

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1. Tinjauan Pustaka

Penelitian tentang clustering sudah banyak dilakukan sebelumnya. Dalam penelitiannya Dwi Noviati Nango (2012) menggunakan clustering untuk mengelompokkan data APBD Kabupaten XYZ. Penelitian yang dilakukan Dwi Noviati Nangi menghasilkan 3 cluster yang masing-masing mewakili nilai pendapatan dari tiap-tiap anggota cluster. Pada penelitian ini menggunakan jumlah k sebanyak 3, yaitu cluster pendapatan, cluster belanja tidak langsung serta cluster belanja langsung. Serta melakukan pengujian 2 parameter nilai centroid yaitu 3 nilai centroid dan 2 nilai centroid. Hasil cluster dengan 3 nilai centroid digunakan untuk mengestimasi nilai belanja tidak langsung dan belanja langsung berdasarkan nilai pendapatan yang sudah diketahui sebelumnya.

Hafilah Hamimi (2014) dalam penelitiannya yang berjudul Analisis Data Anggaran Pendapatan Belanja Daerah Menggunakan *Clustering K-Means dan Forecasting*. Dalam penelitiannya Hafilah Hamimi menggunakan 4 atribut yaitu cluster data pendapatan asli daerah (PAD), cluster data pendapatan transfer (PT), cluster data lain-lain pendapatan yang sah (PYS) dan cluster data jumlah total anggaran belanja. Serta pada clustering pengujian 1 parameter nilai centroid yaitu 3 nilai centroid. Hasil pengelompokkan data APBD (2009-2013) menggunakan clustering k-means membentuk 3 cluster. Hasil forecasting data APBD Kota

Padang dimasa yang akan datang berdasarkan data APBD 5 tahun terakhir (2009-2013) serta keterkaitan masing-masing atribut dengan menggunakan multiple linear regression. Sehingga dengan demikian hasil prediksi data APBD Kota Padang tahun 2014 adalah sebesar 2.352.906,8 juta rupiah.

Rian Saadillah Sukamdi (2016) menggunakan clustering untuk mengelompokkan data mahasiswa dengan melihat pola prestasi akademik mahasiswa melalui nilai UAN dan IPK mahasiswa STMIK AKAKOM. Hasil pengelompokkan data mahasiswa berdasarkan nilai uan dan ipk menjadi 3 kelompok yaitu kelompok 1, kelompok 2, dan kelompok 3. Hasil pengelompokkan data menggunakan formula *euclidean* distance pada angkatan 2008 dengan jumlah data sebanyak 132 menghasilkan 13 iterasi dengan nilai centroid1 5.5 dan 2, centroid2 6.8 dan 3, dan centroid3 7.5 dan 4. Dari hasil pengelompokkan menggunakan nilai uan dan ipk jumlah mahasiswa terbanyak masuk pada kelompok yang kedua, yaitu sebanyak 55 mahasiswa dengan nilai tengah uan 6.8 dan ipk 3.

Vika Indiya Sari (2017) menerapkan metode K-Mean dan KNN ke dalam program berbasis desktop untuk pengklasifikasian kualitas beras. Di dalam penelitiannya Vika Indiya Sari menggunakan objek data mutu giling beras sebagai inputannya yaitu derajat sosoh, kadar air beras, menir, butir kuning rusak, dan butir kapur. Output yang dihasilkan dari penelitian tersebut adalah Mutu I, Mutu II, Mutu III, Mutu IV, dan Mutu V. hasil dari aplikasi mampu membantu pengguna untuk

menentukan kualitas mutu giling beras dengan metode K-Means dan KNN sebesar 94% untuk data pelatihan dan 100% untuk data uji.

Rian Bagus Amrodin (2016) menggunakan metode K-Means untuk mengidentifikasi osteoporosis melalui dataset citra gigi. Tujuan dari penelitian yang dilakukan oleh Rian Bagus Amrodin adalah untuk membantu paramedic dalam mengidentifikasi osteoporosis melalui citra gigi. Hasil pengujian yang diperoleh dari penelitian tersebut adalah nilai akurasi pelatihan dan nilai akurasi pengujian. Hasil pengujian terbaik diperoleh dengan akurasi pelatihan sebesar 84,62% dan akurasi pengujian sebesar 66,67%.

Shintya Bunga Utami (2020) menggunakan metode K-Means Clustering untuk mengelompokkan data penggunaan alat kontrasepsi di masing-masing kecamatan yang ada di kota Yogyakarta. Tujuan dari penelitian yang dilakukan oleh peneliti adalah untuk membantu memberikan informasi pengelompokkan data penggunaan alat kontrasepsi di masing-masing kecamatan yang ada di kota Yogyakarta agar instansi yang terkait dapat dengan tepat menyalurkan permintaan alat kontrasepsi di masing-masing kecamatan. Hasil pengelompokkan data penggunaan alat kontrasepsi di masing-masing kecamatan yang ada di kota Yogyakarta menggunakan clustering k-means membentuk 3 cluster yaitu Rendah, Menengah, dan Tinggi.

Tabel 2.1. Perbandingan Hasil Penelitian

No	Penulis	Objek	Metode	Aplikasi	Hasil
1.	Dwi Noviati Nango 2012	Data APBD kabupaten XYZ	Algoritma K- Means	-	Protipe clusterisasi data APBD Kabupaten XYZ.
2.	Hafilah Hamimi 2014	Data APBD 2009-2013 Kota Padang	Clustering K- Means dan Forecasting	Weka	Prediksi nilai total anggaran belanja dimasa yang akan datang dengan forecasting menggunakan multiple linear regression terhadap data APBD pada Kota Padang.
3.	Rian Saadillah Sukamdi 2016	Data nilai UAN dan IPK Mahasiswa AKAKOM	Clustering menggunakan metode K-Means	NetBeans	Klasifikasi data mahasiswa STMIK AKAKOM
4.	Vika Indiya Sari 2017	Data mutu giling varietas padi lokal Indonesia	K-Means dan K-Nearest Neighbor	Matlab	Membangun sistem untuk menentukan klasifikasi mutu beras.
5.	Rian Bagus Amrodin 2016	Dataset gigi	K-Means	NetBeans	Hasil identifikasi dan hasil akurasi dari dataset gigi yang telah diuji.

6.	Shintya Bunga Utami 2020	Data jumlah peserta kb aktif di kota Yogyakarta tahun 2017.	Clustering K-Means	NetBeans	Klasifikasi data jumlah peserta KB aktif di Yogyakarta berdasarkan penggunaan alat kontrasepsi di masing-masing kecamatan.
----	--------------------------	---	--------------------	----------	--

2.2. Dasar Teori

Dasar teori adalah seperangkat definisi atau konsep yang telah disusun rapi dan sistematis tentang variable-variabel dalam sebuah penelitian. Dasar teori ini akan menjadi dasar yang kuat dalam penelitian yang akan dilakukan.

2.2.1. Keluarga Berencana (KB)

1. Definisi KB

Keluarga berencana merupakan usaha untuk mengukur jumlah anak dan jarak kelahiran anak yang diinginkan. Maka dari itu, pemerintah mencanangkan program atau cara untuk mencegah dan menunda kehamilan (Sulistiyawati, 2013).

2. Tujuan Program KB

Tujuan dilaksanakan program KB yaitu untuk membentuk keluarga kecil sesuai dengan kekuatan sosial ekonomi suatu keluarga dengan cara pengaturan kelahiran anak agar diperoleh suatu keluarga bahagia dan sejahtera yang dapat

memenuhi kebutuhan hidupnya (Sulistiyawati, 2013). Tujuan program KB lainnya yaitu untuk menurunkan angka kelahiran yang bermakna, untuk mencapai tujuan tersebut maka diadakan kebijakan yang dikategorikan dalam tiga fase (menjarangkan, menunda, dan menghentikan) maksud dari kebijakan tersebut yaitu untuk menyelamatkan ibu dan anak akibat melahirkan pada usia muda, jarak kelahiran yang terlalu dekat dan melahirkan pada usia tua (Hartanto, 2012).

2.2.2. Kontrasepsi

1. Definisi Kontrasepsi

Kontrasepsi merupakan usaha-usaha untuk mencegah terjadinya kehamilan. Usaha-usaha itu dapat bersifat sementara dan permanen (Wiknjosastro, 2017). Kontrasepsi yaitu pencegahan terbuahnya sel telur oleh sel sperma (konsepsi) atau pencegahan menempelnya sel telur yang telah dibuahi ke dinding rahim (Nugroho dan Utama, 2014).

2. Efektivitas (Daya Guna) Kontrasepsi

Menurut Wiknjosastro (2017) efektivitas atau daya guna suatu cara kontrasepsi dapat dinilai pada 2 tingkat, yakni :

- a. Daya guna teoritis (*theoretical effectiveness*), yaitu kemampuan suatu cara kontrasepsi untuk mengurangi terjadinya kehamilan yang tidak diinginkan, apabila kontrasepsi tersebut digunakan dengan mengikuti aturan yang benar.
- b. Daya guna pemakaian (*use effectiveness*), yaitu kemampuan kontrasepsi dalam keadaan sehari-hari dimana pemakaiannya dipengaruhi oleh faktor-

faktor seperti pemakaian yang tidak hati-hati, kurang disiplin dengan aturan pemakaian dan sebagainya.

3. Jenis-jenis Alat Kontrasepsi

Jika dilihat dari jangka waktu serta efektivitasnya, alat kontrasepsi dibagi ke dalam dua jenis, yaitu kontrasepsi temporer/sementara dan permanen. Jenis alat kontrasepsi pertama, kontrasepsi temporer umumnya digunakan untuk mencegah dan menjaga jarak antar kehamilan, serta efeknya akan hilang setelah alat terkait tidak digunakan lagi. Sebaliknya, kontrasepsi permanen ditujukan untuk “mensterilkan” pasangan suami istri yang tidak ingin memiliki momongan kembali. Artinya pasangan tersebut dibuat agar tidak lagi menghasilkan keturunan.

1. Jenis Kontrasepsi Temporer/Sementara

a. Kondom pria dan wanita

Jika dibandingkan dengan alat kontrasepsi lainnya maka kondom merupakan alat kontrasepsi yang paling dikenal masyarakat, terutama yang dikhususkan bagi pria. Namun kini telah beredar kondom bagi wanita yang hampir mirip dengan kondom pria, namun harus dipasang di mulut vagina 8 jam sebelum melakukan hubungan seksual.

Kekurangan: jenis alat kontrasepsi ini hanya dapat digunakan sekali, kurang efektif dalam mencegah kehamilan, dapat menimbulkan rasa tidak nyaman pada alat kelamin.

b. Pil KB

Ternyata pil kb dibagi ke dalam dua golongan, yaitu jenis yang mengandung hormone progesterone dan kombinasi progesterone-estrogen (seperti Diane 25 dan Yasmin)

Kekurangan: harus rutin dikonsumsi setiap hari dalam beberapa kasus dapat memicu terganggunya pola menstruasi, kenaikan berat badan, hingga darah tinggi, serta tidak melindungi penggunaanya dari penularan infeksi menular seksual (IMS).

c. Suntik KB

Suntik KB merupakan langkah pencegahan kehamilan dengan menyuntikan hormon progestin pada lengan bagian atas setiap 3 bulan sekali.

Kekurangan: dapat menimbulkan efek serupa penggunaan pil KB, seperti mual dan kenaikan berat badan, tidak melindungi penggunaanya dari IMS, serta dapat menurunkan gairah seksual.

d. IUD/Spiral

IUD atau yang masyarakat kenal dengan spiral, merupakan alat kontrasepsi berbentuk huruf T yang dipasang di dalam rahim. IUD ada yang terbuat dari tembaga (seperti Paragard yang bertahan selama 10 tahun) dan bahan lain yang mengandung hormon (seperti Mirena yang bertahan selama 5 tahun).

Kekurangan: Dapat menimbulkan rasa tidak nyaman, seperti kram; ada risiko tubuh menolak pemasangan IUD, serta memicu ketidakteraturan pola serta volume darah yang dikeluarkan saat menstruasi.

e. Implan

Implan alat kontrasepsi berbentuk batang kecil (40mm) yang dipasang di lengan bagian atas dan berfungsi untuk mencegah kehamilan dengan perlahan melepaskan hormon progestin.

Kekurangan: Dapat memicu iritasi serta rasa tidak nyaman di area lengan yang dipasang implan, meningkatkan risiko menstruasi yang tidak teratur dengan jumlah darah yang berlebih di masa awal penggunaannya, serta tidak dapat digunakan oleh mereka yang menderita diabetes, penyakit liver, serta osteoporosis.

2. Jenis Kontrasepsi Permanen

a. MOW (Metode Operatif Wanita)

Metode operasi wanita merupakan salah satu cara kontrasepsi diikuti dengan tindakan pembedahan pada saluran telur wanita. Tubektomi merupakan tindakan medis berupa penutupan tuba uterine dimaksudkan agar tidak mendapatkan keturunan dalam jangka panjang sampai seumur hidup.

Kekurangan: resiko dan efek samping pembedahan, kadang-kadang sedikit merasakan nyeri pada saat operasi, dan kesuburan sulit kembali.

b. MOP (Metode Operatif Pria)

Metode operasi pria adalah metode sterilisasi dengan cara mengikat saluran sperma. MOP merupakan tindakan pada kedua saluran bibit pria yang mengakibatkan orang atau pasangan yang bersangkutan tidak akan mendapat keturunan lagi. MOP adalah alat kontrasepsi jenis sterilisasi melalui pembedahan dengan cara memotong saluran sperma.

Kekurangan: resiko dan efek samping pembedahan, kadang-kadang menyebabkan komplikasi seperti pendarahan atau infeksi, belum memberikan perlindungan total sampai semua spermatozoa yang sudah ada di dalam sistem reproduksi dikeluarkan.

2.2.3. Clustering

Clustering atau klasterisasi adalah metode pengelompokan data. Menurut Tan, 2006 clustering adalah sebuah proses untuk mengelompokkan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum. Clustering juga dikenal sebagai data segmentasi karena clustering mempartisis banyak data set ke dalam banyak grup berdasarkan kesamaannya.

Metode clustering secara umum dapat dibagi menjadi dua yaitu *hierarchical clustering* dan *partitional clustering* (Tan, 2011).

1. *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 ke dalam cluster.

2. *Partitional Clustering*

Partitional clustering yaitu data dikelompokkan ke dalam sejumlah cluster tanpa adanya struktur hirarki antara satu dengan yang lainnya. Pada metode *partitional clustering* tiap cluster memiliki titik pusat cluster (*centroid*) dan secara umum metode ini memiliki fungsi tujuan yaitu meminimumkan jarak dari seluruh data ke pusat cluster masing-masing.

2.2.4. K-Means Clustering

Metode K-means merupakan metode *clustering* yang paling sederhana dan umum. Hal ini dikarenakan K-means mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang cepat dan efisien. K-means merupakan salah satu algoritma klastering dengan metode partisi yang berbasis titik pusat (*centroid*) selain algoritma yang berbasis objek.

Algoritma ini pertama kali diusulkan oleh MacQueen (1967) dan dikembangkan oleh Hartigan dan Wong tahun 1975 dengan tujuan untuk dapat membagi M data point dalam N dimensi ke dalam sejumlah k cluster dimana proses klastering dilakukan dengan meminimalkan jarak sum squares antara data dengan masing-masing pusat cluster.

K-Means adalah Teknik yang cukup sederhana dan cepat dalam proses clustering objek (clustering). Algoritma K-Means mendefinisikan centroid atau pusat cluster dari cluster menjadi rata-rata point dari cluster tersebut. dalam penerapan algoritma K-Means, jika diberikan sekumpulan data $X = \{x_1, x_2, \dots, x_n\}$ dimana $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ adalah sistem dalam ruang real R_n , maka algoritma K-Means akan menyusun partisi X dalam sejumlah k cluster. Setiap cluster memiliki titik tengah yang merupakan nilai rata-rata dari data-data dalam cluster tersebut. Tahapan awal, algoritma K-Means adalah memilih secara acak k buah objek sebagai centroid dalam data. Kemudian, jarak antara objek dan centroid dihitung menggunakan Euclidian Distance. Berikut merupakan langkah-langkah dalam algoritma K-Means :

1. Menentukan berapa banyak cluster k dari dataset yang akan dibentuk.
2. Menentukan secara acak data k menjadi pusat awal lokasi cluster.
3. Untuk masing-masing data, temukan pusat cluster terdekat. Dengan demikian berarti masing-masing pusat cluster memiliki sebuah subset dari dataset, sehingga mewakili bagian dari dataset. Oleh karena itu, telah terbentuk cluster $k : C_1, C_2, C_3, \dots, C_k$.
4. Untuk masing-masing cluster k , temukan pusat luasan kluster dan perbarui lokasi dari masing-masing pusat cluster ke nilai baru dari pusat luasan.
5. Ulangi langkah ke-3 dan ke-5 hingga data-data pada tiap cluster menjadi terpusat atau selesai.

Distance space digunakan untuk menghitung jarak antara data dan centroid. Adapun persamaan yang dapat digunakan salah satunya yaitu *Euclidean Distance Space*. *Euclidean distance space* sering digunakan dalam perhitungan jarak, hal ini dikarenakan hasil yang diperoleh merupakan jarak terpendek antara dua titik yang diperhitungkan. Adapun persamaannya adalah sebagai berikut :

$$d_{ij} = \sqrt{\sum_{k=1}^p \{x_{ik} - x_{jk}\}^2}$$

.....**Rumus 2.1. Distance Space**

Dimana: :

d_{ij} = Jarak objek antara objek i dan j

P = Dimensi data

x_{ik} = Koordinat dari obyek i pada dimensi k

x_{jk} = Koordinat dari obyek j pada dimensi k

2.2.5. Principal Component Analysis (PCA)

Dalam statistika, analisis komponen utama (*Principal Component Analysis/PCA*) adalah teknik yang digunakan untuk menyederhanakan suatu data dengan cara mentransformasi linier sehingga terbentuk sistem koordinat baru dengan varians maksimum. PCA dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan (Smith, 2002).

Menurut Santosa (2007), *Principal Component Analysis* (PCA) adalah suatu teknik handal untuk mengekstrasi struktur dari suatu set data dengan dimensi yang cukup banyak. Problem dalam PCA adalah menemukan eigenvalue dan eigenvectors. PCA adalah transformasi orthogonal (tegak lurus) dari sistem koordinat dimana data dideskripsikan. Koordinat baru dimana data dideskripsikan dinamakan principal component atau PC. Koordinat tersebut dipilih dimana variansi dari data mencapai maksimum.

Prinsip dasar algoritma PCA adalah memproyeksikan citra ke dalam bidang ruang eigen-nya dengan cara mencari eigenvector yang dimiliki setiap citra dan memproyeksikannya ke dalam ruang eigen yang didapat tersebut. Besar ruang eigen tergantung dari jumlah citra training yang dimiliki (Gökberk , 2005).