

1. Ada banyak metrik untuk mengukur akurasi. Rumus apa yang dipakai untuk mengukur akurasi ini?

Iye, Ditambah di halaman 4 :

Berdasarkan nilai *True Negative* (TN), *False Positive* (FP), *False Negative* (FN), dan *True Positive* (TP) dapat diperoleh nilai akurasi, presisi dan recall. Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan persamaan 4.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (4)$$

Berdasarkan nilai True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP) dapat diperoleh nilai akurasi, presisi dan recall. Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan Persamaan 1. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. Presisi dapat diperoleh dengan Persamaan 2. Sementara itu, recall menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem. Nilai recall diperoleh dengan Persamaan 3.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad \dots (1)$$

$$Presisi = \frac{TP}{FP + TP} \times 100\% \quad \dots (2)$$

$$Recall = \frac{TP}{FN + TP} \times 100 \quad \dots (3)$$

2. Kata Prediksi mau diubah menjadi Prakiraan. perlu diperjelas dalam tulisan, apakah proses prakiraan tersebut didasarkan atas prinsip "time-series" (sekuensial data) atau menggunakan prinsip "klasifikasi" (labelling data)

iye, ditambahkan di halaman 3

Percobaan dari penelitian dapat dilakukan sebuah evaluasi dengan pengukuran nilai akurasi dengan menggunakan *confusion matrix* dengan cara menggunakan tabel **klasifikasi** yang bersifat prediktif. *Confusion matrix* berfungsi untuk menganalisis apakah *classifier* tersebut baik dalam mengenali tuple dari kelas yang berbeda. Nilai dari *true positive* dan *true negative* memberikan informasi ketika *classifier* dalam melakukan **klasifikasi** data bernilai benar, sedangkan *false positive* dan *false negative* memberikan informasi ketika *classifier* salah dalam melakukan **klasifikasi** data[17]. Berikut adalah tabel *confusion matrix* yang digunakan:

3. Tabel 1, mau ditambahkan penjelasan apa itu Y, X1-X4?

Iye, perbaikannya ada dihalaman 4

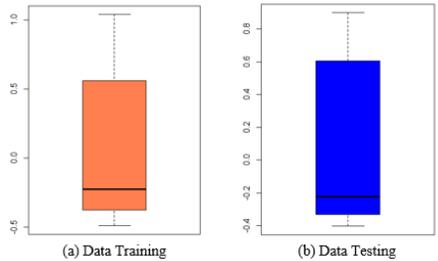
Variabel Y merupakan jumlah kasus TB, X_1 adalah jumlah penduduk, X_2 adalah luas daerah, X_3 adalah bulan, dan X_4 adalah tahun. Variansi data *training* dengan data testing tidak jauh berbeda yang berarti pengelompokkan data *training* cukup mewakili dari dataset agar model yang diperoleh dapat di terapkan pada data *testing*. Box plot (Gambar 3) dari data training (jumlah kasus TB tahun 2017-2018) menunjukkan bahwa tidak terdapat outlier pada dataset dan dari data testing (jumlah kasus TB tahun 2019) juga menunjukkan bahwa tidak terdapat outlier pada dataset.

4

4. Gambar 3, itu boxplot untuk fitur yang mana? tidak ada penjelasan dalam tulisan, hanya dikatakan data training dan data set.

Ditambahkan penjelasan di halaman 5 :

Pada gambar 3, Panjang kotak menggambarkan tingkat penyebaran atau keragaman data pengamatan, sedangkan letak median dan panjang whisker (warna orange pada training dan biru pada testing) menggambarkan tingkat kesimetrisannya⁴



“Boxplots dapat membantu kita dalam memahami karakteristik dari distribusi data. Selain untuk melihat derajat penyebaran data (yang dapat dilihat dari tinggi/panjang boxplot) juga dapat digunakan untuk menilai kesimetrisan sebaran data. Panjang kotak menggambarkan tingkat penyebaran atau keragaman data pengamatan, sedangkan letak median dan panjang whisker (warna orange pada training dan biru pada testing) menggambarkan tingkat kesimetrisannya”.

5. "Hal utama yang dilakukan sebelum melakukan analisis yaitu menormalkan data terlebih dahulu." Tambahkan penjelasan apa yang dimaksud dengan menormalkan data, mungkin bisa ditambahkan juga satu contoh operasi menormalkan data yang dilakukan dalam kerja ini.

Ditambahkan di halaman 2

Sebelum melakukan analisis, hal yang utama dilakukan yaitu normalisasi data. Normalisasi data yaitu melakukan transformasi pada data yang bertujuan untuk mengubah skala pengukuran data asli menjadi bentuk lain sehingga data tersebut dapat memenuhi asumsi-asumsi yang mendasari analisis ragam[20]. Adapun rumus transformasi yang digunakan yaitu:

$$x_i = \frac{x_i - \text{mean}(x)}{sd(x)} \quad \dots (1)$$

Dengan:

$$sd = \sqrt{\text{var}(x)} \quad \dots (2)$$

$$\text{var}(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \text{mean}(x))^2 \quad \dots (3)$$

Contoh data:

No	Kasus TB
1	3
2	0
3	2

Menormalkan data yaitu transformasi data yang bertujuan untuk mengubah skala pengukuran data asli menjadi bentuk lain sehingga data tersebut dapat memenuhi asumsi-asumsi yang mendasari analisis ragam.

Rumus Transformasi yang digunakan yaitu:

$$x_i = \frac{x_i - \text{mean}(x)}{sd(x)} \quad \dots (4)$$

Dengan:

$$sd = \sqrt{\text{var}(x)} \quad \dots (5)$$

$$\text{var}(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \text{mean}(x))^2 \quad \dots (6)$$

Contoh data:

No	Kasus TB
1	3
2	0
3	2
4	2
5	5

$$\begin{aligned} \text{mean}(x) &= \frac{3 + 0 + 2 + 2 + 5}{5} \\ &= \frac{12}{5} \\ &= 2.4 \end{aligned}$$

$$\begin{aligned} \text{var}(x) &= \frac{1}{5-1} \sum_{i=1}^{n=5} (x_i - \text{mean}(x))^2 \\ &= \frac{(3 - 2.4)^2 + (0 - 2.4)^2 + (2 - 2.4)^2 + (2 - 2.4)^2 + (5 - 2.4)^2}{4} \\ &= \frac{0.36 + 5.76 + 0.16 + 0.16 + 6.76}{4} \\ &= \frac{13.2}{4} \\ &= 3.3 \end{aligned}$$

$$\begin{aligned} sd(x) &= \sqrt{\text{var}(x)} \\ &= \sqrt{(3.3)} \\ &= 1.81659 \end{aligned}$$

Proses transformasi:

$$\begin{aligned} x_1 &= \frac{3 - 2.4}{1.81659} = 0.3302891 \\ x_2 &= \frac{0 - 2.4}{1.81659} = -1.321157 \end{aligned}$$

$$x_3 = \frac{2 - 2.4}{1.81659} = -0.220193$$

$$x_4 = \frac{2 - 2.4}{1.81659} = -0.220193$$

$$x_5 = \frac{5 - 2.4}{1.81659} = 1.431253$$

Maka hasil transformasi (menormalan data):

No	Kasus TB
1	0.3302891
2	-1.321157
3	-0.220193
4	-0.220193
5	1.431253

6. "Model yang diperoleh akan diterapkan pada data testing untuk memvalidasi model tersebut." ini adalah prinsip kerja testing untuk klasifikasi. Berarti yang kita lakukan memang Klasifikasi ya? Kalau testing untuk prakiraan time-series, "Model yang diperoleh akan digunakan untuk membangkitkan data masa depan. Data masa depan tersebut akan dibandingkan dengan data real." Mungkin ini pengulangan dari no 2. di atas, perlu secara tegas menentukan melakukan apa, **klasifikasi** atau **prakiraan** berdasarkan time-series, agar apa yang direncanakan konsisten dengan apa yang dilakukan.

Ditambahkan di halaman 3 :

untuk memperjelas, pengklasifikasian yang digunakan yaitu Confusion Matrix.

	Positif	Negatif
TRUE	True Positif	TRUE Negatif
FALSE	False Positif	False Negatif

Dari table diatas, tentunya yang kita harapkan adalah hasil perkiraan yang diperoleh setelah dibandingkan dengan data asli yaitu "true positif" (hasil prediksi positif dan pada data asli juga positif) dan "false negative" (hasil prediksi negative dan pada data asli juga negatif). Jika melihat konsep akurasi yang digunakan (pada persamaan 1) maka dapat dikatakan semakin besar nilai "true positif" dan "false negative" yang diperoleh maka semakin besar pula akurasi yang diperoleh. Contoh:

No	Data Asli	Hasil Perkiraan
1	1	0
2	1	0
3	0	0
4	0	1

5	1	1
6	1	1
7	0	1
8	0	0
9	1	1
10	1	1

Maka Confusion Matrix yang diperoleh:

	Positif	Negatif
TRUE	4	2
FALSE	2	2

Sehingga

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{4 + 2}{4 + 2 + 2 + 2} \times 100\% \\
 &= \frac{6}{10} \times 100\% \\
 &= 0.6 \times 100\% \\
 &= 60\%
 \end{aligned}$$

Karena nilai akurasi yang diperoleh 60% (kurang baik) maka dapat dikatakan bahwa hasil perkiraan yang diperoleh yaitu 60% tepat (yang kita perkirakan sakit betul sakit atau yang kita perkirakan sehat betul sehat) dan 40% kurang tepat (yang kita perkirakan sakit, ternyata sehat atau yang kita perkirakan sehat ternyata sakit)

7. Hasil "akurasi" dalam perbaikan ini (yang memasukkan fitur waktu) malah lebih kecil dari tulisan sebelumnya (yang tidak memasukkan fitur waktu)... artinya, fitur waktu ini malah jadi "noise". Bagaimana pendapat ta ?

Iye pak, fitur waktu ini bukan "noise" melainkan salah satu variable bebas, variable bebas ini akan menjadi model. Misal model regresi linear berganda:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad \dots (6)$$

Dengan:

- \hat{y} : variable dependen (prediksi)
- β_0 : konstanta
- β_1 dan β_2 : koefesien regresi
- x_1 : Jumlah kasus TB
- x_2 : waktu

Dari model regresi (persamaan 6) dapat dilihat bahwa variable sangat mempengaruhi hasil prediksi, anggaplah bahwa x_2 kita hilangkan atau keluarkan maka model akan berubah ($\hat{y} = \beta_0 + \beta_1 x_1$), jika model berubah maka hasil prediksi juga berubah. Jika hasil prediksi berubah maka akurasi juga berubah. Cara kerja model ANN tidak jauh berbeda dengan model regresi linear berganda. Karena ANN memprediksi dengan menggunakan model

tetapi tidak menggunakan 1 model melainkan melakukan percobaan dari sekian model kemudian dipilih model terbaik (nilai akurasi tertinggi).

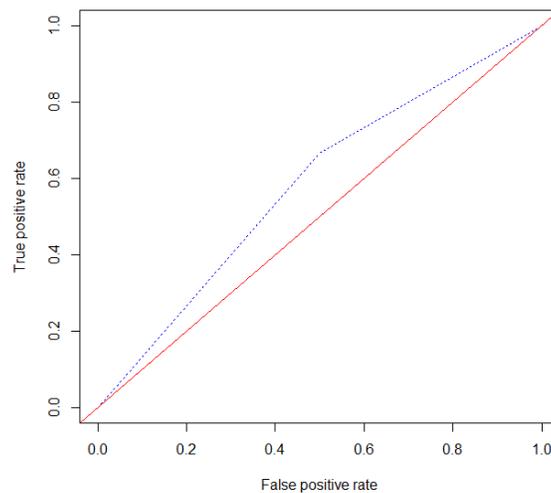
8. Tidak ada penjelasan bagaimana membaca Gambar 4,5,7,8... tambahkan penjelasan, gambar2 tersebut menunjukkan apa? misal, apa itu domain X, apa itu domain Y, bagaimana bisa menentukan apakah pola data prediksi mengikuti pola data asli atau tidak...dll

Iye, Sudah di tambahkan penjelasan gambar

9. Hal yang sama dengan Gambar 6 (plot ROC). Bagaimana membaca gambar ini?

Iye, Sudah di update pada gambar 6

Dari contoh pada no.6 diperoleh kurva ROC berikut:



Garis merah merupakan garis tengah yaitu baseline dan garis biru yaitu hasil perkiraan. Semakin jauh garis biru dari garis merah maka semakin tinggi nilai akurasi yang diperoleh atau semakin luas daerah dibawah garis biru maka semakin tinggi nilai akurasinya dengan kata lain true positif dan false negative yang diperkirakan mendekati data asli.

10. "Artinya hasil uji-t menunjukkan bahwa tidak ada perbedaan yang signifikan antara rata-rata data asli dengan rata-rata hasil prediksi..." ini tiba2 muncul istilah uji-t di tengah2 tulisan tanpa ada penjelasan pendahuluan dan tanpa ada penjelasan lebih lanjut... Sara saya, istilah uji-t ini kalau mau dimasukkan dalam tulisan, harap di-elaborate. Kalau tidak bisa di jelaskan, mending dihapus saja... (hal yang sama berlaku untuk istilah ROC-ANN)

Iye pak Uji-T Sudah di hapus

11. ANN punya akurasi lebih tinggi dari MLP, tapi MSE ANN lebih tinggi dari MSE MLP... tolong dijelaskan dalam tulisan fenomena "kontradiksi" ini...

12. Dalam konteks masalah spasial, sudah tepat disebutkan di awal bahwa nilai jumlah kasus di suatu lokasi dipengaruhi oleh nilai jumlah kasus lokasi2 di sekitarnya. Ini sebuah perspektif yang bagus karena membuka peluang nilai kontribusi tulisan. Sayangnya, dalam tulisan ini saya tidak melihat faktor nilai jumlah kasus lokasi sekitar dipakai untuk melakukan prakiraan

13. Gambar 14, dalam keterangannya di bawah gambar, tambahkan kata "prakiraan"

Iye, Sudah di Perbaiki pada gambar 14