

Penerapan Data Mining Pada Penjualan Kartu Paket Internet Yang Banyak Diminati Konsumen Dengan Metode K-Means

Hendra Nicodemus Turnip¹, Hasanul Fahmi²

^{1,2} STMIK Pelita Nusantara Jl. St. Iskandar Muda No. 1 Medan, Indonesia

¹ hendraturnip96@yahoo.com; ² h.fahmizuhri@gmail.com

INFORMASI ARTIKEL

Kata Kunci:

Kata Kunci : Algoritma K-Means Clustering, Data Mining, Kartu Paket Internet

ABSTRAK

Royal Ponsel saat ini memiliki permasalahan dimana untuk menentukan kartu paket internet mana yang paling banyak diminati konsumen dari penjualan kartu internet Telkomsel, XL, Axis, Im3, 3 (Tri), dan Smartfren. Banyak informasi yang dimiliki namun tidak cukup jika informasi tersebut tidak dimanfaatkan dengan sebaik mungkin, sehingga diperlukan pengelompokan data penjualan untuk mengetahui daya saing produk kartu internet mana saja yang memiliki tingkat penjualan yang paling tinggi berdasarkan penjualan yang ada di usaha Royal Ponsel. Permasalahan ini tentunya dibutuhkan suatu teknologi yang dapat melakukan analisa terhadap data transaksi penjualan kartu paket internet. Salah satunya yaitu dengan menerapkan data mining terhadap penjualan kartu paket internet yang banyak diminati dengan menggunakan perhitungan metode K-Means Clustering. Dengan adanya pengelompokan jenis kartu internet yang digunakan untuk meningkatkan *performance* dari penjualan sehingga dapat menentukan langkah peningkatan stok pada kartu paket internet secara tepat.

Keywords:

Keywords: K-Means Clustering Algorithm, Data Mining, Internet Package Card.

ABSTRACT

Currently, Royal Ponsel has a problem where to determine which internet package card consumers are most interested in from selling internet cards Telkomsel, XL, Axis, Im3, 3 (Tri), and Smartfren. A lot of information is owned but it is not enough if the information is not utilized properly, so it is necessary to group sales data to determine the competitiveness of which internet card products have the highest sales level based on sales in the Royal Ponsel business. Of course, this problem requires a technology that can analyze data on internet package card sales transactions. One of them is by applying data mining to the sale of internet package cards which are in great demand by using the K-Means Clustering method calculation. With the existence of grouping types of internet cards that are used to improve the performance of sales so as to determine the steps to increase stock on internet package cards appropriately.

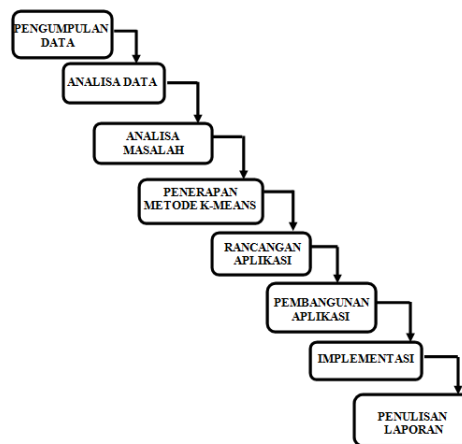
I. Pendahuluan

Data mining adalah suatu konsep yang digunakan untuk menemukan pengetahuan yang tersembunyi didalam database. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstaksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang tersimpan didalam database besar [1]. *Data mining* adalah bagian dari proses KKD (*Knowledge Discovery in Databases*) yang terdiri dari beberapa tahapan seperti pemilihan data, pra pengolahan, transformasi, *data mining*, dan evaluasi hasil [2]. Royal Ponsel saat ini memiliki permasalahan dimana untuk menentukan kartu paket internet mana yang paling banyak diminati konsumen dari penjualan kartu internet Telkomsel, XL, Axis, Im3, 3 (Tri), dan Smartfren sehingga dibutuhkan clustering data untuk menentukan kartu paket internet yang paling banyak diminati hingga yang kurang diminati. Sehingga pihak produsen harus berlomba-lomba menarik minat pelanggan dengan berbagai macam strategi pemasaran agar tidak kalah saing dan tetap konsisten dalam setiap pemasaran produknya agar banyak diminati.

K-Means clustering merupakan salah satu teknik *data mining* yang memberikan deskripsi *cluster* pada sebuah produk. Sehingga hasil simulasi dapat memberikan pengetahuan jenis produk yang akan menjadi stok terbanyak dalam meningkatkan pendapatan. Metode K-Means clustering berusaha mengelompokkan data yang

ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada dalam kelompok yang lain [3]). Sedangkan Algoritma *K-Nearest Neighbor* (K-NN) merupakan metode pengklasifikasian data yang bekerja relatif dengan cara yang lebih sederhana dibandingkan dengan metode pengklasifikasian data lainnya. Algoritma ini berusaha mengklasifikasikan data baru yang belum diketahui class-nya dengan memilih data sejumlah yang letaknya terdekat dari data baru tersebut [4][5] Sehingga *K-Means Clustering* lebih dapat mampu mengclusteringkan data-data dibandingkan dengan algoritma *K-Nearest Neighbor*. Berdasarkan uraian latar belakang diatas, maka penulis melakukan sebuah penelitian dengan judul “Penerapan *Data Mining* Pada Penjualan Kartu Paket Internet Yang Banyak Diminati Konsumen Dengan Metode K-Means[6] ”.

II. Metode



Gambar 1. Metodologi Penelitian dengan Model Waterfall

Berdasarkan kerangka kerja penelitian di atas, maka dapat diuraikan pembahasan masing-masing tahapan dalam peneltiaian dengan model penelitian waterfall yaitu :

1. Pengumpulan Data

Dalam memperoleh data-data, beberapa metode telah dilakukan guna untuk mendapatkan data yang sesuai dengan kebutuhan penelitian. Ada tiga metode dalam penelitian ini yaitu Metode Observasi, Wawancara, dan Literatur.

2. Analisa Data

Teknik analisis data bertujuan menguraikan dan memecahkan masalah yang berdasarkan data yang diperoleh. Analisis yang digunakan analisis data kualitatif.

3. Analisa Masalah

Analisis masalah yang diperoleh selama penelitian pada Royal Ponsel di Jln. Sikambing No.57-B, Sekip, Kecamatan Medan Petisah dari hasil observasi serta wawancara bahwa untuk menentukan Kartu Paket Internet yang paling kanyak di minati Konsumen masih menggunakan sistem manual.

4. Penerapan Metode K-Means

Penerapan metode K-Means dalam penyelesaian masalah Kartu paket Internet yang paling banyak di minati konsumen pada Royal Ponsel dapat dilakukan dengan menerapkan Distance space untuk menghitung jarak antara data dan centroid. Adapun persamaan yang dapat digunakan salah satunya yaitu *Euclidean Distance Space*.

5. Rancangan Aplikasi

Metode perancangan yang digunakan untuk perancangan sistem pembelajaran dengan cara membuat simulasi yang menarik dalam bentuk animasi adalah metode perancangan terstruktur menggunakan UML.

6. Pembangunan Aplikasi

Tahap rancangan aplikasi digunakan untuk mempermudah dan dapat membuat program sehingga lebih terstruktur.

7. Implementasi

Implementasi merupakan suatu tindakan untuk mewujudkan terlaksananya rencana dalam mencapai tujuan tertentu Dan sistem siap untuk di oprasikan.

8. Penulisan Laporan

Pada tahap penulisan laporan mengacu pada sistematika yang telah ditetapkan sehingga laporan tersebut dapat tersaji secara runtut, mudah dipahami, dan mudah dibaca.

III. Hasil dan Pembahasan

Data yang digunakan adalah data hasil penjualan Kartu paket Internet dan Analisis yang digunakan analisis data kualitatif. Analisis data kualitatif adalah upaya yang dilakukan dengan jalan mengumpulkan, memilah-milah, mengklasifikasikan, dan mencatat yang dihasilkan catatan lapangan serta memberikan kode agar sumber datanya tetap dapat ditelusuri. Berikut merupakan sample dari data yang diperoleh yang akan digunakan untuk di *clustering*.

Tabel 1. Data Penjualan Kartu Paket Internet Tahun 2018

BULAN	PENJUALAN					
	Telkomsel	Axis	XL	IM3	Tri (3)	Smartfren
Januari	13.464	8.938	9.346	9.913	11.170	10.771
Februari	15.433	10.115	8.677	9.556	10.876	10.002
Maret	14.221	9.867	8.277	8.029	9.388	13.276
April	11.443	8.009	8.266	8.267	12.278	11.829
Mei	13.588	10.635	9.017	10.282	12.389	8.792
Juni	11.290	11.737	9.281	11.929	10.298	9.373
Juli	13.668	8.465	10.799	9.209	10.380	9.220
Agustus	13.201	8.775	10.367	9.209	9.902	8.998
September	12.700	9.005	8.367	8.381	9.296	10.262
Oktober	14.601	8.746	10.009	8.920	10.290	10.539
November	13.099	10.484	9.739	9.028	11.029	11.939
Desember	13.832	8.498	9.029	10.273	9.269	10.998

Algoritma K-Means *clustering* merupakan algoritma untuk mengelompokkan data yang belum memiliki label atau kelas dan dikelompokkan kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama dan mempunyai karakteristik yang berbeda dengan data yang ada diluar kelompok. Tahapan melakukan teknik *clustering* atau pengelompokan data menggunakan algoritma K-Means *clustering* dapat dijelaskan sebagai berikut :

1. Pilih jumlah *cluster* k .
2. Inisialisasi pusat *cluster* awal sebanyak k , biasa dilakukan dengan cara acak (*random*). Pusat *cluster* tersebut dijadikan sebagai nilai awal *cluster*.
3. Tempatkan semua data/objek ke *cluster* terdekat. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan metode jarak *Euclidean Distance* sebagai berikut:

$$D(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2} \quad (4.1)$$

Dimana :

$D(x_2, x_1)$ adalah nilai jarak data x_1 ke centroid x_2

x_{2j} adalah nilai centroid x_2 indeks ke- j

x_{1j} adalah nilai data x_1 indeks ke- j

4. Hitung kembali jarak pusat *cluster* baru dengan keanggotaan *cluster* yang sekarang. Nilai pusat *cluster* diperoleh dari nilai rata-rata semua keanggotaan *cluster* tertentu.
5. Tentukan kembali keanggotaan *cluster* setiap data berdasarkan 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.

Selanjutnya berdasarkan tabel 1, Maka akan dilakukan proses *clustering* dengan menggunakan algoritma K-Means *clustering*. Adapun tahapannya adalah sebagai berikut:

1. Inisialisasikan pusat *cluster* awal (*initial centroid*) dengan jumlah *cluster* (K) sebanyak 2 *cluster*, maka diambil hanya data bulan Januari dan Februari, adapun nilai pusat *cluster* awal dapat dilihat pada tabel 2, berikut:

Tabel 2. Nilai Pusat Cluster Awal (Iterasi-1)

Bulan	PENJUALAN					
	Telkomsel	Axis	XL	IM3	Tri (3)	Smartfren
Januari	13.464	8.938	9.346	9.913	11.170	10.771
Desember	13.832	8.498	9.029	10.273	9.269	10.998

2. Hitung jarak setiap data ke setiap pusat cluster awal (*initial centroid*) menggunakan persamaan 1. Adapun rincian perhitungan jarak dapat dilihat sebagai berikut:

Tabel 3. Proses Perhitungan Jarak Data ke *Centroid Cluster* (Iterasi-1)

Iterasi-1	C1	C2	Cluster
Januari	$\begin{aligned} &\sqrt{(13.464 - 13.464)^2} \\ &+ (8.938 - 8.938)^2 + (9.364 - 9.346)^2 + (9.913 - 9.913)^2 + (11.17 - 11.17)^2 + (10.771 - 10.771)^2 \\ &= 0 \end{aligned}$	$\begin{aligned} &\sqrt{(13.464 - 13.832)^2} \\ &+ (8.938 - 8.498)^2 + (9.364 - 9.029)^2 + (9.913 - 10.273)^2 + (11.17 - 9.269)^2 + (10.771 - 10.998)^2 \\ &= 2.06 \end{aligned}$	1
Februari	$\begin{aligned} &\sqrt{(15.433 - 13.464)^2} \\ &+ (10.115 - 8.938)^2 + (8.677 - 9.346)^2 + (9.556 - 9.913)^2 + (10.876 - 11.17)^2 + (10.002 - 10.771)^2 \\ &= 2.55 \end{aligned}$	$\begin{aligned} &\sqrt{(15.433 - 13.832)^2} \\ &+ (10.115 - 8.498)^2 + (8.677 - 9.029)^2 + (9.556 - 10.273)^2 + (10.876 - 9.269)^2 + (10.002 - 10.998)^2 \\ &= 3.06 \end{aligned}$	1
Maret	$\begin{aligned} &\sqrt{(14.221 - 13.464)^2} \\ &+ (9.867 - 8.938)^2 + (8.277 - 9.346)^2 + (8.029 - 9.913)^2 + (9.388 - 11.17)^2 + (13.276 - 10.771)^2 \\ &= 3.95 \end{aligned}$	$\begin{aligned} &\sqrt{(14.221 - 13.832)^2} \\ &+ (9.867 - 8.498)^2 + (8.277 - 9.029)^2 + (8.029 - 10.273)^2 + (9.388 - 9.269)^2 + (13.276 - 10.998)^2 \\ &= 3.58 \end{aligned}$	2
April	$\begin{aligned} &\sqrt{(11.443 - 13.464)^2} \\ &+ (8.009 - 8.938)^2 + (8.266 - 9.346)^2 + (8.267 - 9.913)^2 + (12.278 - 11.17)^2 + (11.829 - 10.771)^2 \\ &= 3.34 \end{aligned}$	$\begin{aligned} &\sqrt{(11.443 - 13.832)^2} \\ &+ (8.009 - 8.498)^2 + (8.266 - 9.029)^2 + (8.267 - 10.273)^2 + (12.278 - 9.269)^2 + (11.829 - 10.998)^2 \\ &= 4.51 \end{aligned}$	1
Mei	$\begin{aligned} &\sqrt{(13.588 - 13.464)^2} \\ &+ (10.635 - 8.938)^2 + (9.017 - 9.346)^2 + (10.282 - 9.913)^2 + (12.389 - 11.17)^2 + (8.792 - 10.771)^2 \\ &= 2.92 \end{aligned}$	$\begin{aligned} &\sqrt{(13.588 - 13.832)^2} \\ &+ (10.635 - 8.498)^2 + (9.017 - 9.029)^2 + (10.282 - 10.273)^2 + (12.389 - 9.269)^2 + (8.792 - 10.998)^2 \\ &= 4.38 \end{aligned}$	1
Juni	$\begin{aligned} &\sqrt{(11.29 - 13.464)^2} \\ &+ (11.737 - 8.938)^2 + (9.281 - 9.346)^2 + (11.929 - 9.913)^2 + (10.298 - 11.17)^2 + (9.373 - 10.771)^2 \\ &= 4.40 \end{aligned}$	$\begin{aligned} &\sqrt{(11.29 - 13.832)^2} \\ &+ (11.737 - 8.498)^2 + (9.281 - 9.029)^2 + (11.929 - 10.273)^2 + (10.298 - 9.269)^2 + (9.373 - 10.998)^2 \\ &= 4.84 \end{aligned}$	1
Juli	$\begin{aligned} &\sqrt{(13.668 - 13.464)^2} \\ &+ (8.465 - 8.938)^2 + (10.799 - 9.346)^2 + (9.209 - 9.913)^2 + (10.38 - 11.17)^2 + (9.22 - 10.771)^2 \\ &= 2.43 \end{aligned}$	$\begin{aligned} &\sqrt{(13.668 - 13.832)^2} \\ &+ (8.465 - 8.498)^2 + (10.799 - 9.029)^2 + (9.209 - 10.273)^2 + (10.38 - 9.269)^2 + (9.22 - 10.998)^2 \\ &= 2.95 \end{aligned}$	1
Agustus	$\begin{aligned} &\sqrt{(13.201 - 13.464)^2} \\ &+ (8.775 - 8.938)^2 + (10.367 - 9.346)^2 + (9.209 - 9.913)^2 + (9.902 - 11.17)^2 + (8.998 - 10.771)^2 \\ &= 2.53 \end{aligned}$	$\begin{aligned} &\sqrt{(13.201 - 13.832)^2} \\ &+ (8.775 - 8.498)^2 + (10.367 - 9.029)^2 + (9.209 - 10.273)^2 + (9.902 - 9.269)^2 + (8.998 - 10.998)^2 \\ &= 2.79 \end{aligned}$	1

Iterasi-1	C1	C2	Cluster
September	$\sqrt{(12.7 - 13.464)^2}$ $+ (9.005 - 8.938)^2 + (8.367 - 9.346)^2 + (8.381 - 9.913)^2 + (9.296 - 11.17)^2 + (10.262 - 10.771)^2$ $= 2.77$	$\sqrt{(12.7 - 13.832)^2}$ $+ (9.005 - 8.498)^2 + (8.367 - 9.029)^2 + (8.381 - 10.273)^2 + (9.296 - 9.269)^2 + (10.262 - 10.998)^2$ $= 2.47$	2
Oktober	$\sqrt{(14.601 - 13.464)^2}$ $+ (8.746 - 8.938)^2 + (10.009 - 9.346)^2 + (8.92 - 9.913)^2 + (10.29 - 11.17)^2 + (10.539 - 10.771)^2$ $= 1.89$	$\sqrt{(14.601 - 13.832)^2}$ $+ (8.746 - 8.498)^2 + (10.009 - 9.029)^2 + (8.92 - 10.273)^2 + (10.29 - 9.269)^2 + (10.539 - 10.998)^2$ $= 2.17$	1
November	$\sqrt{(13.099 - 13.464)^2}$ $+ (10.484 - 8.938)^2 + (9.739 - 9.346)^2 + (9.028 - 9.913)^2 + (11.029 - 11.17)^2 + (11.939 - 10.771)^2$ $= 2.20$	$\sqrt{(13.099 - 13.832)^2}$ $+ (10.484 - 8.498)^2 + (9.739 - 9.029)^2 + (9.028 - 10.273)^2 + (11.029 - 9.269)^2 + (11.939 - 10.998)^2$ $= 3.24$	1
Desember	$\sqrt{(13.832 - 13.464)^2}$ $+ (8.498 - 8.938)^2 + (9.029 - 9.346)^2 + (10.273 - 9.913)^2 + (9.269 - 11.17)^2 + (10.998 - 10.771)^2$ $= 2.06$	$\sqrt{(13.832 - 13.832)^2}$ $+ (8.498 - 8.498)^2 + (9.029 - 9.029)^2 + (10.273 - 10.273)^2 + (9.269 - 9.269)^2 + (10.998 - 10.998)^2$ $= 0$	2

3. Tentukan pusat *cluster* baru (*new centroid*) dengan cara menghitung rata-rata nilai atribut setiap data yang termasuk kedalam *cluster* tersebut. Perhitungan nilai rata-rata data dari bulan Januari hingga desember yang termasuk kedalam **Cluster 1 (C1)** sebanyak 9 data pada atribut Telkomsel, yaitu:
- $$= \frac{(13.464 + 15.433 + 11.443 + 13.588 + 11.29 + 13.668 + 13.201 + 14.601 + 13.099)}{9}$$
- = 13.310**

Hal yang sama juga berlaku untuk menghitung rata-rata nilai atribut setiap data yang termasuk kedalam *cluster* berikutnya. Adapun nilai *cluster* yang baru secara lengkap dapat dilihat pada tabel 4. berikut:

Tabel 4. Nilai Pusat Cluster Baru(Iterasi-2)

Cluster	PENJUALAN					
	Telkomsel	Axis	XL	IM3	Tri (3)	Smartfren
C1	13.310	9.545	9.500	9.590	10.957	10.163
C2	13.584	9.123	8.558	8.894	9.318	11.512

4. Kembali kelangkah nomor 2 sampai tidak ada data yang berpindah *cluster* atau sampai batas maksimal iterasi yang ditentukan. Adapun hasil perhitungan jarak data ke *centroid* baru pada iterasi ke-2 adalah sebagai berikut:

Tabel 5. Proses Perhitungan Jarak Data ke *Centroid* Baru (Iterasi-2)

Iterasi-1	C1	C2	Cluster
Januari	0.97	2.38	1
Februari	2.35	3.09	1
Maret	4.13	2.21	2
April	3.69	3.89	1
Mei	2.43	4.61	1
Juni	3.93	5.23	1
Juli	2.09	3.46	1
Agustus	1.99	3.21	1
September	2.49	1.63	2
Oktober	1.90	2.27	1
November	2.11	2.57	1

Iterasi-1	C1	C2	Cluster
Desember	2.37	1.69	2

Berdasarkan hasil perolehan kelompok (*cluster*) pada tabel 3 (iterasi ke-1) dengan tabel 5 (iterasi ke-2), maka tidak adanya terjadi perpindahan *cluster* antara data pada bulan Januari hingga Desember, Sehingga adapun nilai pusat *cluster* akhir yang didapati dapat dilihat pada tabel.6 dan *cluster* dari setiap data dapat dilihat pada tabel 7, sebagai berikut:

Tabel 6.Nilai Pusat *Cluster* Akhir

Cluster	PENJUALAN					
	Telkomsel	Axis	XL	IM3	Tri (3)	Smartfren
C1	13.310	9.545	9.500	9.590	10.957	10.163
C2	13.584	9.123	8.558	8.894	9.318	11.512

IV. Kesimpulan

Setelah melihat dan mengamati hasil analisis yang didapat pada perangkat lunak yang telah diselesaikan, maka dapat diberikan kesimpulan sebagai berikut :

1. Dalam menerapkan metode k-means pada penelitian ini, hanya menerapkan 2 iterasi, dikarenakan tidak terjadi perulangan.
2. Dengan penerapan metode K-means maka dapat diketahui penjualan terbanyak terlihat pada cluster 1.
3. Hasil yang diharapkan telah mampu memenuhi dan menjawab pokok permasalahan dalam penelitian yakni pengelompokan data penjualan.

V. Daftar pustaka

- [1] F. A. Sianturi, "Analisa Decision Tree Dalam Pengolahan Data Siswa," *MEANS (Media Inf. Anal. dan Sist.*, vol. 3, no. 2, pp. 166–172, 2018, [Online]. Available: http://ejournal.ust.ac.id/index.php/Jurnal_Means/.
- [2] M. A. K-means, W. Wahyuni, and H. Fahmi, "Penerapan Data Mining Clustering Pada Siswa-Siswi SMK Swasta Jaya Krama Beringin Dalam Menerima Potongan Biaya Administrasi Sekolah Dengan," vol. 3, no. 2, pp. 1–7, 2020.
- [3] M. S. Fricles Ariwisanto Sianturi, "KOMBINASI METODESIMPLEADDITIVEWEIGHTING (SAW)DENGANALGORITMA NEAREST NEIGHBOR UNTUK REKRUITMEN KARYAWAN," *Mantik Penusa*, vol. 3, no. 2, pp. 1–9, 2019, doi: .1037//0033-2909.I26.1.78.
- [4] M. S. Fricles Ariwisanto Sianturi, "ANALISA PENGARUH LOG TRANSAKSI PADA SISTEM KOMPUTER," *Mantik Penusa*, vol. 2, no. 2, pp. 67–70, 2018, [Online]. Available: <http://ejournal.pelitanusantara.ac.id/index.php/mantik/article/view/422>.
- [5] Fricles Ariwisanto Sianturi, "Implementasi Sistem Pendukung Keputusan Kenaikan Jabatan Guru Dengan Model Profile Matching Pada Sekolah Sma Swasta Raksana Medan," *Mantik Penusa*, vol. 18, no. 2, pp. 44–52, 2015, [Online]. Available: <http://ejournal.pelitanusantara.ac.id/index.php/mantik/article/view/43>.
- [6] F. A. Sianturi *et al.*, "Sistem Pendukung Keputusan Penentuan Kelayakan Perkreditan Anggota Koperasi (Studi Kasus Pada Koperasi Kozero)," *Tek. Inform. Unika St. Thomas*, vol. 02, pp. 88–100, 2017.