

# Perancangan Aplikasi Untuk Memprediksi Seseorang Menderita Penyakit Hipertensi Menggunakan *Data Mining*

Yusran Timur Samuel dan Frengky Simbolon  
Fakultas Teknologi Informasi, Universitas Advent Indonesia

## Abstrak

Hipertensi merupakan salah satu penyakit gangguan pada pembuluh darah. Di Indonesia, penderita penyakit hipertensi mengalami peningkatan sebesar 3,6% setiap tahunnya. Terjadinya peningkatan hipertensi dikarenakan oleh kurang pedulinya masyarakat akan kesehatan, serta karena kurangnya waktu untuk melakukan konsultasi kepada dokter ahli atau pakar. Tujuan dalam penelitian ini (1) Data yang diperoleh dapat digunakan untuk memprediksi seseorang menderita hipertensi. (2) Aplikasi yang dibuat mudah untuk digunakan dan memiliki akurasi yang tinggi untuk memprediksi. Metode yang digunakan dalam penelitian ini adalah menggunakan *data mining* dengan metode klasifikasi *Naïve Bayes Classification* yang merupakan pengklasifikasian dengan metode probabilitas dan statistik. Dalam pengumpulan data yang dibutuhkan untuk *Data Mining* dengan cara memberikan kuesioner kepada pasien yang berkunjung ke rumah sakit X selama 3 minggu dengan atribut yaitu usia, jenis kelamin, berat badan, tinggi badan, perokok, jenis rokok, jumlah rokok, konsumsi minuman beralkohol, aktifitas fisik, jumlah jam tidur, konsumsi daging, konsumsi sayuran, kadar garam, riwayat hipertensi ayah, riwayat hipertensi ibu. Kesimpulan (1) Dengan menggunakan Data Mining pasien dapat langsung mengetahui apakah pasien tersebut menderita hipertensi berdasarkan pola hidup yang dimilikinya. (2) Aplikasi yang telah dibangun dapat digunakan karena memiliki akurasi yang cukup tinggi sebesar 88% serta memiliki sensitivitas sebesar 77%, spesifitas 96%. (3) Penggunaan naïve bayes untuk memprediksi seseorang menderita penyakit hipertensi dapat digunakan karena memiliki akurasi yang tinggi sebesar 88%.

**Kata – kata kunci:** Hipertensi, *Data Mining*, *Naïve Bayes*

## *Design of Application to Predict Someone Suffering Hypertension Using Data Mining*

### Abstract

*Hypertension is one of the diseases of blood vessel disorders. In Indonesia, people with hypertension disease increased by 3.6% every year. The occurrence of hypertension increases due to the lack of care of the public for health, as well as the lack of time to consult a specialist or an expert. The purpose of this study (1) the data obtained can be used to predict a person suffering from hypertension. (2) Applications are made easy to use and have high accuracy prediction. The method used in this research is data mining with Naïve Bayes Classification method which is a classification method using probability and statistic. In collecting data required for Data Mining, researcher giving questionnaires to patients who visited hospital X for 3 weeks with questionnaire attributes that contains age, sex, weight, height, smokers, type of cigarette, the number of cigarettes, the consumption of alcoholic beverages, physical, sleep hours, meat consumption, vegetable consumption, salt level, history of father's hypertension, history of maternal hypertension. Conclusion (1) by using Data Mining, the patient can immediately find out whether the patient is suffering from hypertension based on the lifestyle he/she has. (2) Applications that have been developed can be used and has a fairly high accuracy of 88% and has a sensitivity of 77%, and 96% of specificity. (3) The use of Naïve Bayes to predict a person suffering from hypertension can be used because it has a high accuracy of 88%.*

### Pendahuluan

Hipertensi atau penyakit tekanan darah tinggi adalah gangguan pada pembuluh darah yang diakibatkan oleh suplai oksigen dan nutrisi, yang dibawa oleh darah, terhambat sampai ke jaringan tubuh yang membutuhkannya. Tubuh akan bereaksi lapar, yang mengakibatkan jantung harus bekerja lebih keras untuk memenuhi kebutuhan tersebut. Bila kondisi tersebut berlangsung lama dan menetap, timbullah gejala yang disebut sebagai penyakit tekanan darah tinggi. Hipertensi memiliki beberapa faktor yang dapat meningkatkan risiko terjadinya hipertensi yaitu usia, jenis kelamin, obesitas, kurang

aktivitas fisik, kebiasaan merokok dan konsumsi minimal beralkohol, dan faktor lain seperti riwayat keluarga penderita hipertensi (Prasetyaningrum, 2014).

Indonesia mengalami peningkatan setiap tahunnya untuk masalah hipertensi, menurut riset kesehatan dasar yang dilakukan oleh Dr. Lily S. Sulistyowati, MM, mengatakan peningkatan kasus hipertensi pada tahun 2013 25,8 persen penduduk Indonesia mengidap penyakit hipertensi. pada tahun 2016 Survei indikator Kesehatan Nasional, Indonesia mengalami kenaikan 32,4 persen (Anwar, 2017). Hal di atas disebabkan oleh karena kurang pedulinya masyarakat akan kesehatan dan kurangnya konsultasi kepada pakar atau dokter ahli. Namun dengan kemudahan adanya para pakar atau dokter ahli, terkadang terdapat pula kelemahan seperti jam kerja (praktek) terbatas, jarak antara dokter dan penderita, serta banyaknya pasien sehingga harus menunggu antrian. Dalam hal ini, masyarakat awam selaku pemakai jasa lebih membutuhkan seorang pakar yang bisa memudahkan dalam mendiagnosis penyakit lebih dini agar dapat melakukan pencegahan lebih awal yang sekiranya membutuhkan waktu jika berkonsultasi dengan dokter ahli.

Berdasarkan hal-hal yang telah disebutkan di atas bahwa telah terjadi kenaikan jumlah masyarakat yang menderita hipertensi sebesar 3,6% dari tahun 2013 sampai tahun 2016 sehingga agar dapat menurunkan jumlah penderita hipertensi dibutuhkan suatu aplikasi yang dapat memprediksi seseorang menderita penyakit hipertensi agar dapat melakukan pencegahan lebih awal. Untuk melakukan pencegahan lebih awal penulis akan membangun aplikasi menggunakan metode *data mining* karena metode tersebut dapat digunakan untuk memprediksi berdasarkan data yang dimiliki, data tersebut dapat digunakan sebagai dasar pengetahuan untuk melakukan prediksi sehingga aplikasi dapat melakukan penghitungan probabilitas yang ada berdasarkan data yang telah dimiliki.

Untuk memperkuat pendapat dari penulis maka penulis memberikan salah satu contoh penelitian yang telah dilakukan oleh penelitian yang dilakukan sebelumnya oleh Marco Maureece M, Mulyadi Salim dan Alowisius Yoga Amelga yang mengenai penyakit jantung memiliki akurasi yang tinggi yaitu sebesar 85.90%.

## **Landasan Teori**

### ***Data Mining***

Dalam buku yang ditulis oleh Charu C. Aggarwal (2015: hlm. 63) menjelaskan bahwa *data mining* adalah studi tentang mengumpulkan, pembersihan, pengolahan, analisis, dan mendapatkan wawasan yang berguna dari data. *Data mining* juga merupakan sebuah istilah umum yang digunakan untuk menggambarkan berbagai aspek pengolahan data.

Dalam penelitian yang dilakukan oleh Frista Yulianora, Muchammad Hasbi Latif, dan Rika Jubel Febriana (2014:7) menjelaskan bahwa *data mining* adalah sebuah proses untuk menemukan pola yang menarik dan pengetahuan dari data yang berjumlah besar dan juga pencarian dan analisa dari jumlah data yang sangat besar dan bertujuan untuk mencari arti dari pola dan aturan yang digali datanya sehingga membuat suatu keputusan yang sangat penting.

### **Metode-Metode Dalam *Data Mining***

Dalam data mining memiliki beberapa metode yang dijelaskan dalam skripsi yang diteliti oleh Eka Miranda (2013: hlm. 7) yaitu:

#### ***Naïve Bayes***

*Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan inggris Thomas Bayes. *Naïve Bayes* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari tabel keputusan. *Naïve Bayes* memiliki rumus sebagai berikut:

$$P(Y|X) = \frac{P(X|Y) \times P(Y)}{P(X)}$$

Keterangan :

X: Data dengan class yang belum diketahui

Y: Hipotesis data merupakan suatu class spesifik

P(Y|X): Probabilitas hipotesis berdasar kondisi

P(C): Probabilitas hipotesis

$P(X|Y)$ : Probabilitas berdasarkan kondisi pada hipotesis

$P(X)$ : Probabilitas X

Rumus di atas menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas Y adalah peluang munculnya class Y, dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas Y, dibagi dengan peluang kemunculan karakteristik sampel secara global.

### Decision Tree

Decision tree adalah salah satu metode classification yang paling populer karena mudah untuk diinterpretasi oleh manusia. Decision tree menggunakan model seperti struktur pohon.

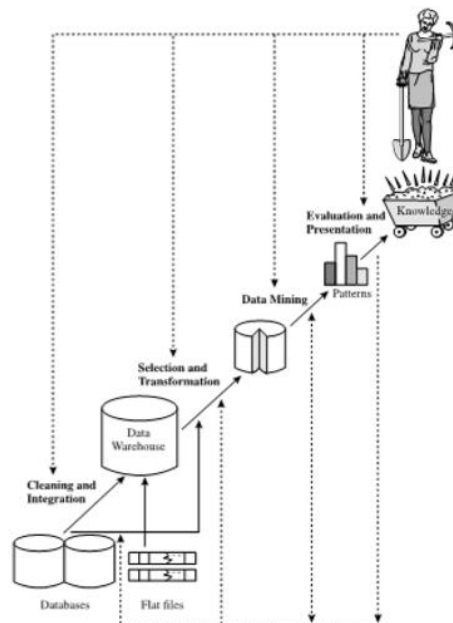
Dalam pembangunan decision tree tidak memerlukan pengaturan domain knowledge atau parameter, karena itu cocok untuk eksplorasi penemuan pengetahuan. Decision tree dapat menangani data multidimensi. Perwakilan dari pengetahuan yang diperoleh dalam bentuk pohon mempermudah untuk dipelajari dan dipahami. Decision tree memiliki akurasi yang cukup baik, namun tingkat keberhasilan penggunaannya tergantung pada data yang ada. Aplikasi klasifikasi decision tree telah digunakan dalam banyak area, seperti kedokteran, manufaktur dan produksi, analisis keuangan, astronomi, dan biologi molekuler.

### Clustering

*Clustering* adalah proses pengelompokan kumpulan data menjadi beberapa kelompok sehingga objek di dalam satu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaan dengan objek di kelompok lain. Perbedaan dan persamaannya biasanya berdasarkan nilai atribut dari objek tersebut dan dapat juga berupa perhitungan jarak. *Clustering* merupakan proses partisi satu set objek data ke dalam himpunan bagian. Setiap himpunan bagian adalah *cluster*, sehingga objek yang ada di dalam *cluster* mirip satu sama dengan yang lainnya, dan mempunyai perbedaan dengan objek dari *cluster* yang lain. Partisi tidak dilakukan dengan manual algoritma *clustering*. Oleh karena itu, *clustering* sangat berguna dan bisa menemukan grup yang tidak dikenal dalam data.

### Tahapan-tahapan Data Mining

Menurut Mujib Ridwan, Hadi Suryono, dan M. Sarosa (2013: hlm. 7) menjelaskan bahwa *data mining* dapat dibagi menjadi beberapa tahap proses yang diilustrasikan pada Gambar 1. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledge base*. Berdasarkan Gambar 2.1 dapat dilihat bahwa terdapat 7 tahapan yang dilalui.



**Gambar 1** Tahap-Tahap *Data Mining*  
(sumber : Mujib Ridwan, Hadi Suryono dan M. Sarosa: hlm. 7)

Tahap-tahap *data mining* adalah sebagai berikut:

1. Pembersihan Data (*Data Cleaning*)  
Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.
2. Integrasi data (*Data Integration*)  
Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.
3. Seleksi Data (*Data Selection*)  
Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.
4. Transformasi Data (*Data Transformation*)  
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*.
5. Proses *Mining*  
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
6. Evaluasi Pola (*Pattern Evaluation*)  
Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.
7. Presentasi Pengetahuan (*Knowledge Presentation*)  
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

### Hipertensi

Menurut Yunita Indah Prasetyaningrum (2014: hlm. 6) menjelaskan bahwa hipertensi merupakan penyakit yang memiliki tekanan darah sistolik lebih besar atau sama dengan 140 mmHg atau diastolik lebih besar atau sama dengan 90 mmHg.

Hipertensi memiliki beberapa jenis yaitu hipertensi esensial, hipertensi primer, atau hipertensi idiopatik yang tidak diketahui sebab pastinya seorang pasien menderita penyakit hipertensi sedangkan untuk hipertensi yang disebabkan oleh penyakit lain disebut sebagai hipertensi sekunder. Sebagian besar kasus hipertensi di dunia terjadi tanpa sebab yang jelas dan hanya berkisar 20% kejadian hipertensi disebabkan oleh penyakit lain.

Penyakit hipertensi bisa disebabkan oleh adanya penyakit, seperti penyakit ginjal kronis, penyakit tiroid, obesitas, atau gangguan tidur. Beberapa jenis obat juga dapat memicu terjadinya hipertensi seperti mengonsumsi pil pengontrol kelahiran, kehamilan, dan terapi hormon. Wanita yang biasanya mengonsumsi pil pengontrol kelahiran akan mengalami peningkatan tekanan darah sistolik ataupun diastolik. Sementara itu, terapi hormon untuk mengurangi gejala menopause bisa menyebabkan sedikit peningkatan tekanan darah sistolik.

Kejadian hipertensi biasanya tidak memiliki tanda dan gejala namun ada beberapa gejala yang sering muncul seperti sakit kepala, rasa panas di tengkuk, atau kepala terasa berat. Namun, gejala tersebut tidak bisa dijadikan patokan ada atau tidaknya hipertensi pada diri seseorang. Satu-satunya cara untuk mengetahuinya adalah dengan melakukan pengecekan tekanan darah. Seorang pasien biasanya tidak menyadari bahwa dirinya mengalami hipertensi hingga ditemukan kerusakan dalam organ, seperti ginjal.

Tekanan darah yang tinggi sangat berbahaya karena dapat memperberat kerja organ jantung. Selain itu, aliran tekanan darah tinggi membahayakan arteri, organ jantung, ginjal dan mata. Penyakit hipertensi sering disebut dengan "*silent killer*" karena tidak memberikan gejala yang khas, tetapi bisa meningkatkan kejadian *stroke*, serangan jantung, penyakit ginjal kronis bahkan kebutaan jika tidak dikontrol dan dikendalikan dengan baik.

### Basis Data

Menurut Robi Yanto (2016: hlm. 2) menjelaskan bahwa basis data adalah pengolahan data baik tercatat dalam lembar kertas maupun menggunakan sistem komputer kemudian ke semua data tersebut akan diarsipkan dengan pola pengarsipan ataupun pola yang sudah terkomputerisasi. Basis data memiliki prinsip utama yaitu mengatur data atau arsip, yang memiliki tujuan utama untuk memberikan kemudahan dan kecepatan dalam pengambilan kembali data atau arsip.

### Metode Sampel

Menurut Sarini Abdullah dan Taufik Edy Sutanto (2015: hlm. 15) menjelaskan bahwa sampel adalah bagian dari populasi yang diambil dengan cara tertentu dan diharapkan dapat mewakili populasi (representatif). Representatif artinya memiliki seluruh sifat-sifat populasi meski jumlahnya lebih sedikit. Sedangkan *Sampling Frame* adalah daftar objek yang akan diambil sampel.

Secara umum, metode pengambilan sampel terbagi sebagai berikut:

1. Simple Random Sampling

Pada metode ini, setiap anggota populasi memiliki kesempatan yang sama untuk terpilih dalam sampel. Misalnya dari populasi berukuran  $N$  akan diambil sampel berukuran  $n$ . Untuk melakukan simple random sampling, kumpulkan semua nama dari anggota populasi, kemudian dipilih secara acak sebanyak  $n$  anggota.

Di dalam simple random sampling memiliki beberapa karakteristik yaitu:

- Setiap objek memiliki kesempatan yang sama untuk terpilih.
- Merupakan bentuk standar-biasanya digunakan sebagai pembandingan dari metode yang lain.
- Cocok ketika populasi relatif kecil, sampling frame lengkap, dan up to date.

2. Systematic Random Sampling

Systematic sampling adalah proses menyeleksi setiap anggota kelipatan ke- $k$  dari daftar anggota populasi. Di dalam systematic random sampling memiliki beberapa prosedur yang ada yaitu:

- Pilih satu elemen dari  $k$  elemen pertama secara acak.
- Pilih setiap elemen ke  $k$  setelahnya.
- Pilih  $k \leq N/n$

3. Stratified Random Sampling

Stratified random sampling diperoleh dengan membagi populasi atas beberapa subgroup. Yang mana dalam subgroup serupa (homogen) tapi antar grup berbeda (heterogen). Dari setiap subgroup diambil sampel secara acak. Banyaknya unit sampel dari setiap strata proporsional atau dalam jumlah yang relative sama untuk setiap strata.

4. Accidental Sampling

Teknik ini juga disebut convenience sampling atau incidental sampling. Dalam teknik ini, peneliti mengumpulkan sampel dari unit sampel yang kebetulan ditemuinya atau mereka yang mudah ditemui dan dijangkau. Setelah jumlahnya mencukupi maka pengambilan sampel dihentikan.

### Cross-Validation

Menurut Rezaul Karim dan Mahedi Kaysar (2016: hlm. 318) menjelaskan bahwa Cross-Validation adalah sebuah teknik validasi untuk menilai kualitas analisis dan hasil statistik. Targetnya adalah membuat model menjadi general pada data test yang independen.

Salah satu kegunaan teknik Cross-Validation yang sempurna adalah membuat prediksi dari model machine learning. Secara teknis, Cross-Validation akan membantu untuk memperkirakan bagaimana model prediktif akan tampil secara akurat dalam praktik saat menerapkan machine learning.

Selama proses Cross-Validatizon, model biasanya dilatih dengan kumpulan data dari jenis yang dikenal, sebaliknya, diuji dengan menggunakan dataset dari jenis yang tidak diketahui. Dalam hal ini, Cross-Validation membantu menjelaskan dataset untuk menguji model dalam tahap pelatihan dengan menggunakan validation set.

### Confusion Matrix

Menurut Aida Indriani (2014: hlm. G-6) menjelaskan bahwa *Confusion Matrix* adalah sebuah tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan. Contoh *confusion matrix* untuk klasifikasi biner ditunjukkan pada tabel berikut:

**Tabel 1** Contoh Tabel *Confusion Matrix*

|                  |   | Kelas Prediksi |    |
|------------------|---|----------------|----|
|                  |   | 1              | 0  |
| Kelas Sebenarnya | 1 | TP             | FN |
|                  | 0 | FP             | TN |

Keterangan untuk tabel tersebut dinyatakan sebagai berikut:

1. *True Positive* (TP), yaitu jumlah dokumen dari kelas 1 yang benar dan diklasifikasikan sebagai kelas 1.
2. *True Negative* (TN), yaitu jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0.
3. *False Positive* (FP), yaitu jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1.
4. *False Negative* (FN), yaitu jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0.

## Analisa dan Perancangan

### Identifikasi dan Analisa Kebutuhan Aplikasi

Aplikasi yang akan di rancang merupakan aplikasi untuk mendeteksi secara dini penyakit hipertensi yang memberikan fasilitas diagnosa di mana terdapat pertanyaan tentang pola hidup, riwayat penyakit orang tua dari pasien dan kondisi fisik dari pasien, yang kemudian disimpulkan menjadi kemungkinan penyakit.

Untuk dapat melakukan diagnosis, aplikasi ini memiliki data gejala dan penyakit yang tersedia dalam basis data berdasarkan kuesioner yang diberikan kepada pasien yang menderita penyakit hipertensi. Untuk pengujian, aplikasi akan dibandingkan dengan data yang dimiliki dari hasil kuesioner yang telah diberikan kepada pasien untuk mengetahui tingkat ketepatan dari hasil diagnosis yang diberikan oleh program.

### Tahap - Tahap Penelitian

**Tabel 2** Tahap Penelitian

| Rentang Waktu            | Kegiatan  |
|--------------------------|---|
| Minggu 1                 | Meminta izin untuk melakukan penelitian dari Fakultas Teknologi Informasi |
| Minggu 2                 | Meminta izin untuk melakukan penelitian di Rumah Sakit X                  |
| Minggu 3 sampai Minggu 7 | Melakukan pengambilan data  |
| Minggu 8                 | Melakukan pengolahan data yang telah diperoleh                            |

Tabel 2 menunjukkan tahap-tahap yang dilalui dalam melakukan penelitian berlangsung yang di mana pada minggu pertama penulis meminta izin untuk melakukan penelitian dari Fakultas Teknologi Informasi Universitas Advent Indonesia sebagai bukti kepada pihak Rumah Sakit X bahwa peneliti memang melakukan penelitian tersebut dengan membawa surat pengantar dari Fakultas Teknologi Informasi. Pada minggu ke-2 peneliti meminta izin kepada pihak rumah sakit untuk melakukan penelitian di rumah sakit X agar dapat memperoleh data yang dibutuhkan dalam penelitian ini. Setelah mendapat izin pada minggu ke-3 sampai ke-7 peneliti melakukan pengambilan data yang akan diberikan kepada pasien rumah sakit X dalam bentuk kuesioner. Pada minggu ke-8 peneliti mengolah data yang telah diperoleh dari kuesioner yang telah disebarakan.

### Akuisisi Pengetahuan

Setelah domain ditentukan kemudian tahap berikutnya adalah akuisisi pengetahuan di mana penggalan, pengumpulan pengetahuan dari sumber-sumber yang tersedia. Pengetahuan diperoleh dari hasil penelitian yang dilakukan oleh orang lain.

Menurut Yunita Indah Prasetyaningrum (2014) memiliki beberapa faktor yang dapat meningkatkan risiko terjadinya hipertensi. Berikut beberapa faktor risiko utama terjadinya hipertensi:

1. Usia  
Kejadian hipertensi cenderung meningkat seiringnya penambahan usia. Sebanyak 65% orang Amerika berusia 60 tahun atau lebih mengalami hipertensi. Meskipun demikian, hipertensi tidak selalu hadir seiring dengan proses penuaan.
2. Jenis Kelamin  
Laki-laki atau perempuan memiliki kemungkinan yang sama untuk mengalami hipertensi selama kehidupannya. Namun, laki-laki lebih berisiko mengalami hipertensi dibandingkan

dengan perempuan saat berusia sebelum 45 tahun. Sebaliknya saat usia 65 tahun ke atas, perempuan lebih berisiko mengalami hipertensi dibandingkan laki-laki. Kondisi ini dipengaruhi oleh hormon wanita yang memasuki masa menopause, lebih berisiko untuk mengalami obesitas sehingga meningkatkan risiko terjadinya hipertensi.

3. **Obesitas**  
Seseorang yang mengalami obesitas atau kegemukan memiliki risiko lebih besar untuk mengalami pre-hipertensi atau hipertensi. Indikator yang biasa digunakan untuk menentukan ada-tidaknya obesitas pada seseorang adalah melalui pengukuran IMT atau lingkaran perut. Meskipun demikian, kedua indikator tersebut bukanlah indikator terbaik untuk menentukan terjadinya hipertensi, tetapi menjadi salah satu faktor risiko yang dapat mempercepat kejadian hipertensi.
4. **Kurang Aktivitas Fisik**  
Aktivitas fisik merupakan pergerakan otot anggota tubuh yang membutuhkan energi atau pergerakan yang bermanfaat untuk meningkatkan kesehatan. Contohnya berkebun, berenang, menari, bersepeda, atau yoga. Aktivitas fisik sangat bermanfaat bagi kesehatan tubuh, khususnya organ jantung dan paru-paru. Aktivitas fisik juga menyehatkan pembuluh darah dan mencegah hipertensi. Usaha pencegahan hipertensi akan optimal jika aktif beraktivitas fisik dibarengi dengan menjalankan diet sehat dan berhenti merokok.
5. **Kebiasaan Merokok dan Konsumsi Minuman Beralkohol**  
Kebiasaan merokok menyebabkan 1 dari 5 kasus kematian di Amerika setiap tahun. Merokok merupakan penyebab kematian dan kesakitan yang paling bisa dicegah. Pasalnya, zat kimia yang dihasilkan dari pembakaran tembakau berbahaya bagi sel darah dan organ tubuh lainnya, seperti jantung, pembuluh darah, mata, organ reproduksi, paru-paru, bahkan organ pencernaan. Selain itu, konsumsi minuman beralkohol juga dapat meningkatkan tekanan darah. Penelitian menunjukkan bahwa risiko hipertensi meningkat dua kali lipat jika mengonsumsi minuman beralkohol lebih dari tiga gelas sehari.
6. **Faktor Lain**  
Riwayat keluarga penderita hipertensi turut meningkatkan risiko kejadian hipertensi. Sementara itu, stres berkepanjangan juga dapat meningkatkan risiko seseorang untuk mengalami hipertensi.

### Kebutuhan Atribut

Kebutuhan atribut yang digunakan oleh penulis merupakan hasil dari penelitian sebelumnya yang dilakukan oleh Anggun Suprihati (2016), Jane A. Kalangi, Adrian Umboh, Vivekenanda Pateda (2015) dan Rinda, Tanto Hariyanto, Vita Maryah Ardiyanti (2017). Penulis menjabarkan sebagai berikut:

**Tabel 3** Tabel Atribut

| No | Atribut          | Keterangan   |
|----|------------------|--|
| 1  | Usia             | Usia pasien (Remaja, Dewasa, Manula)   |
| 2  | Jenis Kelamin    | Jenis kelamin pasien (Laki-laki, Perempuan)  |
| 3  | Perokok          | Pasien perokok atau tidak (Perokok, Tidak)   |
| 4  | Jenis rokok      | Jenis rokok yang digunakan oleh pasien (Filter, Kretek)  |
| 5  | Jumlah rokok     | Jumlah rokok yang digunakan pasien setiap harinya (< 1 bungkus, 1 bungkus, 2 bungkus, > 3 bungkus) |
| 6  | Konsumsi alkohol | Pasien mengonsumsi alkohol atau tidak (konsumsi, tidak konsumsi)                                   |

|    |                           |  |
|----|---------------------------|--|
| 7  | Aktivitas Fisik           | Pasien memiliki kecukupan aktivitas fisik atau tidak (kurang, cukup)                             |
| 8  | Bentuk tubuh              | Pasien gemuk, kurus, sangat gemuk, sangat kurus atau normal (kurus, normal, gemuk)               |
| 9  | Konsumsi Makanan Bergaram | Pasien selalu mengonsumsi makanan yang mengandung garam (sering mengonsumsi, jarang mengonsumsi) |
| 10 | Konsumsi Daging           | Pasien selalu makan daging atau tidak (tidak konsumsi, jarang konsumsi, selalu konsumsi)         |
| 11 | Konsumsi sayur            | Pasien selalu mengonsumsi sayuran atau tidak (tidak konsumsi, jarang konsumsi, selalu konsumsi)  |
| 12 | Jumlah jam tidur          | Jumlah jam tidur dalam 1 hari (< 8 jam, Sekitar 8 jam, > 9 jam)                                  |
| 13 | Riwayat hipertensi ayah   | Ayah dari pasien memiliki riwayat hipertensi atau tidak (memiliki, tidak memiliki)               |
| 14 | Riwayat hipertensi ibu    | Ibu dari pasien memiliki riwayat hipertensi atau tidak (memiliki, tidak memiliki)                |

### Sumber Data

Data yang digunakan dalam penelitian ini didapatkan dari bagian jantung dan pembuluh darah Rumah Sakit X berupa kuesioner kepada pasien yang berkunjung selama 1 bulan. Selanjutnya data tersebut dikonversi ke dalam aplikasi pengolahan data yaitu MySQL. Tabel kuesioner adalah kumpulan data pasien yang telah melakukan pengecekan tekanan darah dan mengisi kuesioner.

Untuk metode pengumpulan data yang digunakan menggunakan penghitungan dari rumus Slovin yang berguna untuk menentukan ukuran sampel minimal (n) jika diketahui ukuran populasi (N) pada taraf signifikansi  $\alpha$  adalah :

$$n = \frac{N}{1 + N\alpha^2}$$

### Instrumen Penelitian

Instrumen merupakan suatu mekanisme untuk mengukur fenomena, digunakan untuk mengumpulkan dan merekam data, pengambilan keputusan, dan pengertian. Suatu instrumen seperti kuesioner biasanya digunakan untuk memperoleh informasi faktual, mendukung observasi, ataupun mengkaji perilaku dan pendapat. Instrumen yang digunakan dalam penelitian ini adalah lembar kaji atau kuesioner yang diadaptasi dari pengetahuan para pakar yang telah melakukan penelitian mengenai pola hidup penderita Hipertensi secara umum.



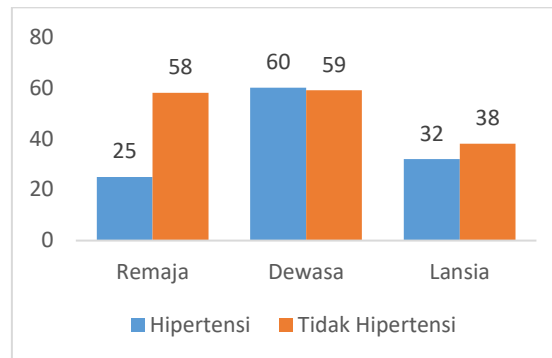
## Hasil dan Pembahasan *Modeling*

### Analisis Awal Pola Persebaran Data Dengan Visual dan Analisis Kluster

Eksplorasi visual dan analisis kluster sangat berguna untuk melakukan analisis awal pengujian terhadap data. Tujuan dari eksplorasi visual dan analisis kluster adalah untuk mengetahui dan mengidentifikasi atribut yang paling berpengaruh dan mengidentifikasi atribut yang paling berpengaruh dalam pola persebaran data yang akan di-mining (*attribute profile*). Berikut adalah *attribute profile* data yang akan di-mining.

#### Usia

Pada Gambar 2 menunjukkan jumlah pasien yang menderita hipertensi dan yang tidak hipertensi berdasarkan kategori yang ada dalam atribut usia. Pada Gambar 5 menunjukkan jumlah yang hipertensi untuk kategori usia remaja sebanyak 25 pasien sedangkan untuk yang tidak hipertensi menunjukkan 58 pasien. Kategori usia dewasa sebanyak 60 pasien yang menderita hipertensi sedangkan 59 pasien yang tidak hipertensi. Pada kategori lansia sebanyak 32 pasien yang hipertensi sedangkan 38 pasien yang tidak hipertensi. Berdasarkan data tersebut dapat diketahui bahwa yang paling banyak menderita hipertensi adalah pada usia dewasa yang memiliki usia antara 30 sampai 45 tahun.

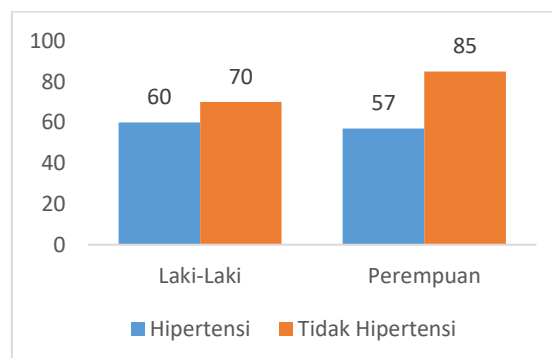


**Gambar 2** *Attribute Profile* Usia

#### Jenis Kelamin

Pada Gambar 3 menunjukkan jumlah pasien yang menderita hipertensi dan yang tidak hipertensi berdasarkan kategori yang ada dalam atribut jenis kelamin. Pada Gambar 6 menunjukkan jumlah yang hipertensi untuk kategori jenis kelamin laki-laki sebanyak 60 pasien menderita hipertensi sedangkan sebanyak 70 pasien yang tidak hipertensi.

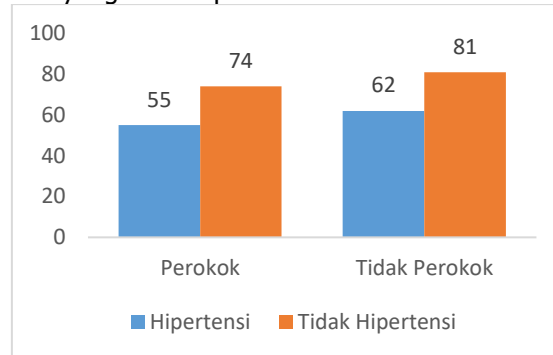
Untuk kategori jenis kelamin perempuan sebanyak 57 pasien menderita hipertensi sedangkan sebanyak 85 pasien yang tidak hipertensi.



**Gambar 3** *Attribute Profile* Jenis Kelamin

**Perokok**

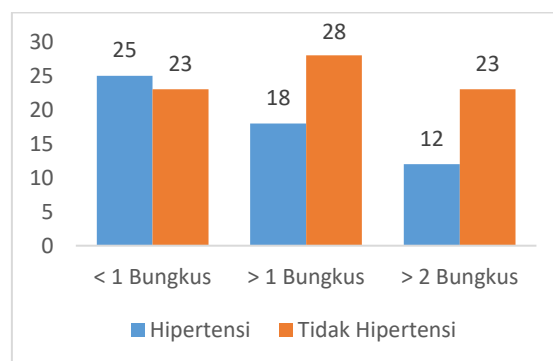
Pada Gambar 4 menunjukkan jumlah pasien yang menderita hipertensi dan yang tidak hipertensi berdasarkan kategori yang ada dalam atribut perokok. Pada Gambar 7 menunjukkan jumlah yang hipertensi untuk kategori perokok sebanyak 55 pasien menderita hipertensi sedangkan sebanyak 74 pasien yang tidak hipertensi. Untuk kategori tidak perokok sebanyak 62 pasien menderita hipertensi sedangkan sebanyak 81 pasien yang tidak hipertensi.



**Gambar 4** Attribute Profile Perokok

**Jumlah Rokok**

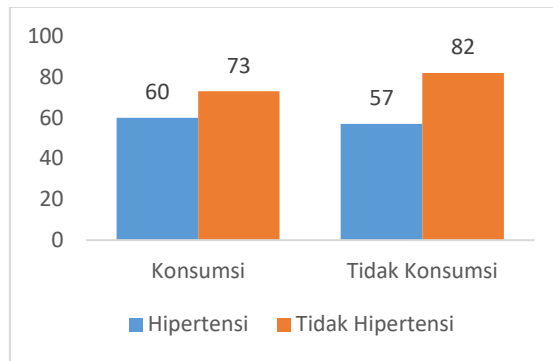
Pada Gambar 5 menunjukkan jumlah pasien yang menderita hipertensi dan yang tidak hipertensi berdasarkan kategori yang ada dalam atribut jumlah rokok. Pada Gambar 9 menunjukkan jumlah yang hipertensi untuk kategori kurang dari 1 bungkus sebanyak 25 pasien menderita hipertensi sedangkan sebanyak 23 pasien yang tidak hipertensi. Untuk kategori lebih dari 1 bungkus sebanyak 18 pasien menderita hipertensi sedangkan sebanyak 28 pasien yang tidak hipertensi. Untuk kategori lebih dari 2 bungkus 12 pasien yang menderita hipertensi sedangkan sebanyak 23 pasien yang tidak hipertensi.



**Gambar 5** Attribute Profile Jumlah Rokok

**Konsumsi Alkohol**

Pada Gambar 6 menunjukkan jumlah pasien yang menderita hipertensi dan yang tidak hipertensi berdasarkan kategori yang ada dalam atribut konsumsi alkohol. Pada Gambar 10 menunjukkan jumlah yang hipertensi untuk kategori konsumsi sebanyak 60 pasien menderita hipertensi sedangkan sebanyak 73 pasien yang tidak hipertensi. Untuk kategori tidak perokok sebanyak 57 pasien menderita hipertensi sedangkan sebanyak 82 pasien yang tidak hipertensi.



**Gambar 6** *Attribute Profile* Konsumsi Alkohol

### **Likelihood**

Nilai dari *likelihood* adalah hasil bagi antara jumlah total kategori atribut dengan *target class* dibagi dengan jumlah total *target class*. Dalam hal ini digunakan fungsi *statement select count* pada *sql* untuk menghitung jumlah total kategori atribut dan jumlah total *target class* (Hipertensi atau Tidak Hipertensi).

SQL *statement* untuk mencari jumlah total kategori atribut sesuai dengan *target class*:

```
select count (attribute) from training where kategori attribute AND hipertensi like 'kategori target class'
```

Berikut adalah tabel untuk total dari jumlah kategori atribut sesuai dengan *target class*:

**Tabel 4** Tabel Jumlah Dari Setiap Kategori *Attribute* Sesuai Dengan *Target Class*

| <i>Attribute</i> | Kategori        | Hipertensi |                  |
|------------------|-----------------|------------|------------------|
|                  |                 | Hipertensi | Tidak Hipertensi |
| Usia             | Remaja          | 25         | 58               |
|                  | Dewasa          | 60         | 59               |
|                  | Lansia          | 32         | 38               |
| Jenis Kelamin    | Laki-Laki       | 60         | 70               |
|                  | Perempuan       | 57         | 85               |
| Perokok          | Perokok         | 55         | 74               |
|                  | Tidak Perokok   | 62         | 81               |
| Jenis Rokok      | Kretek          | 13         | 12               |
|                  | Filter          | 42         | 62               |
| Jumlah Rokok     | < 1 Bungkus     | 25         | 23               |
|                  | > 1 Bungkus     | 18         | 28               |
|                  | > 2 Bungkus     | 12         | 23               |
| Konsumsi Alkohol | Konsumsi        | 60         | 73               |
|                  | Tidak Konsumsi  | 57         | 82               |
| Aktifitas Fisik  | Kurang          | 54         | 70               |
|                  | Cukup           | 63         | 85               |
| Bentuk Tubuh     | Kurus           | 51         | 50               |
|                  | Normal          | 40         | 56               |
|                  | Gemuk           | 26         | 49               |
| Kadar Garam      | Jarang Konsumsi | 54         | 74               |

|                         |                |    |     |
|-------------------------|----------------|----|-----|
|                         | Rajin Konsumsi | 63 | 81  |
| Konsumsi Daging         | Jarang         | 50 | 58  |
|                         | Selalu         | 46 | 55  |
|                         | Tidak Konsumsi | 21 | 42  |
| Konsumsi Sayur          | Jarang         | 52 | 30  |
|                         | Selalu         | 65 | 125 |
| Jam Tidur               | Kurang 8 Jam   | 42 | 61  |
|                         | Sekitar 8 Jam  | 40 | 51  |
|                         | Lebih 9 Jam    | 35 | 43  |
| Riwayat Hipertensi Ayah | Ada            | 62 | 4   |
|                         | Tidak Ada      | 55 | 151 |
| Riwayat Hipertensi Ibu  | Ada            | 56 | 2   |
|                         | Tidak Ada      | 61 | 153 |

SQL statement untuk mencari jumlah total kategori *target class* :  
 select count(hipertensi) from training where hipertensi like 'kategori target class'

**Tabel 5** Tabel Jumlah Dari Total Kategori *Target Class*

|            |                  |
|------------|------------------|
| Hipertensi | Tidak Hipertensi |
| 117        | 155              |

Contoh *likelihood* Usia Remaja tahun untuk kategori *target class* Hipertensi yang menyatakan Hipertensi adalah:

$Likelihood = \text{Total kategori Atribut} / \text{Total kategori target class}$

$Likelihood = 25/117 = 0.213$

Contoh *likelihood* Usia Remaja tahun untuk kategori target class hipertensi yang menyatakan Tidak Hipertensi adalah:

$Likelihood = \text{total kategori atribut} / \text{total kategori target class}$

$Likelihood = 58/155 = 0.374$

**Tabel 6** Tabel *Likelihood*

| Attribute     | Kategori      | Hipertensi |                  |
|---------------|---------------|------------|------------------|
|               |               | Hipertensi | Tidak Hipertensi |
| Usia          | Remaja        | 0.213      | 0.374            |
|               | Dewasa        | 0.512      | 0.380            |
|               | Lansia        | 0.273      | 0.245            |
| Jenis Kelamin | Laki-Laki     | 0.512      | 0.451            |
|               | Perempuan     | 0.487      | 0.548            |
| Perokok       | Perokok       | 0.470      | 0.477            |
|               | Tidak Perokok | 0.529      | 0.522            |
| Jenis Rokok   | Kretek        | 0.111      | 0.077            |
|               | Filter        | 0.358      | 0.4              |
| Jumlah Rokok  | < 1 Bungkus   | 0.213      | 0.148            |
|               | > 1 Bungkus   | 0.153      | 0.180            |

|                         |                 |       |       |
|-------------------------|-----------------|-------|-------|
|                         | > 2<br>Bungkus  | 0.102 | 0.148 |
| Konsumsi Alkohol        | Konsumsi        | 0.513 | 0.470 |
|                         | Tidak Konsumsi  | 0.487 | 0.529 |
| Aktifitas Fisik         | Kurang          | 0.461 | 0.451 |
|                         | Cukup           | 0.538 | 0.548 |
| Bentuk Tubuh            | Kurus           | 0.436 | 0.322 |
|                         | Normal          | 0.342 | 0.361 |
|                         | Gemuk           | 0.222 | 0.316 |
| Kadar Garam             | Jarang Konsumsi | 0.461 | 0.477 |
|                         | Rajin Konsumsi  | 0.538 | 0.522 |
| Konsumsi Daging         | Jarang          | 0.427 | 0.374 |
|                         | Selalu          | 0.393 | 0.354 |
|                         | Tidak Konsumsi  | 0.179 | 0.270 |
| Konsumsi Sayur          | Jarang          | 0.444 | 0.193 |
|                         | Selalu          | 0.555 | 0.806 |
| Jam Tidur               | Kurang 8 Jam    | 0.359 | 0.393 |
|                         | Sekitar 8 Jam   | 0.342 | 0.329 |
|                         | Lebih 9 Jam     | 0.299 | 0.277 |
| Riwayat Hipertensi Ayah | Ada             | 0.53  | 0.025 |
|                         | Tidak Ada       | 0.470 | 0.974 |
| Riwayat Hipertensi Ibu  | Ada             | 0.479 | 0.013 |
|                         | Tidak Ada       | 0.521 | 0.987 |

### Prior

Nilai dari *prior* adalah hasil dari jumlah total kategori *target class* dibagi dengan jumlah total *record data*.

SQL *statement* untuk mencari jumlah total kategori target class

select count(hipertensi) from training where hipertensi like 'kategori target class'

**Tabel 7** Tabel Jumlah Dari Kategori *Target Class*

| Hipertensi | Tidak Hipertensi |
|------------|------------------|
| 117        | 155              |

SQL *statement* untuk mencari jumlah total *record data*

select count(\*) from training → 272

$Prior = \text{Jumlah total kategori } target\ class / \text{Jumlah total record}$

$Prior \text{ untuk } target\ class \text{ Hipertensi} = 117 / 272 = 0.43$

$Prior \text{ untuk } target\ class \text{ Tidak Hipertensi} = 155 / 272 = 0.57$

### Posterior

Nilai dari *posterior* adalah hasil dari *likelihood* dikali dengan *prior*

1. Nilai *posterior* Hipertensi

$Posterior \text{ Hipertensi} = Likelihood \text{ Hipertensi} * Prior \text{ Hipertensi}$

2. Nilai *posterior* Tidak Hipertensi

$Posterior \text{ Tidak Hipertensi} = Likelihood \text{ Tidak Hipertensi} * Prior \text{ Tidak Hipertensi}$

Contoh Perhitungan *Posterior* dengan kategori atribut sebagai berikut:

1. Usia = Remaja
2. Jenis Kelamin = Laki-laki
3. Prokok = Perokok
4. Jenis Rokok = Filter
5. Jumlah Rokok = < 1 Bungkus
6. Konsumsi Alkohol = Tidak Konsumsi
7. Aktivitas Fisik = Kurang
8. Bentuk Tubuh = Gemuk
9. Kadar Garam = >2400 Mg
10. Konsumsi Daging = Selalu Konsumsi
11. Konsumsi Sayur = Selalu Konsumsi
12. Jumlah Jam Tidur = > 9 Jam
13. Riwayat Hipertensi Ayah = Tidak Memiliki
14. Riwayat Hipertensi Ibu = Tidak Memiliki

Nilai *posterior* Hipertensi :

$Posterior\ Hipertensi = (Likelihood\ Hipertensi) * (Prior\ Hipertensi)$

$Posterior\ Hipertensi = (0.231 * 0.512 * 0.47 * 0.358 * 0.213 * 0.487 * 0.461 * 0.222 * 0.538 * 0.393 * 0.555 * 0.299 * 0.47 * 0.521) * 0.43 = 7.20\ e\ -7$

Nilai *posterior* Tidak Hipertensi :

$Posterior\ Tidak\ Hipertensi = (Likelihood\ Tidak\ Hipertensi) * (Prior\ Tidak\ Hipertensi)$

$Posterior\ Tidak\ Hipertensi = (0.374 * 0.451 * 0.477 * 0.4 * 0.148 * 0.529 * 0.451 * 0.316 * 0.522 * 0.354 * 0.806 * 0.277 * 0.974 * 0.987) * 0.57 = 8.12\ e\ -6$

### Perbandingan *Posterior* Hipertensi dan *Posterior* Tidak Hipertensi

Hasil klasifikasi akan menampilkan prediksi berdasarkan nilai *Posterior* yang lebih besar antara *posterior* hipertensi dengan *posterior* tidak hipertensi. Kalau *posterior* hipertensi lebih besar daripada *posterior* tidak hipertensi maka sistem akan menampilkan prediksi yang menunjukkan hipertensi dan juga sebaliknya jika nilai *posterior* tidak hipertensi lebih besar daripada *posterior* hipertensi maka sistem akan menampilkan prediksi yang menunjukkan tidak hipertensi.

### Evaluation

Evaluasi klasifikasi *naïve bayes* menggunakan metode *Cross-Validation* (CV) yang merupakan metode statistik yang dapat digunakan untuk mengevaluasi kinerja model atau algoritma dimana data dipisahkan menjadi dua subset yaitu data proses pembelajaran dan data validasi atau evaluasi. Model atau algoritma dilatih oleh subset pembelajaran dan divalidasi oleh subset validasi. Biasanya CV K-fold digunakan karena dapat mengurangi waktu komputasi dengan tetap menjaga keakuratan estimasi. 10 K-Fold CV adalah salah satu K-Fold yang direkomendasikan untuk pemilihan model terbaik karena cenderung memberikan estimasi akurasi yang kurang bias dibandingkan dengan CV biasa, *leave-one-out* CV dan bootstrap. Dalam 10 K-Fold CV, data dibagi menjadi 10 fold berukuran kira-kira sama, sehingga dimiliki 10 subset data untuk mengevaluasi kinerja model atau algoritma. Untuk masing-masing dari 10 subset data tersebut, CV akan menggunakan 9 fold untuk pelatihan dan 1 fold untuk pengujian (Wibowo, 2017).

Menghitung jumlah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN).

**Tabel 8** Tabel *Confusion Matrix*

|                  | Hipertensi | Tidak Hipertensi |
|------------------|------------|------------------|
| Hipertensi       | 91         | 26               |
| Tidak Hipertensi | 6          | 149              |

Keterangan tabel:

TP = 91

Keterangan dari *data testing* yang menyatakan hipertensi dan sistem menampilkan hipertensi.

TN = 149

Keterangan dari *data testing* yang menyatakan tidak hipertensi dan sistem menampilkan tidak hipertensi.

FN = 26

Keterangan dari *data testing* yang menyatakan hipertensi dan sistem menampilkan tidak hipertensi.

FP = 6

Keterangan dari *data testing* yang menyatakan tidak hipertensi dan sistem menampilkan hipertensi.

P = 117

Keterangan dari *data testing* yang menyatakan hipertensi.

N = 155

Keterangan dari *data testing* yang menyatakan tidak hipertensi.

### **Sensitivitas**

$$\text{Sensitivitas} = \frac{TP}{P}$$

Sensitivitas = 91 / 117

Sensitivitas = 77%

Berdasarkan perhitungan tersebut menyatakan bahwa sistem dalam mengenali *target class* hipertensi secara benar sebanyak 77%.

### **Spesifisitas**

$$\text{Spesifisitas} = \frac{TN}{N}$$

Spesifisitas = 149 / 155

Spesifisitas = 96%

Berdasarkan perhitungan tersebut menyatakan bahwa persentase sistem dalam mengenali *target class* tidak hipertensi secara benar sebanyak 96%

### **Akurasi**

$$\text{Akurasi} = \frac{(TP + TN)}{(P + N)}$$

Akurasi = 240 / 272

Akurasi = 88%

Berdasarkan perhitungan tersebut menyatakan bahwa persentase sistem dalam memberikan hasil pemodelan yang benar sebanyak 88%.

### **Error Rate**

*Error rate* = 1 – akurasi

*Error rate* = 100% - 88% = 12%

Berdasarkan perhitungan tersebut menyatakan bahwa persentase sistem dalam memberikan asli yang salah sebanyak 12%.

### **Supervised Attribute**

*Supervised attribute* merupakan pengertian mengenai atribut yang paling berpengaruh dalam melakukan prediksi. Pada Gambar 19 menunjukkan bahwa dari 15 atribut terdapat 4 atribut berdasarkan aplikasi WEKA yang paling berpengaruh dalam melakukan prediksi yaitu bentuk tubuh, konsumsi sayur, riwayat ayah dan riwayat ibu. Dalam mencari atribut yang paling berpengaruh menggunakan *attribute evaluator* yaitu CSF Subset Evaluator dengan metode pencarian Best First, sehingga diketahuilah bahwa ada 4 atribut yang paling berpengaruh.

```

=== Attribute Selection on all input data ===

Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 81
  Merit of best subset found: 0.414

Attribute Subset Evaluator (supervised, Class (nominal): 15 Keterangan):
  CFS Subset Evaluator
  Including locally predictive attributes

Selected attributes: 8,11,13,14 : 4
  Bentuk Tubuh
  Konsumsi Sayur
  Riwayat Ayah
  Riwayat Ibu
    
```

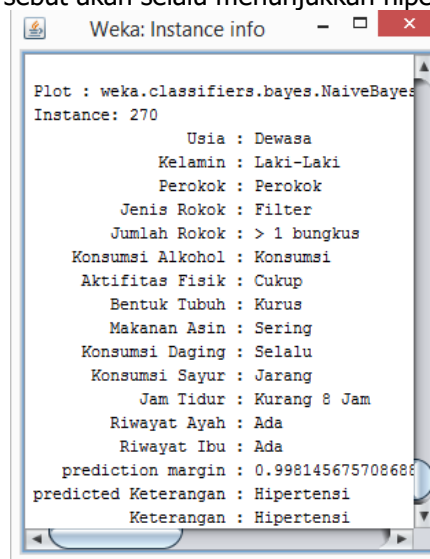
**Gambar 7 Supervised Attribute**

1. Pola Hidup Yang Tidak Menghasilkan Error

Dalam melakukan prediksi selalu ada kesalahan klasifikasi tetapi ada beberapa pola hidup yang ada dalam data yang tidak menghasilkan kesalahan klasifikasi yaitu:

2. Pola Hidup Yang Pasti Menyatakan Hipertensi

Pada Gambar 8 menunjukkan pola hidup hipertensi berdasarkan tempat penelitian dilakukan yang berada di Jakarta Timur dengan isi dalam atribut yaitu usia dewasa, jenis kelamin laki-laki, perokok, jenis rokok filter, jumlah rokok lebih dari 1 bungkus dalam sehari, konsumsi alkohol, aktifitas fisik cukup, bentuk tubuh kurus, selalu mengkonsumsi makanan asin, selalu mengkonsumsi daging, jarang mengkonsumsi sayuran, jam tidur kurang dari 8 jam dalam sehari, memiliki riwayat hipertensi pada ayah, dan memiliki riwayat hipertensi pada ibu. Pada Gambar 8 menunjukkan bahwa pola hidup tersebut memiliki prediction margin sebesar 0.9981 yang merupakan terbesar diantara pola hidup yang lainnya, sehingga pola hidup tersebut akan selalu menunjukkan hipertensi.

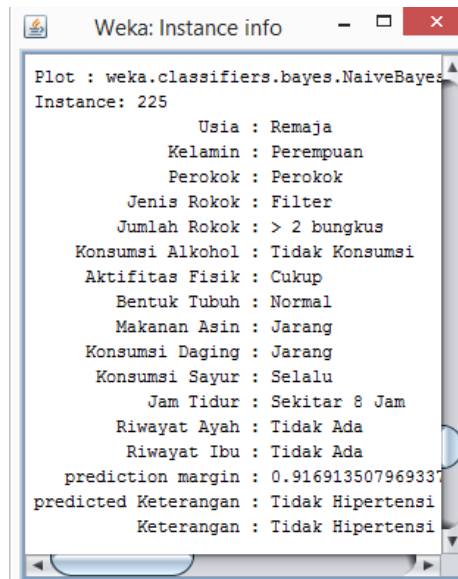


**Gambar 8 Pola Hidup Hipertensi**

3. Pola Hidup Yang Pasti Menyatakan Tidak Hipertensi

Pada Gambar 21 menunjukkan pola hidup tidak hipertensi berdasarkan tempat penelitian dilakukan yang berada di Jakarta Timur dengan isi dalam atribut yaitu usia remaja, jenis kelamin perempuan, perokok, jenis rokok filter, jumlah rokok lebih dari 2 bungkus dalam sehari, tidak konsumsi alkohol, aktifitas fisik cukup, bentuk tubuh normal, sering mengkonsumsi makanan asin, jarang konsumsi daging, selalu konsumsi sayur, jam tidur sekitar 8 jam, riwayat hipertensi ayah tidak ada, dan riwayat hipertensi ibu tidak ada. Pada Gambar 21 menunjukkan *prediction margin* yang dimiliki oleh pola hidup tersebut sebesar 0.9169 sehingga pola hidup tersebut akan selalu menunjukkan tidak hipertensi.





Gambar 9 Pola Hidup Tidak Hipertensi

### Hasil Tampilan Aplikasi

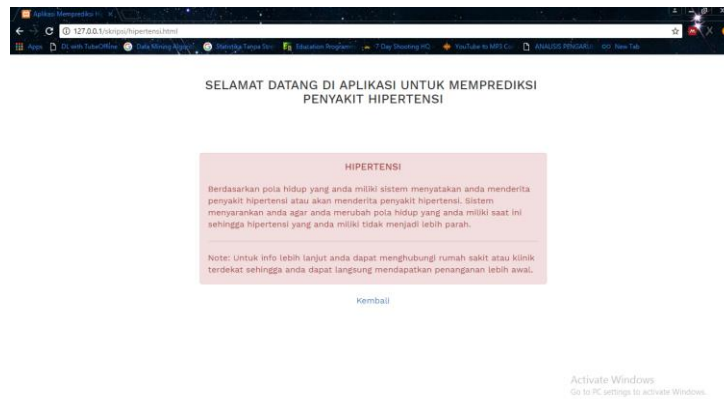
Pada Gambar 10 menunjukkan hasil tampilan dari halaman utama yang akan *user* isi berdasarkan pola hidup yang dimiliki sehingga sistem dapat melakukan prediksi. Pada halaman utama *user* diminta untuk mengisi setiap atribut seperti usia, tinggi badan, berat badan, jenis kelamin dan yang lainnya. Setelah *user* mengisi setiap atribut maka *user* dapat mengklik tombol *process* untuk memproses setiap atribut. Setelah *user* mengklik tombol *process* maka tampilan akan langsung berpindah ke halaman prediksi yang sesuai dengan hasil prediksi berdasarkan perhitungan yang telah dilakukan oleh sistem.



Gambar 10 Tampilan Halaman Utama

### Tampilan Halaman Hasil Prediksi

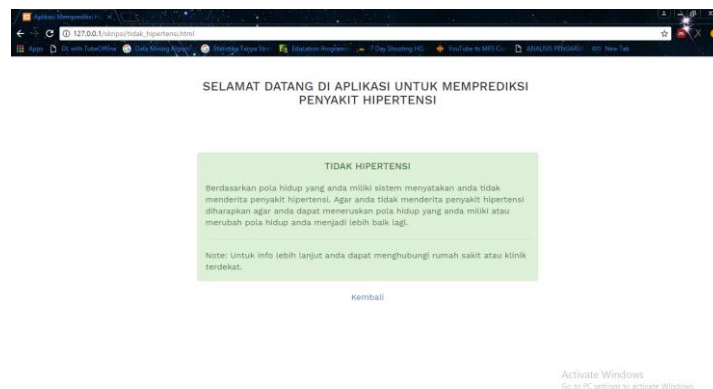
Pada Gambar 11 menunjukkan tampilan dari hasil prediksi yang telah dilakukan oleh sistem dengan pernyataan bahwa *user* menderita penyakit hipertensi atau akan menderita penyakit hipertensi. Pada halaman hasil prediksi yang menyatakan hipertensi berisikan pernyataan bahwa pola hidup yang dimiliki oleh *user* merupakan pola hidup yang harus dirubah agar penyakit hipertensi yang dimiliki tidak menjadi lebih parah. Pada halaman prediksi yang menyatakan hipertensi juga ditampilkan sebuah catatan yang berisikan agar *user* untuk melakukan konsultasi kepada pihak rumah sakit atau klinik terdekat agar apabila hasil prediksi benar *user* dapat langsung menerima penanganan yang lebih awal.



**Gambar 11** Tampilan Halaman Hasil Prediksi Menyatakan Hipertensi

### Halaman Prediksi Menyatakan Tidak Hipertensi

Pada Gambar 12 menunjukkan tampilan dari hasil prediksi yang telah dilakukan oleh sistem dengan pernyataan bahwa tidak menderita penyakit hipertensi. Pada halaman hasil prediksi yang menyatakan tidak hipertensi berisikan pernyataan bahwa pola hidup yang dimiliki oleh *user* merupakan pola hidup yang dapat dipertahankan atau lebih baik lagi apabila *user* meningkatkan kepeduliannya akan kesehatan dengan mengubah pola hidupnya menjadi lebih baik lagi. Pada halaman prediksi yang menyatakan tidak hipertensi juga ditampilkan sebuah catatan yang berisikan agar *user* apabila ingin memperoleh info yang lebih lanjut untuk menghubungi pihak rumah sakit atau klinik terdekat.



**Gambar 12** Tampilan Halaman Hasil Prediksi Menyatakan Tidak Hipertensi

## Kesimpulan dan Saran

### Kesimpulan

Setelah menganalisis sistem dan permasalahan yang terjadi pada pasien yang menderita penyakit hipertensi, maka dapat disimpulkan bahwa:

1. Dengan menggunakan *Data Mining* pasien dapat langsung mengetahui apakah pasien tersebut menderita hipertensi berdasarkan pola hidup yang dimilikinya.
2. Aplikasi yang telah dibangun dapat digunakan karena memiliki akurasi yang cukup tinggi sebesar 88% serta memiliki sensitivitas sebesar 77%, spesifitas 96%.
3. Penggunaan *naïve bayes* untuk memprediksi seseorang menderita penyakit hipertensi dapat digunakan karena memiliki akurasi yang tinggi sebesar 88%.

### Saran

Dari hasil kesimpulan yang diperoleh, maka dapat diberikan beberapa saran untuk pengembangan *data mining* mengenai penyakit hipertensi selanjutnya, yaitu:

1. Melakukan pemodelan klasifikasi lainnya seperti *decision tree*, *rule base classification*, dan lain-lain. Dengan mencari klasifikasi dengan model yang lain sehingga dapat melakukan model *selection* untuk memilih pemodelan yang paling baik hasil tingkat akurasinya.

2. Melakukan pembagian histori menjadi *data training* dan *data testing* dengan teknik *random subsampling* atau *cross validation*. Kedua teknik itu dapat memberikan pembagian *data training* dan *data testing* yang lebih baik dan akurat.

## Referensi

1. Abdullah, Sarini and Taufik Edy Sutanto. 2015. Statistika Tanpa Stres. TransMedia.
2. Anggarwal, Charu C. 2015. Data Mining: The Textbook. illustrate. New York: Springer.
3. Anwar, Firdaus. 2017. "Kemenkes Sebut Kasus Hipertensi Di Indonesia Terus Meningkat." detikHealth. Retrieved (<https://detik.com/health/berita-detikhealth/3503396/kemenkes-sebut-kasus-hipertensi-di-indonesia-terus-meningkat>).
4. Forme, Spark. 2014. PHP: Basics - Professional. Spark Publications.
5. Indriani, Aida. 2014. "Klasifikasi Data Forum Dengan Menggunakan Metode Naive Bayes Clasifier." Seminar Nasional Aplikasi Teknologi Informasi (SNATI).
6. Kalangi, Jane A., Adrian Umboh, and Vivikenanda Pateda. 2015. "HUBUNGAN FAKTOR GENETIK DENGAN TEKANAN DARAH PADA REMAJA." eCI 3(1):66.
7. Karim, Rezaul and Mahedi Kaysar. 2016. Large Scale Machine Learning with Spark. Pact Publishing.
8. Miranda, Eka. 2013. "DATA MINING DENGAN METODE CLASSIFICATION UNTUK TARGETED MARKETING PADA PT. XYZ." Binus University.
9. Prasetyaningrum, Yunita Indah. 2014. Hipertensi Bukan Untuk Ditakuti.
10. Ridwan, Mujib, Hadi Suryono, and M. Sarosa. 2013. "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier." EECCIS 7(1):59.
11. Rinda, Tanto Hariyanto, and Vita Maryah Ardiyani. 2017. "HUBUNGAN KUALITAS TIDUR DENGAN TEKANAN DARAH PADA REMAJA PUTERA DI ASRAMA SANGGAU LANDUNGSARI MALANG." Unitri 2(2):607.
12. Suprihatin, Anggun. 2016. "HUBUNGAN ANTARA KEBIASAAN MEROKOK, AKTIVITAS FISIK, RIWAYAT KELUARGA DENGAN KEJADIAN HIPERTENSI DI WILAYAH KERJA PUSKESMAS NGUTER." Binus University.
13. Yulianora, Frista, Muchammad Hasbi Latif, and Rika Jubel Febriana. 2014. "ANALISIS PERBANDINGAN ALGORITMA DECISION TREE J48 DAN NAIVE BAYES DALAM MENGLASIFIKASIKAN POLA PENYAKIT." Binus University.