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

**Description:**

Kanker payudara adalah kanker paling umum di kalangan wanita di dunia. Ini menyumbang 25% dari semua kasus kanker, dan mempengaruhi lebih dari 2,1 Juta orang pada tahun 2015 saja. Ini dimulai ketika sel-sel di payudara mulai tumbuh di luar kendali. Sel-sel ini biasanya membentuk tumor yang dapat dilihat melalui sinar-X atau dirasakan sebagai benjolan di area payudara.

Tantangan utama untuk mendeteksinya adalah bagaimana mengklasifikasikan tumor menjadi ganas (kanker) atau jinak (non-kanker). Kami meminta Anda untuk menyelesaikan analisis klasifikasi tumor ini menggunakan pembelajaran mesin (dengan SVM) dan Dataset Kanker Payudara Wisconsin (Diagnostik).

|  |  |
| --- | --- |
| Atribut | Arti |
| id | Unique ID |
| diagnosis | The diagnosis of breast tissues (M = malignant, B = benign) |
| Radius\_mean | Radius of Lobes (mean of distances from center to points on the perimeter) |
| Texture\_mean | Mean of Surface Texture (standard deviation of gray-scale values) |
| Perimeter\_mean | Outer Perimeter of Lobes (mean size of the core tumor) |
| Area\_mean | Mean Area of Lobes |
| Smoothness\_mean | Mean of Smoothness Levels (mean of local variation in radius lengths) |
| Compactness\_mean | Mean of Compactness (mean of perimeter^2 / area - 1.0) |
| Concavity\_mean | Mean of Concavity (mean of severity of concave portions of the contour) |
| Concave ponts\_mean | Mean of Cocave Points (mean for number of concave portions of the contour) |
| Symmetry\_mean | Mean of Symmetry |
| Fractal\_dimension | Mean of Fractal Dimension (mean for "coastline approximation" - 1) |
| 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) |

RangeIndex: 569 entries, 0 to 568

Data columns (total 32 columns):

|  |  |
| --- | --- |
| Atribut | Arti |
| id | ID number |
| A diagnosis | The diagnosis of breast tissues (M = malignant, B = benign) |
| Radius\_mean | mean of distances from center to points on the perimeter |
| Texture\_mean | standard deviation of gray-scale values |
| Perimeter\_mean | mean size of the core tumor |
| Area\_mean |  |
| Smoothness\_mean | mean of local variation in radius lengths |
| Compactness\_mean | mean of perimeter^2 / area - 1.0 |
| Concavity\_mean | mean of severity of concave portions of the contour |
| Concave ponts\_mean | mean for number of concave portions of the contour |
| Symmetry\_mean |  |
| Fractal\_dimension | mean for "coastline approximation" - 1 |
| Radius\_se | standard error for the mean of distances from center to points on the perimeter |
| Texture\_se | standard error for standard deviation of gray-scale values |
| Perimeter\_se |  |
| Area\_se |  |
| Smoothness\_se | standard error for local variation in radius lengths |
| Compactness\_se | standard error for perimeter^2 / area - 1.0 |
| Concavity\_se | standard error for severity of concave portions of the contour |
| Concave points\_se | standard error for number of concave portions of the contour |
| Symmetry\_se |  |
| Fractal\_dimension\_se | standard error for "coastline approximation" - 1 |
| Radius\_worst | "worst" or largest mean value for mean of distances from center to points on the perimeter |
| Texture\_worst | "worst" or largest mean value for standard deviation of gray-scale values |
| Perimeter\_worst |  |
| Area\_worst |  |
| Smoothness\_worst | "worst" or largest mean value for local variation in radius lengths |
| Compactness\_worst | "worst" or largest mean value for perimeter^2 / area - 1.0 |
| Concavity\_worst | "worst" or largest mean value for severity of concave portions of the contour |
| Concave points\_worst | "worst" or largest mean value for number of concave portions of the contour |
| Symmetry\_worst |  |
| Fractal\_dimension\_worst | "worst" or largest mean value for "coastline approximation" - 1 |

Link Terkait <https://www.kaggle.com/datasets/reihanenamdari/breast-cancer>

EDA <https://www.kaggle.com/code/kiransheshma/breast-cancer-exploratory-data-analysis>

|  |  |
| --- | --- |
| Atribut | Arti |
| Age | Usia |
| Race | other :American Indian/AK Native, Asian/Pacific Islander |
| Marital Status | Status Pernikahan |
| T Stage | Adjusted AJCC 6th T |
| N Stage | Adjusted AJCC 6th N |
| 6th Stage | Breast Adjusted AJCC 6th Stage |
| differentiate | diferensiasi |
| Grade | Nilai |
| A Stage | Regional — A neoplasm that has extended, Distant — A neoplasm that has spread to parts of the body remote from the primary tumor either by direct extension or by discontinuous metastasis |
| Tumor Size | Each indicates exact size in millimeters. |
| Estrogen Status | Status estrogen |
| Progesterone Status | Status progesteron |
| Regional Node Examined | Reginol Node Examined |
| Reginol Node Positive | Reginol Nodepositive |
| Survival Months | Survival Months |
| Status | Dead or Alive |

data\_shape: (4024, 16)

**Data Description**

Dataset pasien kanker payudara ini diperoleh dari program SEER NCI November 2017 yang memberikan informasi tentang statistik kanker berbasis populasi. Dataset melibatkan pasien wanita dengan duktus infiltrasi dan kanker payudara karsinoma lobular (kutipan primer SIER recode kode histologi NOS 8522/3) yang didiagnosis pada 2006-2010. Pasien dengan ukuran tumor yang tidak diketahui, memeriksa LN regional, LN regional positif, dan pasien yang kelangsungan hidupnya kurang dari 1 bulan dikeluarkan; dengan demikian, 4024 pasien akhirnya dimasukkan.

**Race:** 0 = represent white Race , 1 = represent Black Race and 2 = Represent Other Race .

**Marital Status:** 0 = Married, 1 = Divorced , 2= Single, 3 = Widowed and 4 = Separated

**T Stage:** The T refers to the size and extent of the main tumor. The main tumor is usually called the primary tumor.T1, T2, T3, T4: Refers to the size and/or extent of the main tumor. The higher the number after the T, the larger the tumor or the more it has grown into nearby tissues. T-T0: No evidence of primary tumor. T1 (includes T1a, T1b, and T1c): Tumor is 2 cm (3/4 of an inch) or less across. T2: Tumor is more than 2 cm but not more than 5 cm (2 inches) across. T3: Tumor is more than 5 cm across

**N Stage:** The main tumor is usually called the primary tumor. The N refers to the the number of nearby lymph nodes that have cancer. The M refers to whether the cancer has metastasized. This means that the cancer has spread from the primary tumor to other parts of the body.N1, N2, N3: Refers to the number and location of lymph nodes that contain cancer. The higher the number after the N, the more lymph nodes that contain cancer.

**6th Stage:** 0 = IIA , 1= IIIA, 2 = IIIC , 3=IIB and 4 = IIIB Stage groups for breast cancer,Doctors assign the stage of the cancer by combining the T, N, and M classifications (see above), the tumor grade, and the results of ER/PR and HER2 testing. This information is used to help determine your prognosis (see Diagnosis).Most patients are anxious to learn the exact stage of the cancer. If you have surgery as the first treatment for your cancer, your doctor will generally confirm the stage of the cancer when the testing after surgery is finalized, usually about 5 to 7 days after surgery. When systemic treatment is given before surgery, which is typically with medications and is called neoadjuvant therapy, the stage of the cancer is primarily determined clinically. Doctors may refer to stage I to stage IIA cancer as "early stage" and stage IIB to stage III as "locally advanced."

**Stage IIA** Any 1 of these conditions:

1. There is no evidence of a tumor in the breast, but the cancer has spread to 1 to 3 axillary lymph nodes. It has not spread to distant parts of the body (T0, N1, M0).
2. The tumor is 20 mm or smaller and has spread to 1 to 3 axillary lymph nodes (T1, N1, M0).
3. The tumor is larger than 20 mm but not larger than 50 mm and has not spread to the axillary lymph nodes (T2, N0, M0).

**Stage IIB** Either of these conditions:

1. The tumor is larger than 20 mm but not larger than 50 mm and has spread to 1 to 3 axillary lymph nodes (T2, N1, M0).
2. The tumor is larger than 50 mm but has not spread to the axillary lymph nodes (T3, N0, M0).

**Stage IIIA** The tumor of any size has spread to 4 to 9 axillary lymph nodes or to internal mammary lymph nodes. It has not spread to other parts of the body (T0, T1, T2, or T3; N2; M0). Stage IIIA may also be a tumor larger than 50 mm that has spread to 1 to 3 axillary lymph nodes (T3, N1, M0).

**Stage IIIB** The tumor has spread to the chest wall or caused swelling or ulceration of the breast, or it is diagnosed as inflammatory breast cancer. It may or may not have spread to up to 9 axillary or internal mammary lymph nodes. It has not spread to other parts of the body (T4; N0, N1, or N2; M0).

**Stage IIIC** A tumor of any size that has spread to 10 or more axillary lymph nodes, the internal mammary lymph nodes, and/or the lymph nodes under the collarbone. It has not spread to other parts of the body (any T, N3, M0).

**differentiate** 0 =Poorly differentiated, 1 = Moderately differentiated, 2= Well differentiated and 3 = Undifferentiated

**Grade:**

1. What does cancer grade mean?

Breast cancers are given a grade according to:

1.1 How different the cancer cells are to normal breast cells 1.2 How quickly they are growing

The grade of a cancer is different to the cancer stage. A cancer’s grade is determined when a doctor (pathologist) looks at the cancer cells under a microscope, using tissue from a biopsy or after breast cancer surgery.

1. What are the different grades of breast cancer?

There are three grades of invasive breast cancer:

Grade 1 looks most like normal breast cells and is usually slow growing Grade 2 looks less like normal cells and is growing faster Grade 3 looks different to normal breast cells and is usually fast growing

Or we can say that a low grade number (grade 1) usually means the cancer is slower-growing and less likely to spread. A high grade number (grade 3) means a faster-growing cancer that's more likely to spread. An intermediate grade number (grade 2) means the cancer is growing faster than a grade 1 cancer but slower than a grade 3 cancer.

Sometimes the grade given to a cancer after a biopsy can change after surgery. This is because after surgery there’s more tissue for the pathologist to look at, which can give them more detailed information about the cancer.

**A Stage:** 0 =Regional and 1 = Distant

This parameters shows Stages of Cancer and is a summary of all data , it is an attribute that involves T,N and Grade data.

1. Regional: The cancer has spread outside the breast to nearby structures or lymph nodes.
2. Distant: The cancer has spread to distant parts of the body such as the lungs, liver or bones.

**Estrogen Status:** 0 =Estrogen positive and 1 = Estrogen negative

1. Estrogen positive : - Cancer cells that are ER positive may need estrogen to grow. These cells may stop growing or die when treated with substances that block the binding and actions of estrogen. Also called estrogen receptor positive.
2. Estrogen negative : - negative breast cancers are a group of tumors with poor prognosis and fewer cancer prevention and treatment strategies compared to ER-positive tumors.

**Progesterone Status:** 0 = Progesterone positive and 1 = Progesterone negative

1. Progesterone positive:- This type of breast cancer is sensitive to progesterone, and the cells have receptors that allow them to use this hormone to grow. Treatment with endocrine therapy blocks the growth of the cancer cells.
2. Progesterone negative: - This type of breast cancers have no estrogen or progesterone receptors. Treatment with hormone therapy drugs is not helpful for these cancers. These cancers tend to grow faster than hormone receptor-positive cancers.

**Status:** 0 = Alive and 1 = dead

Hormone receptor-negative (or hormone-negative): - breast cancers have no estrogen or progesterone receptors. Treatment with hormone therapy drugs is not helpful for these cancers. These cancers tend to grow faster than hormone receptor-positive cancers. If they come back after treatment, it’s often in the first few years