tahun 2023 sebesar 25.899 ribu jiwa, yang mengalami penurunan sebesar 262,61 ribu jiwa dibandingkan tahun 2022. Hal ini berbanding terbalik dengan Provinsi Sumatera Barat yang mengalami kenaikan. Pada tahun 2023, jumlah penduduk miskin di Sumatera Barat sebesar 340,37 ribu jiwa, yang mengalami kenaikan sebesar 5,16 ribu jiwa dibandingkan tahun 2022.

Pada tahun 2023, Jumlah penduduk miskin di Kota Padang sebesar 41,37 ribu jiwa yang mengalami penurunan sebesar 400 jiwa dibandingkan tahun 2022. Pada tahun 2022 dan 2023, Kota Padang merupakan kota dengan jumlah penduduk miskin terbanyak di Sumatera Barat. Jumlah penduduk miskin kabupaten/kota di Sumatera Barat tahun 2022 dan 2023 disajikan pada Gambar 1.



Sumber : (Badan Pusat Statistik Sumatera Barat, 2023)

Gambar 1. Jumlah Penduduk Miskin (Ribu Jiwa) Kabupaten/Kota di Sumatera Barat Tahun 2022 dan 2023

Persentase kemiskinan yang terus mengalami kenaikan, akan mengakibatkan kesulitan untuk mendapatkan kesetaraan dan kesejahteraan kehidupan masyarakat. Menurut Undang-undang Republik Indonesia Nomor 11 Tahun 2009, pemerintah berupaya untuk melakukan penyelenggaraan kesejahteraan sosial untuk memenuhi kebutuhan dasar setiap warga negara, yang meliputi rehabilitasi sosial, jaminan sosial, pemberdayaan sosial, dan perlindungan sosial. Berdasarkan informasi publik dari Kementerian Keuangan Republik Indonesia, dalam masa krisis, pemerintah memberikan perlindungan sosial untuk menjaga daya beli masyarakat yang terdampak, terutama masyarakat miskin dan rentan (Kementerian Keuangan,

2022). Perlindungan sosial merupakan usaha pemerintah yang bertujuan untuk memberikan dukungan kepada masyarakat agar mampu mengatasi berbagai kerentanan atau guncangan yang mungkin terjadi sepanjang perjalanan kehidupan. Salah satu bentuk perlindungan sosial adalah program pemberian bantuan yang bersifat tanpa iuran yang bersumber dari APBN dan/atau APBD dengan sasaran masyarakat miskin dan rentan. Program bantuan sosial yang diberikan oleh pemerintah seperti Program Keluarga Harapan (PKH), Kartu Sembako, Program Indonesia Pintar (PIP), Kartu Indonesia Pintar (KIP) Kuliah, dan Bantuan Iuran Jaminan Kesehatan Nasional (JKN).

Berdasarkan situs Kementerian Sosial Republik Indonesia, PKH memberikan bantuan kepada keluarga miskin yang ditetapkan sebagai keluarga penerima manfaat PKH sejak tahun 2007. PKH memberikan manfaat kepada keluarga miskin terutama ibu hamil dan anak untuk memanfaatkan berbagai fasilitas layanan kesehatan dan pendidikan yang tersedia. Manfaat PKH juga mencakup penyandang disabilitas dan lanjut usia dengan mempertahankan taraf kesejahteraan sosialnya sesuai dengan amanat konstitusi dan Nawacita Republik Indonesia. Bantuan PKH dibagikan berdasarkan Data Terpadu Kesejahteraan Sosial (DTKS) yaitu basis data yang digunakan pemerintah untuk menyalurkan bantuan sosial. Berdasarkan publikasi Kementerian Sosial, anggaran bantuan sosial Pemulihan Ekonomi Nasional (PEN) tahun 2021 sebesar Rp 101,4 triliun, anggaran yang peruntukan dalam bantuan PKH sebesar Rp 28,3 triliun (Koesworo, 2022). Tahun 2021, Badan Pemeriksa Keuangan (BPK) menemukan kesalahan dalam penyaluran bantuan sosial dalam program perlindungan sosial pemerintah yang menyebabkan kerugian negara hingga Rp 6,93 triliun

(Sembiring, 2022). Pada penelusuran BPK tersebut ditemukan penyebab terjadinya kesalahan penyaluran bantuan sosial ini terjadi karena berkaitan dengan pembaharuan pada DTKS.

Berdasarkan rekap data DTKS dari Dinas Sosial Kota Padang, sebanyak 31.881 keluarga di Kota Padang menerima bantuan sosial PKH dari Kementerian Sosial pada tahun 2023. Berdasarkan data jumlah penduduk miskin dan data penerimaan bantuan PKH pada tahun 2023, masih banyak terdapat keluarga miskin di Kota Padang yang belum menerima bantuan sosial dari Pemerintah. Penyaluran bantuan yang diamanatkan pemerintah haruslah tepat sasaran sesuai kriteria penilaian, agar tujuan pemberian bantuan PKH dapat tercapai sesuai yang diharapkan. Oleh karena itu, diperlukan metode yang dapat mengklasifikasi rumah tangga yang berhak sebagai penerima PKH, agar pemberian bantuan PKH dari Pemerintah tersampaikan kepada keluarga yang benar-benar membutuhkan.

Pengklasifikasian dapat dilakukan dengan memanfaatkan Statistika. Statistika terus mengalami perkembangan sejalan dengan kemajuan dan pemanfaatan teknologi untuk mempermudah pekerjaan manusia. Pengembangan Statistika yang dapat dimanfaatkan dalam pengklasifikasian data yang berukuran besar adalah *machine learning*. *Machine learning* adalah kecerdasan buatan yang dikembangkan dengan pemanfaatan disiplin ilmu statistika dan matematika untuk memungkinkan komputer belajar dari data dan meningkatkan kemampuannya dalam pengolahan data berukuran besar. Menurut Nelli (2015), *machine learning* adalah disiplin yang berhubungan dengan studi tentang metode untuk pengenalan pola dalam kumpulan data untuk analisis data. Secara khusus, ini berkaitan dengan pengembangan algoritma yang belajar dari data dan membuat prediksi.

Berdasarkan teknik pembelajarannya, *machine learning* dapat dibedakan menjadi *supervised learning* dan *unsupervised learning*.

Menurut Goldstein dkk., (2017), *supervised learning* merupakan metode yang paling populer dalam implementasi algoritma untuk *machine learning*. Dimana mesin dilatih untuk memberikan keluaran yang sudah ditetapkan atau diharapkan sebelumnya. *Supervised learning* terbagi dua yaitu, regresi yang menghasilkan keluaran berupa nilai kontinu dan klasifikasi yang menghasilkan keluaran berupa klasifikasi pada dua atau lebih kategori. Algoritma yang termasuk ke dalam regresi adalah Regresi Linier, Regresi Spasial, Regresi Data Panel, *Multivariate Adaptive Regression Spline* (MARS) dan *Support Vector Regression* (SVR). Algoritma yang termasuk dalam klasifikasi adalah Regresi Logistik, *Decision Tree*, *Naïve Bayes Classifier* (NBC), *Support Vector Machine* (SVM), *Random Forest*, dan *K-Nearest Neighbors* (KNN).

Menurut Igual & Seguí (2017), *unsupervised learning* merupakan teknik yang bertujuan untuk merangkum, menjelaskan dan mengidentifikasi struktur yang terkandung dalam data tanpa menggunakan target. Target merupakan variabel atau output yang diperoleh dari proses pengolahan data. *Unsupervised learning* adalah algoritma pembelajaran yang dapat digunakan untuk pengelompokan data dan reduksi dimensi. Pengelompokan data atau yang lebih dikenal dengan *clustering* bertujuan untuk menemukan kelompok atau kelompok data yang mirip dalam sebuah dataset dan reduksi dimensi bertujuan untuk mengurangi dataset dengan dimensi tinggi menjadi dataset dengan dimensi yang lebih rendah. Algoritma yang termasuk ke dalam *clustering* adalah *K-Mean*, *K-Medoids*,

Cluster Berhierarki. Algoritma yang termasuk ke dalam reduksi dimensi adalah *Principal Component Analysis* (PCA) dan *Linear Discriminant Analysis* (LDA).

Menurut Bhatia (2010), KNN merupakan algoritma yang sederhana dan mudah dipelajari serta efektif jika data pelatihan berukuran besar. Menurut Salvador-Meneses dkk., (2019), KNN adalah metode klasifikasi berbasis jarak dengan prinsip dasar yang sederhana dan dapat diimplementasikan secara luas. Menurut Hutahaean & Wijayanto (2022), label kelas sampel ditentukan berdasarkan kesamaannya. Selain itu, diperlukan jarak terukur untuk perhitungannya. Menurut Tharwat dkk., (2018), KNN merupakan algoritma klasifikasi yang berbasis contoh atau non parametrik dan dianggap teknik paling sederhana dalam data mining. KNN merupakan teknik pengklasifikasian data yang memiliki konsistensi yang kuat, dengan cara menghitung kedekatan antara data baru dengan data lama yang disebut dengan k . Metode pengukuran *euclidian distance* sebagai metode pengukuran untuk menghitung jarak terdekat antar sampel yang sering digunakan dalam algoritma KNN. KNN dapat bekerja dalam mengklasifikasikan data linear atau non-linear dan terdapat *outlier*. KNN dapat bekerja pada data yang memiliki *outlier* karena data tersebut akan diklasifikasikan dalam kelas terdekatnya berdasarkan jarak.

Machine learning banyak digunakan pada data berukuran besar, yang memiliki atribut dan amatan yang banyak. Istilah atribut yang digunakan dalam *machine learning* memiliki makna yang sama dengan variabel. Menurut Bharti & Singh (2014), data berdimensi tinggi mengakibatkan meningkatnya kompleksitas klasifikasi dan mengurangi kemampuan model. Melakukan seleksi pada variabel dapat dilakukan untuk mengurangi kompleksitas dan meningkatkan kemampuan

model. Menurut Okun (2011), melakukan seleksi pada variabel dapat digunakan untuk menghapus variabel yang tidak signifikan terhadap model. Metode untuk melakukan seleksi pada variabel adalah metode seleksi fitur. Istilah fitur yang digunakan dalam seleksi fitur memiliki makna yang sama dengan variabel. Metode untuk seleksi variabel terbagi menjadi tiga yaitu *filter method*, *wrapper method*, dan *embedded method*. Metode seleksi variabel tersebut tergantung pada interaksinya dengan algoritma *machine learning* yang digunakan.

Menurut Tang dkk., (2014), metode *wrapper* merupakan cara yang sederhana dan ampuh dalam seleksi variabel, dalam semua algoritma *machine learning*. Menurut Suppers dkk., (2018), metode *wrapper* merupakan metode dalam seleksi variabel yang bekerja dengan melakukan klasifikasi dengan semua variabel dan menggunakan akurasi model untuk memilih dan mengurangi variabel secara berulang. Metode seleksi variabel yang termasuk dalam metode *wrapper* adalah *Sequential Forward Selection (SFS)*, *Backward Elimination (BE)*, dan *Recursive Feature Elimination (RFE)*. Menurut Senan dkk., (2021), RFE sering digunakan karena kemudahan penggunaan dan konfigurasinya serta efektif dalam pemilihan variabel pada data *training*. RFE bekerja dengan mengurangi satu-persatu variabel dan mempertahankan variabel yang efektif terhadap klasifikasi. Dalam beberapa kasus jumlah variabel yang efektif terhadap klasifikasi belum diketahui. Menurut Artur (2021), *Cross Validation (CV)* dapat digunakan untuk mengatasi ketidak tahuan jumlah variabel yang efektif terhadap akurasi. Penggabungan antara *Recursive Feature Elimination (RFE)* dan *Cross Validation (CV)* disebut dengan *Recursive Feature Elimination Cross Validation (RFECV)*. Berdasarkan penjelasan mengenai kelebihan dari algoritma KNN dan seleksi variabel dengan

RFECV, maka peneliti memilih metode ini untuk digunakan dalam menentukan faktor yang efektif dalam menentukan klasifikasi rumah tangga penerima bantuan PKH.

Berbagai penelitian telah dilakukan pada kelayakan penerima PKH dengan menggunakan perkembangan teknologi terutama di bidang statistika. Menurut Alfiah (2021), dalam melakukan klasifikasi penerima bantuan sosial PKH menggunakan metode NBC variabel yang paling berpengaruh adalah kepemilikan kapal, laptop, perahu motor, telepon, lemari es, cara memperoleh air minum, sumber penerangan, jenis atap, status lahan, jenis kloset, jenis lantai, aset tak bergerak dan jumlah anggota rumah tangga. Menurut Meilaniwati & Fauzan (2022), klasifikasi penduduk miskin penerima PKH menggunakan metode NBC dan KNN memperoleh Metode KNN lebih baik dengan variabel yang paling berpengaruh adalah usia, status kehamilan, pendidikan tertinggi kepala rumah tangga (KRT), kepemilikan aset bergerak, dan kepemilikan aset tidak bergerak. Menurut Dina dkk., (2023), dalam perbandingan algoritma NBC, KNN, dan C4.5 untuk klasifikasi penerima bantuan PKH diperoleh metode algoritma C4.5 lebih baik dengan variabelnya adalah jumlah anggota rumah tangga (ART), fasilitas buang air besar, rumah lain, ada emas, ada lemari es, jumlah kamar, dinding, dan pembuangan tinja.

Faktor-faktor yang mempengaruhi rumah tangga penerima PKH dari beberapa penelitian tersebut terdapat pada variabel-variabel dalam DTKS. Menurut Helmizar dkk., (2021), terdapat 6 kelompok variabel pada DTKS. Kelompok variabel dalam DTKS meliputi identitas rumah tangga, perumahan rumah tangga, aset rumah tangga, program jaring pengaman sosial, demografi

anggota rumah tangga, dan pendidikan anggota rumah tangga. Pada penelitian ini akan digunakan variabel yang terdapat pada kelompok identitas rumah tangga, perumahan rumah tangga, dan aset rumah tangga.

Penelitian yang akan dilakukan membahas mengenai implementasi metode RFECV dalam melakukan seleksi variabel untuk proses klasifikasi menggunakan algoritma KNN dalam mengklasifikasikan rumah tangga yang termasuk ke dalam penerima PKH di Kota Padang. Penelitian ini diharapkan dapat memperoleh variabel yang dapat mempengaruhi keluarga penerima PKH dan metode klasifikasi yang memiliki nilai akurasi yang tinggi. Data yang digunakan adalah data survey sosial dan ekonomi nasional (SUSENAS) tahun 2023 yang nantinya berfokus pada data rumah tangga yang berada di Kota Padang. Berdasarkan penjelasan di atas, penulis melakukan penelitian dengan judul **“Klasifikasi Penerima Program Keluarga Harapan di Kota Padang Menggunakan K-Nearest Neighbors dengan Recursive Feature Elimination Cross Validation”**.

B. Rumusan Masalah

Berdasarkan latar belakang yang dipaparkan, rumusan masalah dalam penelitian ini sebagai berikut.

1. Bagaimana pengaruh implementasi metode RFECV pada klasifikasi rumah tangga penerima bantuan PKH di Kota Padang dengan menggunakan algoritma KNN?
2. Apa faktor-faktor yang menentukan rumah tangga, sehingga dapat diklasifikasikan sebagai rumah tangga yang berhak menerima bantuan PKH di Kota Padang?

C. Batasan Masalah

Berdasarkan latar belakang yang dipaparkan, batasan masalah dalam penelitian ini sebagai berikut.

1. Data yang digunakan adalah Data rumah tangga di Kota Padang tahun 2023, yang bersumber dari Data SUSENAS di Kota Padang.
2. Variabel yang digunakan dalam penelitian ini adalah variabel yang tersedia dalam Data SUSENAS di Kota Padang tahun 2023 dan berpedoman pada variabel yang terdapat dalam DTKS.

D. Tujuan Penelitian

Berdasarkan rumusan masalah yang dipaparkan, tujuan dalam penelitian ini sebagai berikut.

1. Mengetahui performa algoritma KNN dalam mengklasifikasikan rumah tangga penerima bantuan PKH di Kota Padang dengan implementasi metode RFECV.
2. Mengetahui faktor-faktor yang menentukan rumah tangga berhak menerima bantuan PKH di Kota Padang.

E. Manfaat Penelitian

Dari penelitian ini diharapkan manfaat sebagai berikut.

1. Bagi peneliti, penelitian ini bermanfaat untuk menambah wawasan dan pengalaman dalam menganalisis data dengan menerapkan teori yang telah dipelajari. Penelitian ini juga diharapkan dapat memenuhi salah satu syarat lulus Pendidikan Stata I Departemen Statistika Universitas Negeri Padang.
2. Bagi lembaga, penelitian ini dapat membantu lembaga pemerintahan dalam mengklasifikasikan rumah tangga penerima bantuan PKH di Kota Padang

dengan mengetahui faktor-faktor yang menentukan klasifikasi rumah tangga penerima PKH.

3. Bagi pembaca, penelitian ini diharapkan dapat menjadi referensi tambahan dan wawasan dalam kelebihan dan kekurangan masing-masing algoritma dalam pengklasifikasian penerima bantuan PKH.