Lembar Desain COVER BUKU

(desain cover harus original)

PADA COVER Harus Tercantum :

Nama Penulis (nama Pembimbing + Nama Mahasiswa)

+ Logo Penerbit Di bagian bawah

Resa Rianti

Nur Tri Ramadhanti Adiningrum

Cahyo Prianto, S.pd., M.T., CDSP, SFPC

**PENDEKATAN MACHINE LEARNING DALAM PREDIKSI**

**KANKER PAYUDARA MENGGUNAKAN SUPPORT**

**VECTOR MACHINE**

**PENDEKATAN MACHINE LEARNING DALAM PREDIKSI**

**KANKER PAYUDARA MENGGUNAKAN SUPPORT**

**VECTOR MACHINE**

Resa Rianti

Nur Tri Ramadhanti Adiningrum

Cahyo Prianto, S.Pd., M.T., CDSP, SFPC

**CANTUMKAN**

**Logo Penerbit Disini**

**PENDEKATAN MACHINE LEARNING DALAM PREDIKSI KANKER PAYUDARA MENGGUNAKAN SUPPORT VECTOR MACHINE**

©TULISKAN NAMA PENERBIT DISINI

Penulis:

……..

Editor:

(Nama Penguji Sidang)

Cetakan Pertama: Isi dengan Bulan saat upload buku

Cover: Tim Penyusun

Tata Letak: Tim Kreatif Penerbit

Hak Cipta 2023, pada Penulis. Diterbitkan pertama kali oleh:

**ISI NAMA PENERBIT**

ISI ALAMAT PENERBIT

Website: [WEBSITE](http://www.rcipress.rcipublisher.org/) PENERBIT

E-mail: [EMAIL](mailto:rumahcemerlangindonesia@gmail.com) PENERBIT

Copyright © 2023 by NAMA PENERBIT

All Right Reserved

- Cet. I –: NAMA PENERBIT, TAHUN TERBIT

Dimensi : 14,8 x 21 cm

ISBN: KOSONGKAN DULU

Hak cipta dilindungi undang-undang

Dilarang memperbanyak buku ini dalam bentuk dan dengan cara apapun tanpa izin tertulis dari penulis dan penerbit

Undang-undang No.19 Tahun 2002 Tentang

**Hak Cipta Pasal 72**

Undang-undang No.19 Tahun 2002 Tentang Hak Cipta Pasal 72

Barang siapa dengan sengaja melanggar dan tanpa hak melakukan perbuatan sebagaimana dimaksud dalam pasal ayat (1) atau pasal 49 ayat (1) dan ayat (2) dipidana dengan pidana penjara masing-masing paling sedikit 1 (satu) bulan dan/atau denda paling sedikit Rp.1.000.000,00 (satu juta rupiah), atau pidana penjara paling lama 7 (tujuh) tahun dan/atau denda paling banyak Rp.5.000.000.000,00 (lima miliar rupiah).

Barang siapa dengan sengaja menyiarkan, memamerkan, mengedarkan, atau menjual kepada umum suatu ciptaan atau barang hasil pelanggaran hak cipta terkait sebagai dimaksud pada ayat (1) dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/atau denda paling banyak Rp.500.000.000,00 (lima ratus juta rupiah).

# KATA PENGANTAR

Segala puji bagi Allah, Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan buku ini. Tak lupa juga mengucapkan salawat serta salam semoga senantiasa tercurahkan kepada Nabi Besar Muhammad SAW, karena berkat beliau, kita mampu keluar dari kegelapan menuju jalan yang lebih terang.

Kami ucapkan juga rasa terima kasih kami kepada pihak-pihak yang mendukung lancarnya buku ini mulai dari proses penulisan hingga proses cetak, yaitu orang tua kami, rekan-rekan kami, penerbit, dan masih banyak lagi yang tidak bisa kami sebutkan satu per satu.

Adapun, buku kami yang berjudul Bahan Ajar: Pendekatan machine learning dalam prediksi kanker payudara menggunakan support vector machine ini telah selesai kami buat secara semaksimal dan sebaik mungkin agar menjadi manfaat bagi pembaca yang membutuhkan informasi dan pengetahuan mengenai bagaimana memprediksi kanker payudara.

Dalam buku ini, tertulis bagaimana …..

yang menjadi alternatif pegangan bagi mahasiswa dan…

Kami sadar, masih banyak luput dan kekeliruan yang tentu saja jauh dari sempurna tentang buku ini. Oleh sebab itu, kami mohon agar pembaca memberi kritik dan juga saran terhadap karya buku ajar ini agar kami dapat terus meningkatkan kualitas buku.

Demikian buku ajar ini kami buat, dengan harapan agar pembaca dapat memahami informasi dan juga mendapatkan wawasan mengenai bidang sistem informasi manajemen serta dapat bermanfaat bagi masyarakat dalam arti luas. Terima kasih.

Bandung, 02 November 2022

**Penulis**

# DAFTAR ISI

[KATA PENGANTAR vii](#_Toc120884638)

[DAFTAR ISI viii](#_Toc120884639)

[DAFTAR GAMBAR x](#_Toc120884640)

[DAFTAR TABEL xii](#_Toc120884641)

[BAB 1 Apa Itu Kanker Payudara? 1](#_Toc120884642)

[1.1 Apa Itu Kanker? 1](#_Toc120884643)

[1.2 Apa Itu Payudara? 2](#_Toc120884644)

[1.3 Apa itu Kanker Payudara? 3](#_Toc120884645)

[1.4 Penyebab Kanker Payudara 3](#_Toc120884646)

[1.5 Gejala Kanker Payudara 4](#_Toc120884647)

[1.6 Cara Pengobatan Kanker Payudara 5](#_Toc120884648)

[1.7 Cara Mencegah Kanker Payudara 6](#_Toc120884649)

[1.8 Kasus Yang Pernah Terjadi 6](#_Toc120884650)

[BAB 2 Metodologi Pembuatan Aplikasi 8](#_Toc120884651)

[2.1 Algoritma Machine Learning Yang Digunakan 8](#_Toc120884652)

[2.1.1 Support Vector Machine 8](#_Toc120884653)

[2.2 Dataset 11](#_Toc120884654)

[2.2.1 Definisi Dataset 11](#_Toc120884655)

[2.2.2 Tipe Data 11](#_Toc120884656)

[2.2.3 Jenis Dataset 12](#_Toc120884657)

[2.2.4 Kualitas pada dataset 14](#_Toc120884658)

[2.2.5 Dataset Yang Digunakan 17](#_Toc120884659)

[2.3 Metodologi Pengolahan Data 22](#_Toc120884660)

[BAB 3 Apa Saja Perangkat Yang Digunakan Untuk Membuat Aplikasi? 24](#_Toc120884661)

[3.1 Bahasa Pemrograman Yang Digunakan 25](#_Toc120884662)

[3.1.1 Python 25](#_Toc120884663)

[3.1.2 HTML 26](#_Toc120884664)

[3.1.3 CSS 27](#_Toc120884665)

[3.1.4 Bootstrap 28](#_Toc120884666)

[3.1.5 Django 28](#_Toc120884667)

[3.2 Penyimpanan Yang Digunakan 29](#_Toc120884668)

[3.2.1 Database MySQL 29](#_Toc120884669)

[3.3 Perangkat Lunak Yang Dibutuhkan 31](#_Toc120884670)

[3.3.1 Visual Studio Code 31](#_Toc120884671)

[3.3.2 XAMPP 32](#_Toc120884672)

[3.3.3 Jupyter Notebook 33](#_Toc120884673)

[3.3.4 Tableau 35](#_Toc120884674)

[3.3.5 Anaconda 36](#_Toc120884675)

[3.3.6 Kaggle 36](#_Toc120884676)

[3.4 Spesifikasi Perangkat Keras 37](#_Toc120884677)

[DAFTAR PUSTAKA 39](#_Toc120884678)

# DAFTAR GAMBAR

BUATLAH DAFTAR ISI DISINI

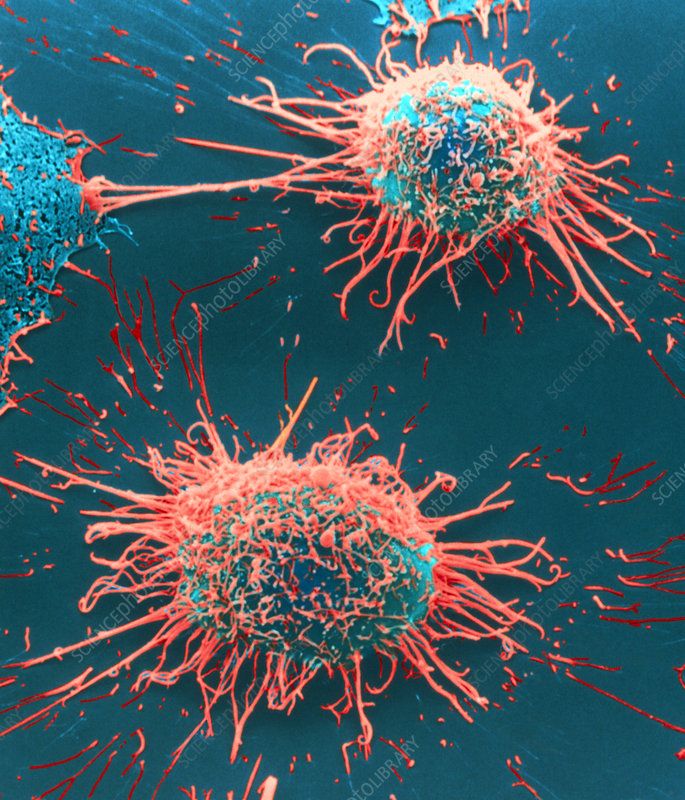
**2 |**

# DAFTAR TABEL

BUATLAH DAFTAR ISI DISINI

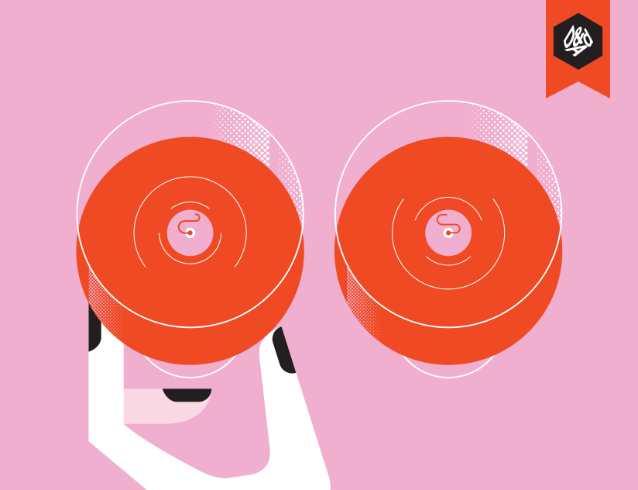
# BAB 1 Apa Itu Kanker Payudara?

## Apa Itu Kanker?



Kanker adalah sel yang tumbuh terus-menerus secara tidak terkendali, tidak terbatas, dan tidak normal (abnormal). Pertumbuhan sel-sel kanker tidak terkoordinasi dengan jaringan lain sehingga berbahaya bagi tubuh. Konteks lain menyebutkan bahwa kanker merupakan tumor ganas yang mengalami pertumbuhan abnormal yang tidak diketahui secara pasti penyebabnya. Dalam kondisi normal, sel hanya akan berkembang biak dengan cara membelah diri jika ada yang mati atau rusak. Sel kanker akan terus mengalami perkembangbiakan meskipun tidak dibutuhkan oleh tubuh. Sel kanker merusak jaringan sel lain yang normal dan menyebar ke organ tubuh lain melalui jaringan ikat, darah, saraf dan jaringan penunjang organ tubuh. Kanker termasuk penyakit yang tidak menular. Penyakit ini timbul akibat kondisi fisik yang tidak normal dan pola hidup yang tidak sehat.

## Apa Itu Payudara?



Payudara dikenal dengan nama mammae. Payudara merupakan organ tubuh bagian atas dada dari spesies mamalia berjenis kelamin betina, termasuk perempuan dari golongan manusia. Biasanya, organ tubuh ini digunakan untuk menyalurkan air susu (ASI) pada bayi mamalia atau manusia yang baru lahir. Keberadaannya yang menonjol ditubuh bagian atas, memiliki nilai estetika yang membuat tubuh Wanita lebih menarik . Fungsi dari payudara itu sendiri adalah sebagai produsen ASI dari lobulus-lobulus, yaitu kelenjar yang dapat menghasilkan ASI dan tubulus atau ductus yang mengantarkan ASI dari kelenjar hingga putting susu (nipple). Maka dari itulahn kelenjar mammae ini menjadi nilai pembeda dari semua mamalia. Bentuknya pun seperti kerucut, namun tidak semua Wanita memiliki ukuran yang sama.

Jadi, payudara merupakan suatu organ vital yang memproduksi ASI dalam jumlah besar.

## Apa itu Kanker Payudara?

Kanker payudara adalah kanker yang paling banyak ditemukan pada wanita dan angka kematiannya masih menempati posisi kedua diantara kanker lainnya[1]. Kanker payudara atau *Carsinoma Mammae* adalah penyakit di mana sel-sel kanker terbentuk di jaringan payudara. Kanker dapat terbentuk di kelenjar yang menghasilkan susu (lobulus) atau saluran yang membawa susu dari kelenjar ke puting (duktus). Kanker juga dapat terbentuk di jaringan lemak atau ikat di dalam payudara. Kanker payudara adalah penyebab kematian kedua pada wanita [3] . Pasien kanker payudara biasanya memiliki ciri psikologis seperti gegar otak, kecemasan, depresi, dan panik. Penyebab kanker payudara masih belum diketahui, namun diyakini multifaktorial. dan berbagai inisiatif lainnya sedang berlangsung. Upaya pencegahan, termasuk pendidikan masyarakat dan remediasi, diperlukan untuk mengatasi masalah kanker payudara, disesuaikan dengan kebutuhan individu. Mencegah kanker payudara sejak dini, yang juga dikenal dengan SADARI (Pemeriksaan Payudara Sendiri), merupakan hal penting yang harus dilakukan setiap orang secara mandiri. Namun, jika kanker payudara terdiagnosis, diperlukan pemeriksaan lebih lanjut oleh dokter.

## Penyebab Kanker Payudara

Faktor penyebab yang memungkinkan untuk timbulnya kanker payudara antara lain :

1. Mengkonsumsi makanan yang mengandung lemak berlebih dan berprotein tinggi, tetapi serat terlalu banyak. Dengan mengkonsumsi makanan seperti itu mengandung zat karsinogen yang dapat merangsang pertumbuhan sel kanker.
2. Pada usia muda telah mengkonsumsi pil kontrasepsi. Penelitian telah membuktikan jika usia dini (remaja) sudah meminum oral (pil) sangat tinggi risikonya terserang kanker payudara
3. Keturunan dari anggota keluarga yang pernah terkena kanker payudara.
4. Mengkonsumsi alkohol terlalu banyak.
5. Wanita merokok memiliki risiko paling besar dapat terserang kanker payudara karena kontaminasi senyawa kimia yang terlalu berlebihan

## Gejala Kanker Payudara

Pada stadium awal, penyakit ini tidak dapat menunjukkan gejala tertentu. Oleh karena itu, sangat penting untuk melakukan pemeriksaan payudara sendiri setiap bulan, sepuluh hari setelah akhir periode perempuan. Rasakan rabaan searah jarum jam dengan hati-hati untuk mendeteksi adanya benjolan atau perubahan pada payudara.

Beberapa gejala kanker payudara yang bisa dialami oleh penderitanya adalah:

1. Benjolan atau pengerasan payudara yang berbeda dari jaringan di sekitarnya.
2. Darah keluar dari puting.
3. Kemerahan atau penggunaan pori-pori kulit payudara yang menyerupai kulit jeruk.
4. Nyeri dan bengkak pada payudara.
5. Pengelupasan kulit di sekitar puting.
6. Perubahan pada kulit payudara, seperti depresi.
7. Perubahan ukuran, bentuk, atau tampilan payudara.
8. Puting ditarik (ditarik atau dibalik) ke dalam.
9. Benjolan atau pembengkakan di ketiak bagian bawah.

Ada banyak penyebab munculnya benjolan di payudara, yang sebagian besar bukan kanker. 90% dari massa payudara tidak bersifat kanker. Gangguan payudara non-kanker termasuk massa jinak seperti fibroadenoma dan kista serta infeksi.

## Cara Pengobatan Kanker Payudara

Perawatan untuk penyakit ini tergantung pada jenis kanker yang Anda derita, stadium kanker, ukuran massa, dan sensitivitas sel kanker terhadap hormon. Pengobatan kanker payudara bisa sangat efektif, mencapai probabilitas kelangsungan hidup 90% atau lebih tinggi, terutama ketika penyakit ini diidentifikasi lebih awal.  Dokter akan menentukan terapi yang paling tepat untuk kondisi pasien berdasarkan hal-hal tersebut. Ada beberapa pilihan pengobatan untuk kanker payudara, antara lain:

1. Pembedahan, yang meliputi pengangkatan kanker atau benjolan (lumpektomi), pengangkatan seluruh payudara (mastektomi), pengangkatan sejumlah kelenjar getah bening (biopsi simpul sentinel), atau pengangkatan beberapa kelenjar getah bening (diseksi kelenjar getah bening aksila).
2. Radioterapi, yang menggunakan sinar-X dan proton untuk membunuh sel kanker.
3. Kemoterapi, yaitu dilakukan dengan menggunakan obat-obatan tertentu untuk membunuh sel kanker.
4. Terapi hormonal, untuk mencegah sel kanker mendapatkan hormon yang mereka butuhkan untuk tumbuh.
5. Terapi biologis, yang bekerja dengan sistem kekebalan pasien untuk membantu melawan sel kanker, atau untuk mengontrol efek samping dari perawatan kanker lainnya.
6. Terapi radiasi menggunakan sinar berenergi tinggi untuk membunuh sel kanker.

## Cara Mencegah Kanker Payudara

Kanker payudara memang sangat mengerikan. Penyakit ini bisa disebabkan oleh banyak faktor seperti faktor keturunan, faktor lingkungan, dan gaya hidup yang tidak sehat. Untuk mencegah kanker payudara, sebaiknya periksa kondisi payudara secara rutin dengan cara menyentuh dan memijatnya. Cara ini cukup efektif untuk mendeteksi pertumbuhan sel kanker payudara sebelum mencapai stadium lanjut. Kanker ini juga dapat dicegah dengan cara-cara yang dijelaskan di bawah ini.

1. Berolahraga minimal 30 menit sehari
2. Berhenti merokok
3. Konsumsi buah dan sayur secara teratur
4. Konsumsi produk kedelai
5. Asupan serat makanan
6. Hindari alkohol
7. Kurangi asupan gula
8. Hindari konsumsi lemak jahat
9. Mengetahui riwayat orang tua
10. Hindari memakan daging merah
11. Mengkonsumsi obat-obatan
12. Operasi

## Kasus Yang Pernah Terjadi

Diperkirakan ada 18,1 juta kasus kanker di seluruh dunia pada tahun 2020. Dari jumlah tersebut, 9,3 juta kasus terjadi pada pria dan 8,8 juta pada wanita. Kanker yang paling umum secara global adalah kanker payudara dengan total 2.261.419 kasus atau sebanyak 12.5% dari total persentase berbagai kanker dan merupakan kanker yang menempati posisi pertama kasus kanker terbayak di dunia.

Pada tahun 2020, terdapat 2,3 juta wanita yang didiagnosis menderita kanker payudara dan 685.000 kematian secara global. Pada akhir tahun 2020, terdapat 7,8 juta wanita hidup yang didiagnosis menderita kanker payudara dalam 5 tahun terakhir, kanker paling umum di dunia. Ada lebih banyak kehidupan yang berakhir akibat oleh wanita karena kanker payudara secara global daripada jenis kanker lainnya. Kanker payudara terjadi di setiap negara di dunia pada wanita pada usia berapa pun setelah pubertas tetapi dengan tingkat yang meningkat di kemudian hari.

Kelangsungan hidup kanker payudara selama minimal 5 tahun setelah diagnosis berkisar dari lebih dari 90% di negara-negara berpenghasilan tinggi, hingga 66% di India dan 40% di Afrika Selatan.

Deteksi dan pengobatan dini telah terbukti berhasil di banyak negara maju dan harus diterapkan di negara-negara dengan sumber daya terbatas di mana beberapa alat standar tersedia. Angka kematian akibat kanker payudara standar usia di negara-negara berpenghasilan tinggi turun 40% antara tahun 1980-an dan 2020. Negara-negara yang berhasil menurunkan angka kematian akibat kanker payudara telah mampu mencapai penurunan angka kematian akibat kanker payudara tahunan sebesar 2-4% per tahun. Jika penurunan kematian tahunan sebesar 2,5% per tahun terjadi di seluruh dunia, 2,5 juta kematian akibat kanker payudara akan dapat dihindari antara tahun 2020 dan 2040.

# BAB 2 Metodologi Pembuatan Aplikasi

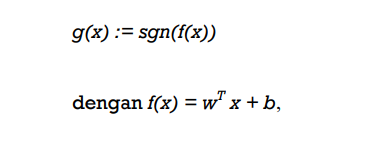
## Algoritma Machine Learning Yang Digunakan

### **Support Vector Machine**

SVM adalah suatu aturan algoritma pembelajaran mesin yang diawasi digunakan untuk setiap klasifikasi dan tantangan regresi. Dimanfaatkan dalam masalah klasifikasi, dalam algoritma SVM, mengarah untuk memplot setiap item pengetahuan sebagai batas tertentu dalam ruang n-dimensi (di mana n adalah banyak fitur yang dimiliki) dengan nilai setiap fitur menjadi nilai koordinat tertentu. Selanjutnya, kami mengkategorikan dengan menemukan level super yang membedakan 2 kategori (kelas) dengan sangat baik. Nilai default cenderung digunakan untuk semua kunci parameter kecuali parameter kernel yang dipilih linier, dan parameter random\_state kami menetapkan nilainya sampai 40, untuk berhasil dalam hasil yang diperlukan [ ] .

Teknik ini menarik orang dalam bidang data mining maupun machine learning karena performansinya yang meyakinkan dalam memprediksi kelas suatu data baru.

Kita akan memulai pembahasan dengan kasus klasifikasi yang secara linier bisa dipisahkan. Dalam hal ini fungsi pemisah yang kita cari adalah fungsi linier. Fungsi ini bisa didefinisikan sebagai :



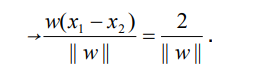
x, w Rn and bR. Masalah klasifikasi ini bisa dirumuskan sebagai berikut: kita ingin menemukan set parameter (w, b) sehingga sgn(f(xi)) =sgn(< w,x > +b) = yi untuk semua i. Fungsi sgn digunakan untuk mengelompokkan semua nilai di atas 0 menjadi +1 dan dibawah 0 menjadi -1. Dalam teknik ini kita berusaha menemukan fungsi pemisah (klasifier/hyperplane) terbaik diantara fungsi yang tidak terbatas jumlahnya untuk memisahkan dua macam obyek. Hyperplane terbaik adalah hyperplane yang terletak di tengah-tengah antara dua set obyek dari dua kelas. Mencari hyperplane terbaik ekuivalen dengan memaksimalkan margin atau jarak antara dua set obyek dari kelas yang berbeda. Jika wx1 + b = +1 adalah hyperplane-pendukung (supporting hyperplane) dari kelas +1, dan wx2 + b = -1 adalah hyperplanependukung dari kelas -1, margin antara dua kelas dapat dihitung dengan mencari jarak antara kedua hyperplane-pendukung dari kedua kelas. Secara spesifik, margin dihitung dengan cara berikut :

wx1 + b = +1

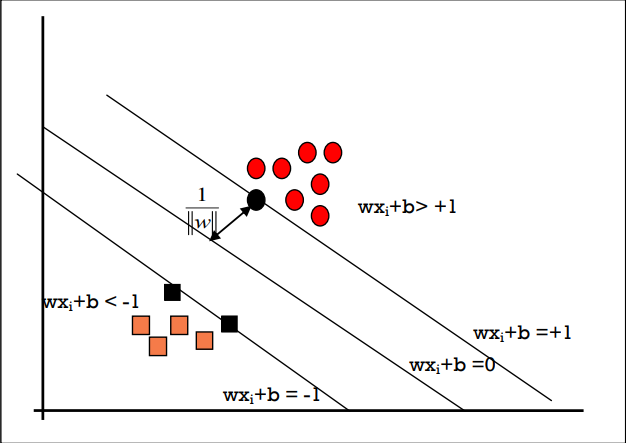
wx2 + b = -1

------------------- -

w(x1 - x2) = 2

→

Gambar tersebut memperlihatkan bagaimana SVM bekerja untuk menemukan suatu fungsi pemisah dengan margin yang maksimal. Untuk membuktikan bahwa memaksimalkan margin antara dua set obyek akan meningkatkan probabilitas pengelompokkan secara benar dari data testing. Pada dasarnya jumlah fungsi pemisah ini tidak terbatas banyaknya. Misalkan dari jumlah yang tidak terbatas ini kita ambil dua saja, yaitu f1(x) and f2(x) (lihat gambar 8.2). Fungsi f1 mempunyai margin yang lebih besar dari pada fungsi f2. Setelah menemukan dua fungsi ini, suatu data baru masuk dengan keluaran −1. Kita harus mengelompokkan apakah data ini ada dalam kelas −1 atau +1 menggunakan fungsi pemisah yang sudah kita temukan.

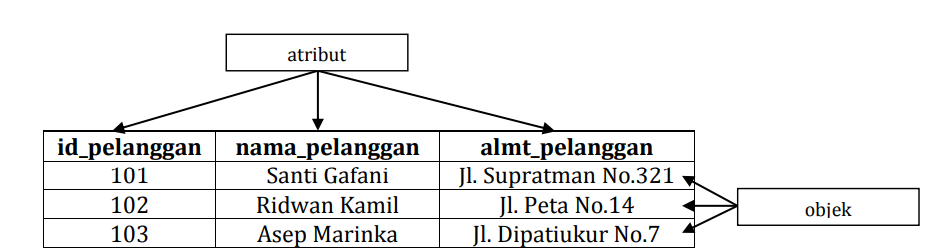


Mencari fungsi pemisah yang optimal untuk obyek yang bisa dipisahkan secara linier Dengan menggunakan f1, kita akan kelompokkan data baru ini di kelas −1 yang berarti kita benar mengelompokkannya. Sekarang kita coba gunakan f2, kita akan menempatkannya di kelas +1 yang berarti salah. Dari contoh sederhana ini kita lihat bahwa memperbesar margin bisa meningkatkan probabilitas pengelompokkan suatu data secara benar. [ ]

## Dataset

### **Definisi Dataset**

Definisi Set Data Set data (data set/himpunan data) merupakan kumpulan objek dan atributnya. Nama lain dari objek yang sering digunakan diantaranya record, point, vector, pattern, event, observation, case, sample, instance, entitas. Objek digambarkan dengan sejumlah atribut yang menerangkan sifat atau karakteristik dari objek tersebut. Atribut juga sering disebut variabel, field, fitur, atau dimensi. Atribut adalah sifat/properti/karakteristik objek yang nilainya bisa bermacam-macam dari satu objek dengan objek lainnya, dari satu waktu ke waktu yang lainnya. Sebagai contoh seorang pelanggan merupakan objek, dimana objek pelanggan tersebut memiliki beberapa atribut seperti id pelanggan, nama, alamat dan lain-lain. Setiap pelanggan memungkinkan memiliki nilai atribut yang berbeda dengan pelanggan lainnya, serta memungkinkan perubahan nilai atribut dari waktu ke waktu. [ ]



### **Tipe Data**

Tipe atribut dapat dibedakan dari nilai beserta sifatnya. Ada empat sifat yang dimiliki atribut secara umum, yaitu:

* + 1. Pembeda (distinctness): = dan ≠
    2. Urutan (order): <, >, ≤, ≥
    3. Penjumlahan, Pengurangan (addition): + dan –
    4. Perkalian, Pembagian (multiplication): \* dan /

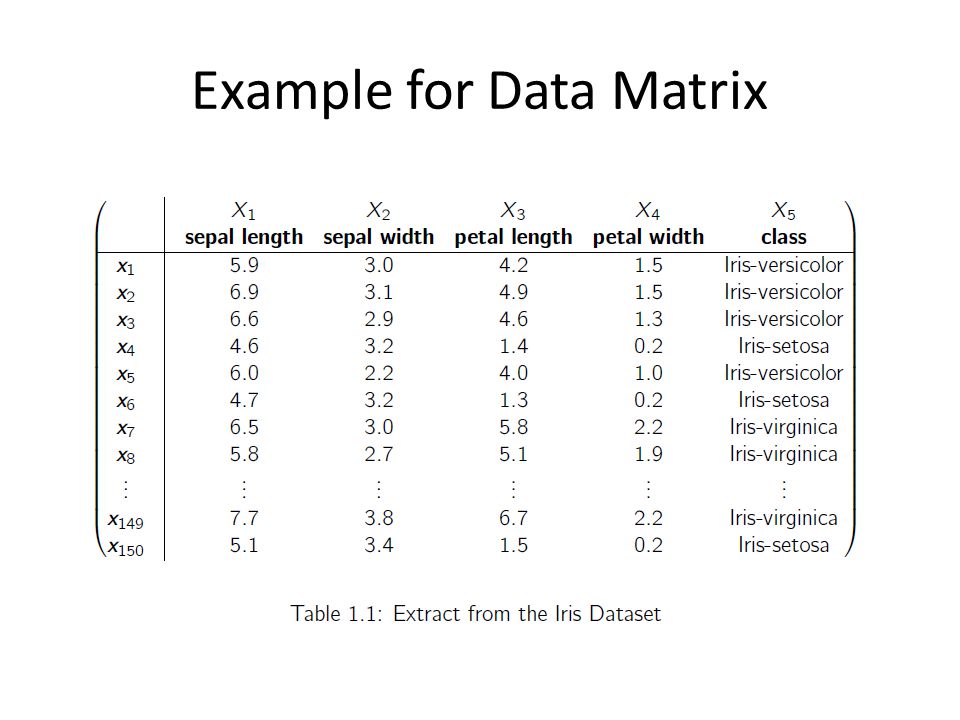
Umumnya tipe atribut ini ada dua yaitu kategori (kualitatif) dan numerik (kuantitatif). Dari kedua tipe tersebut dibagi lagi menjadi beberapa sub tipe yang disesuaikan dengan sifat nilai yang dimilikinya.

### **Jenis Dataset**

Jenis data dapat dibedakan menjadi tiga kelompok, yaitu data record, data berbasis grafik (graph data), dan data terurut (ordered data). Pada perkuliahan ini kita hanya akan menggunakan data record untuk proses data mining. Kebanyakan metode data mining mengasumsikan bahwa set data yang diproses adalah kumpulan baris data (record/entries/objects), dimana setiap barisnya terdiri atas sejumlah fitur (atribut) yang tetap. Dalam set data berbentuk data record, tidak ada hubungan antara baris data dengan baris data yang satu dengan baris data yang lainnya dan juga dengan set data yang lain. Setiap baris data berdiri sendiri sebagai sebuah data individu. Dalam sistem basis data, umumnya ada sejumlah tabel yang saling berhubungan menggunakan suatu kunci (kunci utama, kunci tamu). Akan tetapi dalam set data record, diasumsikan bahwa hanya ada satu tabel yang berisi sejumlah baris data. Oleh karena itu, biasanya set data yang diolah dalam data mining adalah keluaran dari sistem data warehouse yang menggunakan query untuk melakukan pengambilan data dari sejumlah tabel dalam sistem basis data. Ada beberapa contoh set data yang masuk dalam jenis data record, diantaranya data matrik (matrix data), data keranjang belanja (market basket data), dan data dokumen.

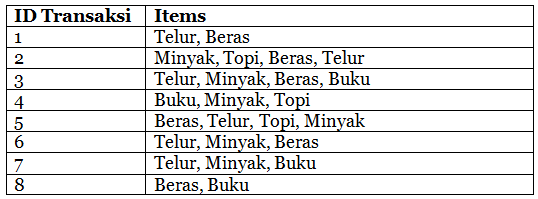
1. Data Matrik

Jika set data berisi kumpulan data yang mempunyai sejumlah atribut (fitur) numerik yang sama, set data tersebut dapat dipandang sebagai vektor (data) dalam wilayah multidimensi, dimana masing-masing dimensi menyatakan satu atribut yang berbeda. Sekumpulan data matrik dapat diinterpretasikan sebagai mantrik M x N, dimana M adalah jumlah baris (satu baris menyatakan satu record/objek) dan N adalah jumlah kolom (dimana satu kolom menyatakan satu atribut/fitur). Contoh data matrik dapat dilihat pada gambar dibawah ini.



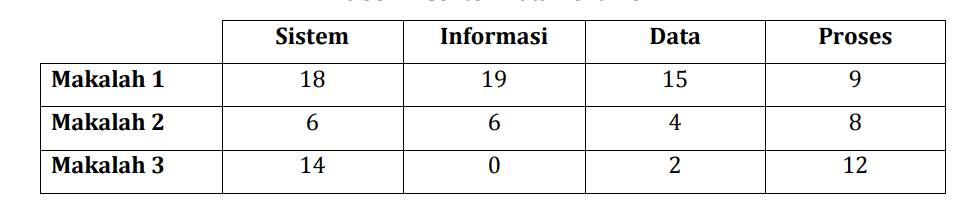
1. Data Keranjang Belanja/Transaksi

Data keranjang belanja (data transaksi) adalah set data yang setiap recordnya berisi sejumlah item dan jumlah item untuk sebuah transaksi bisa berbeda dengan transaksi yang lain. Contohnya bisa dilihat pada kasus keranjang belanja di pasar atau supermarket, dimana setiap pembeli melakukan pembelian barang yang jumlah dan jenisnya bisa berbeda dengan pembeli yang lain. Contoh data transaksi dapat dilihat pada tabel dibawah ini.



1. Data Dokumen

Setiap dokumen merupakan satu vektor ‘term’. Tiap term merupakan satu komponen (atribut) dari vektor tersebut. Nilai dari setiap komponen menyatakan berapa kali kemunculan term tersebut dalam suatu dokumen. Contoh dari data dokumen dapat dilihat pada gambar dibawah ini.



### **Kualitas pada dataset**

Permasalahan kualitas data ditinjau dari aspek pengukuran data dan pengumpulan data.

* 1. Kesalahan Pengukuran (Measurement Error)

Kesalahan ini mengacu pada permasalahan hasil dari proses pengukuran. Problem yang umum terjadi adalah nilai yang dicatat berbeda dari nilai sebenarnya untuk beberapa tingkat. Pada atribut kontinu, beda numerik dari hasil pengukuran dengan nilai sebenarnya disebut dengan error. Yang termasuk dalam jenis kesalahan pengukuran adalah

1. Noise

Merupakan komponen random dari suatu error pengukuran. Noise berkaitan dengan modifikasi dari nilai asli. Contoh: distorsi atau penyimpangan dari suara orang saat berbicara di telepon yang jaringannya buruk.

1. Bias

Suatu variasi pengukuran dari kuantitas yang sedang diukur dengan pengurangan antara mean dan nilai kuantitas yang diketahui.

1. Precision

Kedekatan dari pengukuran berulang (dari kuantitas yang sama) satu dengan yang lainnya. Diukur dengan standar deviasi.

1. Accuracy

Kedekatan pengukuran terhadap nilai sebenarnya dari kuantitas yang sedang diukur.

* 1. Kesalahan Pengumpulan (Collection Error)

Mengacu pada kesalahan-kesalahan (error) seperti hilangnya objek data atau nilai atribut, atau lingkup objek data yang tidak tepat. Yang temasuk dalam kesalahan pengumpulan diantaranya:

1. Outliers

Merupakan objek data dengan sifat yang berbeda sekali dari kebanyakan objek data dalam data-set. Misalkan, terdapat data penelitian tentang tinggi anak siswa SMA yakni 160cm sampai 180cm. Tetapi dalam data tersebut terdapat anak yang mempunyai tinggi 140cm. Data anak dengan tinggi 140cm tersebut yang disebut data outlier, karena berbeda sangat jelas.

Terdapat beberapa hal yang mempengaruhi munculnya data outlier antara lain:

* + 1. Kesalahan dalam pemasukan data
    2. Kesalahan dalam pengambilan sample
    3. Memang ada data-data ekstrim yang tidak dapat dihindarkan keberadaannya.

1. Missing value

Merupakan nilai dari suatu atribut yang tidak ditemukan. Alasan terjadinya missing value adalah:

* + 1. Informasi tidak diperoleh (misal, orang-orang menolak untuk memberikan data umur dan berat bedan)
    2. Atribut yang mungkin tidak bisa diterapkan ke semua kasus (misal, pendapatan tahunan tidak bisa diterapkan pada seseorang yang pengangguran)

Penanganan missing values adalah dengan:

* + 1. Mengurangi objek data
    2. Memperkirakan missing values
    3. Mengabaikan missing values pada saat analisis
    4. Mengganti dengan semua nilai yang mungkin (tergantung probabilitasnya)

1. Duplicate data

Set data mungkin terdiri dari objek data yang ganda (duplikat), atau hampir selalu terjadi duplikasi antara satu dengan yang lainnya. Persoalan utama ketika menggabungkan data dari sumber-sumber yang bervariasi (heterogen).

Contoh: orang yang sama dengan alamat email yang lebih dari satu. Pembersihan data (data cleaning) merupakan proses yang berkaitan dengan permasalahan data yang duplikat. [ ]

### **Dataset Yang Digunakan**

Dalam proses melakukan pengolahan data ini, digunakan suatu material yang dapat mendukung pengolahan data yang dilakukan. Material yang dimaksud ialah berupa data yang dapat digunakan pada proses pengolahan data.

Data yang digunakan dalam penelitian merupakan data pasien kanker payudara. Data tersebut didapatkan melalui situs bernama Kaggle yang dapat dikunjungi melalui *link* yang dicantumkan sebagai berikut:

<https://www.kaggle.com/datasets/yasserh/breast-cancer-dataset>

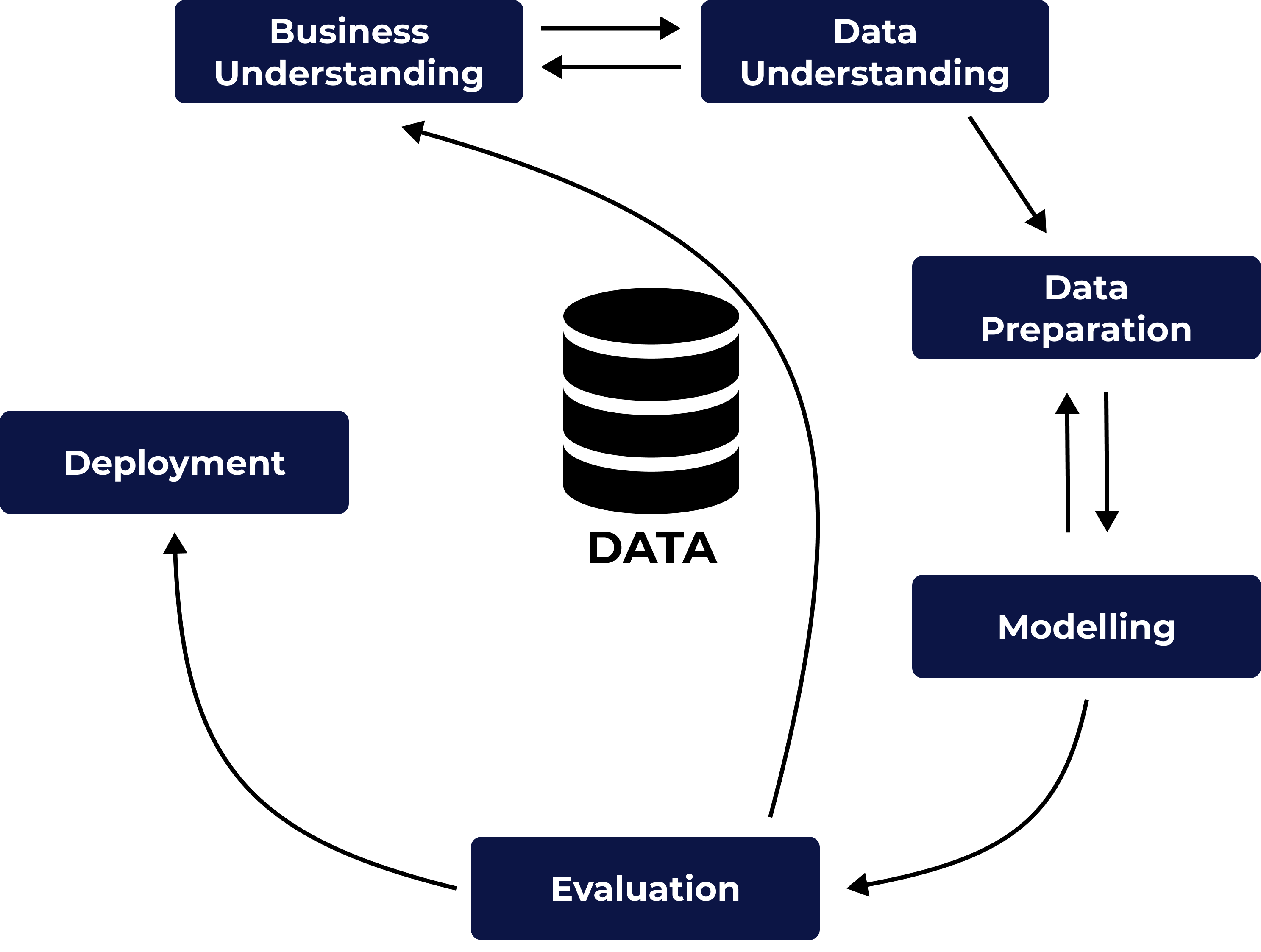
Data bersifat berlabel dan data diolah dengan teknik Supervised Learning untuk mendapatkan hasil model prediksi dengan menggunakan metode Support Vector Machine. memiliki jumlah kolom sebanyak 32 kolom dan jumlah total record sebanyak 569 *record*.

Kolom tersebut diantaranya adalah sebagai berikut.

|  |  |
| --- | --- |
| Atribut | Arti |
| id | Unique ID |
| diagnosis | Diagnosis masalah (M = malignant/ganas, B = benign/jinak) |
| Radius\_mean | Jari-jari lobus/kelenjar (rata-rata jarak dari pusat ke titik-titik di perimeter)  Perimeter: garis keliling inti sel yang diukur sebagai jumlah dari jarak.  Breast Cancer Overview: Causes, Symptoms, Signs, Stages & TypesLobular carcinoma breast cancer Stock Illustration | Adobe Stock |
| Texture\_mean | Rata-rata Tekstur Permukaan (standar deviasi nilai grayscale) |
| Perimeter\_mean | Perimeter Luar Lobus (ukuran rata-rata tumor inti) |
| Area\_mean | Rata-rata Luas Lobus |
| Smoothness\_mean | Rata-rata Tingkat Kehalusan (rata-rata variasi lokal dalam panjang radius)  Myoepithelial cells: good fences make good neighbors | Breast Cancer  Research | Full Text |
| Compactness\_mean | Mean of Compactness (mean of perimeter2 / area - 1.0)  rasio volume dan luas permukaan tumor |
| Concavity\_mean | Mean of Concavity (Rata-rata keparahan bagian cekung dari kontur)  Sebuah cekungan di bagian bawah toraks (dada) yang disebabkan oleh perpindahan tulang rawan xiphoid ke belakang |
| Concave ponts\_mean | Mean of Cocave Points (rata-rata untuk jumlah bagian cekung dari kontur) |
| Symmetry\_mean | Mean of Symmetry  Asimetri payudara mengacu pada ketika satu payudara memiliki ukuran atau bentuk yang berbeda dari yang lain  Kesimetrisan mungkin terkait dengan penyakit payudara, masalah perkembangan, masalah hormonal, sindrom yang tidak diketahui, cedera atau periode perubahan tubuh yang signifikan seperti kehamilan atau menyusui. |
| Fractal\_dimension | Mean of Fractal Dimension (mean for "coastline approximation" - 1)  Analisis fraktal dari gambar spesimen jaringan payudara memberikan deskripsi numerik pola pertumbuhan tumor sebagai angka kontinu antara 1 dan 2. Angka ini, dimensi fraktal, adalah ukuran yang objektif dan dapat direproduksi dari kompleksitas arsitektur jaringan spesimen biopsi. |
| Radius\_se | SE of Radius (standard error for the mean of distances from center to points on the perimeter) |
| Texture\_se | SE of Texture (standard error for standard deviation of gray-scale values) |
| Perimeter\_se | Perimeter of SE |
| Area\_se | Area of SE |
| Smoothness\_se | SE of Smoothness (standard error for local variation in radius lengths) |
| Compactness\_se | SE of compactness (standard error for perimeter^2 / area - 1.0) |
| Concavity\_se | SE of concavity (standard error for severity of concave portions of the contour) |
| Concave points\_se | SE of concave points (standard error for number of concave portions of the contour) |
| Symmetry\_se | SE of symmetry |
| Fractal\_dimension\_se | SE of Fractal Dimension (standard error for "coastline approximation" - 1) |
| Radius\_worst | Worst Radius ("worst" or largest mean value for mean of distances from center to points on the perimeter) |
| Texture\_worst | Worst Texture ("worst" or largest mean value for standard deviation of gray-scale values) |
| Perimeter\_worst | Worst Permimeter |
| Area\_worst | Worst Area |
| Smoothness\_worst | Worst Smoothness ("worst" or largest mean value for local variation in radius lengths) |
| Compactness\_worst | Worse Compactness ("worst" or largest mean value for perimeter^2 / area - 1.0) |
| Concavity\_worst | Worst Concavity ("worst" or largest mean value for severity of concave portions of the contour) |
| Concave points\_worst | Worst Concave Points ("worst" or largest mean value for number of concave portions of the contour) |
| Symmetry\_worst | Worst Symmetry |
| Fractal\_dimension\_worst | Worst Fractal Dimension ("worst" or largest mean value for "coastline approximation" - 1) |

## Metodologi Pengolahan Data

Diagram alur metodologi pengolahan data yang digunakan ialah menggunakan metodologi CRISP-DM (*Cross Industry Standard Process for Data Mining*). Metodologi penelitian CRISP-DM digunakan dalam penelitian karena metodologi ini mampu menjabarkan setiap langkah pengolahan data dengan jelas serta mudah dipahami untuk diimplementasikan.



Berdasarkan diagram alur metodologi penelitian diatas, dapat dipaparkan sebagai berikut.

1. *Business* *Understanding*

Pada tahap ini merupakan tahap untuk menentukan seperti apa proses bisnis yang akan dibangun untuk melakukan hal apa yang ingin dicapai. Selain itu, dalam tahap ini juga diperlukan analisis apa saja alat dan bahan yang akan diperlukan untuk setiap fase proyek. Dalam hal ini, proses bisnis yang dilakukan pada penelitian adalah memprediksi gaji berdasarkan faktor-faktor spesifik. Prediksi gaji tersebut akan menggunakan data kepegawaian. Alat yang digunakan adalah berupa perangkat keras berupa laptop. Perangkat lunak yang digunakan ialah Jupyter Notebook, Anaconda, serta Microsoft Excel.

1. *Data* *Understanding*

Pada tahap *data understanding*, hal yang dilakukan adalah melakukan pemahaman bisnis yang berfokus untuk mengidentifikasi, mengumpulkan dan menganalisis kumpulan data untuk mencapai tujuan bisnis. Dalam hal ini, proses yang dilakukan ialah mengumpulkan data awal yang berasal dari Kaggle. Setelah data ditemukan, kemudian hal yang dilakukan adalah menjelajahi data dan melakukan verifikasi kualitas data (seberapa bersih atau otornya data yang digunakan).

1. *Data* *Preparation*

Tahap *data preparation* dilakukan untuk menyiapkan data akhir untuk dilakukannya pemodelan dengan cara memperbaiki masalah yang terdapat pada data. Hal yang dilakukan dalam penelitian ini adalah membersihkan data, serta memilih data yang akan digunakan dalam pemodelan. Sehingga nantinya akan didapatkan data akhir yang siap dimodelkan.

1. *Modelling*

Tahap *modelling* atau pemodelan adalah menentukan algoritma teknik pemodelan yang akan digunakan untuk menyelesaikan masalah yang ada. Dalam penelitian ini, model yang digunakan adalah menggunakan model regresi linier berganda. Karena model ini adalah model yang cocok untuk digunakan dalam melakukan prediksi terutama jika variabel yang mempengaruhi sangat banyak.

1. *Evaluation*

Tahap *evaluation* lebih berfokus untuk melihat apakah pemodelan yang dilakukan memiliki hasil yang baik atau malah sebaliknya. Pada tahap ini, evaluasi model dilakukan dengan menggunakan pengujian asumsi untuk melihat hasil-hasil evaluasi dari model.

1. *Deployment*

Yang terakhir ialah tahap *deployment*. Dimana pada tahap ini, hasil model penelitian yang telah dilakukan akan dikembangkan dan didokumentasikan. Pada tahap ini, hal yang dilakukan dalam penelitian yaitu mengembangkan model ke dalam bentuk aplikasi sehingga lebih mudah digunakan dalam melakukan proses prediksi. Selain itu, penelitian ini juga dibuatkan dalam bentuk laporan akhir untuk didokumentasikan.

# BAB 3 Apa Saja Perangkat Yang Digunakan Untuk Membuat Aplikasi?

## Bahasa Pemrograman Yang Digunakan

### **Python**



Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Pyhton disebutkan sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standat yang besar serta kompreherensif. Python mendukung multi paradigma pemrograman, utamanya; namun tidak dibatasi; pada pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional. Salah satu fitur yang tersedia pada python adalah sebagai bahasa pemrogaman dinamis yang dilengkapi dengan manajamen memori otomatis.

Seperti halnya bahasa pemrograman dinamis lainnya, python umumnya digunakan sebagai bahasa *script* meski pada praktiknya penggunaan bahasa ini lebih luas mencakup konteks pemanfaatan yang umumnya tidak dilakukan dengan menggunakan bahasa *script*. Python dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform sistem operasi. Saat ini kode python dapat dijalankan di berbagai platform sistem operasi, beberapa diantaranya adalah :

1. Linux/Unix
2. Windows
3. Max OS X
4. Java Virtual Machine
5. Amiga
6. Palm
7. Symbian (untuk produk-produk Nokia)

Python didistribusikan dengan beberapa lisensi yang berbeda dari beberapa versi. Namun pada prinsipnya Python dapat diperoleh dan dipergunakan secara bebas, bahkan untuk kepentingan komersial. Lisensi Python tidak bertentangan baik menurut definisi Open Source maupun General Public License (GPL).

Bahasa pemrograman Python tentunya tidak dapat diemplementasikan atau digunakan secara langsung pada *software* teks editor ataupun terminal pada *device* seperti bahasa pemrograman PHP. Bahasa pemrograman ini akan dapat digunakan jika telah ter-install pada *device* yang digunakan.

### **HTML**



HTML (*Hyper Text Mark Up Language*) ialah bahasa yang dipergunakan dengan tujuan untuk mendeskripsikan struktur dar sebuah halaman web. HTML memiliki fungsi untuk mempublikasikan dokumen secara online. *Tags* adalah statement dasar dari HTML. Sebuah *tag* dinyatakan dalam kurung siku (<>). *Tags* yang ditujukan untuk sebuah dokumen atau bagian dari suatu dokumen harus dibuat dalam bentuk berpasangan. *Tag* terdiri dari dua macam yaitu *tag* pembuka dan *tag* penutup. Dimana *tag* penutup menggunakan tambahan tanda garis miring (/) di awal nama *tag*(Pahlevi, Mulyani and Khoir, 2018).

### **CSS**



CSS memiliki kepanjangan *Cascading Style Sheet* adalah bahasa-bahasa yang dapat merepresentasikan halaman web. Representasi tersebut diantaranya adalah warna, *layout*, dan *font*. Dengan digunakannya CSS dalam membangun halaman web, seorang web developer dapat membuat halaman web yang dapat beradaptasi dengan berbagai macam ukuran layar. Pembuatan skrip CSS biasanya terpisah dengan halaman HTML. Meskipun skrip CSS dapat disisipkan di dalam halaman HTML. Hal ini ditujukan untuk memudahkan pengaturan halaman HTML yang memiliki rancangan yang sama (Henderson, 2009:72) (Pahlevi, Mulyani and Khoir, 2018).

### **Bootstrap**



Bootstrap adalah suatu *framework* untuk CSS (*Cascading Style Sheet*) dan berupa produk bentuk opensource (terbuka untuk siapapun) yang dibuat oleh Mark Otto dan Jacob Thornton. Pada awalnya Bootstrap ini dibuat untuk membuat standarisasi dalam front end untuk seluruh programmer yang ada di perusahaannya. Bootsrap telah berubah dari yang sebelumnya adalah CSS-Driven proyek ke sebuah host dari JavaScript plugins dan ikon yang dapat dengan mudah digunakan untuk *form*ulir dan tombol.

### **Django**



Django ialah sebuah web *framework* berbasis bahasa pemrograman Python yang didesain untuk membuat suatu aplikasi web yang dinamis, kaya fitur dan aman. (Saputra and Fathoni Aji, 2018) Django yang dikembangkan oleh Django *Software* Foundation terus mendapatkan perbaikan sehingga membuat web *framework* yang satu ini menjadi pilihan utama bagi banyak pengembang aplikasi web. (Saputra and Fathoni Aji, 2018)

*Framework* Django bertujuan untuk memudahkan pengembangan situs web dan basis data yang kompleks (Django *Software* Foundation, 2017). Salah satu keunggulan django yaitu *framework* ini telah merepresentasikan model ORM (*Object Relational Mapper*) sehingga tidak perlu lagi untuk menyesuaikan *query* jika terjadi perubahan *database* yang digunakan. Berikut ini adalah beberapa keuntungan menggunakan Django:

1. *Object-Relational Mapping (ORM) Support.*
2. *Automatic Admin Interface.*
3. *Elegant URL Design.*
4. *Template System.*
5. *Cache system.*
6. *Internationalization.*
7. *A light weight web server for development and test*(Rahimah, Rusdianto and Ananta, 2019).

## Penyimpanan Yang Digunakan

### **Database MySQL**



MySQL adalah sebuah perangkat lunak sistem manajemen basis data SQL (Bahasa Inggris: *database management system*) atau DBMS yang *multihead*, *multi-user*, dengan sekitar 6 juta instalasi di seluruh dunia. MySQL adalah DBMS yang *open source* dengan dua bentuk lisensi, yaitu *Free Software* (perangkat lunak bebas) dan *Shareware* (perangkat lunak berpemilik yang penggunaannya terbatas). Jadi MySQL adalah *database* *server* yang gratis dengan lisensi GNU *General Public License* (GPL) sehingga dapat dipakai untuk keperluan pribadi atau komersil tanpa harus membayar lisensi yang ada. Istilah semacam baris, kolom, tabel, dipakai pada MySQL. SQL sendiri merupakan suatu Bahasa yang dipakai di dalam pengambilan data pada *relational database* atau *database* yang terstruktur. Jadi MySQL adalah *database management system* yang menggunakan Bahasa SQL sebagai Bahasa penghubung antara perangkat lunak aplikasi dengan *database server*. MySQL sendiri sebenarnya merupakan lanjutan dari proyek UNIREG yang dikerjakan oleh Michael Montry Widenius dan TcX (perusahaan perangkat lunak asal Swedia). Sayangnya, UNIREG belum terlalu kompatibel dengan *database* dinamis yang dipakai di *website*. TcX kemudian mencari alternatif lain dan menemukan perangkat lunak yang dikembangkan oleh David Hughes, yaitu miniSQL atau mSQL. Namun, ditemukan masalah lagi karena mSQL tidak mendukung *indecing* sehingga belum sesuai dengan kebutuhan TcX. MySQL *server* mendukung beberapa macam sistem operasi (*cross-platform*), khususnya Linux dan Windows. Proses instalasinya pun sangat mudah. Jika menggunakan Windows, bisa menginstall XAMPP untuk menjalankan MySQL *server* yang di dalamnya sudah terdapat juga modul untuk menjalankan Apache, PHP, FileZilla, dan Tomcat. Sedangkan di Linux, dapat menginstall MySQL secara terpisah atau menginstall LAMp (Linux, Apache, MySQL, PHP) yang sudah ada pada modul Apache dan PHP. Selain itu, XAMPP juga sudah tersedia di Linux(Hasan Bisri Isa Alfaris, Choirul Anam and Ali Masy’an, 2013).

## Perangkat Lunak Yang Dibutuhkan

### **Visual Studio Code**



Visual Studio Code (VS Code) adalah sebuah teks editor ringan dan handal yang dibuat oleh Microsoft untuk sistem operasi *multiplatform*, artinya tersedia juga untuk versi Linux, Mac, dan Windows. Teks editor ini secara langsung mendukung bahasa pemrograman dengan bantuan plugin yang dapat dipasang via *marketplace* Visual Studio Code(Yudo Ardianto Agung, 2020).

Berikut fitur-fitur unggulan Visual Studio Code :

1. *Cross platform*

Dapat berjalan di Windows, Linux, dan MacOS. Artinya dapat bekerja pada sistem operasi manapun dengan menggunakan Visual Studio Code.

1. *Lightweight*

Tak perlu menunggu lama untuk memulai. Bahasa, tema, debugger, commands, dan lain-lainnya dapat dikontrol sepenuhnya sesuai keinginan.

1. *Powerfull Editor*

*Source code editing* yang sangat produktif, seperti membuat code *snippets*, *IntelliSense*, *auto correct*, dan format*ting*.

1. *Code Debugging*

Salah satu fitur bagus yang ditawarkan Visual Studio Code adalah membantu untuk melakukan debug pada kode dengan cara mengawasi kode, variabel, *call stack* dan *expression* yang mana saja.

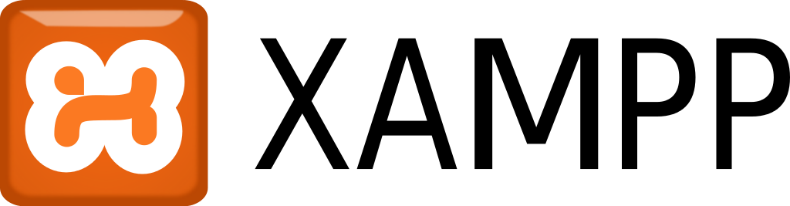
1. *Source Control*

Visual Studio Code memiliki *intgrated source control* termasuk *Git support in-the-box* dan penyedia *source code control* lainnya di pasaran. Ini meningkatkan siklus rilis proyek secara signifikan.

1. *Intergrated* *Terminal*

Tidak ada lagi *multiple* Windows dan Alt-Tabs. Perintah tersebut dapat dilakukan dengan *command-line task* sekejap dan membuat banyak terminal di dalam editor.

### **XAMPP**



XAMPP adalah sebuah perangkat lunak komputer yang sistem penamaannya diambil dari akronim kata Apache, MySQL/MariaDB, PHP, dan Perl. Sementara imbuhan huruf “X“ yang terdapat pada awal kata berasal dari istilah *cross platform* sebagai simbol bahwa aplikasi ini bisa dijalankan di empat sistem operasi berbeda, seperti Linux, Windows, MacOS, dan juga Solaris. Jika dijabarkan, masing-masing huruf yang ada di dalam nama XAMPP menurut para ahli memiliki arti sebagai berikut ini:

1. X (*Cross Platform*)

Merupakan kode penanda untuk *software cross platform* atau yang bisa berjalan di banyak sistem operasi. Jadi, ada XAMPP untuk Windows, MacOS, Linux. Semua itu bersifat *free* download XAMPP.

1. A (*Apache*)

Apache adalah aplikasi *web server* yang bersifat gratis dan bisa dikembangkan oleh banyak orang (*open source*).

1. M (*MySQL/MariaDB*)

MySQL atau MariaDB merupakan aplikasi *database server* yang dikembangkan oleh orang yang sama. MySQL berperan dalam mengolah, mengedit, dan menghapus daftar melalui *database*.

1. P (*PHP*)

Huruf “P“ yang pertama dari akronim kata XAMPP adalah inisial untuk menunjukkan eksistensi bahasa pemrograman PHP. Bahasa pemrograman ini biasanya digunakan untuk membuat *website* dinamis, contohnya dalam *website* berbasis CMS WordPress.

1. P (*Perl*)

Sementara itu, huruf P selanjutnya merupakan singkatan dari bahasa pemrogaman Perl yang kerap digunakan untuk memenuhi berbagai macam kebutuhan. Perl ini bisa berjalan di dalam banyak digunakan.

Program aplikasi XAMPP berfungsi sebagai *server* lokal untuk mengampu berbagai jenis data *website* yang sedang dalam proses pengembangan. Dalam prakteknya, XAMPP digunakan untuk menguji kinerja fitur ataupun menampilkan konten yang ada di dalam *website* kepada orang lain tanpa harus terkoneksi dengan internet, cukup akses melalui XAMPP *Control panel*, atau istilahnya *website offline*. XAMPP bekerja secara *offline* layaknya *web hosting* biasa namun tidak bisa diakses oleh banyak orang.

### **Jupyter Notebook**



Jupyter Notebook adalah lingkungan komputasi interaktif berbasis *web* untuk membuat dokumen *notebook* jupyter. Jupyter notebook mendukung beberapa bahasa seperti Python (Ipython), Julia, R, dan lain-lain. Sebagian besar digunakan untuk analisis data, visualisasi data dan komputasi eksplorasi interaktif lebih lanjut. Awalnya dikembangkan untuk aplikasi *data science* yang ditulis dengan Python, R, dan Julia, Jupyter Notebook berguna dalam segala macam cara untuk semua jenis proyek :

1. Visualisasi Data

Kebanyakan orang memiliki eksposur pertama mereka ke Notebook Jupyter melalui visualisasi data, sebuah *notebook* bersama yang mencakup *rendering* dari beberapa data yang ditetapkan sebagai grafik. Jupyter Notebook memungkinkan untuk membuat visualisasi, tetapi juga membagikannya dan memungkinkan perubahan interaktif pada kode dan kumpulan data yang dibagikan.

1. Berbagi Kode

Layanan *cloud* seperti GitHub dan Pastebin menyediakan cara untuk berbagi kode, tetapi sebagian besar non-interaktif. Dengan Jupyter Notebook dapat dilihat kode, menjalankannya, dan menampilkan hasilnya langsung di *browser web*.

1. Interaksi Langsung dengan Kode

Kode Notebook Jupyter tidak statis, itu dapat diedit dan dijalankan kembali secara bertahap, dengan umpan balik yang diberikan langsung di *browser*. Notebook juga dapat menyematkan kontrol pengguna (Misalnya Slider atau bidang input teks) yang dapat digunakan sebagai sumber input untuk kode.

1. Mendokumentasikan Contoh Kode

Jika memiliki sepotong kode dan ingin menjelaskan secara bertahap cara kerjanya, dengan umpan balik langsung di sepanjang jalan, dapat menyematkannya dalam Jupyter Notebook. Kode akan tetap berfungsi penuh, dapat menambahkan interaktivitas bersama dengan penjelasan, yang ditampilkan dan diceritakan pada saat yang sama(small-business-tracker.com, 2022).

### **Tableau**

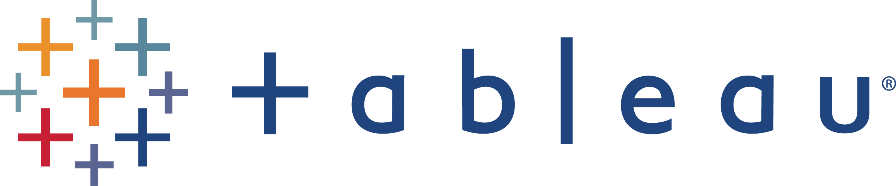


Tableau adalah *platform business intelligence* untuk membantu dalam melihat dan memahami data. *Platform* ini merupakan hasil dari proyek ilmu komputer di Stanford yang didirikan pada tahun 2003. Tujuan dibuatnya Tableau adalah untuk meningkatkan aliran analisis dan membuat data lebih dapat diakses oleh orang-orang melalui visualisasi. Berikut keunggulan menggunakan Tableau :

1. Visualisasi yang mudah dan interaktif

Keunggulan yang membuat populer Tableau adalah visualisasinya mudah, interaktif, dan menarik. Jika *tools* lain hanya menyajikan grafik biasa, *tools* ini bisa menyajikan grafik bergerak, sehingga tampilan lebih menarik.

1. *User-friendly*

Jika membuat grafik perlu adanya *coding*, Tableau tidak perlu ada *coding*. Tableau diperuntukan bagi siapapun yang bermain dengan data tapi ingin lebih sederhana dalam menyajikannya.

1. Banyak sumber data yang mampu diolah

Banyak data kompleks yang beraturan maupun tidak beraturan, dan juga berbagi sumber, Tableau bisa mengolah dari berbagai sumber. Bahkan bisa menggabungkan dari berbagai *cloud*, *spreadsheet*, dan *big data*.

1. *Dashboard mobile friendly*

Jika tidak sempat membuka laptop, Tableau bisa diakses melalui *smarthphone*. Meski lewat smartphone, *user interface*nya mudah untuk dioperasikan.

1. Terintegrasi dengan bahasa srkip(Institut Teknologi Bandung, 2022).

### **Anaconda**



Anaconda merupakan aplikasi gratis yang memiliki banyak package dan tools atau dengan kata lain merupakan *one place solution platform* untuk *data scientist* agar memudahkan dalam mengerjakan proyek karena di dalamnya terdapat berbagai macam *software*. Anaconda bisa dikatakan sebagai paket distribusi python dari *continuum analytics* yang berisi paket *Python* ditambah beberapa paket tambahan untuk keperluan pemrograman *data science*, matematika, hingga teknik dalam satu distribusi yang *user friendly*.

### **Kaggle**

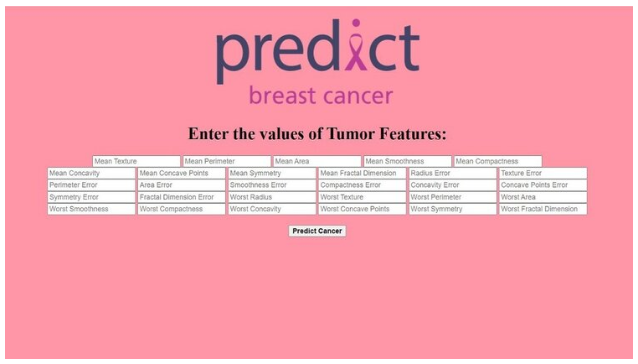


Kaggle adalah situs dan platform untuk berlomba membuat model terbaik untuk menganalisa dan memprediksi suatu dataset. Dataset ini diberikan oleh suatu perusahaan, dengan suatu deskripsi masalah tertentu. Misalnya, diberikan data rumah beserta fitur-fiturnya dan harga jualnya, dan deskripsi masalahnya adalah buatlah model untuk memprediksi harga jual berdasarkan fitur-fitur tersebut.

## Spesifikasi Perangkat Keras

|  |  |  |  |
| --- | --- | --- | --- |
| No | Nama Perangkat | Spesifikasi | Keterangan |
| 1 | Hardisk | 200 GB | Media untuk menyimpan data aplikasi yang dibuat |
| 2 | Memory | 4 GB | *Memory System* yang digunakan |
| 3 | Processor | AMD A4-9120 RADEON R3, 4 COMPUTE CORES 2C+2G  2.20 GHz | Untuk kecepatan transfer data dari sistem yang sangat bergantung pada kecepatan prosesor komputer |
| 4 | Infrastruktur jaringan | PC | Mengolah, menginput serta menghasilkan *output*  data ataupun informasi yang sesuai dengan keinginan pengguna (*user*) |

<https://www.jietdat.ac.in/acclaimed-projects/predictionofbreastcancer>



Kanker Payudara adalah penyakit paling utama yang mempengaruhi 2,1 juta wanita setiap tahun yang menyebabkan jumlah kematian terbesar di kalangan wanita. Perawatan dini tidak hanya membantu menyembuhkan kanker tetapi juga membantu dalam pencegahan kekambuhannya. Dan karenanya sistem ini terutama berfokus pada prediksi kanker payudara di mana ia menggunakan algoritma pembelajaran mesin yang berbeda untuk membuat model seperti pohon keputusan, regresi logistik, dan hutan acak yang diterapkan pada data pra-proses yang mencurigai akurasi prediksi yang lebih besar. Di antara semua model, Klasifikasi Hutan Acak menghasilkan akurasi terbaik. Teknik-teknik ini dikodekan dalam Python dan menggunakan perpustakaan NumPy, Pandas, Seaborn.

# DAFTAR PUSTAKA

Blogs.reed.edu. 2017. Text analysis using Voyant Tools. [online] Available at: <https://blogs.reed.edu/ed- tech/2017/03/text-analysis-using-voyant-tools/> [Accessed 11 September 2022].

Calado, F., 2018. Using Voyant-Tools to Formulate Research Questions for Textual Data - GC Digital Fellows. [online] GC Digital Fellows. Available at:

<https://digitalfellows.commons.gc.cuny.edu/2018/11/01

/using-voyant-tools-to-formulate-research-questions-for- textual- data/#:~:text=One%20of%20the%20benefits%20of,text

%20is%20displayed%20for%20reading.> [Accessed 11

August 2022].

En.wikipedia.org. 2022. Voyant Tools - Wikipedia. [online] Available at:

<https://en.wikipedia.org/wiki/Voyant\_Tools#:~:text=Ran ge%20of%20Uses-

,History,Lisa%20Goddard%2C%20and%20Mark%20Turc ato.> [Accessed 10 August 2022].

GitHub. 2016. *GitHub - voyanttools/Voyant*. [online] Available at:

<https://github.com/voyanttools/Voyant> [Accessed 10

August 2022].

-oo00oo-

DIHALAMAN INI, SILAHKAN ANDA ISI DENGAN PROFIL PENULIS, BAIK ITU PEMBIMBING INTERNSHP 1 DAN MAHASISWA

BUAT COVER BELAKANG BUKU

YANG BERISI :

SINOPSIS DARI BUKU YANG ANDA TULIS, SINOPSIS DAPAT DIAMBIL DARI ABSTRAK PADA JURNAL YANG ANDA BUAT

CATATAN :

DESAIN HARUS ORIGINAL