Pertemuan ke-14 Kecerdasan Buatan (Bagian Pertama)

Mathematics Basics

Probability and Statistics

Presented by:
Prof. Dr. Ir. R. Rizal Isnanto, S.T., M.M., M.T., IPU, ASEAN Eng.

Foreword

- This document introduces the mathematics basics used in AI, including linear algebra, probability and statistics as well as optimization problem.

Contents

Statistical Basics

- Population: The set of data (numeric or otherwise) corresponding to the entire collection of units about which information is sought.
- Sample: A subset of the population data that are actually collected in the course of a study.
- Example:
  - Population--Blood pressure readings of ALL people in China.
  - Sample--Blood pressure readings of 1000 randomly selected people in China.
- In most studies, it is difficult to obtain information about the whole population. That is why we rely on samples to make estimates and inferences related to the whole population.

Parameters vs Statistics

- A parameter is a number that describes a population.
- A statistic is a number that describes a sample.
- Parameters are usually denoted using Greek letters $\{\mu, \sigma\}$ while statistics are usually denoted using Roman letters $\{x, s\}$.
- A parameter is a fixed number (usually unknown). A statistic is a variable whose value varies from sample to sample.

Descriptive Statistics

- Many methods are available for summarising data in numeric form.
- Numeric
  - Measure of location

    Mean, Median, Mode
  - Measure of spread

    Standard deviation, MAD (median absolute deviation), IQR
  - Others:

    Min, Max, Quartile, Five number summaries (used later in boxplot)

Measure of Location

- Consider n samples of data drawn from population. $\{x_{1,} x_2, \dots, x_n\}$
- Sample mean: the sum of all the observations divided by the number of observations. It is written in symbols as:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Sample median:
  - The $\frac{(n+1)}{2}^{th}$ largest observation if n is odd.
  - The average of the $(\frac{n}{2})^{th}$ and $(\frac{n}{2}+1)^{th}$ largest observation if n is even.
- Sample mode: the most frequently occurring value among all the observations in a sample.

Median or Mean?

- Both the median and the mean are measures of location, but which is preferable?
- For symmetric data, the mean is usually less variable from sample to sample than the median.
- For skewed data, the median is a better measure of location.
- The median does not reacted as much as the mean by outliers. This property of the median is known as ‘robustness’.
- The mean is easier to compute than the median and is much easier to handle theoretically.

Measure of Spread-Standard deviation

- The **standard deviation (SD)** or **variance** measures how spread out the data are around the mean.
- Steps to calculate sample variance
  1. Find the mean of the data.
  2. Make a list of **deviations** from the mean.
  3. Calculate the average of the squares of deviations.

$$var = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}, \qquad SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

- The sample SD is defined as: $SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$

MEDIAN ABSOLUTE DEVIATION (MAD)

In statistics, the **median absolute deviation** (**MAD**) is a robust measure of the variability of a univariate sample of quantitative data. It can also refer to the population parameter that is estimated by the MAD calculated from a sample.[1]

For a univariate data set X_1, X_2, ..., X_n, the MAD is defined as the median of the absolute deviations from the data's median $\tilde{X} = \operatorname{median}(X)$:

$$\operatorname{MAD} = \operatorname{median}(|X_i - \tilde{X}|)$$

that is, starting with the residuals (deviations) from the data's median, the MAD is the median of their absolute values.

Example [edit]

Consider the data (1, 1, 2, **2**, 4, 6, 9). It has a median value of 2. The absolute deviations about 2 are (1, 1, 0, 0, 2, 4, 7) which in turn have a median value of 1 (because the sorted absolute deviations are (0, 0, 1, **1**, 2, 4, 7)). So the median absolute deviation for this data is 1.

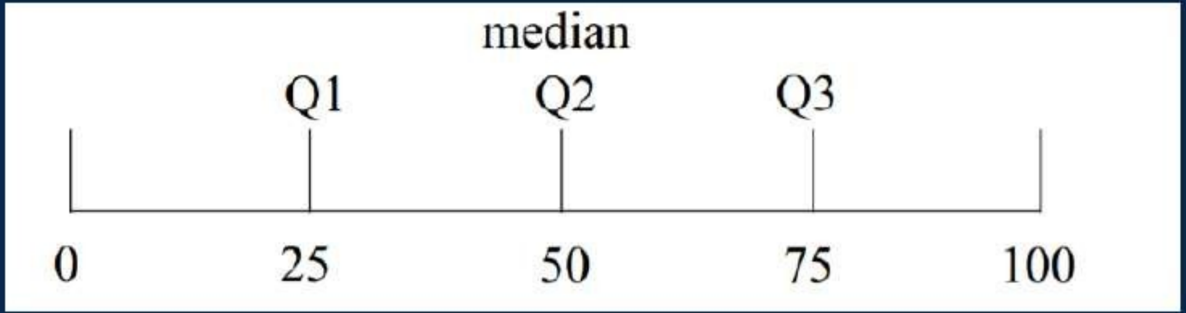
Uses [edit]

The median absolute deviation is a measure of statistical dispersion. Moreover, the MAD is a robust statistic, being more resilient to outliers in a data set than the standard deviation. In the standard deviation, the distances from the mean are squared, so large deviations are weighted more heavily, and thus outliers can heavily influence it. In the MAD, the deviations of a small number of outliers are irrelevant.

Because the MAD is a more robust estimator of scale than the sample variance or standard deviation, it works better with distributions without a mean or variance, such as the Cauchy distribution.

Measure of Spread – IQR

- The three quartiles, Q1, Q2, and Q3, approximately divide an ordered data set into four equal parts.



- The **Inter-quartile range (IQR)** is defined as the upper quartile (Q3 ;75th percentile) minus the lower quartile (Q1; 25th percentile). It is the width of the interval that contains the middle 50% of the data

$$IQR = Q3 - Q1$$

SD vs IQR

- Similar to median vs mean
- Sample standard deviations and the IQR $(= Q3 - Q1)$ are both measures of spread. The IQR is robust, like the median, but it is harder to handle theoretically than the standard deviation.

Probability Basics: Experiment, Sample Space, Probability

- A **random experiment** must satisfy the following properties:
  - can be repeated under the same conditions with more than one possible outcome.
  - The outcome of each trial cannot be predicted with certainty.
- **Sample space** S: The set of all the possible outcomes of a random experiment.
- **Sample point**: each possible outcome of a random experiment.
- **Random event A**: a set of possible outcomes of the experiment, thus A is the subset of S. $P(A) = \sum_{i \in A} p_i$
- Example:
  - Random experiment E_1: Observe the possible outcomes of numbers when a dice is rolled.
  - Sample space: $S = \{1, 2, 3, 4, 5, 6\}$
  - Sample point: $e_i = 1, 2, 3, 4, 5, 6$
  - Random event A_1: "The outcome is 5" can be denoted as $A_1 = \{x | x = 5\}$.

Conditional Probability and Bayes' Theorem

- Conditional probability of Y given X has occurred, assuming $P(X) \neq 0$

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

- **Bayes' Theorem**:

$$P(Y|X) = \frac{P(YX)}{P(X)} ;\ P(X|Y) = \frac{P(XY)}{P(Y)}$$

$$\because P(YX) = P(XY)$$

$$\therefore P(Y|X)P(X) = P(X|Y)P(Y)$$

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

$P(X|Y)$ indicates $Posterior$; $P(Y|X)$ indicates $Likelihood$; $P(X)$ indicates $Prior$; $P(Y)$ indicates $Normalizer$.

$$P(hypothesis|data) = \frac{P(data|hypothesis)P(hypothesis)}{P(data)}$$

Conditional Probability and Bayes' Theorem

- If X is a probability space $\{X_1, X_2, \ldots, X_n\}$ consisting of mutually independent events, then $P(Y)$ can be calculated by the **total probability theorem**

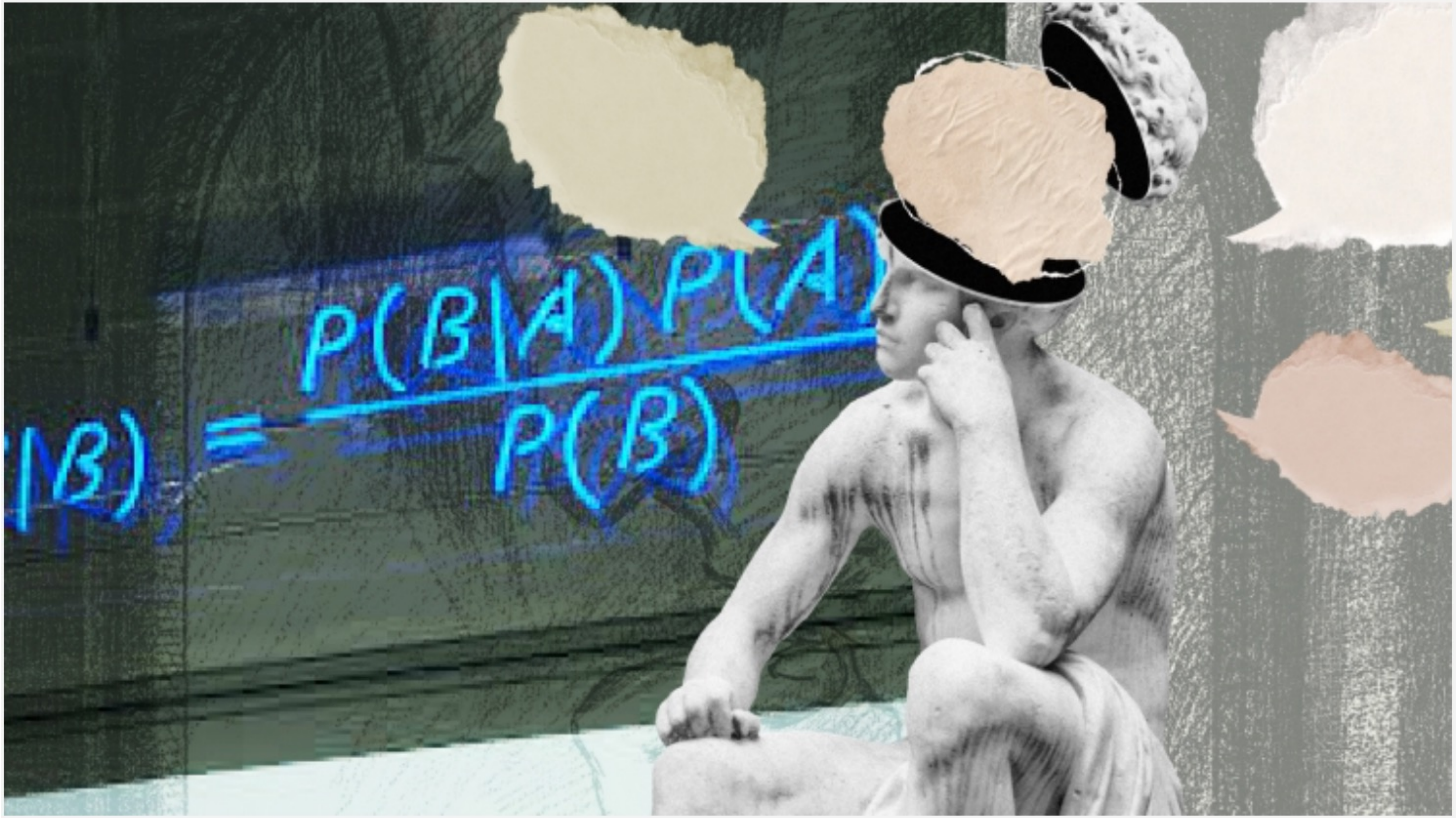
$$P(Y) = P(Y|X_1)P(X_1) + P(Y|X_2)P(X_2) + \cdots + P(Y|X_n)P(X_n)$$

- **Bayes' Theorem**:

$$P(X_i|Y) = \frac{P(Y|X_i)P(X_i)}{\sum_{i=1}^{n} P(Y|X_i)P(X_i)}$$

- Application of Bayes' Theorem: statistical machine translation, Bayesian network, and others.

Dari Teorema Bayes ke Aplikasi Dunia Nyata: Memahami Klasifikasi Naive Bayes

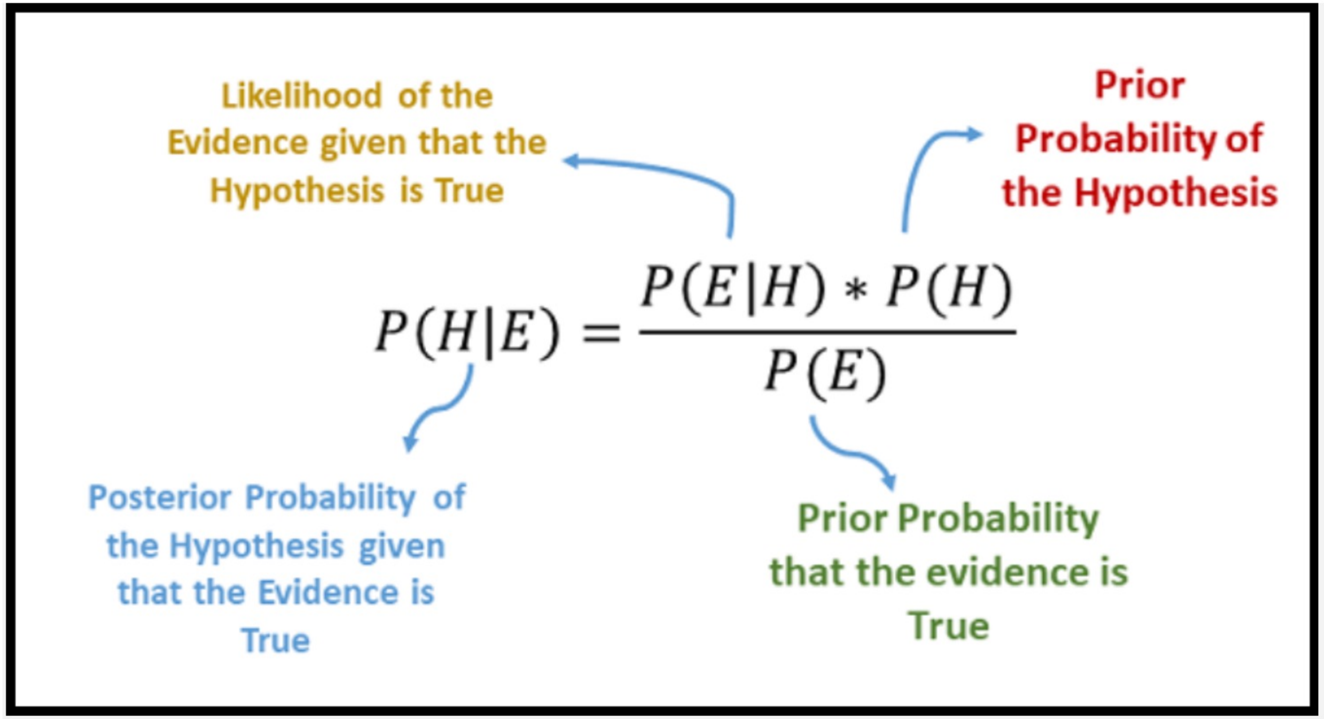


Supervised Learning merupakan salah satu pilar utama dalam machine learning yang membantu memberikan prediksi yang tepat. Terutama, klasifikasi melakukan prediksi berdasarkan fitur-fitur tertentu. ***Naive Bayes*** adalah salah satu metode klasifikasi yang cukup populer karena sederhana dan peran efektivitasnya dalam beberapa **penerapannya pada dunia nyata**, seperti *email spam detection*, *weather prediction*, *medical diagnosis*, *text classification*, dan *recommendation system*. Artikel ini secara khusus akan membahas terkait ***Naive Bayes*** dari dasar hingga penerapan praktis

Pendahuluan tentang Naive Bayes

Mari kita jelajahi **algoritma *Naive Bayes***, sebuah teknik klasifikasi yang berdasarkan pada Teorema Bayes dengan asumsi bahwa semua fitur yang memprediksi nilai target adalah **independen satu sama lain**. Algoritma ini menghitung probabilitas setiap kelas dan kemudian memilih yang memiliki probabilitas tertinggi. Algoritma ini berasal dari karya **Thomas Bayes** dan **prinsip-prinsip Bayesian**, terutama probabilitas kondisional. "Naive" di sini berarti asumsi bahwa fitur-fitur bersifat independen dan tidak berpengaruh terhadap satu sama lain. Untuk memahami cara kerja Naive Bayes, kita akan melihat penerapannya dalam kehidupan nyata. Misalnya, kita akan menggunakan contoh sederhana: ***email* spam detection**. Melalui ilustrasi ini, kita akan memahami proses langkah demi langkah bagaimana Naive Bayes menghitung probabilitas dan menggunakan Teorema Bayes untuk klasifikasi.

Teorema Bayes yang ditemukan oleh ***Thomas Bayes***, adalah dasar dari algoritma Naive Bayes. Teorema ini membantu kita dalam proses **menghitung probabilitas (peluang)** suatu peristiwa berdasarkan informasi yang sudah kita ketahui sebelumnya. Secara sederhana, Teorema Bayes menyatakan bahwa kita bisa memperbarui keyakinan kita tentang suatu peristiwa dengan menggunakan bukti baru. Secara matematis, Teorema Bayes ditulis sebagai:



$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Di sini, P(H|E) adalah probabilitas H terjadi diberikan E, P(E|H) adalah probabilitas E terjadi diberikan H, P(H) adalah probabilitas H terjadi, dan P(E) adalah probabilitas E terjadi. Dalam konteks deteksi email spam, kita bisa menggunakan teorema ini untuk menghitung probabilitas sebuah email merupakan spam berdasarkan kata-kata yang ada di dalamnya. Misalnya, kita ingin mengetahui apakah email dengan kata "gratis" adalah spam atau bukan. Dengan Teorema Bayes, kita bisa menghitung probabilitas email itu adalah spam jika mengandung kata "gratis" dengan melihat seberapa sering kata "gratis" muncul di email spam dibandingkan dengan email non-spam.

Mari kita bayangkan ada banyak email yang sudah kita labeli sebagai spam atau bukan. Kita bisa menghitung P(spam), yaitu probabilitas suatu email adalah dikategorikan sebagai spam, dan **P(gratis|spam)**, yaitu probabilitas kata "gratis" muncul dalam email spam. Begitu juga kita bisa menghitung **P(non−spam)** dan **P(gratis|non−spam)**.

Dengan berdasarkan data ini, **kita bisa menghitung apakah email baru dengan kata "gratis" kemungkinan besar adalah spam atau bukan,** atau dengan kata lain soal dituliskan sebagai:

Seberapa peluang email merupakan spam jika ada kata “gratis”-nya, yang dinotasikan sebagai: **P(spam|gratis)**?

Proses ini menjadi dasar dari banyak aplikasi Naive Bayes lainnya, seperti filter spam di email, prediksi cuaca, dan diagnosis medis. Untuk memahami lebih dalam bagaimana persamaan matematika ini bekerja dalam proses perhitungannya, mari kita lanjut ke poin berikutnya!

Bagaimana cara kerjanya? Sebagai contoh, jika kita memiliki 1000 email, di mana 500 adalah spam dan 500 sisanya adalah non-spam. Dari 500 email spam, 100 diantaranya mengandung kata "gratis", sedangkan dari 500 email non-spam, hanya ditemukan 10 yang mengandung kata "gratis". Berdasarkan itu, kita bisa menghitung dengan cara berikut:

P(spam) = 500/1000 = 0.5

P(non-spam) = 500/1000 = 0.5

P(gratis|spam) = 100/500 = 0.2

P(gratis|non-spam) = 10/500 = 0.02

Dengan pemanfaatan Teorema Bayes, kita dapat melakukan proses perhitungan terhadap P(spam|gratis) dengan sebagai berikut:

P(spam|gratis) = P(gratis|spam).P(spam) / P(gratis)

Untuk menghitung P(gratis), kita gunakan:

P(gratis) = P(gratis|spam).P(spam) + P(gratis|non-spam).P(non-spam) P(gratis) = (0.2*0.5) + (0.02*0.5) = 0.11. Akhirnya, kita bisa menghitung P(spam|gratis) sebagai berikut.

P(spam|gratis) = (0.2*0.5) / 0.11 = 0.91

Jadi, probabilitas bahwa email dengan kata "gratis" adalah spam sekitar 91%.

Independence

- Two events A and B are independent if

$$P(YX) = P(Y)P(X)$$

- If X and Y are independent then Independence

$$P(Y \mid X) = \frac{P(YX)}{P(X)} = \frac{P(\mathrm{Y})P(X)}{P(X)} = P(Y)$$

Expectation and Variance

- **Mathematical expectation:** the sum of: [(each of the possible outcomes) × (the probability of the outcome occurring)]. The expectation is one of the most basic mathematical characteristics of a probability distribution. It represents the average value of a random variable.
  - For a discrete random variable, $E(X) = \sum_{k=1}^{\infty} x_k p_k, k = 1, 2, \ldots.$
  - For a continuous random variable, $E(X) = \int_{-\infty}^{\infty} xf(x)dx.$
- **Variance:** measures the dispersion of random variables or a set of numbers. This means that it measures the gap between random variables and their expectations.

$$D(X) = Var(X) = E\{[X - E(X)]^2\}$$

In addition, $\sqrt{D(X)}$ is often represented by $\sigma(X)$, and is called the standard deviation or mean square deviation. $X^* = \frac{X-E(X)}{\sigma(X)}$ is called the standardized variable of X.

Covariance, Correlation Coefficient, and Covariance Matrix

- **Covariance**: somewhat measures the linear relationship between two variables.

$$Cov(X,Y) = E\left[\left(X - E(X)\right)\left(Y - E(Y)\right)\right]$$

- **Correlation coefficient**: sometimes called linear correlation coefficient, which measures the linear relationship between two variables.

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

- The **covariance matrix** of a random variable (X_1, X_2):

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

where $c_{ij} = Cov\left(X_i, X_j\right) = E\left\{[X_i - E(X_i)]\left[X_j - E\left(X_j\right)\right]\right\}, i, j = 1, 2, \cdots, n$. The elements along the diagonal of the covariance matrix are the variance of X_1, X_2, and the rest are the covariance of X_1, X_2.

TUGAS MANDIRI
MATAKULIAH KECERDASAN BUATAN
“MATHEMATICS BASIC”

Pertemuan ke-14

Prof. Dr. Ir. R. Rizal Isnanto, S.T., M.M., M.T., IPU, ASEAN Eng.

TUGAS KE-3

Departemen Teknik Komputer
Universitas Diponegoro
Semarang, Indonesia

Mathematics Basics 2

Statistika

1. Rerata (*mean*)

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Notice the symbol $\bar{X}$ (said "X bar") to indicate the mean of the set X. All this formula says is "Add up all the numbers and then divide by how many there are".

Unfortunately, the mean doesn't tell us a lot about the data except for a sort of middle point. For example, these two data sets have exactly the same mean (10), but are obviously quite different:

$$[0\ 8\ 12\ 20]\ and\ [8\ 9\ 11\ 12]$$

So what is different about these two sets? It is the *spread* of the data that is different. The Standard Deviation (SD) of a data set is a measure of how spread out the data is.

2. Simpangan Baku (Standard Deviation)

3

How do we calculate it? The English definition of the SD is: "The average distance from the mean of the data set to a point". The way to calculate it is to compute the squares of the distance from each data point to the mean of the set, add them all up, divide by $n-1$, and take the positive square root. As a formula:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}}$$

Set 1:

X	$(X - \bar{X})$	$(X - \bar{X})^2$
0	-10	100
8	-2	4
12	2	4
20	10	100
Total		208
Divided by (n-1)		69.333
Square Root		8.3266

Set 2:

X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
8	-2	4
9	-1	1
11	1	1
12	2	4
Total		10
Divided by (n-1)		3.333
Square Root		1.8257

3. Varians

4

Variance is another measure of the spread of data in a data set. In fact it is almost identical to the standard deviation. The formula is this:

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}$$

You will notice that this is simply the standard deviation squared, in both the symbol (s^2) and the formula (there is no square root in the formula for variance). s^2 is the usual symbol for variance of a sample. Both these measurements are measures of the spread of the data. Standard deviation is the most common measure, but variance is also used. The reason why I have introduced variance in addition to standard deviation is to provide a solid platform from which the next section, covariance, can launch from.

Exercises PR no. 1, 2, dan 3

Find the mean (a), standard deviation (b), and variance (c) for each of these data sets.

1 • [12 23 34 44 59 70 98]

2 • [12 15 25 27 32 88 99]

3 • [15 35 78 82 90 95 97]

1\. a.
b.
c.

2\. a.
b.
c.

3\. a.
b.
c.

4. Kovarians (covariance) (1) 5

The last two measures we have looked at are purely 1-dimensional. Data sets like this could be: heights of all the people in the room, marks for the last COMP101 exam etc. However many data sets have more than one dimension, and the aim of the statistical analysis of these data sets is usually to see if there is any relationship between the dimensions. For example, we might have as our data set both the height of all the students in a class, and the mark they received for that paper. We could then perform statistical analysis to see if the height of a student has any effect on their mark.

Standard deviation and variance only operate on 1 dimension, so that you could only calculate the standard deviation for each dimension of the data set *independently* of the other dimensions. However, it is useful to have a similar measure to find out how much the dimensions vary from the mean *with respect to each other*.

Kovarians (covariance) (2)

6

Covariance is such a measure. Covariance is always measured *between* 2 dimensions. If you calculate the covariance between one dimension and *itself*, you get the variance. So, if you had a 3-dimensional data set (x, y, z), then you could measure the covariance between the x and y dimensions, the x and z dimensions, and the y and z dimensions. Measuring the covariance between x and x, or y and y, or z and z would give you the variance of the x, y and z dimensions respectively.

The formula for covariance is very similar to the formula for variance. The formula for variance could also be written like this:

$$var(X) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})}{(n-1)} = cov(x,x)$$

where I have simply expanded the square term to show both parts. So given that knowledge, here is the formula for covariance:

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

Kovarians (covariance) (3)

You might ask "is $cov(X, Y)$ equal to $cov(Y, X)$"? Well, a quick look at the formula for covariance tells us that yes, they are exactly the same since the only difference between $cov(X, Y)$ and $cov(Y, X)$ is that $(X_i - \bar{X})(Y_i - \bar{Y})$ is replaced by $(Y_i - \bar{Y})(X_i - \bar{X})$. And since multiplication is commutative, which means that it doesn't matter which way around I multiply two numbers, I always get the same number, these two equations give the same answer.

	Hours(H)	*Mark(M)*
Data	9	39
	15	56
	25	93
	14	61
	10	50
	18	75
	0	32
	16	85
	5	42
	19	70
	16	66
	20	80
Totals	167	749
Averages	$\boxed{13.92}$	$\boxed{62.42}$

Covariance:

H	M	$(H_i - \bar{H})$	$(M_i - \bar{M})$	$(H_i - \bar{H})(M_i - \bar{M})$
9	39	-4.92	-23.42	115.23
15	56	1.08	-6.42	-6.93
25	93	11.08	30.58	338.83
14	61	0.08	-1.42	-0.11
10	50	-3.92	-12.42	48.69
18	75	4.08	12.58	51.33
0	32	-13.92	-30.42	423.45
16	85	2.08	22.58	46.97
5	42	-8.92	-20.42	182.15
19	70	5.08	7.58	38.51
16	66	2.08	3.58	7.45
20	80	6.08	17.58	106.89
Total				1149.89
Average				104.54

5. Matriks Kovarians (1)

Recall that covariance is always measured between 2 dimensions. If we have a data set with more than 2 dimensions, there is more than one covariance measurement that can be calculated. For example, from a 3 dimensional data set (dimensions x, y, z) you could calculate $cov(x, y)$, $(cov(x, z)$, and $cov(y, z)$. In fact, for an n-dimensional data set, you can calculate $\frac{n!}{(n-2)!*2}$ different covariance values.

A useful way to get all the possible covariance values between all the different dimensions is to calculate them all and put them in a matrix. I assume in this tutorial that you are familiar with matrices, and how they can be defined. So, the definition for the covariance matrix for a set of data with n dimensions is:

$$C^{n \times n} = (c_{i,j}, c_{i,j} = cov(Dim_i, Dim_j)),$$

where $C^{n \times n}$ is a matrix with n rows and n columns, and Dim_x is the xth dimension. All that this ugly looking formula says is that if you have an n-dimensional data set, then the matrix has n rows and columns (so is square) and each entry in the matrix is the result of calculating the covariance between two separate dimensions. Eg. the entry on row 2, column 3, is the covariance value calculated between the 2nd dimension and the 3rd dimension.

Matriks Kovarians (2)

An example. We'll make up the covariance matrix for an imaginary 3 dimensional data set, using the usual dimensions x, y and z. Then, the covariance matrix has 3 rows and 3 columns, and the values are this:

$$C = \begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix}$$

Some points to note: Down the main diagonal, you see that the covariance value is between one of the dimensions and itself. These are the variances for that dimension. The other point is that since $cov(a,b) = cov(b,a)$, the matrix is symmetrical about the main diagonal.

PR No. 4 dan 5

10

Work out the covariance between the x and y dimensions in the following 2 dimensional data set, and describe what the result indicates about the data.

4

Item Number:	1	2	3	4	5
x	10	39	19	23	28
y	43	13	32	21	20

cov (x,y).

Calculate the covariance matrix for this 3 dimensional set of data.

5

Item Number:	1	2	3
x	1	-1	4
y	2	1	3
z	1	3	-1

$$\boxed{C = \begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix}}$$

Kecerdasan Buatan

Pertemuan 14 (Bagian kedua)

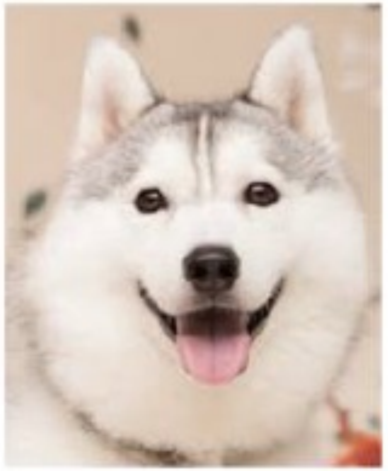
Metode Klasifikasi menggunakan *Confusion Matrix*

Prof. Dr. Ir. R. Rizal Isnanto, S.T., M.M., M.T., IPU, ASEAN Eng.

Departemen Teknik Komputer
Universitas Diponegoro

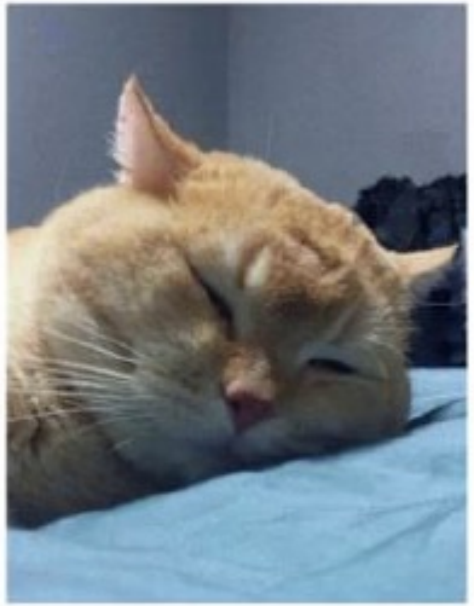
Image Classification (1)

- Image classification is a basic research topic in the AI field and a core issue in the computer vision field.
- The image classification operation is performed based on image processing technologies to determine the category of an input image.
- Classify dogs by breeds:



- Husky, Golden retriever, and Border collie

Image Classification (2)



- Confidence: used to measure the reliability of the image classification result.
- Whether a husky or a border collie is shown in the left figure?

  HUAWEI

Classification Performance Measurement (1)

- Ground truth
  - In image processing, the actual label in a dataset is called the ground truth (GT).
- Accuracy
  - Ratio of the number of correct samples to the total number of samples.
- Error rate
  - Ratio of the number of incorrect samples to the total number of samples.

Classification Performance Measurement (2)

- Terms and Definitions
  - P: condition positive, indicating the number of real positive cases in the data.
  - N: condition negative, indicating the number of real negative cases in the data.
  - TP: true positive, indicating the number of positive cases that are correctly classified by the classifier.
  - TN: true negative, indicating the number of negative cases that are correctly classified by the classifier.
  - FP: false positive, indicating the number of positive cases that are incorrectly classified by the classifier.
  - FN: false negative, indicating the number of negative cases that are incorrectly classified by the classifier.

Actual \ Prediction	Yes	No	Total
Yes	TP	FN	P
No	FP	TN	N
Total	P'	N'	$P+N$

Confusion matrix

- Confusion matrix: at least an $m \times m$ table. $CM_{i,j}$ indicates the number of cases that actually belong to class i but are classified into class j by the classifier.
  - Ideally, for a high accuracy classifier, most prediction values should be located in the diagonal of the table (from $CM_{1,1}$ to $CM_{m,m}$) while values outside the diagonal are 0 or close to 0. That is, the sum of FP and FN is close to 0.

Classification Performance Measurement (3)

Indicator	Formula
Accuracy and recognition rate	$\frac{TP+TN}{P+N}$
Error rate and misclassification rate	$\frac{FP+FN}{P+N}$
Sensitivity, true positive rate, and recall rate	$\frac{TP}{P}$
Specificity and true negative rate	$\frac{TN}{N}$
Precision	$\frac{TP}{TP+FP}$
F, F_1, F-score (harmonic mean of the recall rate and precision)	$\frac{2 \times precision \times recall}{precision + recall}$
F_β, where β is a non-negative real number.	$\frac{(1+\beta^2) \times precision \times recall}{\beta^2 \times precision + recall}$

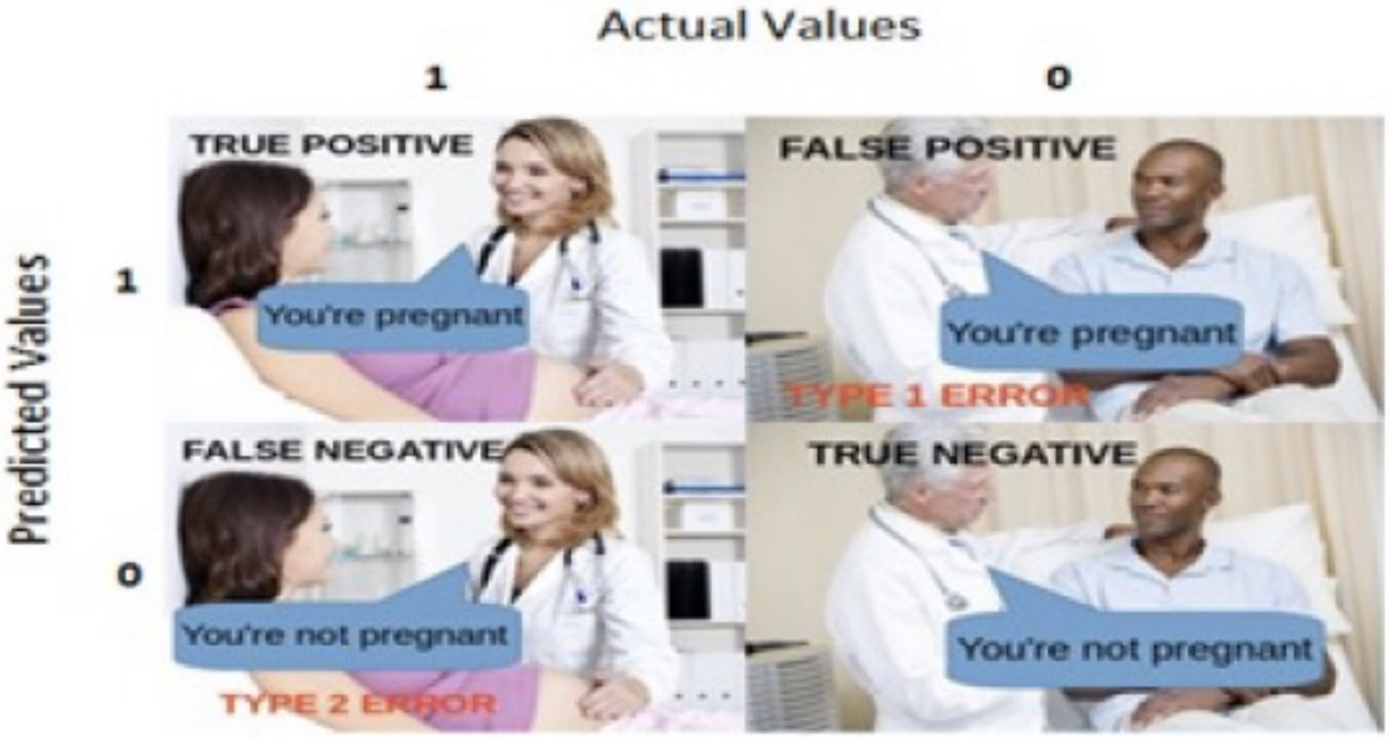
HUAWEI

Confusion Matrix

Confusion Matrix

Machine Learning merupakan salah satu cabang dari disiplin ilmu kecerdasan buatan (*artificial intelligence*) yang membahas bagaimana sistem dibangun berdasarkan pada data. Jadi *machine learning* merupakan proses komputer untuk belajar dari data (*learn from data*). Jika tidak ada data, komputer tidak akan bisa belajar. Salah satu teknik aplikasi pada *machine learning* adalah *supervised learning*. Klasifikasi merupakan *supervised learning*, yang merupakan model prediksi dimana hasil prediksinya bersifat diskrit. Bagaimana mengukur performa dari model klasifikasi yang digunakan ? Jawaban sederhana nya adalah membandingkan nilai aktual dengan nilai prediksi. *Confusion Matrix* adalah pengukuran performa untuk masalah klasifikasi *machine learning* dimana keluaran dapat berupa dua kelas atau lebih.

Confusion Matrix adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Ada empat istilah yang merupakan representasi hasil proses klasifikasi pada *confusion matrix* yaitu True Positif, True Negatif, False Positif, dan False Negatif. Mari kita pahami apa itu True Positif, False Positif, False Negatif, dan True Negatif dalam analogi kehamilan (dapat dilihat pada Gambar 1).



Gambar 1. Contoh Confusion Matrix

- True Positive (TP) :
  Interpretasi: Anda memprediksi positif dan itu benar.
  Anda memprediksikan bahwa seorang wanita hamil dan wanita tsb memang benar hamil.

- True Negative (TN):
  Interpretasi: Anda memprediksi negatif dan itu benar.
  Anda memprediksikan bahwa seorang pria tidak hamil dan benar ya pria kan tidak mungkin hamil :D.

- False Positive (FP): (Kesalahan Tipe 1)
  Interpretasi: Anda memprediksi positif dan itu salah.
  Anda memprediksikan bahwa seorang pria hamil tetapi tidak mungkin pria bisa hamil :D.

- False Negative (FN): (Kesalahan Tipe 2, kesalahan tipe 2 ini sangat berbahaya)
  Interpretasi: Anda memprediksi negatif dan itu salah.
  Anda memperkirakan bahwa seorang wanita tidak hamil tetapi sebenarnya wanita tsb hamil.

Nilai Prediksi vs Nilai Aktual

Dari contoh di atas dapat digambarkan bahwa:

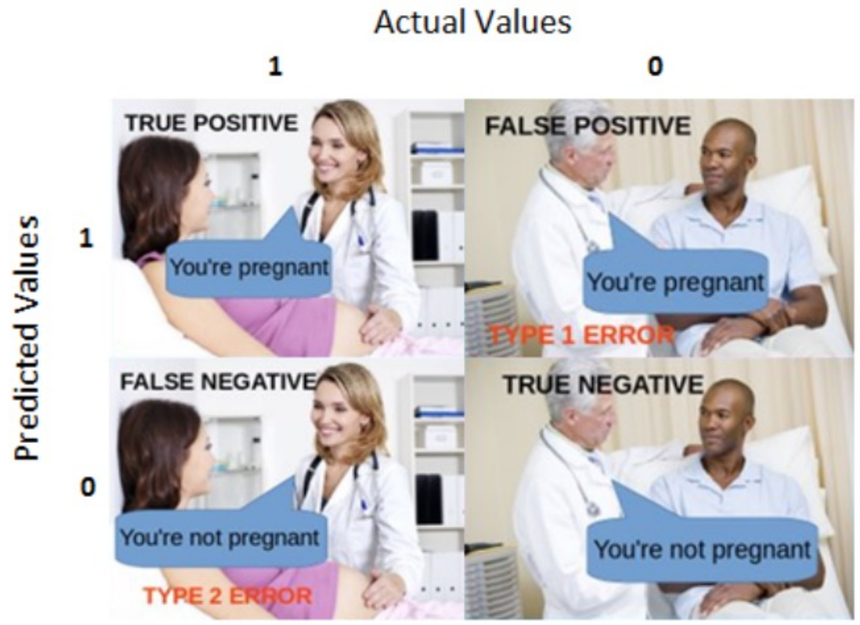
- Nilai Prediksi adalah keluaran dari program dimana nilainya Positif dan Negatif
- Nilai Aktual adalah nilai sebenarnya dimana nilainya True dan False

Contoh Confusion Matrix :

Suatu perusahaan membuat sebuah model yang dilatih untuk memprediksi apakah seorang karyawan di perusahaan tsb terkena covid-19 atau tidak. Dengan asumsi perusahaan tsb mempunyai 175 karyawan. Dari model *classifier* memprediksi karyawan positif covid-19 sebanyak 145 dan karyawan negatif covid-19 sebanyak 30 tetapi pada kenyataannya, karyawan positif covid-19 sebanyak 150 dan karyawan negatif covid-19 sebanyak 25 (dapat dilihat pada Tabel 1).

Tabel 1 Contoh Confusion Matrix

n= 175	Aktual : Positif (1)	Aktual : Negatif (0)
Prediksi : Positif (1)	TP: 125	FP: 20
Prediksi : Negatif (0)	FN: 25	TN: 5
	150	25



Confusion Matrix [Image 3] (Image courtesy: My Photoshopped Collection)

Pada kasus ini :

- True Positive (TP): kita memprediksi karyawan positif covid-19 dan memang benar karyawan tsb positif covid-19
- True Negative (TN): kita memprediksi karyawan negative covid-19 dan memang benar karyawan tsb negatif covid-19.
- False Positive (FP):kita memprediksi karyawan positif covid-19 dan ternyata prediksi salah, ternyata karyawan negatif covid-19
- False Negative (FN) : kita memprediksi karyawan negatif covid-19 dan ternyata prediksi salah, ternyata karyawan tsb positif covid-19. Seperti telah dijelaskan di atas bahwa FN merupakan kesalahan tipe 2 dimana kesalahan ini sangat berbahaya. Contoh : karyawan di prediksi negatif covid-19 padahal ternyata karyawan positif covid-19 maka karyawan tersebut terlambat mengetahui keadaan sebenarnya sehingga tidak segera dilakukan tindakan pencegahan pengobatan dan isolasi mandiri. Dimana karyawan dapat menularkan virus covid-19 ke banyak karyawan lainnya dalam perusahaan dan dapat menyebabkan kematian untuk karyawan lain juga karyawan tsb.

Tabel 1 Contoh Confusion Matrix

n= 175	Aktual : Positif (1)	Aktual : Negatif (0)
Prediksi : Positif (1)	TP: 125	FP: 20
Prediksi : Negatif (0)	FN: 25	TN: 5
	150	25

maka dapat dihitung nilai *accuracy*, *precision*, *recall* dan F-1 score.

- *Accuracy*mennggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar

Accuracy = (TP+TN) / (TP+FP+FN+TN)

= (125+5)/ (125+20+25+5)

= 0.742

= 0.742 *100 % = 74.2 %

- *Precision* menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model. *Precision* = (TP) / (TP + FP)

= 125 / (125 + 20)

= 0.86

= 0.86 * 100% = 86%

- *Recall* atau *sensitivity*: menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi. *Recall* = TP / (TP + FN)

Tabel 1 Contoh Confusion Matrix

n= 175	Aktual : Positif (1)	Aktual : Negatif (0)
Prediksi : Positif (1)	TP: 125	FP: 20
Prediksi : Negatif (0)	FN: 25	TN: 5
	150	25

= 125 /(125+25)

= 0.83

= 0.83 * 100% = 83%

- F-1 Score menggambarkan perbandingan rata-rata precision dan recall yang dibobotkan. Accuracy tepat kita gunakan sebagai acuan performansi algoritma jika dataset kita memiliki jumlah data False Negatif dan False Positif yang sangat mendekati (*symmetric*). Namun jika jumlahnya tidak mendekati, maka sebaiknya kita menggunakan F1 Score sebagai acuan.

F-1 Score = (2 * *Recall* * *Precision*) / (*Recall* + *Precision*)

= (2* 0.83 * 0.86) / (0.83 + 0.86)

= 1.4276 / 1.69

= 0.84 *100%

= 84%

Referensi :

1. Sarang Narkhede. Understanding Confusion Matrix. May 2018. https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62.

2. Jianfeng Xu, Yuanjian Zhang, Duogian Miao.Three-way confusion matrix for classification: A measure driven view. Information Sciences Volume 507, January 2020. Elsevier. https://doi.org/10.1016/j.ins.2019.06.064

PR no. 6

Sebuah program studi di suatu universitas membuat sebuah model untuk memprediksi apakah seorang lulusan prodi bisa diterima di bidang yang sesuai dengan kompetensi yang diajarkan di Prodi tersebut. Dengan asumsi Prodi tersebut memiliki 225 alumni, model *classifier* memprediksi alumni yang bekerja sesuai kompetensi sebanyak 175 orang dan alumni yang bekerja tidak sesuai kompetensi sebanyak 50 orang. Namun, pada kenyataannya, alumni yang bekerja sesuai kompetensi sebanyak 200 orang dan yang tidak sesuai kompetensi sebanyak 25 orang. Data rinci dapat dilihat pada tabel di balik ini.

n = 225	Aktual Sesuai kompetensi (1)	Aktual Tidak sesuai kompetensi (0)
Prediksi Sesuai kompetensi (1)	160	15
Prediksi Tidak sesuai kompetensi (0)	40	10
	200	25

Pertanyaan:

1. Berapakah nilai TP, TN, FP, dan FN-nya?
2. Carilah nilai *Accuracy*-nya
3. Carilah nilai *Precision*-nya
4. Carilah nilai *Recall* atau *Sensitivity*-nya
5. Carilah nilai *F-1 score*-nya

Klasifikasi *Confusion Matrix* untuk kasus 3 Kelas.

Nah, untuk penjelasan yang lebih sederhana dan mudah di mengerti, saya akan mencoba memberi contoh kasus pada 3 kelas dengan nama yang berbeda yaitu *jeruk, pepaya dan anggur.*

Confusion Matrix 3x3

	Prediksi		
Aktual	Jeruk	Pepaya	Anggur
Jeruk	7	8	9
Pepaya	1	2	3
Anggur	3	2	1

Tabel Confusion Matrix 3x3

Pada contoh di samping adalah kasus penyelesaian 3 kelas yang sudah terisi nilainya. Tidak seperti pada klasifikasi biner diatas, untuk penyelesaian daripada 3 kelas ini adalah kamu harus mencari nilai **TP, TN, FP** dan **FN** pada masing-masing kelas. Oke, daripada akan semakin bingung , kita akan mencoba menghitung nilai ***Accuracy*** dari tabel disamping.

Confusion Matrix 3x3

	Prediksi		
Aktual	Jeruk	Pepaya	Anggur
Jeruk	7	8	9
Pepaya	1	2	3
Anggur	3	2	1

Rumus ***Accuracy*** kali ini sedikit berbeda dengan rumus pada kasus binari klasifikasi. Pada kasus ini rumus yang digunakan adalah ***(TP / Jumlah Data).***
Untuk mencari nilai **TP** kamu harus mencari pada masing-masing kelas yang pada kondisi kelas **aktual** mampu di prediksi dengan benar. Contoh di atas adalah pada tabel yang bewarna **Biru**.

- **TP =** 7 + 2 + 1
- Jumlah Data = 36
- ***Accuracy* = TP /** Jumlah Data
- 10 / 36 = **0,277**

Menghitung *Precision*.

Untuk menghitung nilai presisi, kamu harus menghitung nilai presisi dari masing masing kelas kemudian menjumlahkan dan mencari nilai reratanya. Contoh penyelesaiannya ;

Rumus ***Precision = TP / (TP + FP)***

	Jeruk	Pepaya	Anggur
TP	7	2	1
FP	8+9	1 + 3	3 + 2
Precission	7 / (7 + 17) = **0,29**	2 / (2 + 4) = **0,33**	1 / (1 + 5) = **0,16**

selain jeruk selain pepaya selain anggur

Tabel disamping adalah hasil perolehan nilai *precision* dari masing-masing tiap kelasnya. Kemudian kamu tambahkan semua nilai terus **dibagi** jumlah kelas. Nah, ini sama hal nya dengan mencari nilai rerata dari semua presisi untuk dijadikan **satu** nilai presisi.

Precision Precision

- ***All Precision = Precision A + B + C / Jumlah Kelas***
- **(0,29 + 0,33 + 0,16) / 3 = 0, 67.**

Menghitung *Recall*

Sama halnya dengan menghitung nilai presisi tadi, untuk mencari nilai *recall* juga harus di tentukan nilai dari masing-masing kelasnya, yang kemudian akan ditentukan nilai rerata dari masing *recall* tersebut. Contoh penyelesaian ;

Rumus ***Recall*** / ***Sensitivity*** = ***TP / (TP + FN)***

	Jeruk	Pepaya	Anggur
TP	7	2	1
FN	1+3	8+2	9+3
Recall	7 / (7 + 4) = **0,63**	2 / (2 + 10) = **0,16**	1 / (1 + 11) = **0,08**

Setelah sudah ditentukan nilai masing-masing *recall* dari setiap kelas nya, seperti tadi kamu harus mencari nilai rerata dari semua nilai dengan menjumlahkan dan **dibagi** jumlah kelas.

Untuk F-1 Score bisa dicari sendiri.

Rumusnya adalah

Recall 1   Recall

- ***All Recall = Recall A + B + C / Jumlah Kelas***
- **(0,63 + 0,16 + 0,08) / 3 = 0,29.**

PR no. 7

Sebuah program studi di suatu universitas membuat sebuah model untuk memprediksi apakah seorang lulusan prodi bisa diterima di bidang yang sesuai dengan kompetensi yang diajarkan di Prodi tersebut. Dengan asumsi Prodi tersebut memiliki 300 alumni, model *classifier* diprediksi alumni yang bekerja sesuai kompetensi sebanyak 175 orang, alumni yang bekerja tidak sesuai kompetensi sebanyak 50 orang dan yang tidak bekerja 75 orang. Namun, pada kenyataannya, alumni yang bekerja sesuai kompetensi sebanyak 200 orang, yang tidak sesuai kompetensi sebanyak 30 orang, dan yang tidak bekerja 70 orang. Data rinci dapat dilihat pada tabel di balik ini.

n = 300	Aktual Sesuai kompetensi	Aktual Tidak sesuai kompetensi	Aktual Tidak bekerja
Prediksi Sesuai kompetensi	150	15	10
Prediksi Tidak sesuai kompetensi	30	10	10
Prediksi Tidak bekerja	20	5	50
	200	**30**	**70**

Sebuah program studi di suatu universitas membuat sebuah model untuk memprediksi apakah seorang lulusan prodi bisa diterima di bidang yang sesuai dengan kompetensi yang diajarkan di Prodi tersebut. Dengan asumsi Prodi tersebut memiliki 300 alumni, model *classifier* diprediksi alumni yang bekerja sesuai kompetensi sebanyak 175 orang, alumni yang bekerja tidak sesuai kompetensi sebanyak 50 orang dan yang tidak bekerja 75 orang. Namun, pada kenyataannya, alumni yang bekerja sesuai kompetensi sebanyak 200 orang, yang tidak sesuai kompetensi sebanyak 30 orang, dan yang tidak bekerja 70 orang. Data rinci dapat dilihat pada tabel di balik ini.

Pertanyaan:

1. Berapakah nilai TP, TN, FP, dan FN-nya?
2. Carilah nilai *Accuracy*-nya
3. Carilah nilai *Precision*-nya
4. Carilah nilai *Recall* atau *Sensitivity*-nya
5. Carilah nilai *F-1 score*-nya

PENGUMPULAN TUGAS

- Tugas Minggu ke-3 dikumpulkan di Google Classroom paling lambat Kamis 5 Juni 2025 pukul 23.59)
- Tugas-tugas yang belum dikumpulkan harap dikumpulkan meskipun ada pengurangan nilai
- Nilai tugas, PR, dan Kuis (jika ada) --> Total 70%, UTS dan UAS → Total 30%
- Terima kasih

Tentang UAS

- **Pelaksanaan UAS: Menunggu jadwal**
- **Sifat: Luring (datang ke kampus), buka catatan dan kalkulator, tutup laptop**
- Sifat Essay, materi: *Generative AI, Linear Algebra for AI, Statistics and Probability for AI*
- Waktu: 45 menit (di luar waktu UAS dari Pak Yudi Eko Windarto)
- Mahasiswa diminta sebelum mengumpulkan tugas ke Pengawas untuk:
  1. mengambil foto dari jawaban UAS
  2. melakukan merge-PDF dari butir 1
  3. unggah butir 2 ke Google Classroom
- Tambahan waktu untuk melakukan butir 1-3 → 10 menit

Penutup

- Demikian paparan saya sampaikan
- Ada pertanyaan?
- Terima kasih...