**PENDEKATAN SUPERVISED LEARNING UNTUK DIAGNOSA KEHAMILAN**

**Fahira1, Zian Asti Dwiyanti2, Roni Habibi3**

1,2,3Jurusan Teknik Informatika, Universitas Logistik dan Bisnis Internasional

Bandung, Indonesia

e-mail: 1fahira1678@gmail.com, 2ziandwiasti23@gmail.com, 3xxxxx@gmail.com

|  |
| --- |
| Artikel Info : Diterima : 00-00-0000 | Direvisi : 00-00-0000 | Disetujui : 00-00-0000 |

**Abstrak** - Template Paper ini digunakan sebagai panduan untuk penulisan paper (makalah ilmiah) untuk publikasi di Jurnal-jurnal yang dikelola oleh Bina Sarana Informatika. Penulis paper harus mengikuti instruksi yang diberikan di paper ini. Abstrak dibatasi mulai dari 100 - 250 kata dalam Bahasa Indonesia dan Bahasa Inggris (di cetak miring), tidak boleh mengandung persamaan, gambar, dan table. Ukuran huruf untuk abstrak, kata kunci, dan badan makalah adalah 10 point menggunakan font Times New Roman. Pada abstrak memuat prolog dari permasalahan yang diangkat, metode yang diusulkan serta hasil-hasil yang diinginkan (sesuai hipotesa). Jumlah halaman dalam naskah minimal 5 halaman dan maksimal 10 halaman menggunakan ukuran kertas A4 dengan spasi berjarak 1 (single).

Kata Kunci : 3 kata atau frase

***Abstracts*** *- The paper's abstract templates are used as a guide for writing paper for publications in journals Bina Sarana Informatika. The paper writers must follow the instructions given in this paper. Abstract is restricted from 100-250 words in Indonesian Language and English Language (in italics), should not contain equations, pictures, and table. The letter size for abstract, keywords, and body papers is 10 points using Times New Roman font. In the abstract they contain prologue of the raised problem, the proposed method and the desired results (hypothesis). The number of pages in the text is a minimum of 5 and maximum of 10 pages using a paper A4 with a space of 1 (single).*

*Keywords : 3 words or phrases*

**PENDAHULUAN**

Data merupakan sekumpulan fakta atau deskripsi tentang sesuatu yang diperoleh melalui observasi atau mencari sumber tertentu. Informasi tersebut dapat berupa asumsi atau fakta yang belum diolah lebih lanjut, tetapi setelah dianalisis melalui penelitian atau eksperimen, informasi tersebut dapat menjadi lebih kompleks seperti database, informasi, atau bahkan solusi untuk menyelesaikan masalah tertentu. Data juga dibutuhkan dalam dunia bisnis untuk diolah dan disajikan dengan tepat.

Berdasarkan data yang tersedia, Indonesia merupakan salah satu negara dengan jumlah dan pertumbuhan penduduk yang besar. Menurut laporan Kompas.com pada tanggal 3 Februari 2022, jumlah penduduk di Indonesia terus bertambah setiap tahun. Hasil sensus tahun 1980 menunjukkan bahwa jumlah penduduk di Indonesia sebanyak 147 juta jiwa dengan tingkat pertumbuhan sebesar 2,34 persen per tahun. Kondisi pertumbuhan penduduk di Indonesia tentu saja membutuhkan perhatian yang serius untuk mengurangi jumlah kematian. Salah satu cara yang dapat dilakukan adalah dengan memberikan kemudahan bagi ibu hamil untuk mengetahui kondisi kehamilannya.

Profesi baru yang dikenal sebagai data science telah muncul di era ini. Data science merupakan gabungan dari inferensi data, pengembangan algoritma, dan teknologi yang digunakan untuk menyelesaikan masalah analitik yang kompleks. Data science juga termasuk analisis prediktif data untuk menyaring dan menemukan data yang benar agar menghasilkan informasi yang akurat sesuai dengan kondisi sebenarnya. Data science atau ilmu data adalah seperangkat prinsip fundamental yang mendukung dan memandu ekstraksi informasi dan pengetahuan dari data. Ada ratusan jenis metode atau algoritma yang dapat digunakan dalam data science, salah satunya yang akan dibahas oleh penulis dalam buku ini adalah pengklasifikasian.

**TINJAUAN PUSTAKA**

1. **Kehamilan**

Kehamilan adalah istilah yang digunakan untuk menggambarkan periode di mana janin berkembang di dalam rahim atau rahim wanita. Kehamilan biasanya berlangsung sekitar 40 minggu, atau lebih dari 9 bulan, diukur dari periode menstruasi terakhir hingga melahirkan. Penyedia layanan kesehatan mengacu pada tiga segmen kehamilan, yang disebut trimester.

Indikasi awal kehamilan adalah tidak adanya periode menstruasi, mual di pagi hari, dan payudara penuh dan nyeri; tetapi tanda-tanda kehamilan yang positif dan pasti adalah suara detak jantung janin, yang terdengar dengan stetoskop antara minggu ke-16 dan ke-20 kehamilan; gambar ultrasonografi janin yang sedang tumbuh, yang dapat diamati selama kehamilan; dan gerakan janin, yang biasanya terjadi pada minggu ke-18 hingga ke-20 kehamilan.

Selama bulan-bulan awal kehamilan, wanita mungkin memperhatikan bahwa mereka sering buang air kecil, karena tekanan rahim yang membesar pada kandung kemih; merasa lelah dan mengantuk; tidak menyukai makanan yang sebelumnya enak; memiliki rasa berat panggul; dan tunduk pada muntah (yang bisa parah) dan rasa sakit di sisi perut, karena rahim yang tumbuh meregangkan ligamen bundar yang membantu menopangnya, sendiri-sendiri atau bersama-sama. Sebagian besar gejala ini mereda saat kehamilan berlanjut. Tanda dan gejala kehamilan sangat jelas pada minggu ke-12.

1. **Supervised Learning**

Machine learning dapat diartikan sebuah computer yang memiliki kemampuan belajar tanpa deprogram secara explisit. Program tersebut memanfaatkan data untuk membangun model dan mengambil keputusan berdasarkan model yang telah dibangun. Machine learning dapat dikelompokkan berdasarkan bagaimana cara belajar. Sehingga dapat melakukan tugasnya. Pembagian machine learning berdasarkan cara belajarnya dibagi menjadi tiga kelompok yaitu supervised learning, unsupervised learning, dan reinforcement learning.

Secara bahasa, supervised learning merupakan pembelajaran terarah/terawasi. apabila kita analogikan dalam proses pembelajaran, komputer atau mesin akan mempelajari data training yang berisi label. apabila dianalogikan ke murid & pengajar, dimana komputer menjadi murid yang belajar maka pengajar akan meminta murid buat belajar berdasarkan soal yang telah mempunyai solusi & kunci jawaban. Pada dasarnya saat melakukan proses pembelajaran, sistem diberikan data yang berisi solusi atau output yang akan diinginkan atau kasus yang akan diselesaikan oleh sistem.

Selain itu supervised learning juga bisa diartikan sebagai pengklasifikasian. Adapun contoh-contoh algoritma yang termasuk ke dalam kategori supervised learning adalah Linear Regression (Regresi Linear), Logistic Regression (Regresi Logistik), Linear Discriminant Analysis, k-Nearest Neighbors, Support Verctor Machines (SVMs), Random Forest, Decision Tree, dan Naïve Bayes.

Langkah pertama pada proses supervised learning merupakan mengumpulkan data pelatihan berlabel. Label merupakan keluaran & memberikan umpan balik buat algoritma. Asalkan tersedia data yang cukup, langkah selanjutnya adalah membagi data berlabel ini menjadi 3 set: pelatihan, pengujian, & validasi. Algoritma memakai set pelatihan buat menyesuaikan model buat meminimalkan kesalahan. Misalnya set pelatihan mungkin berisi banyak sekali gambar fauna dengan label yang terkait dengan setiap gambar, memungkinkan algoritma untuk membandingkan label yg diprediksi menggunakan label yg benar.

Set validasi terpisah dari set pelatihan & memungkinkan seseorang mengukur kemajuan algoritma pembelajaran secara mandiri. Ukuran ini bisa dipakai untuk memilih titik potong pada algoritma pelatihan untuk menyeimbangkan akurasi model yang dipelajari versus overfitting.

Set tes merupakan set terakhir & dimaksudkan untuk dipakai hanya saat model sudah ditemukan optimal dalam set validasi. Kumpulan ini memberikan evaluasi ‘dunia nyata' dari kinerja model dalam data yang belum pernah dicermati sebelumnya. Data uji adalah semacam 'ujian akhir' untuk model yg sudah mempelajari data pelatihannya secara efektif & bisa digeneralisasikan ke data baru.

1. **Decision Tree**

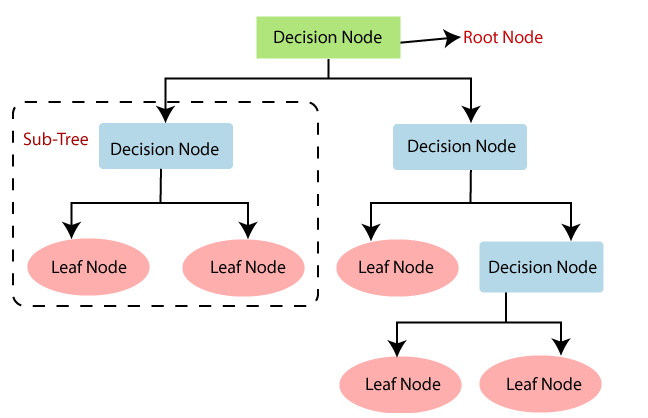
Decision Tree adalah salah satu algoritma yang paling bagus, karena berbagai fitur yang menarik, sederhana, lengkap, tanpa parameter, dan mampu menangani tipe data campuran. Decision Tree adalah teknik supervised learning yang dapat digunakan untuk masalah klasifikasi dan Regresi, tetapi sebagian besar lebih banyak digunakan untuk menyelesaikan masalah Klasifikasi. Ini adalah pengklasifikasi berstruktur pohon, dimana simpul internal mewakili fitur kumpulan data, cabang mewakili aturan keputusan dan setiap simpul daun mewakili hasilnya.

Pada pohon keputusan terdapat dua simpul yaitu simpul keputusan dan simpul daun. Simpul keputusan digunakan untuk membuat keputusan apa pun dan memiliki banyak cabang, sedangkan simpul Daun adalah keluaran dari keputusan tersebut dan tidak mengandung cabang lebih lanjut.

Keputusan atau pengujian dilakukan berdasarkan fitur dari kumpulan data yang diberikan. Ini adalah representasi grafis untuk mendapatkan semua kemungkinan solusi untuk masalah/keputusan berdasarkan kondisi yang diberikan.

Ini disebut pohon keputusan karena, mirip dengan pohon, dimulai dengan simpul akar, yang meluas ke cabang lebih lanjut dan membangun struktur seperti pohon.

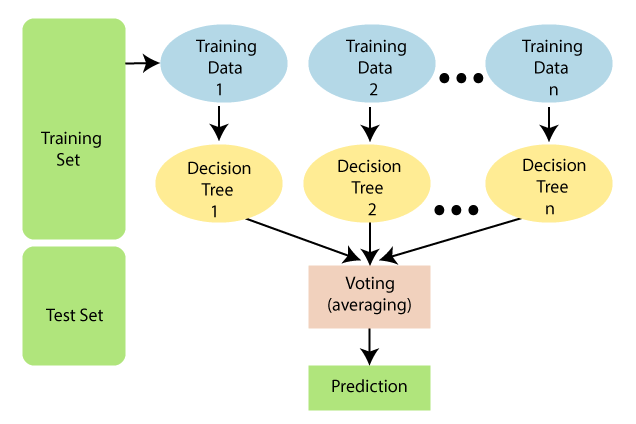
Sebuah pohon keputusan hanya mengajukan sebuah pertanyaan, dan berdasarkan jawaban (Ya/Tidak), selanjutnya membagi pohon menjadi subpohon. Diagram di bawah menjelaskan struktur umum pohon keputusan:



1. **Random Forest**

Random Forest adalah algoritma machine learning populer yang termasuk dalam teknik supervised learning. Algoritma ini dapat digunakan untuk masalah Klasifikasi dan Regresi di ML. Ini didasarkan pada konsep pembelajaran ensemble, yang merupakan proses menggabungkan beberapa pengklasifikasi untuk memecahkan masalah yang kompleks dan untuk meningkatkan kinerja model.

Seperti namanya, "Random Forest (Hutan Acak) adalah pengklasifikasi yang berisi sejumlah pohon keputusan pada berbagai himpunan bagian dari kumpulan data yang diberikan dan mengambil rata-rata untuk meningkatkan akurasi prediktif dari kumpulan data tersebut." Alih-alih mengandalkan satu pohon keputusan, hutan acak mengambil prediksi dari setiap pohon dan berdasarkan suara mayoritas prediksi, dan memprediksi hasil akhir. Semakin banyak jumlah pohon di hutan mengarah pada akurasi yang lebih tinggi dan mencegah masalah overfitting.



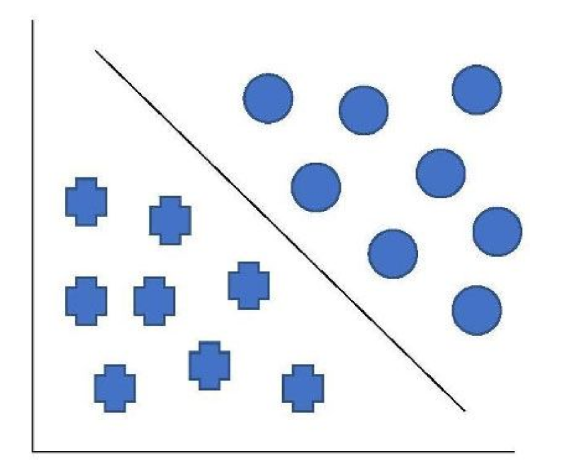
1. ***Classification***

Klasifikasi (Classification) dalam Supervised Learning adalah saat kita dapat menetapkan label khusus untuk pengamatan kita. Contoh klasifikasi adalah memprediksi apakah gejala seseorang akan menyebabkan kanker payudara atau tidak, berdasarkan data sebelumnya yang dikumpulkan dalam gejala dan hasil yang diharapkan (Ya/Tidak).

Gambar dibawah adalah contoh visual bagaimana klasifikasi dapat mengklasifikasikan dua kelas terpisah. menggunakan fungsi linier. Pada gambar ini, kita memiliki dua kelas, salah satunya diwakili oleh Solid Plus (+) dan yang lainnya diwakili oleh lingkaran (o).

Kedua kelas dipisahkan oleh fungsi linier, yaitu garis yang menyatakan bahwa beberapa pengamatan termasuk dalam kategori (+), dan beberapa pengamatan termasuk dalam kategori (o), dan masing-masing memiliki karakteristik tertentu.

Dengan menggunakan pelabelan ini, kita dapat memprediksi label masa depan untuk data yang tidak diketahui dengan menilai fitur dan menerapkan fungsi linier berikut ke fitur untuk menemukan hubungan antara label dan data yang diberikan.

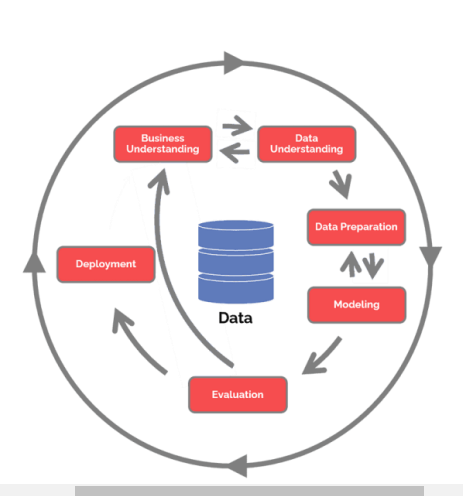


**METODE PENELITIAN**

Dalam penelitian ini, metode yang digunakan adalah analisis data secara kualitatif. Metode ini mengembangkan deskripsi berdasarkan "kejadian" yang diperoleh saat kegiatan lapangan berlangsung. Karenanya, proses pengumpulan data dan analisis data tidak dapat dipisahkan satu sama lain. Sehingga kami menyelesaikan masalah dengan mengacu pada kejadian yang terjadi secara langsung dan proses bisnis di sebuah klinik.

Data dikumpulkan dengan cara melakukan pengamatan langsung terhadap klinik dan menganalisis proses bisnis yang terjadi. Penelitian yang digunakan adalah penelitian deskriptif analitik. Peneliti akan menganalisis data yang terkumpul untuk mencari hubungannya dengan masalah yang berkaitan.

Sebagai referensi, penulis mengikuti tahapan model Cross-Industry Standard Process for Data Mining (CRISP-DM), terdiri dari Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment.



Tahapan – Tahapan CRISP-DM adalah :

1. Business Understanding

Business understanding adalah pemahaman terhadap tujuan bisnis, penilaian situasi dan menerjemahkan tujuan bisnis ke dalam tujuan data mining. Dalam penelitian ini membutuhkan pengetahuan untuk mendapatkan data dari beberapa sumber untuk diagnosa kehamilan dan tidak lupa pula melakukan analisa untuk membuat model prediksi tersebut.

1. Data Understanding

Pada tahap ini dilakukan proses pengumpulan data, kemudian melakukan analisa data serta melakukan evaluasi kualitas data yang digunakan dalam pengerjaan internship. Sumber data yang digunakan terdapat dari berbagai sumber dan penelitian terdahulu, serta melakukan interview ke beberapa expert. Maka sumber literatur banyak didapatkan dari buku, paper atau jurnal, karya ilmiah dan situs-situs penunjang.

1. Data Preparation

Data Preparation atau bisa disebut juga dengan data preprocessing adalah suatu proses/langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas (input yang baik untuk data mining tools).

1. Modelling

Dalam tahap ini, berbagai macam metode pemodelan dipilih dan diterapkan ke dataset yang sudah disiapkan untuk mengatasi kebutuhan bisnis yang sesuai. Adapun metode yang digunakan yaitu Decision Tree sebagai metode utama klasifikasi.

1. Evaluation

Pada tahap terakhir ini, model yang sudah dibuat diuji dan dievaluasi keakuratan. Tahap ini mengukur sejauh mana model yang sudah dipilih memenuhi sasaran-sasaran bisnis dan bila demikian, sejauh manakah itu (apakah perlu lebih banyak model untuk dibuat dan diukur).

1. Deployment

Deployment adalah membuat sistem dari model yang telah dibuat dan dikembangkan ke sistem yang telah ada sebelumnya. Pastinya akan melihat evaluasi terlebih dahulu, jika evaluasi sudah baik, maka deployment dilakukan.

**HASIL DAN PEMBAHASAN**

Pada bagian ini, dijelaskan hasil penelitian dan pada saat yang sama diberikan pembahasan yang komprehensif. Hasil dapat disajikan dalam angka, grafik, tabel, dan lain-lain yang membuat pembaca memahami dengan mudah. Pada bagian ini ditekankan nilai baru dari penelitian yang memuat inovasi, serta implikasinya. Pembahasan dapat dibuat dalam beberapa sub-bab.

1. Sub Bab 1

xx

2. Sub Bab 2

yy

**KESIMPULAN**

Memberikan pernyataan bahwa apa yang diharapkan, seperti yang dinyatakan dalam "Pendahuluan" akhirnya dapat mengakibatkan "Hasil dan Diskusi", sehingga ada kompatibilitas. Selain itu dapat juga ditambahkan prospek pengembangan hasil penelitian dan prospek penerapan studi lanjutan. Hindari Data Statistik dan Sampaikan pula rekomendasi untuk penelitian berikutnya berdasarkan sumber.

**REFERENSI**

Referensi utama adalah jurnal internasional dan jurnal nasional atau buku ber-ISBN. Semua referensi harus yang paling relevan dan merupakan sumber terbaru. Referensi ditulis dalam gaya (*style*) APA. Setiap kutipan harus ditulis dalam urutan Ascending (A-Z). Tidak perlu dikelompokkan berdasarkan buku, jurnal, dan lain lain. Referensi dari website/url internet dapat diperoleh dari sumber yang terpercaya. Referensi harus ditulis menggunakan bantuan *software reference manager* seperti Mendeley atau yang lainnya dengan jumlah minimal 10 referensi.

Contoh penulisan referensi menggunakan style APA. Minimal 10 referensi:

Birtha, Arifudzaki; Soemantri, Maman; Abdian, F. (2010). Aplikasi Sistem Informasi Persediaan Barang pada Perusahaan Export Hasil Laut Berbasis Web. *Transmisi*, *12*(1), 1.

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. *Soft Computing* (Vol. 54). https://doi.org/10.1007/978-3-642-19721-5

Liza, Y., & Yupinti. (2012). Sistem Informasi persediaan barang pada PT.Surya Nusa Bhaktindo Bengkulu. *Media Infotama*, *8*(1), 90–117.

Marcoulides, G. a. (2005). *Discovering Knowledge in Data: an Introduction to Data Mining:Discovering Knowledge in Data: An Introduction to Data Mining*. *Journal of the American Statistical Association* (Vol. 100). https://doi.org/10.1198/jasa.2005.s61

Aggarwal, C. C. (2015). *Data Mining*. Cham: Springer International Publishing. Https://Doi.Org/10.1007/978-3-319-14142-8

Alfiah, F., Susanti, E., Kristinna, J., Ardiansyah, O. R., & Pradipta, D. (2015). Manfaat Menganalisis Pengaruh Sosial Media, 6–8.

Balahur, A., Mihalcea, R., & Montoyo, A. (2014). Computational Approaches To Subjectivity And Sentiment Analysis: Present And Envisaged Methods And Applications. *Computer Speech And Language*, *28*(1), 1–6. Https://Doi.Org/10.1016/J.Csl.2013.09.003

Basari, A. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J. (2013). Opinion Mining Of Movie Review Using Hybrid Method Of Support Vector Machine And Particle Swarm Optimization. *Procedia Engineering*, *53*, 453–462. Https://Doi.Org/10.1016/J.Proeng.2013.02.059

Chou, J. S., Cheng, M. Y., Wu, Y. W., & Pham, A. D. (2014). Optimizing Parameters Of Support Vector Machine Using Fast Messy Genetic Algorithm For Dispute Classification. *Expert Systems With Applications*, *41*(8), 3955–3964. Https://Doi.Org/10.1016/J.Eswa.2013.12.035

Das, T. K., Acharjya, D. P., & Patra, M. R. (2014). Opinion Mining About A Product By Analyzing Public Tweets In Twitter. *2014 International Conference On Computer Communication And Informatics: Ushering In Technologies Of Tomorrow, Today, Iccci 2014*, 3–6. Https://Doi.Org/10.1109/Iccci.2014.6921727

Haddi, E., Liu, X., & Shi, Y. (2013). The Role Of Text Pre-Processing In Sentiment Analysis. *Procedia Computer Science*, *17*, 26–32. Https://Doi.Org/10.1016/J.Procs.2013.05.005

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts And Techniques*. *San Francisco, Ca, Itd: Morgan Kaufmann*. Https://Doi.Org/10.1016/B978-0-12-381479-1.00001-0

Hashimi, H., Hafez, A., & Mathkour, H. (2015). Selection Criteria For Text Mining Approaches. *Computers In Human Behavior*, *51*, 729–733. Https://Doi.Org/10.1016/J.Chb.2014.10.062

Hidayat, Andi Nurul. (2015). Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online Menggunakan Algoritma Support Vector Machine Dan Naive Bayes. *Jurnal Elektronik Sistim Informasi Dan Komputer (Jesik)*, *1*(1), 1–7.