

Perbaikan Kinerja Clustering K-Means pada Data Ekonomi Nelayan dengan Perhitungan Sum of Square Error (SSE) dan Optimasi nilai K cluster

Improvement of K-Means Clustering Performance on Fishermen's Economic Data with Sum of Square Error (SSE) Calculations and Optimization of K cluster values

Leonardo Petra Refialy¹, Hervin Maitimu², Melgusen Soyano Pesulima³

^{1,2,3}Fakultas Ilmu Komputer, Universitas Kristen Indonesia Maluku

E-mail: ¹leo.refialy@gmail.com, ²hervinmaitimu0312@gmail.com, ³boypesulima7@gmail.com

Abstrak

Clustering adalah jenis metode statistik yang dipakai untuk mengelompokkan banyak data atau objek ke dalam beberapa kelompok (*cluster*), sehingga objek-objek yang berada dalam satu kelompok akan mempunyai kemiripan yang tinggi. Clustering K-Means merupakan metode pembentukan cluster berbasis data, dimana objek secara acak dalam cluster pertama yang terbentuk dijadikan sebagai titik tengah/titik pusat (centroid). K-means merupakan jenis cluster yang memiliki tingkat ketelitian dan efisiensi yang baik. Kelemahan dalam algoritma k-Means yaitu dalam menganalisa dan menentukan Nilai K dalam mengklaster data pada suatu dataset yang tidak optimal akan menghasilkan cluster yang buruk. Sum of Square Error (SSE) merupakan hasil penjumlahan dari seluruh jarak masing-masing data dengan titik pusat clusternya. Semakin kecil nilai SSE yang didapat, semakin seragam data yang ada didalam masing-masing cluster, semakin baik cluster yang dihasilkan. Penelitian ini melakukan analisis cluster dengan K-means untuk menghasilkan kelompok cluster serta perhitungan nilai Sum of Square Error untuk setiap data dengan nilai K yang berbeda. Proses perhitungan nilai K dalam mencari nilai SSE yang minimum sehingga dapat dilakukan perhitungan selisih nilai SSE dari setiap nilai K cluster. Hasil perhitungan selisih tersebut sebagai penentu jumlah K cluster yang optimal dengan tingkat akurasi yang lebih baik.

Kata kunci: *Clustering, K-Means, Sum of Square Error.*

Abstract

Clustering is a type of statistical method used to group a lot of data or objects into several groups (clusters), so that the objects in one group will have high similarities. Clustering K-Means is a data-based cluster formation method, where random objects in the first cluster formed serve as the center point (centroid). K-means is a type of cluster that has a good level of accuracy and efficiency. The weakness in the k-Means algorithm is that in analyzing and determining the K value in clustering data in a dataset that is not optimal, it will result in a bad cluster. Sum of Square Error (SSE) is the sum of all distances for each data to the center point of the cluster. The smaller the SSE value obtained, the more uniform the data in each cluster, the better the resulting cluster. This study performs cluster analysis using K-means to produce cluster groups and calculates the Sum of Square Error value for each data with a different K value. The process of calculating the K value in finding the minimum SSE value so that the SSE value difference from each K cluster value can be calculated. The result of this difference calculation determines the optimal number of K clusters with a better level of accuracy.

Keywords: *Clustering, K-Means, Sum of Square Error.*

1. PENDAHULUAN

Clustering adalah jenis metode statistik yang dipakai untuk mengelompokkan banyak data atau objek ke dalam beberapa kelompok sesuai karakteristik yang dimiliki objek atau data tersebut. Objek tersebut akan dikelompokkan ke dalam satu atau beberapa kelompok (*cluster*) sehingga objek-objek yang berada dalam satu kelompok akan mempunyai kemiripan satu dengan yang lain [1]. Salah satu ciri clustering yang baik atau optimal adalah jika menghasilkan cluster yang berisi data dengan tingkat kemiripan yang tinggi pada cluster yang sama dan tingkat kemiripan rendah pada cluster yang berbeda [2].

Clustering K-Means merupakan salah satu metode pembentukan cluster berbasis data, dimana objek secara acak dalam cluster pertama yang terbentuk dijadikan sebagai titik tengah/titik pusat (*centroid*) [3]. Nilai akhir dari metode *k-means* sendiri adalah memaksimalkan kemiripan data di dalam satu cluster dan meminimalkan kemiripan data antar cluster. Ukuran kemiripan dalam cluster dengan fungsi jarak yang berarti jarak yang terdekat merupakan data yang mirip dengan titik pusat cluster [4].

Langkah-langkah dalam analisis K-Means adalah sebagai berikut: [5]

1. Pilih jumlah K Cluster
2. Inisialisasi k pusat Cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat Cluster diberi nilai awal dengan angka-angka random.
3. Alokasikan semua data/ objek ke Cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut
4. Tugaskan lagi setiap objek memakai pusat Cluster yang baru. Jika pusat Cluster tidak berubah lagi maka proses Clustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat Cluster tidak berubah lagi.

Hal yang menjadi keunggulan k-means yaitu mudah diimplementasikan dan memiliki tingkat ketelitian yang cukup baik terhadap ukuran objek sehingga metode ini relatif lebih terukur dan efisien. Selain itu metode ini juga mudah dijalankan, relatif cepat dan mudah beradaptasi [2]. Hal yang menjadi kelemahan dalam algoritma k-Means yaitu dalam menganalisa dan menentukan jumlah k terbaik dalam mengkluster data pada suatu dataset yang tidak optimal akan menghasilkan cluster yang buruk. [6]. Pada algoritma K-Means, objek data dalam proses cluster nantinya akan menuju beberapa titik pusat (*centroid*) dimana titik tersebutlah yang digunakan sebagai acuan dalam pengelompokan dan data-data yang dilakukan clusterisasi numeric [7]. Pemilihan titik pusat (*centroid*) awal merupakan hal yang sangat sensitif untuk algoritma clustering K-Means. Hal itu nampak pada perbedaan hasil cluster yang seringkali berbeda pada beberapa perulangan percobaan dengan titik awal yang berbeda juga. Hasil cluster mempunyai variasi dan terkadang percobaan lebih banyak menghasilkan hasil yang lebih baik pula [8]. Fungsi tujuan dari K-means adalah menemukan solusi dimana tidak ada solusi lain yang memiliki nilai SSE (*Sum Squared Error*) lebih kecil dari solusi yang ditemukan [9]. SSE merupakan hasil penjumlahan dari seluruh jarak masing-masing data dengan titik pusat clusternya. Semakin kecil nilai SSE yang didapat, semakin seragam data yang ada didalam masing-masing cluster, semakin baik cluster yang dihasilkan [6]. Tujuan dari *K-means* adalah untuk meminimalkan hasil jumlah *squared error* dari seluruh *k* cluster [10]

Pada penelitian yang dilakukan oleh [7] menghasilkan k-means clustering pada data penggolongan keluarga pada golongan kurang mampu, sederhana dan kaya menggunakan variabel pemasukan, pengeluaran dan total asset sebagai pembanding yang nantinya dipakai sebagai analisa cluster dengan menggunakan algoritma k-means. Pada penelitian tersebut, data yang dipakai dalam variabel yaitu data pemasukan, pengeluaran dan total asset tidak menggunakan standar/kriteria tertentu yang nantinya menghasilkan variasi nilai yang banyak,

untuk itu peneliti dalam penelitian ini menggunakan standarisasi nilai pada variabel sehingga lebih mudah dalam perhitungan dan analisa data, selain itu kekurangan pada penelitian tersebut mempunyai tingkat akurasi yang hanya mempunyai akurasi sebesar 69%. Sementara pada penelitian [11] menggunakan SSE (Sum of Square Error) pada metode elbow untuk mengoptimalkan kinerja dari algoritma K-means, dalam penelitiannya melakukan pencarian pusat cluster yang paling optimum berbasis Sum Of Squared Error (SSE), menghasilkan beberapa cluster, dimana antar anggota cluster memiliki tingkat kemiripan yang tinggi. Penelitian tersebut Perbaikan performa cluster K-Means akan diterapkan pada Analisis Penilaian Online Customer Review dan Rating pada Online Marketplace. Penelitian tersebut menggunakan data bertipe string yang bersumber dari ulasan produk online yang menghasilkan 2 cluster, yang masing-masing cluster mempunyai hubungan kemiripan yang sangat tinggi, hal itu menunjukkan bahwa kelompok kluster dengan nilai SSE minimum menghasilkan cluster yang sangat baik dan mempunyai akurasi yang juga baik. pada penelitian [6] menganalisis perhitungan jumlah cluster terbaik pada algoritma k-means, menghasilkan perhitungan selisih nilai SSE pada setiap K dan mengambil selisih nilai SSE pada objek dalam dataset untuk penentuan nilai K yang optimal. Dalam penelitian tersebut menggunakan metode elbow yang mana dapat dilihat pada grafik nilai SSE dan selisih pada tiap nilai K dalam pencarian K yang optimal. Berdasarkan hasil penelitian [6] [7] dan [11] maka penelitian ini melakukan metode K-means dalam analisis clustering data tingkat ekonomi nelayan kota ambon dengan perbaikan kinerja clustering k-means pada data dengan menggunakan perhitungan dan pengujian metode SSE (Sum of Square Error) pada setiap perhitungan selisih nilai SSE dengan nilai K yang berbeda untuk menentukan nilai K mana yang optimal dalam perhitungan cluster k-means.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Penggunaan jenis data ekonomi nelayan dipilih berdasarkan penelitian [7] yang menggunakan data ekonomi masyarakat dalam proses clustering K-means. Penelitian ini menggunakan instance dan variabel serupa namun ada modifikasi dalam penentuan value standar (standarisasi nilai). Dalam penelitian ini, data diperoleh dari survey lapangan pada nelayan kecil di daerah pesisir di kota ambon, data didapat dengan proses wawancara dan pengisian kuisisioner oleh responden. Penelitian melibatkan 100 data nelayan kecil yang diolah menggunakan metode clustering K-Means yang dimodifikasi. Proses awal pada penelitian ini dengan melakukan proses penghimpunan data, kemudian akan dilanjutkan dengan proses transformasi data sehingga semua objek dalam dataset menggunakan data bertipe numerik dianalisis dengan K-Means clustering.

2.2 Proses Transformasi Data

Proses selanjutnya adalah proses transformasi data dari kuisisioner nelayan menjadi dataset numerik untuk dianalisis. Variabel clustering dalam K-Means untuk proses analisis cluster untuk melakukan pemetaan data nelayan dari 3 yaitu:

Pada bagian ini proses pengisian nilai variabel disesuaikan dengan nilai standar 1 dan 2, sesuai dengan kriteria di bawah ini :

1. Pengeluaran. Contoh = < Rp.600.000 = 1, >Rp.600.000 = 2
2. Pengeluaran. Contoh = < Rp.600.000 = 1, >Rp.600.000 = 2
3. Total Aset. Contoh = <Rp.5.000.000 = 1, >Rp5.000.000 = 2

transformasi data yang dipakai dalam analisis k-means clustering dapat dilihat pada tabel 1.

Tabel 1 Contoh Tabel Dataset Nelayan

Data_id	Pendapatan	Pengeluaran	Total Aset
D1	1	2	2
D2	2	1	2
D3	1	2	2

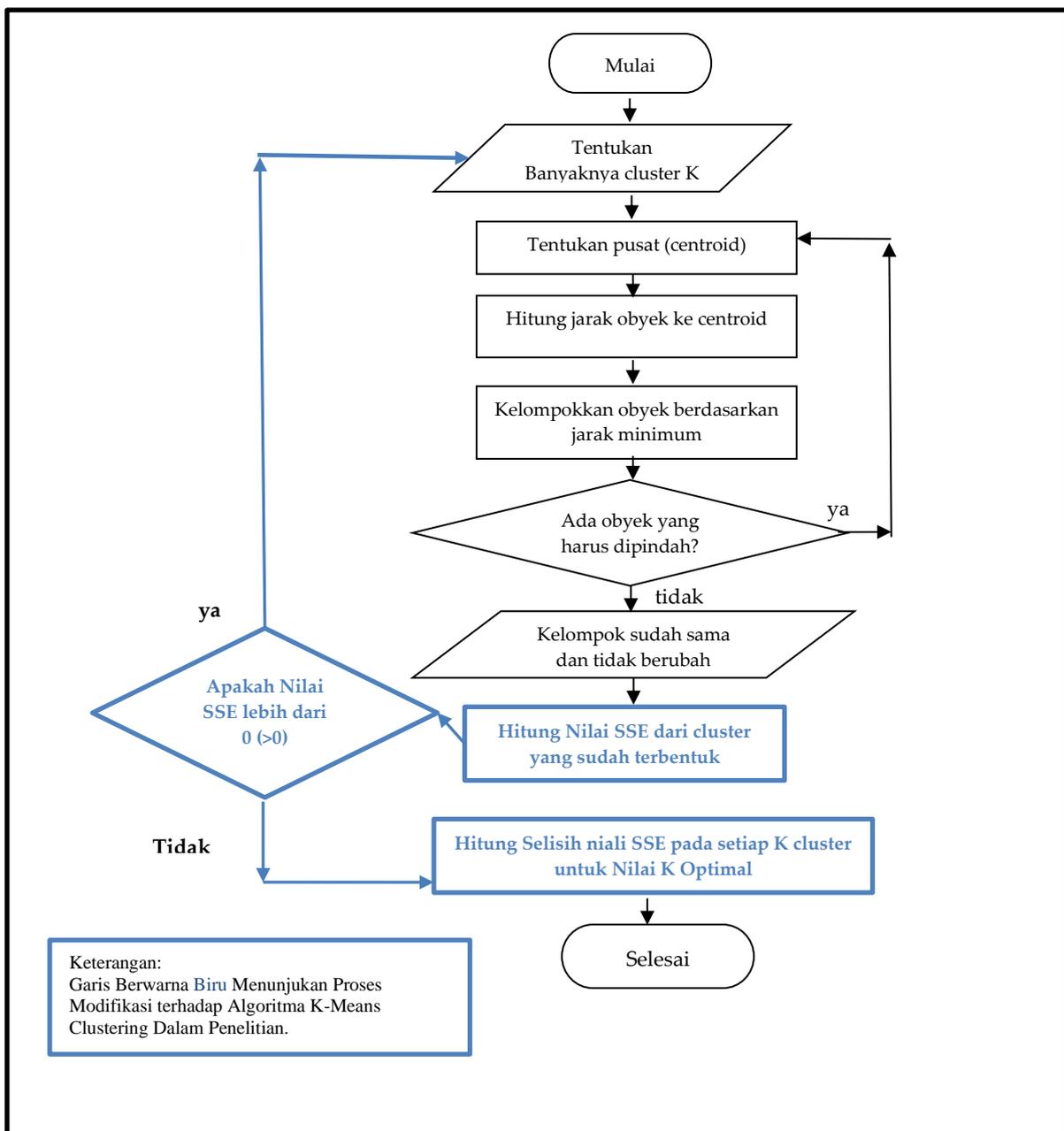
2.3 Proses Analisis Clustering K-Means dengan perhitungan SSE

Algoritma k-means menetapkan nilai-nilai cluster atau kelompok secara acak (random) untuk menjadi pusat dari cluster yang disebut centroid[12]. Centroid dipakai sebagai titik perhitungan jarak antara objek pada data dengan menggunakan perhitungan jarak euclidean.

Jarak *euclidean* mengukur jumlah kuadrat perbedaan nilai masing-masing variabel. Jarak *euclidean* didefinisikan dengan rumusan seperti pada Persamaan 1 [13].

$$d_{ik} = \sqrt{\sum_{j=1}^p (x_{ij} - c_{kj})^2} \quad (1)$$

Keterangan rumus: n : banyak cluster ; X_i : atribut data ; Y_i : atribut centroid



Gambar 1. Flowchart algoritma K-Means dengan SSE dan optimasi Nilai K Cluster

Penjelasan flowcart algoritma proses analisis *k-means clustering* dengan perhitungan *Sum of Square Error (SSE)* dan *Optimasi nilai K cluster* sebagai berikut:

1. Tentukan Banyaknya Cluster. Dalam perhitungan, dimulai dengan menggunakan nilai K (banyaknya kelompok cluster) adalah 2 (K=2)
2. Tentukan titik pusat cluster (centroid). Centroid merupakan titik pusat yang digunakan sebagai dasar dalam menghitung pengelompokkan data. Penentuan pusat awal diambil dari secara random/acak dari dataset yang dipakai dalam analisa k-means.
3. Perhitungan jarak pusat cluster (centroid). Pengukuran jarak antara data dengan pusat cluster digunakan euclidian distance, dapat dilihat pada Persamaan 1.
4. Setelah melalui perhitungan jarak dan penetapan kelompok cluster yang sudah tidak mengalami perpindahan objek, dataset yang sudah melalui proses clustering dengan menggunakan 2 kluster akan di hitung nilai SSE (Sum of Square Error) untuk melihat apakah nilai SSE sudah minimum atau belum, jika nilai SSE masih belum mencapai nilai minimum (0) maka akan kembali ke langkah penentuan nilai K (atau K+1) dst, sampai menemukan nilai SSE yang minimum (0).
5. Hitung semua selisih nilai SSE pada setiap K cluster. Nilai selisih SSE yang paling besar (penurunan signifikan) adalah merupakan nilai K cluster yang paling optimal dengan akurasi yang paling baik.

Dalam Perhitungan Nilai Sum of Square Error menggunakan rumus 3 [13]

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|x_i - c_k\|_2^2 \quad (3) \quad \text{Keterangan rumus:}$$

K : jumlah cluster ; X_i : *attribut data* ; X_{ci} : *attribut centroid* ; C_i : *centroid*

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengelompokan Clustering K-Means dengan perhitungan SSE

Pada bagian ini akan dilihat hasil perhitungan jarak dengan Euclidian Distance pada objek dataset dan centroid. Nilai K awal yang akan dipakai adalah 2 (K=2). Dalam perhitungan jarak tersebut, setelah mencapai keadaan dimana tidak ada perpindahan objek pada dataset. Maka dilanjutkan dengan perhitungan nilai SSE di pada setiap data dengan nilai K yang berbeda, prinsip perhitungan nilai SSE secara trial and error dan akan diulang sampai nilai K mencapai nilai minimum, dengan nilai 0 dipakai sebagai nilai minimum.

Tabel 1 Perhitungan K-Means Cluster dengan Nilai K=2

K=2 (Jumlah Kelompok Cluster)	Centroid (titik pusat)			Jumlah Dataset Pada Kluster	Jumlah Iterasi (Max Iteration)	Nilai Sum of Square Error (SSE)
C0	1	1	1	64	2	18.74305555555557
C1	2	2	2	36		

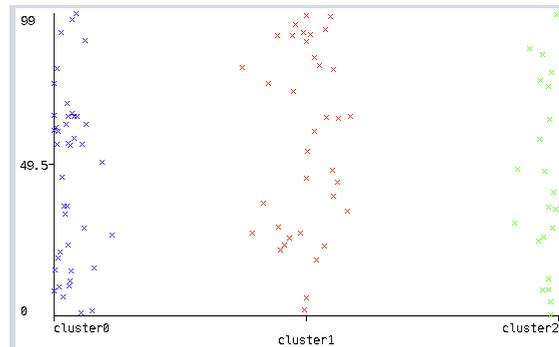


Gambar 2 Visualisasi K-Means dengan percobaan K=2 pada dataset .

Selanjutnya akan dilihat hasil perhitungan jarak dengan Euclidian Distance pada setiap objek dataset dan centroid. Nilai K selanjutnya ialah (K+1) sehingga, nilai K berikut yang akan dipakai dalam perhitungan adalah (2+1) Jadi K = 3. Hasil yang diperoleh dari perhitungan jarak serta memperoleh kelompok cluster sebagai berikut :

Tabel 2 Perhitungan K-Means Cluster dengan Nilai K=3

K=3 (Jumlah Kelompok Cluster)	Centroid (titik pusat)			Jumlah Objek Dataset Pada Kluster	Jumlah Iterasi (Max Iteration)	Nilai Sum of Square Error (SSE)
C0	1	1	1	42	2	4.305555555555555
C1	2	2	2	36		
C2	1	1	2	22		

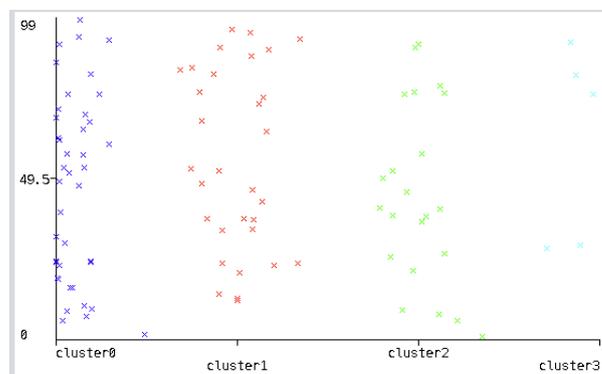


Gambar 3 Visualisasi K-Means dengan percobaan K=3 pada dataset

Selanjutnya akan dilihat hasil perhitungan jarak dengan Euclidian Distance pada dataset dan centroid. Nilai K selanjutnya ialah (K+1) sehingga, nilai K berikut yang akan dipakai dalam perhitungan adalah (3+1) Jadi K = 4. Hasil yang diperoleh dari perhitungan jarak serta memperoleh kelompok cluster pada Tabel.3 dibawah ini:

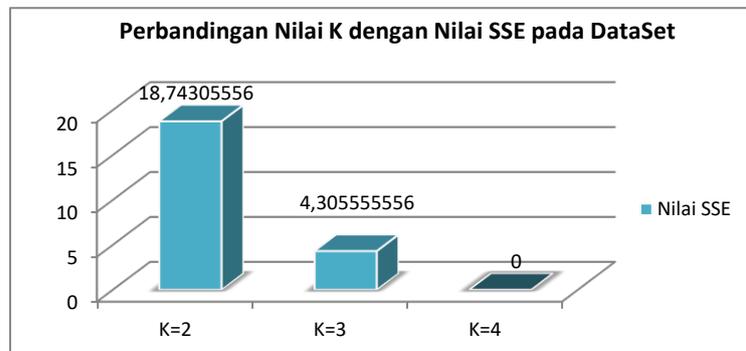
Tabel 3 Perhitungan K-Means Cluster dengan Nilai K=3

K=3 (Jumlah Kelompok Cluster)	Centroid (titik pusat)			Jumlah Dataset Pada Kluster	Jumlah Iterasi (Max Iteration)	Nilai Sum of Square Error (SSE)
C0	1	1	1	42	2	0.0
C1	2	2	2	31		
C2	1	1	2	22		
C3	1	2	2	5		



Gambar 4 Visualisasi K-Means dengan percobaan K=4 pada dataset

Setelah proses pencarian nilai Sum of Square Error (SSE) secara trial and error mencapai nilai minimum dengan menggunakan nilai minimum adalah 0, Maka hasil perhitungan pada K-means dengan menggunakan K=2 dengan penentuan centroid dengan cluster 0 : 1,1,1 dan cluster 1: 2,2,2 dengan keadaan kesetimbangan pada iterasi ke 2 (2 iterasi), menghasilkan 64 dataset yang masuk pada cluster 0 dan 36 dataset masuk pada cluster 1. Dengan nilai Sum of Square Error (SSE) = 18.74305555555557 (dapat dilihat pada gambar 2). Sementara pada perhitungan pada K-means dengan menggunakan K=3 dengan penentuan centroid dengan cluster 0 : 1,1,1, cluster 1: 2,2,2 dan cluster 3 : 1,1,2 dengan keadaan kesetimbangan pada iterasi ke 2 (2 iterasi), menghasilkan 64 dataset yang masuk pada cluster 0 dan 36 dataset masuk pada cluster 1 dengan nilai Sum of Square Error (SSE) = 4.30555555555555 (dapat dilihat pada gambar 3). perhitungan ketiga dengan menggunakan K=4, dengan penentuan centroid pada cluster 0 : 1,1,1, cluster 1: 2,2,2, cluster 2 : 1,1,2, dan cluster 3 = 1,2,2, menghasilkan 42 dataset masuk pada cluster 0, 31 dataset masuk pada cluster 1, 22 dataset masuk pada cluster 2, 5 dataset masuk pada cluster 3 dengan nilai SSE 0. Karena pada perhitungan k-means dengan perhitungan nilai Sum of Square Error telah mencapai nilai minimum (0) maka pencarian nilai K baru tidak lagi dilanjutkan. Dengan demikian hasil perhitungan setelah mendapat nilai SSE minimum yaitu pada nilai K=4 dengan nilai SSE 0 Lebih Jelasnya dapat digambarkan pada Gambar 5.



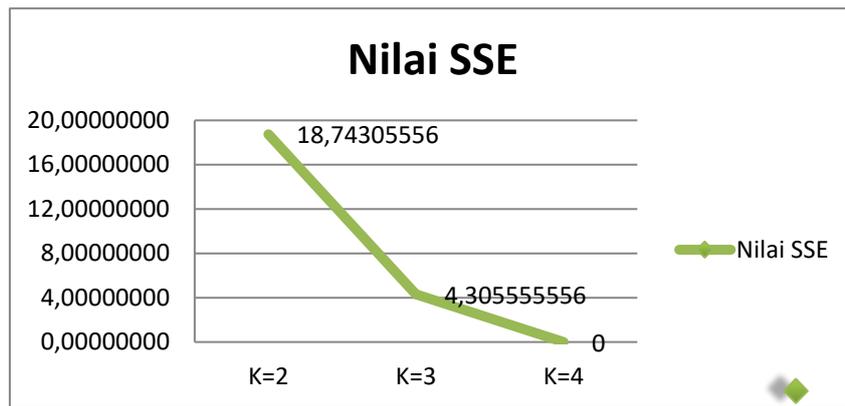
Gambar 5 Grafik Nilai K Dengan Nilai SSE dari DataSet

Setelah perhitungan nilai SSE telah mencapai nilai minimum maka selanjutnya melakukan pengujian dengan menghitung selisih dari Sum of Square Error (SSE) pada K awal dan nilai K mengalami peningkatan, maka akan dilihat nilai selisih SSE mana yang mempunyai penurunan yang drastis/signifikan.

Tabel 4 Perbandingan Nilai K, Nilai SSE dan Selisih nilai SSE pada K lama dan K baru

Nilai K	Nilai SSE	Selisih SSE K sebelumnya
2	18,74305556	0
3	4,30555556	14,437500004
4	0	4.305555556

Perhitungan selisih nilai SSE pada setiap K cluster, menghasilkan nilai K cluster yang Optimal dapat dilihat pada tabel 4. Nilai K yang optimal ada pada nilai K=3 dimana selisih SSE turun secara drastis, pada grafik nilai SSE terlihat adanya pola elbow sesuai pada penelitian [6]. yang mana menyatakan suatu cluster mempunyai nilai selisih terbesar antara nilai SSE pada setiap data K merupakan K Klaster yang optimal. Hal tersebut dapat dilihat pada gambar 6.



Gambar 6 Grafik Nilai SSE dengan nilai K=3 sebagai Nilai K Optimal

4. KESIMPULAN DAN SARAN

Berdasarkan hasil perhitungan jarak antara objek dataset dengan menggunakan 3 variabel pada data nelayan dengan centroid menggunakan k-means clustering dengan menerapkan modifikasi algoritma K-Means dengan menambahkan pencarian nilai Sum of Square Error yang minimum, serta perhitungan K secara trial and error untuk mendapatkan nilai K yang optimal. Setelah nilai SSE pada cluster dengan K cluster tertentu sudah mencapai nilai minimum (minimum pada nilai K=4), maka tidak ada lagi penentuan nilai K yang baru, karena nilai SSE antara objek dan cluster sudah mencapai titik 0. Hasil minimum dapat mencerminkan tingkat keseragaman data dalam kelompok kluster yang sudah semakin baik [11]. Selanjutnya untuk mencapai nilai K yang optimal digunakan selisih nilai SSE pada setiap keadaan K cluster dalam perhitungan sampai nilai minimum. Nilai optimal yaitu nilai K yang mengalami penurunan nilai SSE secara drastis (paling signifikan) sehingga membentuk grafik elbow sehingga jumlah cluster (nilai K) paling baik ada pada nilai K=3 dengan nilai selisih yang paling besar (penurunan nilai SSE yang paling drastis), sesuai dengan penelitian [6] dengan metode elbow dalam penentuan nilai k yang optimal pada k means cluster. Kelebihan dalam perhitungan k-means dalam penelitian ini adalah tingkat akurasi kelompok cluster yang mempunyai keterkaitan yang lebih seragam dan lebih baik, dikarenakan mendapat nilai Sum of Square Error (SSE) yang minimum dengan menggunakan mencari nilai K yang optimal. Nilai K yang optimal menunjukkan jumlah cluster terbaik yang menghasilkan kelompok cluster yang memiliki kinerja yang baik. Kekurangan pada penelitian ini adalah pada proses standarisasi nilai variabel menjadi 2 memudahkan dalam proses perhitungan tetapi hasilnya bisa saja berbeda-beda jika standarisasinya mempunyai lebih dari 2 nilai standar, sehingga saran untuk penelitian ke depan ialah metode yang lebih baik dalam melakukan perhitungan standard nilai pada variabel yang dipakai dalam analisis k-means serta menambahkan metode pengujian lain dalam menunjang metode perhitungan SSE yang dalam hal ini menghasilkan performa cluster yang lebih baik lagi.

UCAPAN TERIMA KASIH

Terima Kasih yang sebesar-besarnya Kepada Lembaga Penelitian Universitas Kristen Indonesia Maluku yang sudah mensponsori, mendanai serta memberikan motivasi terhadap penulis dalam pelaksanaan penelitian dan penulisan jurnal ini. Tak lupa juga penulis mengucapkan terima kasih untuk semua pihak yang sudah membantu dalam penelitian ini.

DAFTAR PUSTAKA

- [1] L. Ramadhani, I. Purnamasari, and F. D. T. Amijaya, "Penerapan Metode Complete Linkage dan Metode Hierarchical Clustering Multiscale Bootstrap," *J. Ekspansional*, vol. 9, no. 1, pp. 1–10, 2018.
- [2] M. Anggara, H. Sujaini, and H. Nasution, "Pemilihan distance measure Pada K-Means clustering Untuk Pengelompokkan member di Alvaro fitness," *JUSTIN (Jurnal Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 186–191, 2016.
- [3] K. Kameshwaran and K. Malarvizhi, "Survey on clustering techniques in data mining," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 2, pp. 2272–2276, 2014.
- [4] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018.
- [5] E.-J. Lee and S. Y. Shin, "When do consumers buy online product reviews? Effects of review quality, product type, and reviewer's photo," *Comput. Human Behav.*, vol. 31, pp. 356–366, 2014.
- [6] N. P. E. Merliana and A. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering," 2015.
- [7] E. Fammaldo and L. Hakim, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Kesejahteraan Keluarga Untuk Program Kartu Indonesia Pintar," *J. Ilm. Teknol. Infomasi Terap.*, vol. 5, no. 1, pp. 23–31, 2018.
- [8] I. P. A. Pratama and A. Harjoko, "Penerapan algoritma invasive weed optimnization untuk penentuan titik pusat klaster pada K-Means," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 9, no. 1, pp. 65–76, 2017.
- [9] X. Cui, T. E. Potok, and P. Palathingal, "Document clustering using particle swarm optimization," in *Proceedings 2005 IEEE Swarm Intelligence Symposium, 2005. SIS 2005.*, 2005, pp. 185–191.
- [10] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010.
- [11] R. Nainggolan and E. N. Purba, "Perbaikan Performa Cluster K-Means Menggunakan Sum Squared Error (SSE) Pada Analisis Online Customer Review Terhadap Produk Toko Online," *J. TIMES*, vol. 8, no. 2, pp. 1–8, 2019.
- [12] R. T. Vuldari, *Data Mining: Teori dan Aplikasi Rapidminer*. Penerbit Gava Media, 2017.
- [13] Y. Agusta, "K-Means–Penerapan, Permasalahan dan Metode Terkait," *J. Sist. dan Inform.*, vol. 3, no. 1, pp. 47–60, 2007.