



**PENERAPAN ALGORITMA *K – MEANS CLUSTERING* PADA
PENGELOMPOKAN INVESTOR INVESTASI REKSADANA**

SKRIPSI

STEPHEN KURNIA

1910511001

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2023



**PENERAPAN ALGORITMA *K – MEANS CLUSTERING* PADA
PENGELOMPOKKAN INVESTOR INVESTASI REKSADANA**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar
Sarjana**

STEPHEN KURNIA

1910511001

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2023

PERNYATAAN ORISINALITAS

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Stephen Kurnia
NIM : 1910511001
Tanggal : 17 Juli 2023
Judul Skripsi : **Penerapan Algoritma *K* – *Means Clustering* Pada Pengelompokkan Investor Investasi Reksadana**

Bilamana pada kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 17 Juli 2023

Yang Menyatakan,

A 10,000 Rupiah Indonesian banknote is shown with a signature written over it. The banknote features the Garuda Pancasila emblem and the number 10000. The serial number 25BE9AKX55072788 is visible at the bottom.

Stephen Kurnia

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademika Universitas Pembangunan Nasional “Veteran” Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Stephen Kurnia
NIM : 1910511001
Fakultas : Ilmu Komputer
Program Studi : S1 Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan karya ilmiah saya kepada Universitas Pembangunan Nasional “Veteran” Jakarta Hak Bebas Royalti Non-Eksklusif (*Non-Exchange Royalty Free Right*) untuk dipublikasikan dengan judul:

Penerapan Algoritma K – Means Clustering Pada Pengelompokkan Investor Investasi Reksadana

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalih media atau memformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis atau pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 17 Juli 2023

Yang Menyatakan,



Stephen Kurnia


LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Yang bertanda tangan di bawah ini:

Nama : Stephen Kurnia
NIM : 1910511001
Program Studi : S-1 Informatika
Judul Skripsi/TA : Penerapan Algoritma *K – Means Clustering* Pada Pengelompokkan Investor Investasi Reksadana

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



(Yuni Widiastiwati, S.Kom, M.Si.)

Penguji 1



(Iin Ernawati S.Kom., M.Si.)

Penguji 2



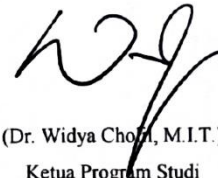
(Dr. Ermatita, M.Kom.)

Pembimbing



(Dr. Ermatita, M.Kom.)

Dekan



(Dr. Widya Chofri, M.I.T.)

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 6 Juli 2023



PENERAPAN ALGORITMA *K – MEANS CLUSTERING* PADA PENGELOMPOKKAN INVESTOR INVESTASI REKSADANA

Stephen Kurnia

ABSTRAK

Di era teknologi seperti sekarang ini, banyak peran penggunaan teknologi di segala bidang kehidupan masyarakat, khusus nya dalam aktivitas finansial dan perekonomian. Salah satu produk digital investasi yang beredar sekarang ini adalah Reksadana. Reksadana adalah salah satu produk investasi digital yang minim resiko serta dapat digunakan oleh masyarakat yang masih pemula untuk belajar dan memulai investasi. Ada beberapa jenis Reksadana yang saat ini dijual kepada masyarakat, seperti Reksadana pendapatan tetap, Reksadana pasar uang, Reksadana campuran, Reksadana syariah dan Reksadana saham. Karna begitu banyaknya jenis reksadana yang ada saat ini, investor kerap bingung dalam menentukan investasi reksadana dengan profit yang baik serta resiko yang paling aman. Apalagi dengan Investor yang masih pemula dalam melakukan kegiatan berinvestasi. Maka dari itu dilakukan penelitian yang berjudul “Penerapan Algoritma *K – Means Clustering* Pada Pengelompokkan Investor Investasi Reksadana” dengan tujuan agar para penyedia Reksadana dapat mengetahui produk Reksadana yang memiliki profil risiko paling aman dan juga keuntungan yang didapat dengan kebutuhan para Investor dalam mencari produk reksadana yang cocok untuk dirinya dan juga Investor dapat mengetahui profil risiko dari investasi Reksadana yang mereka ambil. Metode yang digunakan dalam penelitian ini menggunakan Algoritma *K - Means Clustering* dengan dataset yang berasal dari OJK Reksadana dengan jumlah data sebanyak 1529 data yang bertujuan untuk mengelompokkan tipe profil investor dengan menggunakan nilai $K = 3$ dengan jenis kluster adalah jenis Investor dengan profil risiko mereka dalam berinvestasi, kemudian dilakukan proses evaluasi kluster dengan menggunakan metode Silhouette score. Hasil yang didapatkan dari penelitian ini adalah adanya visualisasi data berupa grafik hasil kluster kelompok profil risiko investor Reksadana dan kemudian dilakukan perhitungan rata - rata untuk memudahkan proses analisis dan dilakukan proses evaluasi nilai kluster dengan Silhouette score yang menghasilkan skor untuk klasterisasi $k = 3$ adalah 0.8283326733985543

Kata Kunci : Reksadana, Investasi, masyarakat, *K – Means Clustering*

APPLICATION OF *K - MEANS CLUSTERING* ALGORITHM ON MUTUAL FUND INVESTOR GROUPING

Stephen Kurnia

ABSTRACT

In the current technological era, there are many roles in the use of technology in all areas of people's lives, especially in financial and economic activities. One of the investment digital products currently circulating is Mutual Funds. Mutual funds are a digital investment product that has minimal risk and can be used by people who are beginners to learn and start investing. There are several types of Mutual Funds that are currently being sold to the public, such as Fixed Income Mutual Funds, Money Market Mutual Funds, Mixed Mutual Funds, Sharia Mutual Funds and Stock Mutual Funds. Because there are so many types of mutual funds that exist today, investors are often confused about choosing a mutual fund investment with good profits and the safest risk. Especially with Investors who are still beginners in investing activities. Therefore, a study was conducted entitled "Application of the K - Means Clustering Algorithm in Grouping Mutual Fund Investment Investors" with the aim that Mutual Fund providers can find out the Mutual Fund products that have the safest risk profile and also the benefits obtained according to the needs of Investors in finding mutual fund products that are suitable for themselves and also Investors can find out the risk profile of the Mutual Fund investments they take. The method used in this study uses the K - Means Clustering Algorithm with a dataset originating from OJK Mutual Funds with a total of 1529 data which aims to classify investor profile types using a value of $K = 3$ with cluster types being types of Investors with their risk profiles in investing , then the cluster evaluation process is carried out using the Silhouette score method. The results obtained from this study are data visualization in the form of a graph of the cluster results of the mutual fund investor risk profile group and then calculating the average to facilitate the analysis process and evaluating the cluster value with the Silhouette score which produces a score for clustering $k = 3$ which is 0.8283326733985543

Keywords : Mutual Funds, Investment, Community, K – Means Clustering

KATA PENGANTAR

Segala puji syukur kehadirat Tuhan Yang Maha Esa, dengan rahmat serta pimpinannya sehingga penulis dapat menyelesaikan Skripsi yang berjudul “**Penerapan Algoritma K – Means Clustering Pada Pengelompokkan Investor Investasi Reksadana**” dengan baik. Penyusunan skripsi ini diajukan adalah sebagai syarat dalam memperoleh gelar sarjana ilmu komputer pada program studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional veteran Jakarta.

Dalam melakukan penulisan serta penyusunan skripsi ini, penulis mendapatkan banyak bantuan serta dukungan yang sangat positif dari beberapa pihak yang telah memberikan dukungan serta memberikan masukan yang positif sehingga penulisan skripsi ini dapat berjalan dengan baik. Untuk itu, penulis menyampaikan terima kasih kepada :

1. Dr. Ermatita, M.Kom., selaku dekan dari Fakultas Ilmu Komputer serta dosen pembimbing yang sangat berjasa dalam memberikan bimbingan untuk menyelesaikan pembuatan skripsi ini.
2. Dr. Widya Cholil, M.I.T, selaku Ketua Program Studi Sarjana Jurusan Informatika.
3. Orang tua yang senantiasa memberikan dukungan dan doa.
4. Saudara Timothy Kurnia yang senantiasa memberikan masukan agar skripsi ini dapat terselesaikan dengan baik.
5. Rekan - Rekan Laurenza, Hozana, Daffy, Ajeng dan rekan - rekan lain yang tidak bisa saya sebutkan satu persatu yang selalu bersabar mendengarkan keluh kesah dan motivasi.

Namun penulis masih menyadari bahwa skripsi ini masih memiliki beberapa kekurangan, baik dari segi materi ataupun dengan teknik penyajian. Maka dari itu, adanya kritik serta saran yang membangun sangat penulis harapkan untuk menyempurnakan kekurangan yang ada pada penyusunan skripsi ini.

Jakarta, 28 Mei 2023

Penulis

DAFTAR ISI

PERNYATAAN ORISINALITAS	i
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	ii
LEMBAR PENGESAHAN	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Ruang Lingkup	4
1.6 Luaran yang Diharapkan	4
1.7 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Reksadana	6
2.2 Data Mining	6
2.2.1 Pengertian Data Mining	6
2.2.2 Tahapan Data Mining	7
2.3 Clustering	9
2.4 K – Means Clustering	10
2.5 Profil Risiko Investor	12
2.6 Python	13
2.7 Silhouette Score	13
2.8 Review Penelitian Terdahulu	14
BAB III METODOLOGI PENELITIAN	16

3.1 Alur Penelitian Penelitian	16
3.2.1 Mengidentifikasi Masalah	18
3.2.2 Perumusan Masalah	18
3.2.2 Studi Literatur	18
3.2.3 Pengumpulan Data	19
3.2.3 Melakukan Pra - Proses Data	19
3.2.5 Melakukan Proses Evaluasi kluster	21
3.2 Perangkat Penelitian	21
3.3 Waktu Dan Jadwal Penelitian	22
BAB IV HASIL DAN PEMBAHASAN	23
4.1 Data	23
4.2 Praproses Data	25
4.2.1 Integrasi Data	25
4.2.2 Mengubah Tipe Data	30
4.2.3 Melakukan <i>Cleaning Data</i>	31
4.3 Proses Klasterisasi dengan Algoritma K - Means <i>Clustering</i>	33
4.4 Melakukan Proses Visualisasi Data dengan Grafik	44
4.5 Melakukan Proses Evaluasi kluster	47
BAB V PENUTUP	50
5.1 Kesimpulan	50
5.2 Saran	50
DAFTAR PUSTAKA	51
RIWAYAT HIDUP	53
LAMPIRAN	54

DAFTAR TABEL

Tabel 2. 1 Jadwal Penelitian	22
Tabel 2. 2 data NAB awal 30 Desember 2022	23
Tabel 2. 3 Table numerik untuk perhitungan atribut baru.....	25
Tabel 2. 4 Atribut NAB per Unit 30 Desember 2022	26
Tabel 2. 5 Atribut NAB per Unit untuk setiap bulan	27
Tabel 2. 6 tabel dengan atribut nilai Return per tahun	28
Tabel 2. 7 Data sebelum pengubahan tipe data menjadi float.....	29
Tabel 2. 8 Hasil Pengubahan tipe data menjadi float	30
Tabel 2. 9 Tabel data dengan atribut missing value	31
Tabel 2. 10 contoh data dengan nilai NaN	32
Tabel 2. 11 tabel perbandingan jumlah data sebelum dan sesudah dilakukan proses drop	33
Tabel 2. 12 Data Sample	34
Tabel 2. 13 data centroid random.....	35
Tabel 2. 14 data perhitungan jarak ke tiap centroid	37
Tabel 2. 15 data hasil kelompok berdasarkan posisi terkecil dari nilai centroid..	38
Tabel 2. 16 pencarian nilai centroid baru berdasarkan rata - rata penjumlahan klaster	39
Tabel 2. 17 data centroid lama	39
Tabel 2. 18 Data centroid baru	40
Tabel 2. 19 Data hasil perhitungan ke setiap centroid	40
Tabel 2. 20 data hasil pengelompokkan klaster berdasarkan nilai centroid terkecil	41
Tabel 2. 21 data hasil perhitungan rata - rata pada anggota cluster untuk pencarian nilai centroid baru	42
Tabel 2. 22 Data centroid lama.....	42
Tabel 2. 23 Data centroid baru	43
Tabel 2. 24 Data hasil perhitungan rata - rata data tiap atribut dan klaster.....	45
Tabel 2. 25 Tabel hasil perhitungan nilai Silhouette score untuk setiap k klaster	48

DAFTAR GAMBAR

Gambar 3. 1 Tahapan Penelitian.....	17
Gambar 4. 1 Program Mengubah tipe data.....	30
Gambar 4. 2 fungsi mengganti nilai 0,0 menjadi NaN.....	32
Gambar 4. 3 fungsi untuk membuang data NaN	33
Gambar 4. 4 grafik hasil klasterisasi	44
Gambar 4. 5 fungsi untuk memunculkan grafik hasil klasterisasi.....	44
Gambar 4. 6 fungsi untuk mengimport hasil klaster kedalam excel	45
Gambar 4. 7 Program menghitung nilai skor Silhouette	48
Gambar 4. 8 Program untuk menampilkan hasil Skor Silhouette pada setiap k klaster	48