

## DAFTAR ISI

Halaman Judul	i
Halaman Pengesahan Pembimbing	ii
Pernyataan	iii
Kata Pengantar	iv
Daftar Isi	vi
Daftar Gambar	viii
Daftar Tabel	x
Daftar Lampiran	xi
Arti Lambang dan Singkatan	xii
Intisari	xiv
Abstract	xvi
<b>BAB I. PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang Masalah .....	1
1.2. Permasalahan.....	4
1.3. Perumusan Masalah.....	4
1.4. Tujuan Penelitian.....	5
1.5. Keaslian dan Temuan Penelitian.....	5
<b>BAB II. TINJAUAN PUSTAKA.....</b>	<b>6</b>
2.1. Klasterisasi K-Means .....	7
2.2. Klasterisasi K-Means Data Nonlinear Separable.....	9
2.3. Kernel Principal Component Analysis (KPCA) .....	10
2.4. Bandwidth KPCA.....	10
2.5. Peta Penelitian.....	13
<b>BAB III. LANDASAN TEORI .....</b>	<b>14</b>
<b>3.1. Matriks dan Ruang Vektor .....</b>	<b>14</b>
<b>3.2. Jarak Pengukuran.....</b>	<b>20</b>
<b>3.3. Klasterisasi K-Means .....</b>	<b>20</b>
3.3.1. Penentuan Banyaknya Klaster .....	22
3.3.2. Seleksi Initial Sentroid .....	22
3.3.3. Sifat-sifat Algoritma Klasterisasi K-Means .....	22
3.3.4. Validasi Klasterisasi K-Means .....	23
<b>3.4. Trik Kernel .....</b>	<b>26</b>
3.4.1. Fungsi Kernel .....	27
3.4.2. Trik Kernel .....	27
<b>3.5. Kernel Principal Component Analysis .....</b>	<b>30</b>
3.5.1. Data Linearly Separable .....	30
3.5.2. Principal Component Analysis (PCA).....	32
3.5.3. PCA di Ruang Fitur.....	35
3.5.4. Kernel Principal Component Analysis (KPCA).....	36

<b>BAB IV. OPTIMISASI KLASTERISASI K-MEANS.....</b>	<b>39</b>
4.1. Jarak Euclid Pada Klasterisasi K-Means.....	39
4.2. Standarisasi Data Pada Klasterisasi K-Means.....	40
4.3. Pemetaan Data Nonlinear Separable.....	41
4.4. Reduksi Dimensi dengan Principal Component Analysis (PCA).....	44
4.5. Bandwidth Kernel Gaussian.....	47
4.6. Validitas dan Konsistensi Hasil Klasterisasi.....	48
4.7. Algoritma Optimisasi Klasterisasi K-Means Berdasarkan Ukuran Bandwidth KPCA.....	49
<b>BAB V. SIMULASI OPTIMISASI UKURAN BANDWIDTH.....</b>	<b>59</b>
5.1. Deskripsi Fisher's <i>Iris</i> Data.....	59
5.2. Standarisasi Data.....	62
5.2.1. Entropy dan SSW pada Klasterisasi K-Means Terhadap Data Original dan Data Terstandar.....	63
5.2.2. Pengaruh Standarisasi Terhadap Validitas Klasterisasi K-Means di Ruang Fitur.....	63
5.3. Banyaknya PC Pada Ruang Fitur.....	65
5.4. Hasil Dan Pembahasan.....	66
5.4.1. Hasil Simulasi $n = 30$ .....	66
5.4.2. Interpolasi Besaran Bandwidth KPCA untuk Optimisasi Klasterisasi K-Means .....	68
5.4.3. Besaran Bandwidth KPCA untuk Optimisasi Klasterisasi K-Means.....	71
<b>BAB VI. KESIMPULAN DAN SARAN.....</b>	<b>75</b>
6.1. Kesimpulan.....	75
6.2. Saran.....	76
6.3. Masalah Terbuka.....	76
<b>DAFTAR PUSTAKA.....</b>	<b>77</b>
<b>LAMPIRAN .....</b>	<b>85</b>

## DAFTAR GAMBAR

Gambar 2.1. Peta Penelitian.....	13
Gambar 3.1. Ilustrasi perpindahan data (●) dan sentroid (+) antar kluster pada algoritma klasterisasi K-Means.....	22
Gambar 3.2. Entropy : kecocokan label kluster dan label kelas .....	24
Gambar 3.3. Sum of Squared Within (SSW) dan Sum of Squared Between (SSB) .....	25
Gambar 3.4. Plot data awal di $\mathbb{R}^2$ (kiri) dan hasil pemetaan di $\mathbb{R}^3$ (kanan).....	27
Gambar 3.5. Plot data awal di ruang input $\mathbb{R}^2$ .....	29
Gambar 3.6. Data linearly separable.....	31
Gambar 3.7. Data nonlinearly separable.....	31
Gambar 3.8. Pemetaan data nonlinearly separable pada ruang input menjadi linearly separable di ruang fitur.....	32
Gambar 3.9. Plot data awal di $\mathbb{R}^2$ (kiri) dan hasil pemetaan KPCA (kanan) dengan $K_{ij} = K(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ ^2}{2\gamma^2}\right), \gamma = 1;$ menggunakan 3 PC pertama.....	36
Gambar 4.1. Global maximum/minimum dan lokal maximum/minimum suatu fungsi.....	42
Gambar 4.2. Ilustrasi dua kluster pada ruang 2 dimensi, berbentuk bulat dengan densitas yang sama a.terpisahkan dengan baik, b. tidak terpisahkan dengan baik.....	43
Gambar 4.3. Pemetaan dari ruang input ke ruang fitur yang berdimensi lebih tinggi dan solusinya pada ruang input.....	44
Gambar 4.4. Rotasi sistem koordinat dari variabel asal $X_1$ dan $X_2$ ke variabel baru $Y_1$ dan $Y_2$ sebagai kombinasi linear dari $X_1$ dan $X_2$ pada PCA.....	45
Gambar 4.5. Plot varians dari Iris dataset pada tiap Principal Component .....	46
Gambar 4.6. Flowchart penentuan besaran bandwidth kernel Gaussian $\gamma$ untuk optimisasi hasil klasterisasi K-Means.....	55
Gambar 4.7. Flowchart penentuan banyaknya PC yang digunakan pada klasterisasi K-Means di ruang fitur $\mathcal{F}$ .....	58
Gambar 5.1. Deskripsi petal dan sepal kelas tanaman Iris: Setosa, Versicolor dan Virginica .....	59
Gambar 5.2. Plot densitas kelas Iris dataset vs Sepal.Length, Sepal.Width, Petal.Length dan Petal.Width.....	60
Gambar 5.3. Plot 3D Iris dataset $X_1$ = Sepal.Length, $X_2$ = Sepal.Width dan $X_3$ = Petal.Length.....	60
Gambar 5.4. Plot 3 PC pertama dengan varian kumulatif : 99.48%.....	61

Gambar 5.5. Plot 3 PC pertama dengan bandwidth kernel Gaussian $\gamma = 0.1$ .....	61
Gambar 5.6. (a) Entropy, (b) Percentage SSW dari klasterisasi K-Means pada data original dan data terstandar .....	63
Gambar 5.7. Persentase SSW dari klasterisasi K-Means menggunakan 1,2 & 3 PCs untuk $\gamma = 0.2$ pada data original dan terstandar.....	64
Gambar 5.8. Entropy dari klasterisasi K-Means menggunakan 2 PCs dan $\gamma = 0.2$ pada data original dan data terstandar .....	64
Gambar 5.9. Plot Rata-rata SSW pada klasterisasi K-Means di ruang fitur pada beberapa PC untuk nilai bandwidth kernel Gaussian $\gamma=0.1, 0.2, 0.3$ dan $0.4$ .....	65
Gambar 5.10. Plot nilai Entropy pada klasterisasi K-Means di ruang Fitur pada beberapa PC untuk nilai bandwidth kernel Gaussian $\gamma=0.1, 0.2, 0.3$ dan $0.4$ .....	66
Gambar 5.11. Boxplots dari Entropy hasil klasterisasi K-Means pada beberapa besaran bandwidth .....	67
Gambar 5.12. Boxplots dari persentase SSW hasil klasterisasi K-Means pada beberapa ukuran bandwidth KPCA.....	68
Gambar 5.13. Plot fungsi polinomial nilai Entropy derajat 4 sampai dengan derajat 10 pada ukuran bandwidth $\gamma=0.1, 0.2, \dots, 1.0$ .....	69
Gambar 5.14. Plot fungsi polinomial nilai entropy derajat 9 pada ukuran bandwidth $\gamma=0.1, 0.2, \dots, 1.0$ .....	70
Gambar 5.15. Plot fungsi polinomial persentase SSW derajat 4 sampai dengan derajat 10 pada ukuran bandwidth $\gamma=0.1, 0.2, \dots, 1.0$ .....	70
Gambar 5.16. Plot fungsi polinomial persentase SSW derajat 9 pada ukuran bandwidth $\gamma=0.1, 0.2, \dots, 1.0$ .....	71
Gambar 5.17. Plot perbandingan rata-rata nilai entropy pada data original, Data terstandar dan pemetaan KPCA untuk bandwidth kernel Gaussian $\gamma= 0.20, 0.24, 0.25, 30$ .....	73
Gambar 5.18. Plot perbandingan rata-rata nilai SSW pada data original, data terstandar dan pemetaan KPCA untuk bandwidth kernel Gaussian $\gamma= 0.20, 0.24, 0.25, 30$ .....	73

## DAFTAR TABEL

Tabel 2.1. Perbandingan Penelitian.....	11
Tabel 3.1. Matriks kernel hasil pemetaan data $(x, y) = \{(5,13), (10,12), (7,8), (7,7), (10,2), (5,1), (8,8), (8,7)\}$ ke ruang Fitur $\mathcal{F}$ .....	30
Tabel 3.2. Hasil Pemetaan Nonlinear $\Phi(\mathbf{x}) = e^{- (x-t) ^2}$ pada Ruang Input.....	32
Tabel 3.3. Skor Komponen 8 PC dari Matriks Kernel di Ruang Fitur.....	35
Tabel 5.1. Rata-rata dan Varians Perhitungan Entropy dan SSW Pada Beberapa Besaran Bandwidth.....	67
Tabel 5.2. Residual Fungsi Polinomial Nilai Entropy, derajat 4 sampai dengan derajat 10 pada besaran Bandwidth $\gamma=0.1, 0.2, \dots, 1.0$ .....	69
Tabel 5.3. Residual Fungsi Polinomial Persentase SSW, derajat 4 sampai dengan derajat 10 pada besaran Bandwidth $\gamma=0.1, 0.2, \dots, 1.0$ .....	71
Tabel 5.4. Rata-rata dan Varians Perhitungan Entropy dan SSW (%) Pada Hasil klasterisasi K-Means data original, data terstandar, dan pemetaan KPCA dengan bandwidth kernel Gaussian $\gamma=0.20, 0.24, 0.25$ dan $0.30$ , pengulangan prosedur sebanyak 100 kali.....	72

## DAFTAR LAMPIRAN

Lampiran 1. Fisher's <i>Iris</i> Dataset (Fisher 1936) .....	85
Lampiran 2. Deskripsi Fisher's <i>Iris</i> Dataset.....	89
Lampiran 3. R Script untuk K-Means di ruang Fitur dengan pemetaan KPCA, data terstandarisasi, $\sigma = \gamma$ , banyaknya Features = #PC, running n kali.....	93
Lampiran 4. Matlab Script untuk SSW: Menentukan model terbaik melalui nilai residual terkecil SSW.....	93
Lampiran 5. Matlab Script untuk Entropy: Menentukan model terbaik melalui nilai residual terkecil Entropy.....	96
Lampiran 6. Data hasil simulasi perhitungan validitas eksternal dan internal pada beberapa format data (#running =100 x).....	98
Lampiran 7. Program Dan Hasil Perhitungan Skor Komponen Pada PCA Terhadap Data Pada Ruang Fitur Hasil Pemetaan Kernel Gaussian Dengan Bandwidth $\gamma=0.1$ .....	101

## ARTI LAMBANG DAN SINGKATAN

$\rightarrow$	: Pemetaan
$k$	: Banyaknya klaster
$p$	: Banyaknya Principal Component (PC)
$\mathbb{R}^d$	: Ruang dimensi- $d$
$\mathcal{F}$	: Ruang fitur
$\gamma$	: Ukuran bandwidth
$\Phi(x_i)$	: Pemetaan $x_i$ menggunakan fungsi kernel
$a_{ik}$	: Keanggotaan data ke- $k$ ke klaster ke- $i$
$v_i$	: Nilai sentroid klaster ke- $i$
$X_{min}$	: Nilai minimum $X_i$
$X_{max}$	: Nilai maksimum $X_i$
$Z_i$	: Hasil pemetaan normal baku data ke- $i$
$N_i$	: Hasil pemetaan normal data ke- $i$
$X_i$	: Data ke- $i$
$\mu$	: Rata-rata pada populasi
$\bar{x}$	: Rata-rata pada sampel
$\sigma$	: Simpangan baku populasi
$s$	: Simpangan baku sampel
$d(x,y)$	: Jarak Euclid antara $x$ dan $y$
$D^{\text{Mahalanobis}}(x,y)$	: Jarak Mahalanobis antara $x$ dan $y$
$\Sigma_{p \times p}$	: Matriks kovarian ukuran $p \times p$
$\Sigma_{p \times p}^{-1}$	: Invers matriks kovarian berukuran $p \times p$
$\log \sum_{i=1}^m a_i$	: Logaritma dari jumlahan $a_i, i=1,2,\dots,m$
$\ x\ $	: Norm vektor $x$
$x \cdot y$	: Dot Product antara $x$ dan $y$
$K(x,y)$	: Kernel Gaussian antara $x$ dan $y$
SSB	: Sum Square Between
SSE	: Sum Square Error
SSW	: Sum Square Within
MSE	: Mean Square Error
AMSE	: Asymptotic Mean Square Error
RBF	: Radial Basis Function
PCA	: Principal Component Analysis
KPCA	: Kernel Principal Component Analysis
SVM	: Support Vector Machines
SVC	: Support Vector Clustering