

Aplikasi Pohon Keputusan Dalam Memprediksi Kanker Payudara

Rehagana Kevin Christian Sembiring - 13519117
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13519117@std.stei.itb.ac.id

Abstract—Kanker merupakan salah satu penyakit paling mematikan di dunia. Kanker dapat menyerang berbagai bagian tubuh manusia dan belum ada obat yang dapat menyembuhkan penyakit ini. Dari begitu banyaknya kanker, kanker payudara adalah salah satu jenis kanker yang paling banyak di derita oleh manusia terutama perempuan. Makalah kali ini akan membahas prediksi apakah seseorang menderita kanker payudara berdasarkan data hasil pengecekan kondisi tubuh si penderita menggunakan pohon keputusan.

Keywords—Kanker, Kanker Payudara, video game, Pohon Keputusan

I. PENDAHULUAN

Kanker merupakan nama yang diberikan kepada koleksi dari beberapa penyakit yang berkaitan. Kanker sendiri terjadi pada saat sel kanker sudah berada pada sel normal dan menyebabkan sel tersebut terus bertumbuh tak terkendali. Kanker dapat terjadi dimana saja pada tubuh manusia yang terdiri dari triliunan sel. Normalnya pada tubuh manusia sel akan membelah saat dibutuhkan dan juga ketika sel yang sudah tua dan rusak mati, akan digantikan oleh sel baru. Namun, ketika kanker berkembang, sel yang sudah tua atau rusak tetap hidup dan sel baru akan terbentuk saat tidak dibutuhkan (cancer.gov).

Pada tahun 2017, kanker telah menyebabkan kematian bagi 9.6 juta jiwa dan kanker merupakan penyebab kematian dari setiap 6 kematian di dunia. Hal ini membuat kanker menjadi penyakit paling mematikan setelah penyakit jantung (Ritchie, Hannah 2018). Pada tahun 2020, tercatat bahwa 5 kanker paling mematikan adalah kanker paru-paru, kanker kolorektal, kanker payudara, kanker pancreas dan kanker prostat (Naveed Saleh, 2019).

Kanker payudara sendiri tipe kanker yang dimulai di payudara. Kanker dimulai ketika sel mulai untuk tumbuh tak terkendali. Sel kanker payudara biasanya membentuk tumor yang dapat dilihat dengan sinar x-ray atau dapat dirasakan seperti gumpalan. Kanker payudara umumnya menyerang wanita, tetapi pria juga dapat terkena kanker ini. Seperti yang sudah disebutkan di atas, kanker ini merupakan kanker ke tiga paling mematikan setelah kanker paru-paru dan juga kanker kolorektal karena menyumbang 41760 kematian pada 2019 dari 268600 kasus yang terjadi (Naveed Saleh, 2019).

Dari data diatas dapat dilihat bahwa sebenarnya angka kematian dan kasus yang terjadi akibat kanker payudara berbanding sangat jauh. Kanker payudara sendiri merupakan

salah satu jenis kanker yang dapat di cegah jika sudah di ketahui pada tahap awal. Sudah banyak cara-cara yang dilakukan untuk melakukan preferensi terhadap kanker ini.

Pada makalah ini akan dibahas tentang cara melakukan preferensi ini dengan menggunakan pohon keputusan dari data-data pemeriksaan terhadap orang yang memiliki gejala kanker payudara ini.

II. LANDASAN TEORI

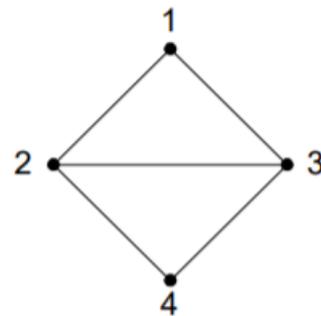
A. Graf

Graf adalah kumpulan simpul dan busur yang secara sistematis dinyatakan sebagai $G = (V, E)$ dimana :

G = Graf

V = Simpul atau Vertex

E = Busur atau edge



Gambar 2.1 Contoh Graf

(Sumber : <https://informatika.stei.itb.ac.id/~rinaldi-munir/Matdis/2020-2021/Graf-2020-Bagian1.pdf>)

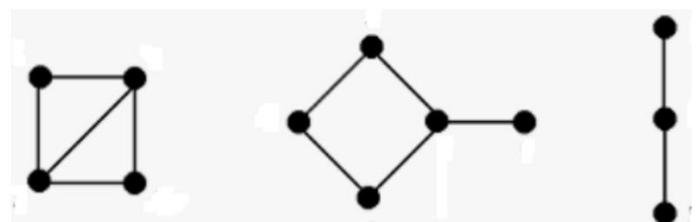
Pada gambar diatas terdapat kumpulan simpul yaitu $\{1,2,3,4\}$ dan juga sisi $\{(1,2), (1,3), (2,3), (2,4), (3,4)\}$.

B. Jenis-jenis Graf

Berdasarkan tidak adanya gelang atau sisi ganda pada graf, graf dibedakan atas :

1. Graf Sederhana

Graf yang tidak mengandung gelang maupun sisi ganda



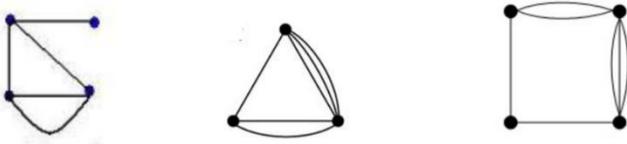
Gambar 2.2 Graf Sederhana

(Sumber : <https://informatika.stei.itb.ac.id/~rinaldi-munir/Matdis/2020-2021/Graf-2020-Bagian1.pdf>)

2. Graf Tak Sederhana

Graf yang mengandung sisi ganda atau gelang, dibedakan lagi menjadi :

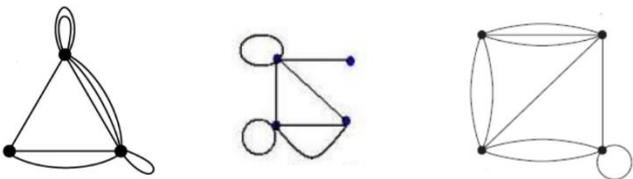
1. Graf ganda → Graf mengandung sisi ganda



Gambar 2.3 Graf Ganda

(Sumber : <https://informatika.stei.itb.ac.id/~rinaldi-munir/Matdis/2020-2021/Graf-2020-Bagian1.pdf>)

2. Graf semu → Graf yang mengandung sisi gelang



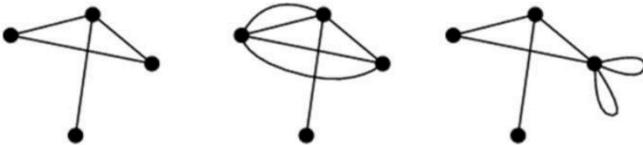
Gambar 2.4 Graf Semu

(Sumber : <https://informatika.stei.itb.ac.id/~rinaldi-munir/Matdis/2020-2021/Graf-2020-Bagian1.pdf>)

Berdasarkan orientasi arah pada sisi graf, graf dibedakan menjadi :

1. Graf tak berarah

Graf yang sisinya tidak mempunyai orientasi arah

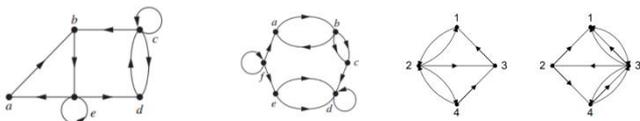


Gambar 2.5 Graf tak Berarah

(Sumber : <https://informatika.stei.itb.ac.id/~rinaldi-munir/Matdis/2020-2021/Graf-2020-Bagian1.pdf>)

2. Graf Berarah

Graf yang setiap sisinya diberikan orientasi arah.



Gambar 2.6 Graf Berarah

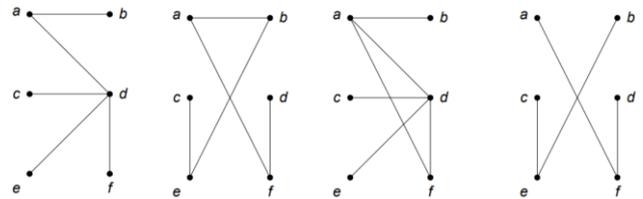
(Sumber : <https://informatika.stei.itb.ac.id/~rinaldi-munir/Matdis/2020-2021/Graf-2020-Bagian1.pdf>)

C. Pohon

Pohon adalah graf tak berarah yang tidak memiliki sirkuit. Pohon memiliki beberapa sifat unik, misalkan $G = (V,E)$ adalah graf tak berarah yang memiliki simpul sebanyak n buah, maka

pernyataan dibawah ini ekuivalen :

- G adalah pohon
- Setiap simpul di dalam G terhubung dengan lintasan tunggal
- G terhubung dan memiliki $n-1$ sisi
- G tidak mengandung sirkuit dan mengandung $n-1$ sirkuit
- Penambahan satu sisi pada graf hanya akan membuat satu sirkuit
- Semua sisi G adalah jembatan



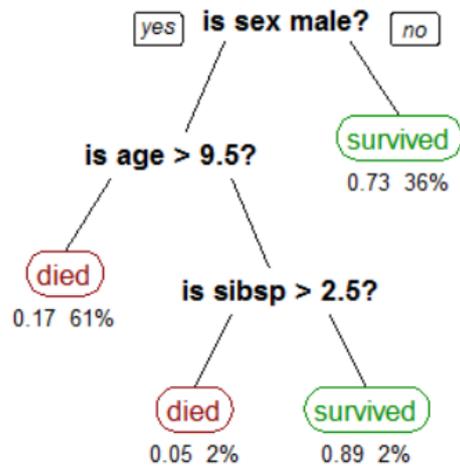
pohon pohon bukan pohon bukan pohon

Gambar 2.7 Pohon

(Sumber : <https://informatika.stei.itb.ac.id/~rinaldi-munir/Matdis/2020-2021/Graf-2020-Bagian1.pdf>)

D. Pohon Keputusan (Decision Tree)

Pohon keputusan adalah salah satu aplikasi dari struktur data pohon yang digunakan untuk mengambil keputusan berdasarkan indikator-indikator yang ada. Dari gambar di bawah, teks yang tebal merepresentasikan kondisi, yang menjadi dasar ke cabang mana pohon akan bergerak. Akhir dari pohon yang tidak membelah lagi disebut leaf/daun yang menjadi keputusan (Prashant Gupta,2017).



Gambar 2.8 Contoh Pohon Keputusan

(sumber : <https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052>)

III. PENERAPAN POHON KEPUTUSAN DALAM MEMREDIKSI KANKER PAYUDARA

A. Deskripsi Data

Sebelum memulai analisis tentang penyakit kanker payudara, maka hal yang harus dilakukan pertama adalah mengumpulkan data-data yang berhubungan dengan penelitian. Dalam kasus ini,

penulis sudah memperoleh data yang sudah terstruktur dan dapat langsung digunakan.

[Grab your reader's attention with a great quote from the document or use this space to emphasize a key point. To place this text box anywhere on the page, just drag it.]

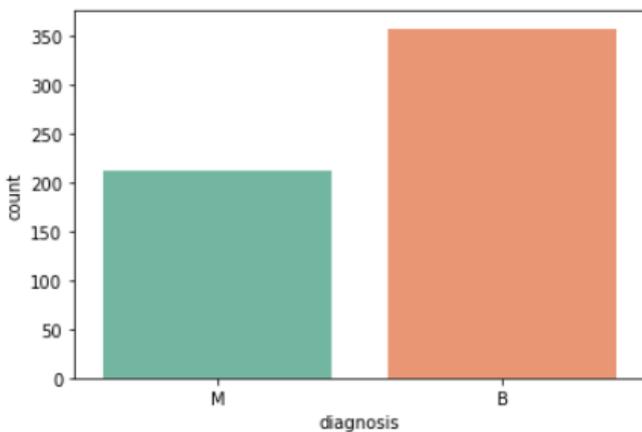
Data yang digunakan berasal dari data hasil komputasi dari gambar yang diperoleh dari fine needle aspirate(FNA) pada massa payudara. Fine needle aspirate itu sendiri adalah teknik untuk mengambil sampel jaringan menggunakan jarum halus.

Setelah dikomputasi data menghasilkan 30 fitur yang dapat dirangkum menjadi :

Id	Nomor unik dari data
Diagnosis	M = Berbahaya B = Tidak Berbahaya
Radius	Radius dari sel
Texture	Standard deviasi dari nilai <i>gray-scale</i>
Perimeter	Keliling sel
Area	Luas sel
Smoothness	Variasi dari Panjang radius
Compactness	(Perimeter ² /area - 1.0)
Concavity	Kecekungan
Concave points	Titik kecekungan
Symmetry	
Fractal Dimension	Aproksimasi garis pinggir - 1

Data sesungguhnya dapat diakses melalui platform Kaggle yaitu : <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>.

Perbandingan antara diagnosa Berbahaya dan Tidak Berbahaya :

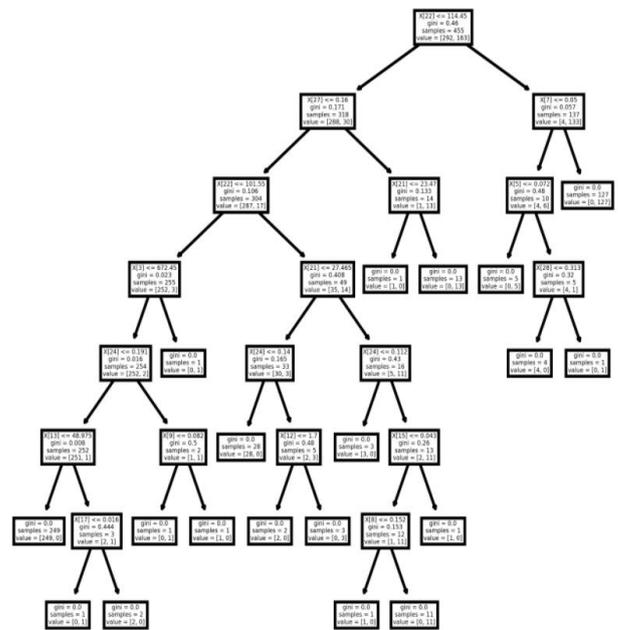


B. Pemodelan Data

Sebelum data diberikan ke model, maka ada beberapa hal yang harus dilakukan terhadap data tersebut. Pertama kolom-kolom yang tidak relevan seperti id dan juga kolom yang tidak bernama akan di drop(dihapus).

Data yang telah dihapus kolom-kolom nya lalu akan dibagi menjadi 2 bagian , yaitu data train dan data test. Data train adalah data yang nantinya akan diberikan ke model dan berisi 80% dari total data, sedangkan data test akan menjadi data untuk menguji akurasi dari model tersebut dan berisi 20% dari total data. Karena kolom label masih berupa string dan string tidak dapat diproses oleh model, maka kita akan mengganti label M menjadi 1 dan label B menjadi 0. Lalu, kolom target yaitu diagnosis akan di pisahkan menjadi sebuah data target.

Setelah data selesai diolah maka model yang disini menggunakan Decision Tree Classifier akan ditraining menggunakan data train dan diuji kebenarannya menggunakan data test. Pohon Keputusan yang diperoleh sebagai berikut :



Gambar 3.1 Pohon Keputusan

C. Hasil Penelitian

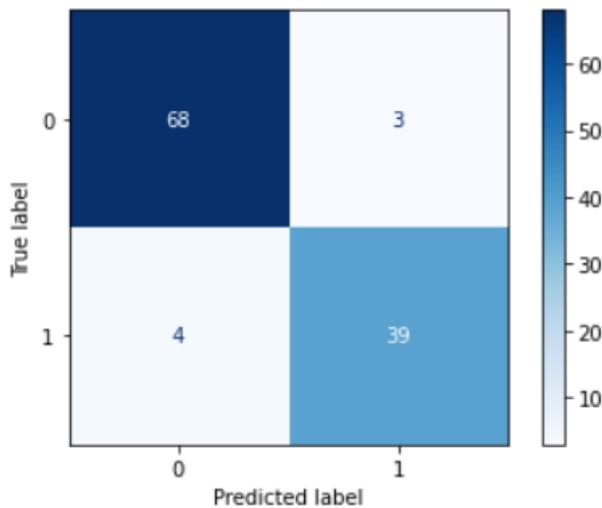
Dari data yang berjumlah 568 buah, dengan menggunakan pohon keputusan diperoleh hasil berikut :

	precision	recall	f1-score	support
0	0.94	0.96	0.95	71
1	0.93	0.91	0.92	43
accuracy			0.94	114
macro avg	0.94	0.93	0.93	114
weighted avg	0.94	0.94	0.94	114

0.9385964912280702

Gambar 3.2 Akurasi Model

Diperoleh akurasi yang lumayan baik dari model dengan menggunakan Decision Tree Classifier yaitu 0.9385 atau sekitar 94%. Dan juga diperoleh confusion matrix sebagai berikut :



Dari confusion matrix tersebut dapat terlihat bahwa error yang rendah yaitu kurang dari 10, hal ini membuktikan bahwa hasil prediksi menggunakan decision tree sudah termasuk bagus. Jumlah error yang hampir sama untuk label 0 dan 1 juga membuktikan bahwa confusion dari model termasuk kecil.

IV. KESIMPULAN

Pohon keputusan adalah salah satu aplikasi dari konsep pohon pada matematika diskrit dan juga merupakan salah satu aplikasi dari konsep pohon yang paling aplikatif. Pohon keputusan banyak sekali digunakan dalam machine learning seperti contoh yang ada di makalah ini.

Banyak juga masalah-masalah dunia nyata seperti penyakit kanker payudara ini dan lainnya yang dapat dibantu penyelesaiannya menggunakan konsep pohon keputusan. Mungkin juga konsep ini dapat digunakan untuk membantu mengatasi masalah COVID-19 saat ini

V. PENUTUP

Pertama-tama puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas berkat dan rahmatNya penulis dapat menyelesaikan makalah ini sebagai tugas mata kuliah Matematika Diskrit. Penulis juga mengucapkan terima kasih kepada kedua orang tua dan teman-teman yang selalu mendukung secara moral maupun doa. Kemudian penulis juga turut mengucapkan terima kasih kepada para dosen pengampu mata kuliah Matematika Diskrit ini atas ilmu dan bimbingan yang diberikan. Terakhir, penulis ingin memohon maaf apabila terdapat banyak kekurangan dan kesalahan kata pada makalah ini. Penulis juga berharap makalah ini dapat digunakan sebaik-baiknya untuk kebutuhan-kebutuhan kedepannya.

REFERENCES

- [1] <https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2020-2021/Pohon-2020-Bag1.pdf> – Pohon(Bagian 1).
- [2] <https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2020-2021/Pohon-2020-Bag2.pdf> – Pohon(Bagian 2).
- [3] <https://www.kaggle.com/aryansrivatsava/logisticregression-vs-decisiontree-vs-randomfores> – Fork of predicting breast cancer using decision tree
- [4] <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data> - Breast Cancer Wisconsin (Diagnostic) Data Set.
- [5] <https://towardsdatascience.com/visualizing-decision-trees-with-python-scikit-learn-graphviz-matplotlib-1c50b4aa68dc> - Visualizing Decision Trees with Python.
- [6] <https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052> - Decision Trees in Machine Learning
- [7] <https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2019-2020/Makalah2019/13518138.pdf> - Penerapan Pohon Keputusan Sebagai Pendeteksi Penipuan pada Transaksi Online (William, 2019).

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Medan, 5 Desember 2020

Nama dan NIM