

# Pengembangan Strategi Pembelajaran Mesin untuk Membangun Visualisasi Cerdas Berdasarkan Preferensi Pengguna

Evans Winanda Wirga, Johan Harlan, I Made Wiryana dan Ernianti Hasibuan

Teknik Informatika, Teknologi Industri, Universitas Gunadarma  
Jl. Margonda raya no 100, Pondok Cina - Depok  
E-mail: evansww@staff.gunadarma.ac.id, harlan\_johan@staff.gunadarma.ac.id,  
mwiryana@staff.gunadarma.ac.id, ernianti@staff.gunadarma.ac.id

## Abstrak

Presentasi dan visualisasi yang didukung pola interaksi yang berdasarkan preferensi pengguna sebagai dasar pengambil keputusan serta berdasarkan pertimbangan data yang tepat akan mendukung *Decision Support System* yang baik. Penelitian ini memaparkan pengembangan strategi pembelajaran mesin dengan model dataset visualisasi dalam membangun visualisasi yang berdasarkan preferensi pengguna. Data diklasifikasi menjadi data pengontrol visualisasi (Dc) dan data visualisasi (Dv), Selanjutnya dibuat model dataset visualisasi data sebagai bahan untuk pembelajaran mesin dalam menentukan hasil visualisasi. Penerapan *machine learning* menggunakan model dataset ini memungkinkan untuk menghasilkan smart visualization berdasarkan preferensi pengguna saat memilih visualisasi yang sesuai dengan kebutuhannya. Data visualisasi yang digunakan adalah data Indeks Keluarga Sehat dari Kementerian Kesehatan Republik Indonesia. .

**Kata kunci** : visualisasi, *machine learning*, preferensi pengguna, data, IKS

## Pendahuluan

Pemanfaatan *machine learning* dalam beragam upaya penyelesaian permasalahan terkait dengan besarnya data yang perlu dipertimbangkan untuk sampai pada satu pengambilan keputusan tertentu. Jumlah informasi yang dihasilkan saat ini sangat banyak dengan keragaman bidang terkait domain kerja pengguna. Wawasan yang memiliki arti dan makna dari sekumpulan data mentah membutuhkan beberapa metode untuk diterapkan sebelum bisa memberikan arti bagi penggunanya. Pengguna suatu sistem informasi sangat dipengaruhi oleh representasi visual dalam memaknai sekumpulan data yang dihadapi, terutama bila kondisi mengharuskan pengguna mengambil satu keputusan berdasar data yang ada.

Penggunaan metode visualisasi data dengan tingkat visualisasi dan data yang signifikan membutuhkan keterampilan analitik. Sayangnya, rata-rata pengguna tidak memiliki keterampilan analitik visual yang optimal untuk kondisi pengamatan data tertentu. Bahkan ahli pada domain tertentu kebanyakan memiliki sedikit atau bahkan tidak terlalu memiliki landasan pengetahuan khusus visualisasi. Kekurangan tersebut berdampak negatif, karenanya perusahaan atau lembaga terpaksa mem-

pekerjakan analis data dengan keterampilan visualisasi dan analisis, tetapi semakin kompleksnya domain pengetahuan akan meningkatkan kompleksitas dan biaya eksplorasi data.

Metode visualisasi data telah banyak dikembangkan dengan berbagai pendekatan seiring dengan perkembangan teknologi yang ada. Metode visualisasi data yang berorientasi pada karakteristik data bertujuan untuk meningkatkan pemahaman terhadap data dari berbagai hubungan yang berbeda pada data tersebut dan dari prosedur yang merepresentasikan data tersebut. Metode visualisasi data yang berorientasi tugas bertujuan merancang teknik yang berbeda dalam menyimpulkan tujuan representasional atau niat pengguna, untuk menemukan visualisasi, pengguna harus membagi masalah menjadi submasalah. Sehingga untuk setiap submasalah dimungkinkan untuk menemukan entri dalam matriksnya. Sebuah representasi masalah kompleks dapat ditemukan dengan menggabungkan metode representasi kandidat untuk submasalah.

Pemanfaatan *machine learning* untuk melacak perilaku pengguna dan merekomendasikan interaksi yang terjadi dalam tugas analisis dengan menggunakan sekumpulan data tertentu untuk memodelkan perilaku pengguna dalam sistem visual-

isasi. Hal ini dapat dilakukan dengan melakukan evaluasi online dan melacak perilaku pengguna saat melakukan tugas analisis data secara visual dan melakukan interaksi selanjutnya terhadap data hingga memutuskan satu simpulan terhadap data tersebut.

Informasi yang dikumpulkan digunakan untuk memodelkan pola pengguna dalam setiap tugas analisisnya. Model yang dihasilkan akan dapat digunakan untuk merekomendasikan model visualisasi yang paling optimal untuk pengguna dengan tugas dan kompleksitas data tertentu. Model menghasilkan rekomendasi berdasarkan informasi yang diperoleh melalui pemantauan perilaku pengguna pada periode waktu tertentu.

Penelitian ini, dilakukan peninjauan kumpulan publikasi akademik dan artikel online yang mengevaluasi visualisasi berdasarkan orientasi data dan orientasi tugas dari pengguna. Selanjutnya dari metode yang telah ditinjau berkontribusi dalam pengembangan metode visualisasi lain yang berdasarkan preferensi pengguna. Metode ini yang menggabungkan pendekatan data, tugas, dan referensi visual pengguna.

Dalam penelitian ini menggunakan data Program Indonesia Sehat dengan Pendekatan Keluarga (PISPK) dengan Kementerian Kesehatan Republik Indonesia. Data yang digunakan adalah Indeks Keluarga Sehat (IKS), yaitu data yang menyatakan bahwa suatu keluarga sehat, dengan keluarga adalah suatu kesatuan inti (ayah, ibu, dan anak) sebagaimana dinyatakan dalam kartu keluarga. Suatu keluarga sehat atau tidak dinyatakan dengan sejumlah penanda atau indikator. Program Indonesia Sehat telah disepakati adanya 12 indikator utama untuk penanda status kesehatan sebuah keluarga.

## Metode

Sebelum melakukan penelitian ini, telah dilakukan review terhadap beragam hasil riset sebelumnya yang terkait kurangnya peran ruang pengetahuan dan preferensi user pada sistem rekomendasi visualisasi data. Berdasarkan hal tersebut, salah satu yang akan menjadi fokus penelitian ini adalah pada sistem berorientasi karakteristik data dan task.

## Visualisasi Berorientasi Karakteristik Data

**BHARAT** dikenal sebagai salah satu sistem yang mengawali digunakannya “rules” (aturan) untuk menentukan tipe visualisasi data yang sesuai dengan attribute data tertentu. Penelitian yang dilakukan pada tahun 1981 ini, menghasilkan visualisasi data yang mungkin dilakukan tidak sebanyak apa yang tersedia pada saat ini. Pada sistem ini lebih mengandalkan Diagram Garis, Diagram Pie

dan Diagram Batang, yang berdasar pada desain algoritma sederhana. [1]

**APT** merupakan sistem yang dikembangkan pada tahun 1986 oleh Mackinlay yang mencoba memformalisir dan mengkodifikasi spesifikasi desain dari tampilan diagram ini secara otomatis. Sistem ini didasari dari penelitian lain yang memanfaatkan semiologi pada grafik dimana ditentukan beberapa visual variabel yang spesifik seperti posisi, ukuran, nilai, warna, orientasi, dan lain sebagainya yang kemudian diklasifikasikan berdasar fitur yang dapat diwakili dengan cara terbaik. [2]

**VizML** merupakan sebuah platform dengan pendekatan berbasis ML untuk rekomendasi visualisasi menggunakan kumpulan besar data dan visualisasi terkait. Pada tahap awal dijelaskan bahwa visualisasi sebagai proses membuat pilihan desain yang memaksimalkan efektivitas, yang bergantung pada dataset, tugas, dan konteks. Kemudian dirumuskan rekomendasi visualisasi sebagai masalah pengembangan model yang belajar membuat pilihan desain. Proses ini melatih dan menguji model pembelajaran mesin menggunakan satu juta pasangan visualisasi set data yang unik dari *feedback* Komunitas Plotly. Proses pengumpulan dan pembersihan korpus ini, mengekstrak fitur dari setiap kumpulan data, dan mengekstrak lima pilihan desain utama dari visualisasi yang sesuai. [3]

**Tableau & Fitur Show Me.** Tableau adalah tonggak sejarah di dunia visualisasi data. Pada tahun 2007 Tableau memperkenalkan fitur yang disebut Show Me. Fungsi Show Me membutuhkan beberapa kelebihan dari VizQL untuk secara otomatis menyajikan data. Di bagian utama dari fitur ini adalah bahwa sistem rekomendasi berorientasi karakteristik data. Pengguna memilih atribut data yang menarik baginya dan Tableau merekomendasikan visualisasi yang sesuai. Tableau menentukan jenis visualisasi yang tepat untuk digunakan dengan melihat jenis atribut dalam data. [4]

**ManyEyes.** Viegas dkk, membuat situs web publik pertama yang mana pengguna dapat mengunggah data dan membuat visualisasi interaktif secara kolaboratif: ManyEyes. Alat ini dibuat untuk non-ahli, yang dapat diakses oleh siapa saja tanpa memandang pengetahuan dan pelatihan sebelumnya. Pilihan desain dibuat untuk mencerminkan upaya menemukan keseimbangan antara kekuatan kemampuan analisis data dan aksesibilitas ke pengguna visualisasi yang non-ahli. Visualisasi dibuat dengan mencocokkan kumpulan data dengan salah satu dari 13 jenis visualisasi data yang diimplementasikan dalam alat ini. [5]

**VizDeck.** Pada tahun 2012 Key dkk. mengembangkan alat yang disebut VizDeck. Alat berbasis web ini merekomendasikan visualisasi data berdasarkan properti statistik data. Cara ini mengadopsi metafora permainan kartu untuk mengatur beberapa visualisasi ke dalam aplikasi dasbor visual interaktif. Vizdeck diciptakan Key et al. yang men-

emukan bahwa ilmuwan tidak dapat melatih diri dengan cepat dalam menggunakan alat yang lebih canggih seperti Tableau. Alat ini mendukung Diagram Spread, histogram, Diagram Batang, Diagram lingkaran atau pie, Diagram deret waktu, Diagram Garis, dan peta. [7]

**SEEDB.** Pada tahun 2015 Vartak dkk. mengusulkan mesin yang disebut SEEDB. Pada penelitian ini, visualisasi dilihat berdasarkan teori bahwa visualisasi data cenderung akan menarik penggunaannya jika dapat menampilkan penyimpangan besar dari beberapa referensi (misalnya kumpulan data lain, data historis, atau data lainnya). Ini akan membantu untuk mengidentifikasi visualisasi yang paling menarik dari serangkaian besar visualisasi yang potensial. Penelitian ini mengidentifikasi bahwa ada lebih banyak aspek yang menentukan daya tarik sebuah visualisasi, seperti estetika, preferensi pengguna, metadata, dan tugas pengguna. [8]

**Voyager.** Pada tahun 2016, Wongsuphasawat dkk. mengembangkan aplikasi web rekomendasi visualisasi yang disebut Voyager, berdasarkan mesin rekomendasi Compas dan menggunakan bahasa spesifikasi tingkat tinggi yang disebut Vegalite. Konsep ini memasangkan penelusuran dengan rekomendasi visualisasi untuk mendukung eksplorasi multivariat, dalam bentuk data tabular. [9]

## Visualisasi Berorientasi *Task*

Berikut adalah beberapa alat bantu dan teknik visualisasi data dengan pendekatan tugas (*task*).

**BOZ.** merupakan alat desain grafis dan presentasi otomatis yang mendesain grafik berdasarkan analisis tugas dan kemudian akan didukung oleh presentasi grafik. Sistem akan menganalisis deskripsi logis dari tugas yang akan dilakukan dan merancang persepsi dari tugas yang setara. BOZ menghasilkan grafik dengan prosedur persepsi yang menjelaskan bagaimana menggunakan grafik untuk menyelesaikan tugas. Alat ini mampu merancang presentasi yang berbeda dari informasi yang sama, disesuaikan untuk persyaratan tugas yang berbeda. [10]

**IMPROVISE.** Pada penelitian sebelumnya, daftar task dari user umumnya akan dibuat secara manual dalam bentuk daftar task atau daftar tugas. Pada tahun 1998, Zhou dan Feiner memperkenalkan teknik linguistik tingkat lanjut untuk mengotomatisasi derivasi tugas pengguna dari permintaan bahasa alami. Mereka memperkenalkan taksonomi tugas visual untuk mengotomatisasi proses untuk mendapatkan maksud presentasi dari teks. Antar-muka taksonomi antara level tinggi tugas yang dapat diselesaikan dengan teknik visualisasi tingkat rendah. Misalnya, tugas visual "Fokus" menyiratkan bahwa teknik visual seperti "Memperbesar" atau "Sorotan" dapat digunakan. Taksonomi ini diimplementasikan dalam IMPROVISE. Contoh kasus penggunaan IMPROVISE adalah alat

ini akan menyajikan gambaran umum informasi pasien rumah sakit kepada perawat. Untuk mencapai tujuan ini, ia membangun sebuah struktur diagram yang mengatur berbagai informasi (misalnya jalur IV) di sekitar komponen inti (tubuh pasien). Dengan cara desain top-down, IMPROVISE membuat diagram struktur 'kosong' dan kemudian mengisinya dengan komponen. Kemudian mempartisi dan mengkodekan informasi pasien ke dalam kelompok yang berbeda. [11]

**HARVEST.** Pada tahun 2009, Gotz dan Wen memperkenalkan pendekatan baru yang didorong oleh perilaku pengguna. Alih-alih membutuhkan deskripsi tugas eksplisit, mereka menggunakan tugas implisit informasi yang diperoleh dengan memantau perilaku pengguna untuk membuat rekomendasi lebih efektif. Rekomendasi Visualisasi Berbasis Perilaku (BVDR), dimana pendekatan ini memiliki dua fase. Pada fase pertama BDVR, akan mendeteksi empat pola yang telah ditentukan dari aktivitas pengguna. Pada fase kedua, alat akan memberi inputan dalam bentuk pola yang terdeteksi menjadi algoritma rekomendasi, yang menyimpulkan maksud pengguna dalam hal tugas visual umum (misalnya perbandingan) dan menyarankan visualisasi yang lebih baik mendukung kebutuhan pengguna. [11]

**DataSlicer.** Sebuah studi oleh Alborzi et al. mengambil pendekatan lain. Hipotesisnya adalah bahwa untuk banyak kumpulan data dan tugas analisis umum, ada relatif sedikit "irisan data" yang menghasilkan visualisasi yang efektif. Data irisan adalah subset data yang berbeda. Tujuannya adalah untuk meningkatkan pengalaman pengguna dengan menyarankan irisan data, yang ketika divisualisasikan, akan menyajikan solusi yang benar untuk tugas pengguna dengan cara yang efektif. Pada waktu tertentu dalam mengerjakan tugas, pengguna dapat meminta sistem untuk menyarankan visualisasi yang akan berguna untuk menyelesaikan tugas. [12]

## Visualisasi Berdasarkan Preferensi Pengguna

Visualisasi data adalah alat untuk mendukung kemudahan melakukan analisa dan pemikiran terhadap data yang tampak dalam bentuk visual. Hal ini dapat digunakan untuk mengeksternalisasi pengetahuan tentang tugas atau domain analitis yang kompleks.

Preferensi visualisasi data tidak dapat sepenuhnya dipahami tanpa memahami cara berpikir penggunaannya. Pemahaman ini bukanlah hal yang mudah, mengingat begitu beragamnya jenis pengguna visualisasi pada data. Pada akhirnya, memahami visualisasi data membutuhkan pemahaman tentang bagaimana pengguna merespon terhadap ragam visualisasi yang diterima dan apa alasan pengguna memiliki preferensi terhadap visualisasi data tersebut.

Setiap pengguna tentunya berpikir dengan cara yang berbeda. Ada beberapa aspek yang membedakan satu pengguna dengan yang lainnya. Pengalaman, kepribadian, dan kemampuan kognitif pengguna akan mempengaruhi pendekatan pengguna untuk menyelesaikan tugas dan pemahamannya tentang domain dari masalah yang dihadapi. Menurut psikologi kognitif, menunjukkan bahwa perbedaan individu dapat berdampak signifikan pada kemudahan pengguna menggunakan antarmuka atau alat untuk memecahkan masalah melalui visualisasi data.

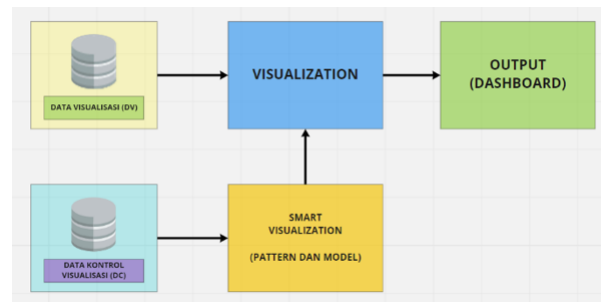
Pertimbangan pemanfaatan visualisasi data sangat berbeda berdasar pengalaman, latar belakang, kepribadian, dan kemampuan kognitif dari penggunaannya, namun visualisasi data, seperti banyak perangkat lunak lainnya, terus dirancang untuk satu pengguna ideal. Namun disadari bahwa mendesain setiap visualisasi untuk setiap jenis pengguna individu adalah hal yang sangat sulit. Pengetahuan tentang perbedaan yang ada antara kelompok pengguna dapat digunakan untuk memandu desain untuk domain tertentu dan untuk menyarankan beberapa mode analisis atau opsi penyesuaian dalam satu sistem yang menyajikan visualisasi data ini. Kini terdapat pula bidang penelitian baru yang menggunakan pendekatan yang berlawanan dengan metode tradisional dari desain yang menyatakan "satu ukuran cocok untuk semua". Penelitian ini menunjukkan bahwa gaya kognitif individu pengguna serta desain visual dapat menentukan nilai visualisasi. Selain itu, perbedaan individu ini tampak lebih jelas dalam tugas yang lebih kompleks. Temuan ini menunjukkan bahwa preferensi pengguna terhadap visualisasi data telah memiliki latar belakang yang menjadi awal penelitian selanjutnya dalam konteks perbedaan di antara penggunaannya. Perlu dipahami bagaimana kemampuan kognitif dan faktor kepribadian mempengaruhi penggunaan visualisasi, dan apa yang masih perlu dilakukan lebih jauh dalam menentukan visualisasi data yang ideal untuk pengguna.

## Tahapan Penelitian

Visualisasi data merupakan salah satu pendekatan untuk mendapatkan informasi dan wawasan terhadap sekumpulan data. Dunia menghasilkan lebih dari 2,5 exabyte data setiap hari, sehingga saat ini data menjadi sesuatu yang sangat bernilai, salah satu faktor menjadikan data bernilai yaitu dengan visualisasi data yang baik, yang memungkinkan pengguna dapat melihat tren dan pola yang tidak mudah diidentifikasi dalam baris dan kolom angka suatu tabel.

Data yang diperlukan untuk membentuk suatu visualisasi data tidak hanya dataset yang akan divisualisasikan. Untuk menghasilkan suatu jenis diagram yang sesuai dengan jenis data, diperlukan data properti dari diagram tersebut. Pada penelitian ini digunakan dua jenis data untuk meng-

hasilkan suatu model visualisasi data, seperti pada Gambar 1



Gambar 1: Model Data Visualisasi

Pada Gambar 1 dijelaskan dalam penelitian ini menggunakan alur penggunaan data pada suatu visualisasi data. Data diklasifikasikan menjadi dua yaitu:

- **Data Pengontrol Visualisasi (Dc)**, merupakan data yang digunakan sebagai pengontrol visualisasi yang akan dihasilkan. Data pengontrol dibentuk berdasarkan preferensi pengguna dengan metode asesmen pengguna.
- **Data Visualisasi (Dv)**, yaitu data yang digunakan untuk menghasilkan suatu bentuk visualisasi, data visualisasi menggunakan data Indeks Keluarga Sehat (IKS) Wilayah dari Kementerian Kesehatan Republik Indonesia.

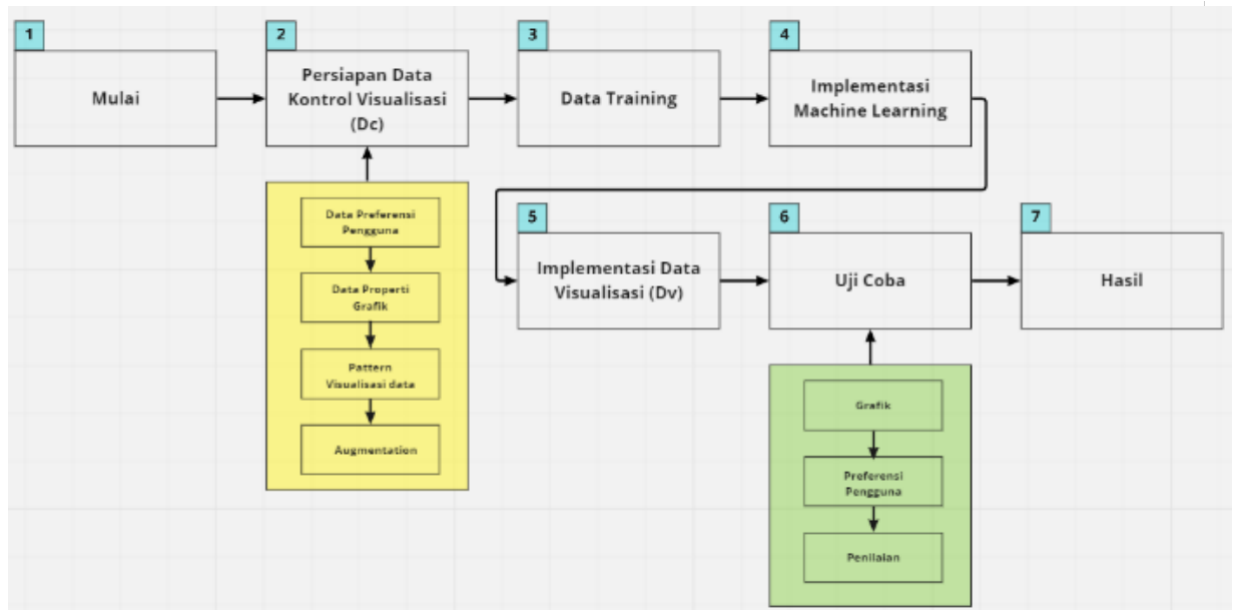
Pada Gambar 2 merupakan gambaran umum tahapan penelitian yaitu terdiri dari:

1. Persiapan Data Kontrol Visualisasi Data : proses persiapan data dilakukan yaitu menentukan dan menginisialisasi jenis data untuk visualisasi yaitu data visualisasi dan data pengontrol visualisasi.
2. Data training : pada tahap ini dilakukan proses persiapan data dan metode training untuk digunakan pada pada algoritma *machine learning*. Pada tahapan ini dilakukan pembuatan pattern visualisasi yang berdasarkan data pengontrol visualisasi. Setelah itu dilakukannya proses *Augmentation* untuk memperbanyak data training.
3. Menentukan Algoritma *Machine Learning* untuk Rekomendasi : pada tahap ini dilakukan inisialisasi dan analisis dataset untuk menentukan penggunaan algoritma *Machine Learning* untuk sistem rekomendasi
4. Implementasi *Machine Learning* : pada tahap ini dilakukan implementasi Algoritma *Machine Learning* untuk rekomendasi yang sebelumnya telah di tentukan pada langkah 3
5. Implementasi Data Visualisasi : pada tahap ini dilakukan implementasi visualisasi dengan menggunakan data indeks keluarga sehat (IKS).

6. Testing (Uji Coba) pada tahap ini dilakukan uji coba terhadap hasil visualisasi dari *Machine Learning*. Proses ujicoba ini akan melibatkan pengguna yaitu dengan mengeluarkan

grafik, preferensi pengguna serta pendapat pengguna terhadap hasil visualisasi.

7. Hasil : tahap ini merupakan hasil visualisasi yang dihasilkan oleh sistem.



Gambar 2: Tahapan Penelitian

## Pembahasan

### Data Pengontrol Visualisasi

Data Pengontrol Visualisasi Data (Dc) diantaranya adalah data preferensi dari pengguna untuk menghasilkan visualisasi data yang berdasarkan preferensi dan tujuan (goals) dari pengguna. Data preferensi pengguna diklasifikasikan kedalam beberapa kategori yaitu :

**Data Pengguna (S)**, Data preferensi pengguna sangat berkaitan dengan data personal dari pengguna, data personal dapat memperkaya demografi pengguna yang bisa digunakan untuk menentukan visualisasi yang akan ditampilkan. Data personal pengguna diantaranya :

- Nama,
- Jenis Kelamin, data informasi jenis kelamin dari pengguna.
- Umur, merupakan data informasi umur dari pengguna.
- Pendidikan, merupakan data pendidikan terakhir dari pengguna.
- Daerah Asal, merupakan data daerah asal pengguna.
- Pekerjaan, merupakan data pekerjaan pengguna.

- Bidang Pekerjaan, merupakan bidang pekerjaan dari pengguna.
- Jabatan, merupakan data jabatan pengguna
- Provinsi, daerah tempat pengguna bekerja.
- Warna Kesukaan untuk Diagram, merupakan data referensi warna yang lebih disukai pengguna berkerja dengan diagram.
- Durasi Menggunakan Komputer, merupakan data durasi atau lama pengguna menggunakan komputer dalam sehari.
- Durasi Menggunakan HP, merupakan data durasi atau lama pengguna menggunakan handphone/smartphone dalam sehari.
- Durasi Menggunakan Internet, merupakan data durasi atau lama pengguna menggunakan internet dalam sehari.

**Tipe Tugas (V)** Tipe tugas dari suatu visualisasi data diklasifikasikan menjadi empat yaitu :

- **Perbandingan**, yaitu suatu bentuk visualisasi data dengan memisahkan satu set variabel dengan yang lain dan menampilkan bagaimana kedua variabel tersebut berinteraksi. sebagai contoh, jumlah pengunjung 5 situs web dalam satu bulan.
- **Hubungan**, yaitu suatu bentuk visualisasi data yang menunjukkan hubungan atau korelasi antara dua atau lebih variabel melalui

data yang disajikan. seperti kapitalisasi pasar saham tertentu dari waktu ke waktu versus tren pasar secara keseluruhan.

- **Komposisi**, yaitu suatu bentuk visualisasi data yang mencoba mengumpulkan berbagai jenis informasi yang membentuk keseluruhan dan menampilkannya bersama-sama.
- **Distribusi**, yaitu suatu bentuk visualisasi data yang mencoba menyusun kumpulan informasi terkait atau tidak terkait secara sederhana untuk melihat bagaimana korelasinya, jika ada, dan untuk memahami apakah ada interaksi antara variabel, seperti jumlah bug yang dilaporkan selama setiap bulan beta.
- **Deskriptif**, yaitu suatu bentuk analisis data yang menjawab pertanyaan tentang apa yang terjadi. Proses analisis dengan tujuan deskriptif adalah dengan menggabungkan data mentah dari berbagai sumber data untuk memberikan informasi, tetapi informasi hanya sebagai signal bahwa ada sesuatu tentang benar atau salah tanpa menjelaskan alasannya.
- **Diagnosis** yaitu suatu bentuk analisis data dengan menggunakan data historis yang diukur terhadap data lain untuk menjawab pertanyaan mengapa sesuatu dapat terjadi. Suatu analisis dengan dapat memberikan wawasan mendalam tentang masalah tertentu.
- **Prediktif** yaitu suatu bentuk analisis data yang memungkinkan pengguna untuk mengetahui apa yang mungkin terjadi. Metode ini menggunakan data pada analisis deskriptif dan diagnosis untuk mendeteksi cluster dan pengecualian dan untuk memprediksi tren masa depan, dan menjadikan suatu cara untuk memprediksi.
- **Preskriptif**, yaitu suatu bentuk analisis yang bertujuan untuk memberikan suatu saran atau masukan tindakan yang harus diambil untuk mencegah masalah yang mungkin terjadi di masa depan atau mengambil suatu keuntungan dari suatu tren.

## Data Visualisasi

Data visualisais merupakan data pembentuk keluarga visualisasi. Data yang digunakan untuk implementasi visualisasi data adalah data indeks keluarga sehat (IKS) dari Kementerian Kesehatan Republik Indonesia. Data IKS sebelumnya telah dilakukan perhitungan sehingga menghasilkan nilai dan kesimpulan status keluarga sehat keluarga di Indonesia. Data Keluarga Sehat [permenkes-2016] menyatakan bahwa suatu keluarga sehat atau tidak digunakan sejumlah penanda atau indikator. Dimana, keluarga adalah satu kesatuan keluarga

inti (ayah, ibu, dan anak) sebagaimana dinyatakan dalam Kartu Keluarga. Jika dalam satu rumah tangga terdapat kakek dan atau nenek atau individu lain, maka rumah tangga tersebut dianggap terdiri lebih dari satu keluarga.

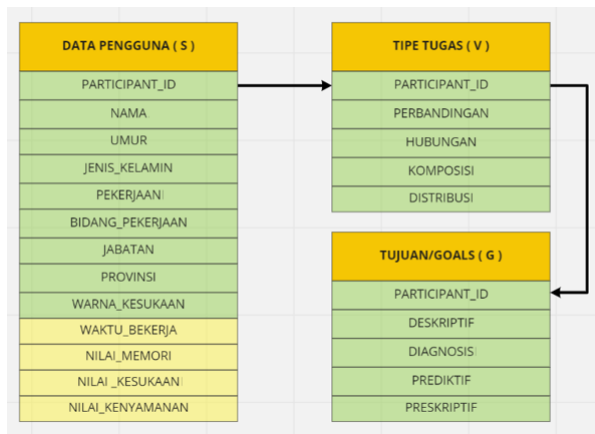
Suatu keluarga sehat atau tidak dinyatakan dengan sejumlah penanda atau indikator. Program Indonesia Sehat telah disepakati adanya 12 indikator utama untuk penanda status kesehatan sebuah keluarga. Kedua belas indikator utama tersebut adalah sebagai berikut. [15]

- Indikator 1, Keluarga mengikuti program Keluarga Berencana (KB)
- Indikator 2, Ibu melakukan persalinan di fasilitas kesehatan
- Indikator 3, Bayi mendapat imunisasi dasar lengkap
- Indikator 4, Bayi mendapat air susu ibu (ASI) eksklusif
- Indikator 5, Balita mendapatkan pemantauan pertumbuhan
- Indikator 6, Penderita tuberkulosis paru mendapatkan pengobatan sesuai standar
- Indikator 7, Penderita hipertensi melakukan pengobatan secara teratur
- Indikator 8, Penderita gangguan jiwa mendapatkan pengobatan dan tidak ditelantarkan
- Indikator 9, Anggota keluarga tidak ada yang merokok
- Indikator 10, Keluarga sudah menjadi anggota Jaminan Kesehatan Nasional (JKN)
- Indikator 11, Keluarga mempunyai akses sarana air bersih
- Indikator 12, Keluarga mempunyai akses atau menggunakan jamban sehat

Berdasarkan indikator tersebut, dilakukan penghitungan Indeks Keluarga Sehat (IKS) dari setiap keluarga. Sedangkan keadaan masing-masing indikator, mencerminkan kondisi PHBS dari keluarga yang bersangkutan.

Penggunaan data IKS sebagai Data Visualisasi yaitu :

- Data IKS per wilayah kecamatan di Indonesia.
- Data diambil per – Januari 2020
- Jumlah Data Individu = > 150 juta penduduk Indonesia
- Jumlah Provinsi = 34
- Jumlah Kecamatan = 6,390
- Jumlah Kelurahan = 55,344



Gambar 3: Model Dataset Visualisasi Data

## Visualisasi Dengan Pendekatan *Machine Learning*

Pembangunan metode visualisasi data berdasarkan preferensi pengguna dilakukan dengan pendekatan solusi teknologi dari *machine learning* (pembelajaran mesin). Upaya penerapan *machine learning* dalam visualisasi data dilakukan dengan merancang metode dan model dataset visualisasi data dari data pengontrol visualisasi (Dc) seperti pada

Gambar 3. Model dataset visualisasi digunakan sebagai data untuk pembelajaran mesin agar dapat menghasilkan visualisasi data berdasarkan preferensi pengguna. Model dataset visualisasi selanjutnya dilakukan augmentasi atau memperbanyak data pattern.

Proses Augmentasi pada Model dataset visualisasi data menggunakan model augmentasi yang dirancang menggunakan kategori yang telah diklasifikasikan pada Gambar 2 yang datanya diperoleh dari asesmen preferensi pengguna. Model augmentasi data dilakukan dengan input dari preferensi pengguna yaitu data pengguna (S), Tipe Tugas (V) dan Tujuan/Goal pengguna (G) dengan output yaitu visualisasi (V).

$$\langle S_n, V_n, G_n \rangle \rightarrow P \quad (1)$$

Dengan :

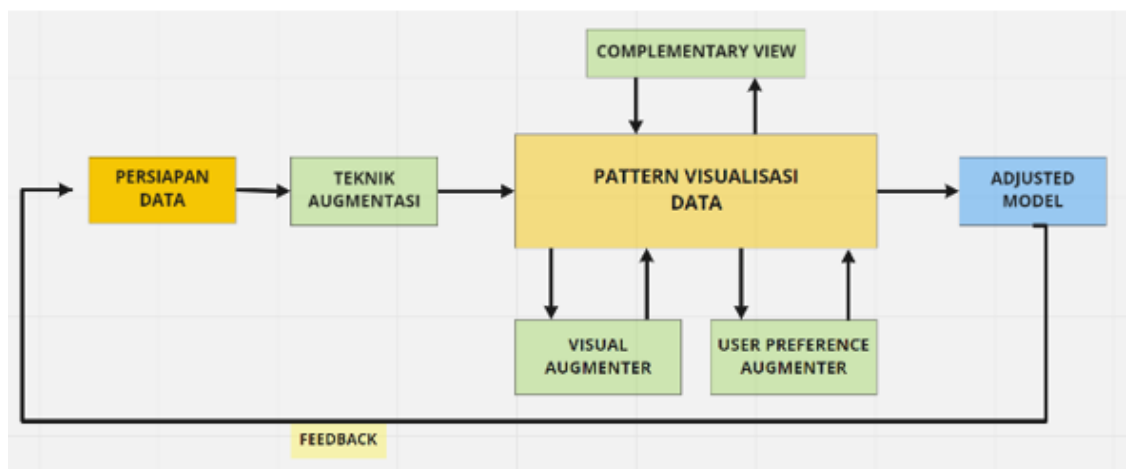
$S_n$  =Data pengguna  $S_n = [n_1 | n_2 | n_3 | n_{..}]$

$V_n$  =Tipe Tugas  $V_n = [n_1 | n_2 | n_3 | n_{..}]$

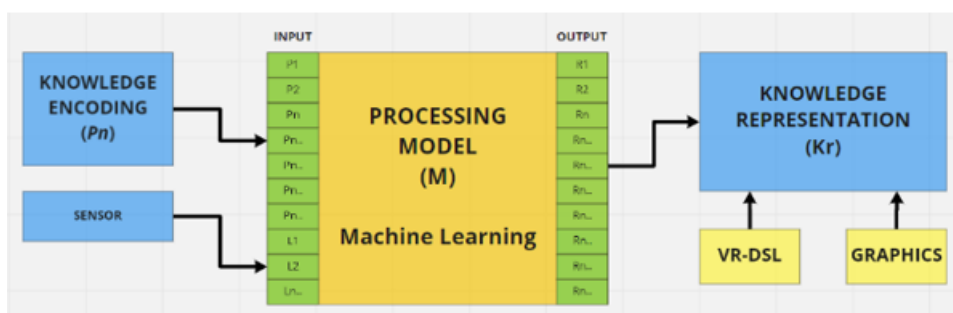
$G_n$  =Tujuan/Goal  $G_n = [n_1 | n_2 | n_3 | n_{..}]$

$P_n$  =Output Pattern

Pada model augmentasi (1), nilai n merupakan jenis parameter dari preferensi pengguna dimana setiap parameter dapat menentukan jenis keluaran visualisasi data.



Gambar 4: Augmentasi Pattern Visualisasi Data



Gambar 5: Alur Proses Implementasi *Machine Learning*

## Fomula atau Rumus

Proses implementasi *machine learning* dilakukan setelah semua data telah siap digunakan. Pada Gambar 3 menjelaskan bahwa penggunaan data training yang dibentuk dari data preferensi pengguna dan data properti diagram yang selanjutnya disebut sebagai knowledge encoding. Knowledge encoding menjadi input yang disesuaikan dengan model visualisasi yang dibentuk dengan metode *machine learning*.

Selanjutnya machine learning menghasilkan jenis output dalam bentuk knowledge representation yang selanjutnya digunakan sebagai pembentuk visualisasi data baik menggunakan Vr-DSL maupun grafik, lihat Gambar 4 dan 5.

## Penutup

Pembangunan visualisasi data yang berdasarkan dengan preferensi pengguna dilakukan dengan mengklasifikasi data visualisasi yaitu data pengontrol visualisasi (Dc) dan data visualisasi (Dv). Selanjutnya dibuat suatu model dataset visualiasi data sebagai bahan untuk pembelajaran mesin dalam menentukan hasil visualiasi. Penerapan *machine learning* dengan menggunakan dataset visualisasi data berdasarkan preferensi pengguna memungkinkan untuk menghasilkan smart visualization berdasarkan preferensi pengguna saat memilih visualisasi yang sesuai dengan kebutuhan pengguna.

## Daftar Pustaka

- [1] S. Gnanamgari, "Information presentation through default displays", Ph.D. Dissertation, University of Pennsylvania, 1981.
- [2] J. Mackinlay, "Automating the design of graphical presentations of relational information", ACM Trans. Graph. (TOG) 5(2), 110–141, 1986.
- [3] P. Hanrahan, "VizQL: a language for query, analysis and visualization", In: Proceedings of the 2006 ACM SIGMOD international conference on Management of data. ACM, 2006.
- [4] J. Mackinlay, P. Hanrahan and C. Stolte, "Show me: automatic presentation for visual analysis", IEEE Trans. Vis. Comput. Graph. 13(6), 1137–1144, 2007.
- [5] F. Viegas, M. Wattenberg, F. van Ham, J. Kriss and M. McKeon, "ManyEyes: a site for visualization at internet scale", IEEE Trans. Vis. Comput. Graph. 13(6), 1137–1144, 2007.
- [6] Watson, "Smart data analysis and visualization", diakses daring pada <https://www.ibm.com/watson-analytics>, Aug 2017.
- [7] A. Key, D. Perry, B. Howe and C. Aragon, "VizDeck: self-organizing dashboards for visual analytics", In: Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, 2012.
- [8] M. Vartak, S. Madden, A. Parameswaran and N. Polyzotis, "SeeDB: supporting visual analytics with data-driven recommendations", VLDB 8, 2182–2193, 2015.
- [9] K. Wongsuphasawat, D. Moritz, J. Mackinlay, B. Howe and J. Heer, "Voyager: exploratory analysis via faceted browsing of visualization recommendations", IEEE Trans. Vis. Comput. Graph. 22(1), 649–658, 2016.
- [10] S. Casner and J.H. Larkin, "Cognitive efficiency considerations for good graphic design", Carnegie-Mellon University Artificial Intelligence and Psychology Project, Pittsburgh, 1989.
- [11] M.X. Zhou and S.K. Feiner, "Visual task characterization for automated visual discourse synthesis", In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM Press/Addison-Wesley Publishing Co., 1998.
- [12] D. Gotz and Z. Wen, "Behavior-driven visualization recommendation", In: Proceedings of the 14th International Conference on Intelligent User Interfaces. ACM, 2009.
- [13] F. Alborzi, Rada Chirkova, Pallavi Deo, Christopher Healey, Gargi Pingale, Vaira Selvakani, J. Reutter and S. Chaudhuri, "DataSlicer: task-based data selection for visual data exploration", arXiv preprint, 2017.
- [14] M. Tory and T. Moller, "Human factors in visualization research", IEEE Transactions on Visualization and Computer Graphics, 10(1):72–84, jan 2004.
- [15] Anonim, "Peraturan menteri kesehatan republik indonesia nomor 39 tahun 2016 tentang pedoman penyelenggaraan program indonesia sehat dengan pendekatan keluarga sehat, Kementerian Kesehatan Indonesia, 2016.