Klasterisasi Data *Wholesale Customers* Menggunakan *K-Means++ Clustering*

SKRIPSI

untuk memenuhi persyaratan dalam menyelesaikan program sarjana Strata-1 Statistika

Oleh

ERIEN SYARIF

NIM. 1811017320024

PROGRAM STUDI STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS LAMBUNG MANGKURAT

BANJARBARU

JUNI 2023

Klasterisasi Data *Wholesale Customers* Menggunakan *K-Means++ Clustering*

SKRIPSI

untuk memenuhi persyaratan dalam menyelesaikan program sarjana Strata-1 Statistika

Oleh

ERIEN SYARIF

NIM. 1811017320024

PROGRAM STUDI STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS LAMBUNG MANGKURAT

BANJARBARU

JUNI 2023

SKRIPSI

KLASTERISASI DATA *WHOLESALE CUSTOMERS* MENGGUNAKAN *K-MEANS++ CLUSTERING*

Oleh:

ERIEN SYARIF
1811017320024

telah dipertahankan di depan Dosen Penguji pada tanggan 09 Juni 2023

Susunan Dosen Penguji:

Pembimbing I

Yuana Sukmawaty, S.Si., M.Si
NIP 198810152015042002

Penguji I:

Oni Soesanto, S.Si., M.Si
NIP 197301262005011003

Pembimbing II

Irwan Budiman, S.T., M.Kom.
NIP 197703252008121001

Penguji II

I Putu Edy Suardiyana Putra, M.Kom., Ph.D

Banjarbaru, 23 Juni 2023
Mengetahui,
Koordinator Program Studi Statistika

Dewi Anggraini, S. Si., M.App.Sci., Ph.D
NIP 198303282005012001

PERNYATAAN

Saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sepanjang pengetahuan serta kesadaran Saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, Juni 2023

Erien Syarif
NIM 1811017320024

ABSTRAK

Klasterisasi Data *Wholesale Customers* Menggunakan *K-Means++ Clustering* (Oleh: Erien Syarif; Pembimbing: Yuana Sukmawaty, S.Si., M.Si. dan Irwan Budiman, S.T., M.Kom., 2023; 56 halaman)

Segmentasi pelanggan merupakan proses pengelompokan pelanggan berdasarkan karakteristik dan perilaku pelanggan. Segmentasi dilakukan dengan menganalisis berbagai jenis informasi tentang pelanggan seperti demografis, geografis, dan perilaku. Manfaat segmentasi pelanggan, yaitu mempelajari dan memahami perilaku pelanggan. Penelitian ini bertujuan mengetahui jumlah klaster (*k*) optimal dari hasil klasterisasi dan menganalisis karakteristik pelanggan dari klaster yang terbentuk. Data yang digunakan adalah data *Wholesale Customers* yang didapatkan dari *UCI Machine Learning Repository.* Data tersebut berisi pengeluaran tahunan pelanggan dari distributor grosir untuk berbagai barang. Metode klasterisasi dalam penelitian ini menggunakan algoritma *K-Means++ Clustering* dan jumlah klaster (*k*) optimal didapatkan menggunakan metode *Silhouette Coefficient.* Hasil penelitian menunjukan bahwa *k* optimal yang didapatkan adalah *k* = 2 dengan nilai *Silhouette Coefficient* sebesar 0.7390. Penelitian ini menghasilkan 2 klaster, yaitu "Pelanggan dengan Pengeluaran Besar" dan "Pelanggan dengan Pengeluaran Kecil". Jumlah pelanggan pada klaster 1 dan klaster 2 sebanyak 14 dan 426 pelanggan. Klaster 1 memiliki rata-rata pengeluaran lebih besar pada setiap atribut dibandingkan dengan klaster 2. Pengeluaran terbesar pada klaster 1 dan klaster 2 terdapat pada atribut *Milk* dengan rata-rata pengeluaran sebesar 33847.79 dan atribut *Fresh* dengan rata-rata pengeluaran sebesar 11531.44.

Kata kunci: *K-Means++ Clustering,* Segmentasi Pelanggan, *Silhouette Coefficient*

ABSTRACT

WHOLESALE CUSTOMERS CLUSTERING USING *K-MEANS++ CLUSTERING*
(By Erien Syarif; Supervisor: Yuana Sukmawaty, S.Si., M.Si. and Irwan Budiman, S.T., M.Kom., 2023; 56 pages)

Customer segmentation is the process of grouping customers based on customer characteristics and behavior. Segmentation is done by analyzing different types of information about customers such as demographic, geographic, and behavioral. The benefits of customer segmentation are learning and understanding customer behavior. This study aims to determine the optimal number of clusters (*k*) from the clustering results and analyze the customer characteristics of the clusters formed. The data used is *Wholesale Customers* obtained from the *UCI Machine Learning Repository.* The data contains the annual spending of customers of wholesale distributors for various goods. The clustering method in this study uses the *K-Means++ Clustering algorithm* and the optimal *k* is obtained using the *Silhouette Coefficient method.* The results showed that the optimal *k* obtained was *k* = 2 with a *Silhouette Coefficient* value of 0.7390. This research resulted in 2 clusters, namely "Customers with High Spend" and "Customers with Low Spend". The number of customers in cluster 1 and cluster 2 is 14 and 426 customers. Cluster 1 has a higher average spend on each attribute compared to cluster 2. The largest spending in cluster 1 and cluster 2 is in the *Milk* attribute with an average spending of 33847.79 and *the Fresh* attribute with an average spending of 11531.44.

Keywords: *Customer Segmentation*, *K-Means++ Clustering, Silhouette Coefficient*

PRAKATA

Alhamdulilah puji syukur dipanjatkan kepada Allah Subhanahu wa Ta'ala yang telah memberikan kesehatan, kekuatan dan kemudahan kepada peneliti, sehingga dapat menyelesaikan penelitian ini. Shalawat serta salam juga tidak lupa kita panjatkan kepada Nabi Muhammad Shallallahu 'alaihi wa sallam serta keluarga dan para sahabat yang telah membawa zaman kegelapan menuju zaman yang penuh cahaya.

Penelitian yang berjudul "**Klasterisasi Data *Wholesale Customers* Menggunakan *K-Means++ Clustering***" ini telah diselesaikan untuk memenuhi salah satu persyaratan dalam menyelesaikan program sarjana di Program Studi Statistika Fakultas MIPA Universitas Lambung Mangkurat.

Penulisan skripsi ini tidak akan tercapai tanpa bantuan dan dukungan dari banyak pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada berbagai pihak, yaitu:

1. Orang tua yang telah memberikan segalanya;
2. Ibu Yuana Sukmawaty, S.Si., M.Si selaku dosen pembimbing 1 dan dosen pembimbing akademik serta Bapak Irwan Budiman, S.T., M.Kom. selaku dosen pembimbing 2 yang telah membantu, mengarahkan, memberikan masukan dan membimbing dalam menyelesaikan skripsi;
3. Koordinator Program Studi beserta seluruh jajaran dosen dan staf Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lambung Mangkurat (FMIPA ULM);
4. Alpi, Dini, Kethy, Lalu, Umi, Nela, Inal, Nawiw, Aldi, Tomi, Geo, Ardi dan juga teman-teman angkatan 2018 FMIPA ULM;
5. Rekan-rekan pengurus BEM FMIPA ULM Periode 2020 yang telah menjadi tempat untuk mengembangkan diri;
6. Berbagai pihak yang telah turut serta membantu yang tidak dapat disebut satu persatu.

Penelitian ini jauh dari kata sempurna. saran, masukan dan kritik sangat diharapkan untuk perbaikan selanjutnya. Akhir kata penulis berharap semoga skripsi ini dapat memberikan manfaat bagi pembacanya.

Banjarbaru, Juni 2023

Erien Syarif

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN