Penerapa Metode *Hierarchical Clustering* dan *K-means Clustering* untuk Mengelompokkan Potensi Lokasi Penjualan LINKAJA.

Putra Prima Arhandi
Jurusan Teknologi Informasi
Politeknik Negeri Malang
Malang
putraprima@polinema.ac.id

Annisa Tufika Firdausi
Jurusan Teknologi Informasi
Politeknik Negeri Malang
Malang
annisa.taufika@polinema.ac.id

Ivon Fadhila Rahma
Jurusan Teknologi Informasi
Politeknik Negeri Malang
Malang
ivon.rahma@gmail.com

***Abstrak*— Tempat atau lokasi yang tepat merupakan salah satu faktor peningkatan strategi bisnis, maka dari itu suatu perusahaan haruslah mengelompakkan mana saja lokasi yang berpotensi bisnis bagi penjualannya. Dan jika harus dikelompokkan secara manual membutuhkan waktu yang sangat lama. Oleh sebab itu, Clustering menjadi solusi untuk mengatasi masalah tersebut. Clustering akan mengelompokkan lokasi yang menguntungkan, cukup, dan kurang. Clustering memiliki dua metode, yaitu partisi dan hierarki. Dua metode ini memiliki kelebihan dan kekurangan masing-masing, dan dengan menggabungkan keduanya dapat diperoleh hasil cluster yang lebih baik. Dari hasil cluster dengan menggunakan data penjualan LINKAJA di PT Telekomunikasi, maka diperoleh hasil bahwa gabungan metode Hierarchical Clustering dan K-means memberikan hasil cluster yang lebih baik dalam pengelompokan lokasi penjualan LINKAJA dengan parameter saldo penjaualan dan umur.**

Kata kunci— Clustering, Hierarchical Clustering, K-Means Clustering.

I. PENDAHULUAN

LINKAJA adalah wajah baru dari TCASH, layanan keuangan digital (e-wallet) milik Telkomsel, anak usaha PT Telekomunikasi Indonesia (Persero) Tbk atau Telkom. Selain itu, LinkAja merupakan gabungan dari layanan keuangan elektronik milik PT Bank Mandiri (Persero) Tbk dengan E-Cash, UnikQu dari PT Bank Negara Indonesia (Persero) Tbk, dan PT Bank Rakyat Indonesia (Persero) Tbk dengan merek T-Bank. memiliki fungsi yang sama dengan uang tunai sebagai alat pembayaran yang sah, di mana nilainya setara dengan nilai uang tunai yang disetorkan terlebih dahulu ke rekening LINKAJA dan uang yang disetorkan bukanlah bersifat simpanan sebagaimana diatur dalam peraturan perundang-undangan perbankan dan oleh karenanya LINKAJA tidak memberikan bunga serta tidak dijamin oleh

Lembaga Penjamin Simpanan [1]. Stiker LINKAJA dan saldo LINKAJA bisa dapatkan di banyak tempat salah satunya seperti, Grapari, Alfamart, Indomart, tempat yang berlogo Bang LINKAJA atau Stand LINKAJA yang berada di lokasi – lokasi tertentu. Stand LINKAJA banyak terdapat di beberapa tempat yang sudah bekerja sama dengan Telkomsel. Hal Tersebut untuk mempermudah pengguna LINKAJA dalam menjangkau Stiker LINKAJA atau Top Up Saldo tanpa harus membayar pajak seperti saat bertransaksi di Alfamart, Indomart, maupun transaksi melalui Bank.

Saldo LINKAJA pada setiap Stand khusus di PT Telekomunikasi Selular disediakan pada bagian Divisi Broadband Digital Sales (BDS). Namun kebutuhan yang berbeda di setiap lokasi Stand LINKAJA dan proses penambahan saldo saat saldo habis pada setiap Stand, harus melalui penanggung jawab di Divisi BDS. Waktu respon yang tidak menentu mengakibatkan pengisian saldo LINKAJA yang sudah habis pada setiap Stand LINKAJA cukup membutuhkan waktu dan berdampak pada pelanggan yang sedang membeli saldo. Maka dari itu perlu penambahan jumlah saldo pada lokasi Stand LINKAJA yang berpotensi.

Menurut Kotler (2009) place merupakan segala sesuatu yang menunjukkan berbagai kegiatan bisnis untuk membuat produk agar mudah diperoleh oleh pelanggan dan selalu tersedia bagi pelanggan. Apabila berada dalam kondisi bisnis yang memiliki persaingan ketat, faktor pemilihan lokasi menjadi komponen utama yang penting agar usaha yang dijalankan juga dapat bersaing secara efektif, maka perlu adanya lokasi usaha yang strategis dan mudah dijangkau oleh konsumen. Ketepatan pemilihan lokasi merupakan salah satu faktor yang menentukan kesuksesan sebuah usaha. Para pengusaha selalu memiliki pertimbangan yang matang mengenai lokasi sebelum membuka usahanya[2].

Metode K-means merupakan metode clustering yang paling sederhana dan umum [3]. Hal ini dikarenakan K-means mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien [4]. Namun, K-means mempunyai mempunyai kelemahan yang diakibatkan oleh penentuan pusat awal cluster. Hasil cluster yang terbentuk dari metode K-means ini sangatlah tergantung pada inisiasi nilai pusat awal cluster yang diberikan [3]. Hal ini menyebabkan hasil clusternya berupa solusi yang sifatnya local optimal.

Oleh karena itu dalam rangka meningkatkan strategi bisnis, peran PT Telekomunikasi Selular yaitu mengelompokkan lokasi Stand LINKAJA agar dapat

mengoptimalkan strategi bisnis dengan cara penambahan saldo LINKAJA, stiker LINKAJA dan penambahan kontrak kerja disetiap tempat yang berpotensi. Hal ini tidak semata mengutamakan keuntungan pihak PT Telekomunikasi Selular saja melainkan mempermudah konsumen dalam mendapatkan pelayanan, fasilitas, dan jangkauan dari PT Telekomunikasi Selular sebagai operator selular terbesar di Indonesia.

Pengelompokan tersebut dapat menggunakan metode pengelompokan dengan algoritma K-Mean dikolaborasikan oleh metode hierarki untuk penentuan pusat awal cluster. Penggabungan antara metode K-means Clustering dan Hierarchical clustering dikarenakan setiap metode memiliki kekurangan dan kelebihannya masing masing. Untuk metode K-means Clustering kelebihannya dapat melakukan analisis sampel dalam ukuran yang lebih besar dengan lebih efisien. Selain itu, hanya memiliki sedikit kelemahan pada data outlier, ukuran jarak yang digunakan, dan variabel tak relevan atau variabel yang tidak tepat. Sedangkan kelemahannya adalah untuk titik bakal random lebih buruk dari pada metode hirarkhi. Jika Hierarchical Clustering memiliki kelebihan mempercepat pengolahan dan menghemat waktu karena data yang diinputkan akan membentuk hierarki atau membentuk tingkatan tersendiri sehingga mempermudah dalam penafsiran, namun kelemahan dari metode ini adalah seringnya terdapat kesalahan pada data outlier, perbedaan ukuran jarak yang digunakan, dan terdapatnya variabel yang tidak relevan[11]. Jadi penggabungan keduanya untuk mendapat hasil yang lebih baik.

Dengan data yang sudah di kelompokan menggunakan algoritma K-Means, diharapkan dapat mempermudah PT Telekomunikasi Selular dalam meningkatkan strategi pemasaran dengan penambahan saldo LINKAJA, stiker LINKAJA, dan perpanjangan kontrak di setiap tempat berdasarkan tempat yang menguntungkan, cukup, dan kurang.

II. LANDASAN TEORI

2.1 LINKAJA

LINKAJA adalah uang elektronik yang diselenggarakan oleh Telkomsel yang telah terdaftar dan diawasi oleh Bank Indonesia, memiliki fungsi yang sama dengan uang tunai sebagai alat pembayaran yang sah, di mana nilainya setara dengan nilai uang tunai yang disetorkan terlebih dahulu ke rekening LINKAJA dan uang yang disetorkan bukanlah bersifat simpanan sebagamana diatur dalam peraturan perundang-undangan perbankan dan oleh karenanya LINKAJA tidak memberikan bunga serta tidak dijamin oleh Lembaga Penjamin Simpanan.

Nomor Rekening LINKAJA adalah Nomor Telepon selular dari Pemegang LINKAJA yang tercatat pada saat registrasi, dimana 1 (satu) Rekening LINKAJA hanya berlaku untuk 1 (satu) Nomor Telepon selular, dan juga sebaliknya.

Pengisian Saldo LINKAJA (Cash In/Top Up) dapat dilakukan di titik pelayanan Telkomsel atau Mitra LKD yang telah bekerjasama dengan Telkomsel untuk menyelenggarakan fasilitas Pengisian Saldo (Cash In/Top Up) sesuai dengan ketentuan yang berlaku, atau melalui transfer dana dari rekening uang atau bank lainnya. Transaksi menggunakan LINKAJA hanya dapat dilakukan selama saldo LINKAJA mencukupi [1].

Untuk kemudahan Pemegang LINKAJA, dari waktu ke waktu Telkomsel dapat menyediakan media seperti tetapi tidak terbatas pada Aplikasi (LINKAJA Wallet) dan NFC Tag (LINKAJA TAP) penggunan non Telkomsel, USSD (*800#) khusus untuk pengguna Telkomsel, SMS (2828) yang seluruhnya merupakan satu kesatuan yang tidak terpisahkan dari produk LINKAJA dan informasi dari media akan dijelaskan lebih lanjut dalam website LINKAJA [1].

$$R(x,y) = \begin{cases} 1 & jika\ f(x,y) >= T \\ 0 & sebaliknya \end{cases} \quad (1)$$

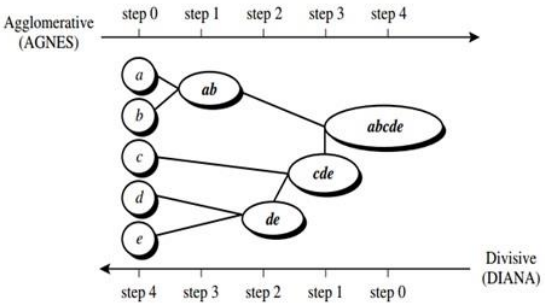
2.2 Clustering

Analisis kelompok (cluster analysis) adalah mengelompokkan data (objek) yang didasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya.

Clustering merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan cluster. Objek yang di dalam cluster memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan cluster yang lain. Partisi tidak dilakukan secara manual melainkan dengan suatu algoritma clustering. Oleh karena itu, clustering sangat berguna dan bisa menemukan group atau kelompok yang tidak dikenal dalam data. Clustering banyak diguakan dalam berbagai aplikasi seperti misalnya pada business inteligance, pengenalan pola citra, web search, bidang ilmu biologi, dan untuk keamanan (security)[5].

2.3 Hierarchical Clustering

Pada hierarchical clustering data dikelompokkan melalui suatu bagan yang berupa hirarki, dimana terdapat penggabungan dua grup yang terdekat disetiap iterasinya ataupun pembagian dari seluruh set data kedalam cluster.



Gambar 2.1 Hierarchical Clustering

Langkah melakukan Hierarchical clustering [6]:

a. Identifikasi item dengan jarak terdekat
b. Gabungkan item itu kedalam satu cluster
c. Hitung jarak antar cluster
d. Ulangi dari awal sampai semua terhubung

2.4 Clustering K-Means

K-Means adalah suatu teknik pengelompokan data yang mana keberadaan tiap-tiap titik data dalam suatu cluster ditentukan oleh derajat keanggotaan. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981[7].

K-Means merupakan algoritma clustering yang berulangulang. Algoritma K- Means dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya cluster yang ingin dibentuk. Kemudian tetapkan nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari cluster atau bisa disebut dengan centeroid menggunakan rumus hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasikan setiap data berdasarkan kedeketannya dengan centroid. Lakukan langkah tersebut hingga nilai centroid tidak berubah (stabil)[8].

Algoritma k-means adalah algoritma yang mempartisi data ke dalam cluster

– cluster sehingga data yang memiliki kemiripan berada pada satu cluster yang sama dan data yang memiliki ketidaksamaan berada pada cluster yang lain. Sarwono mengemukakan secara lebih detail, algoritma K-Means adalah sebagai berikut[9] :

a. Menentukan k sebagai jumlah kluster yang ingin di bentuk.
b. Membangkitkan nilai random untuk pusat cluster awal (centroid) sebanyak k.
c. Menghitung jarak setiap data input terhadap masing – masing centroid menggunakan rumus jarak Euclidean (Euclidean Distance) hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Berikut adalah persamaan Euclidian Distance:

$$d(xi, \mu j) = \sqrt{\sum (xi - \mu j)^2} \quad (1)$$

Dimana :
xi : Data kriteria,
µj : Centroid pada cluster ke-j

d. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid (jarak terkecil).
e. Memperbaharui nilai centroid. Nilai centroid baru di peroleh dari rata-rata cluster yang bersangkutan dengan menggunakan rumus:
Dimana:

$$\mu j(t+1) = \frac{1}{Nsj} \sum_{j \in sj} xj \quad (2)$$

µj(t+1) : Centroid baru pada iterasi ke (t +1)
Nsj : Banyak data pada cluster Sj.

f. Melakukan perulangan dari langkah 2 hingga 5, sampai anggota tiap cluster tidak ada yang berubah.
Jika langkah 6 telah terpenuhi, maka nilai p usat cluster (µj) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data (Sarwono).

2.5 Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek. Dalam pengklasifikasian data terdapat dua proses yang dilakukan yaitu:

a. Proses training
Pada proses training digunakan training set yang telah diketahui label-labelnya untuk membangun model atau fungsi.
b. Proses testing
Untuk mengetahui keakuratan model atau fungsi yang akan dibangun pada proses training, maka digunakan data yang disebut dengan testing set untuk memprediksi label-labelnya.

Klasifikasi dokumen adalah pemberian kategori yang telah didefinisikan kepada dokumen yang belum memiliki kategori [12] Mengklasifikasi dokumen merupakan salah satu cara untuk mengorganisasikan dokumen. Dokumen-dokumen yang memiliki isi yang sama akan dikelompokkan ke dalam kategori yang sama. Dengan demikian, orang-orang yang melakukan pencarian informasi dapat dengan mudah melewatkan kategori yang tidak relevan dengan informasi yang dicari atau yang tidak menarik perhatian.

Peningkatan jumlah dokumen yang begitu pesat telah mendorong berkembangnya metode pengklasifikasian secara otomatis. Metode ini dapat melakukan klasifikasi dengan cara belajar dari sekumpulan contoh dokumen yang telah diklasifikasi sebelumnya. Keuntungan dari metode ini adalah dapat menghemat waktu kerja dan memperoleh hasil yang lebih baik. Metode ini adalah metode support vector machine.

III. METODOLOGI PENELITIAN

Dalam penelitian ini, dilakukan beberapa langkah, yaitu:

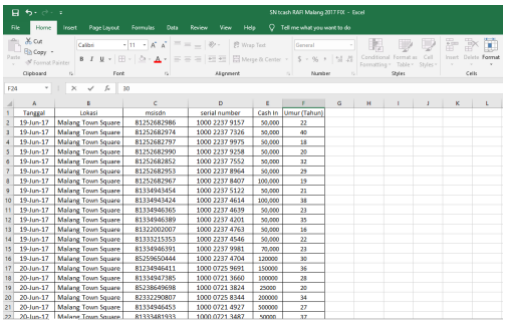
3.1 Data Set

Metode pengambilan data pada penyusunan penulisan ini, adalah :

a. Studi literatur dengan tujuan :
- Mempelajari dan memahami algoritma Hierarchical Clustering dan K- Means dalam data mining dengan mengumpulkan literatur yang berkaitan, yang didapat dari beberapa referensi - referensi, jurnal – jurnal, melalui buku ataupun website.
- Mempelajari hal yang dapat meningkatkan strategi bisnis.
- Mengetahui data transaksi LINKAJA di beberapa tempat.

b. Pengumpulan data melalui GraPARI Telkomsel Malang.
- Pengumpulan data transaksi LINKAJA yang sudah tercatat di GraPARI Telkomsel Malang.
- Pengumpulan data dengan wawancara untuk mendapatkan data seakurat mungkin. Pengumpulan

data dilakukan sesuai dengan sumber dan jenis data yang diperlukan.

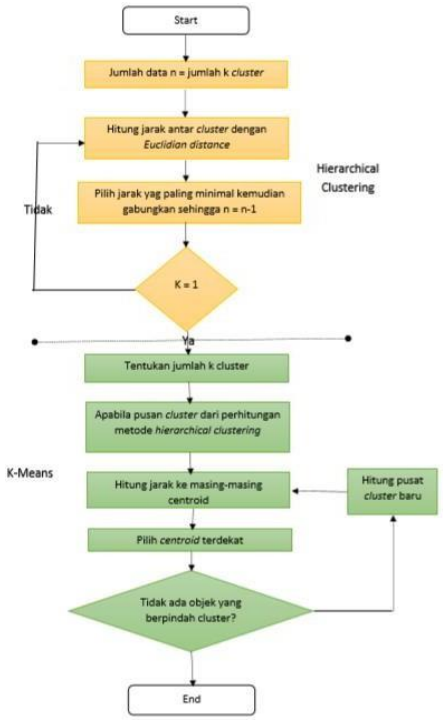
- 

Tanggal	Lokasi	msisdn	serial number	Cash In	Umur (Tahun)
19-Jun-17	Malang Town Square	81252682986	1000 2237 9157	50,000	22
19-Jun-17	Malang Town Square	81252682974	1000 2237 7326	50,000	40
19-Jun-17	Malang Town Square	81252682797	1000 2237 9975	50,000	18
19-Jun-17	Malang Town Square	81252682990	1000 2237 9258	50,000	20
19-Jun-17	Malang Town Square	81252682852	1000 2237 7552	50,000	32
19-Jun-17	Malang Town Square	81252682953	1000 2237 8964	50,000	29
19-Jun-17	Malang Town Square	81252682967	1000 2237 8407	100,000	19
19-Jun-17	Malang Town Square	81334943454	1000 2237 5122	50,000	21
19-Jun-17	Malang Town Square	81334943424	1000 2237 4614	100,000	38
19-Jun-17	Malang Town Square	81334946365	1000 2237 4639	50,000	23
19-Jun-17	Malang Town Square	81334946389	1000 2237 4201	50,000	35
19-Jun-17	Malang Town Square	81322002007	1000 2237 4763	50,000	16
19-Jun-17	Malang Town Square	81333215353	1000 2237 4546	50,000	22
19-Jun-17	Malang Town Square	81334946391	1000 2237 9981	70,000	23
19-Jun-17	Malang Town Square	85259650444	1000 2237 4704	120000	30
20-Jun-17	Malang Town Square	81234946411	1000 0725 9691	150000	36
20-Jun-17	Malang Town Square	81334947385	1000 0721 3660	100000	28
20-Jun-17	Malang Town Square	85238649698	1000 0721 3824	25000	20
20-Jun-17	Malang Town Square	82332290807	1000 0725 8344	200000	34
20-Jun-17	Malang Town Square	81334946453	1000 0721 4927	500000	27
20-Jun-17	Malang Town Square	81333481933	1000 0721 3487	50000	37

Gambar 3(a) Data Sampel

3.2 Proses Clustering

Berikut ini adalah tahapan pengolahan data pada penerapan metode hierarchical clustering dan k- means clustering untuk mengelompokkan potensi lokasi penjualan LINKAJA mulai dari start untuk menghitung titik pusat dengan menggunakan hierarchical clustering hingga menemukan titik pusat pada tiga kelompok dan di lanjutkan menghitungan dengan metode k-means dengan pusat yang telah dihitung sebelumnnya, hingga menghasilkan anggota sesuai dengan kelompoknya yaitu kelompok menguntungkan, cukup, dan kurang. Proses tersebut ditunjukkan seperti Gambar 3):



Gambar 3.1 Flowchart Kombinasi metode Hierarchical clustering dan Kmeans clustering.

IV. KESIMPULAN DAN SARAN

A. *Kesimpulan*

Adapun kesimpulan yang dapat diambil dari hasil penelitian yang dilakukan mengenai penerapan metode hierarchical clustering dan k-means clustering untuk mengelompokkan potensi lokasi penjualan linkaja adalah sebagai berikut.

a. Hasil clustering menggunakan metode K-means clustering dengan metode hierarchical clustering untuk menentukan titik pusatnya menghasilnya iterasi yang lebih sedikit sehingga mempercepat proses.
b. Dari hasil pengujian akurasi penerapan metode hierarchical clustering dan k-means clustering untuk mengelompokkan potensi lokasi penjualan linkaja pada data tahun 2017 lokasi yang menguntungkan adalah Malang Town Square, Jatim Park 3, Batu Night Spectaculer dan lokasi yang cukup adalah Museum Angkut dan lokasi yang kurang adalah Kantin Fakultas Ilmu Sosial dan Ilmu Politik Universitas Brawijaya, Kantin Perpustakaan Universitas Brawijaya, Batu Town Square, Universitas Negeri Malang, KFC Kawi, KFC Sarina, Mc Donald Watugong, Mc Donald Kayutangan, Jatim Park 2, Alun – alun Kota Batu. Sedangkan data tahun 2018 lokasi yang menguntungkan adalah Malang Town Square, Jatim Park 3 dan lokasi yang cukup adalah Museum Angkut dan lokasi yang kurang adalah Kantin Fakultas Ilmu Sosial dan Ilmu Politik Universitas Brawijaya, Kantin Perpustakaan Universitas Brawijaya, Batu Town Square, Batu Night Spectaculer, Universitas Negeri Malang, KFC Kawi, KFC Sarina, Mc Donald Watugong, Mc Donald Kayutangan, Jatim Park 2, Alun – alun Kota Batu.
c. Hasil klasterisasi metode hierarchical clustering dan k-means clustering untuk mengelompokkan potensi lokasi penjualan linkaja memperoleh tingkat keakurasian system 93,33% pada tahun 2017 dan memperoleh tingkat keakurasian system 90,90% pada tahun 2018 berdasarkan 3 kelompok yang diuji.

B. *Saran*

Berdasarkan penelitian hal yang disarankan yaitu :

a. penelitian ini masih dapat dikembangkan lagi menjadi sistem yang lebih lengkap lagi karena masih banyak data pendukung responden yang dapat dianalisis dari hasil clustering.
b. Sistem disarankan menggunakan desktop karena proses clustering nya yang lama jika menggunakan php.

DAFTAR PUSTAKA

[1] Telekomunikasi Selular. (2017). Syarat Dan Ketentuan Layanan LINKAJA [Online] Tersedia: https://www.LinkAja.id/terms_conditions

[2] Kotler, Philip dan Kevin Lane Keller. 2009. Manajemen Pemasaran. Terjemahan oleh Benyamin Molan. Edisi Keduabelas. Jilid 1. Jakarta: Indeks

[3] B. Santosa. 2007 Data Mining. Teknik Pemanfaatan Data untuk Keperluan Bisnis, First Edition ed. Yogyakarta: Graha Ilmu
[4] K. Arai and A. R. Barakbah. 2007. "Hierarchical K-means: an algorithm for centroids initialization for Kmeans,"
[5] Edy Irwansyah, Muhammad Faisal. “Advanced Clustering Teori dan Aplikasi”
[6] Han, J., Kamber, M., Pei, J. 2012 : Data Mining Concept and Techniques, 3rd ed Morgan Kaufmann-Elsevier, Amsterdam
[7] Kusumadewi dan Purnomo. 2010. Aplikasi Logika Fuzzy untuk mendukung keputusan. Yogyakarta: Graha Ilmu
[8] Rismawan, T. 2008. Aplikasi K-Means Untuk Pengelompokkan Mahasiswa
[9] Sarwono, Y. T. (n.d.). Aplikasi Model Jaringan Syaraf Tiruan Dengan Radial Basis Function Untuk Mendeteksi Kelainan Otak (Stroke Infark). Jurnal Sistem Informasi, 34.