

## Tugas 6 Metode Kecerdasan Buatan – 22 Mei 2022

Nama : Rosa Amalia Nursinta

NIM : 11190940000041

Kelas : Matematika 6B

1. Jelaskan apa yang dimaksud dengan dimensi acting rationally pada konsep intelligent agent!

Intelligent agent adalah sebuah system agen pintar yang dirancang untuk bekerja secara otomatis pada setiap aplikasinya, dengan sensornya yaitu menerima pesan dari lingkungan kemudian memberikan respon atau tindakan sesuai dengan apa yang sudah diprogram oleh pembuat guna mempermudah tugas manusia.

Artificial Intelligence dibagi kedalam 2 bagian utama, yaitu yang berkonsentrasi pada proses pemikiran dan alasan (Thinking) dan yang kedua berkonsentrasi pada tingkah laku (Acting). Perbedaannya adalah thinking hanya akan menghasilkan model, sementara acting berujung pada keputusan agen untuk melakukan aksi tertentu. Kemudian untuk pemrosesan suatu sistem dibagi menjadi dua yaitu, bagian yang berfokus pada keberhasilan dalam hal ketepatan untuk kinerja manusia (Humanly) dan yang kedua berfokus terhadap kinerja yang ideal (Rationally). Sebuah sistem dapat dikatakan rasional jika sistem tersebut melakukan “hal yang benar”, mengingat apa yang ia tahu.

Acting Rationally yaitu kemampuan agen AI untuk bertindak secara logis dan rasional dengan memanfaatkan penalaran logika. Agen yang melakukan pendekatan acting rationally diharapkan untuk melakukan suatu tindakan yang beroperasi sendiri, dapat memperhatikan lingkungan, bertahan untuk waktu yang lama, mudah beradaptasi dengan perubahan, mencapai suatu tujuan. Agen rasional adalah salah satu yang bertindak untuk mencapai hasil yang terbaik, dan ketika tidak ada kepastian tetap akan menampilkan hasil yang terbaik.

2. Berikan contoh permasalahan supervised learning classification maupun regression!

- a. Contoh permasalahan supervised learning classification: bisa digunakan dalam dunia perbankan, yaitu dalam penentuan kelayakan kredit nasabah. Dari beberapa variabel prediktor yang berkaitan seperti jenis pekerjaan, penghasilan perbulan, jumlah tanggungan, dan lain-lain. Dengan metode klasifikasi dapat menentukan apakah nasabah layak mendapatkan kredit atau tidak. Dengan adanya prediksi seperti itu dapat menguntungkan pihak perbankan dalam melakukan kredit kepada nasabah sehingga meminimalisir kerugian pihak perbankan.
- b. Contoh permasalahan supervised learning regression: pengaruh luas tanah terhadap harga rumah.

3. Berikan contoh permasalahan supervised learning classification untuk kasus multilabel - multiclass!

- a. Klasifikasi multiclass artinya tugas klasifikasi dengan lebih dari dua kelas, misalnya dalam mengklasifikasikan satu set gambar buah-buahan yang mungkin mangga, apel, atau alpukat. Klasifikasi multi-kelas membuat asumsi bahwa masing-masing sampel ditugaskan untuk satu dan hanya satu label: buah dapat berupa apel atau pir tetapi tidak keduanya sekaligus.
- b. klasifikasi multilabel artinya memberikan kepada masing-masing sampel satu set label target. Ini dapat dianggap sebagai memprediksi properti titik data yang tidak saling eksklusif, seperti topik yang relevan untuk dokumen. Sebuah teks mungkin tentang agama, politik, keuangan atau pendidikan pada saat yang sama atau tidak sama sekali.

4. Jelaskan dan berikan ilustrasi metode holdout pada tahapan validasi model!

Dengan kondisi terbatasnya data yang digunakan untuk training dan testing, diperlukan metode untuk mendapatkan hasil tingkat akurasi dari sebuah metode pada machine learning. Metode holdout adalah metode yang akan menyediakan sejumlah data untuk digunakan sebagai data testing, dan sisanya sebagai data training.

Dataset training adalah data yang digunakan untuk pelatihan model. Sedangkan dataset testing digunakan untuk melihat seberapa baik performa model tersebut pada data yang baru (data yang tidak ada di data training). Secara umum, pembagian metode holdout ini adalah 80% dari keseluruhan data untuk data training dan sisanya sebesar 20% digunakan untuk data testing.



Ilustrasi (2 sub himpunan)

Asumsikan terdapat data training dan data testing dibangun dengan distribusi yang sama untuk setiap kelas, agar proporsi setiap kelas sama.

Langkah-langkah:

1. Membagi dataset menjadi 2, biasanya  $\frac{2}{3}$  data training dan  $\frac{1}{3}$  data testing (misal 60/40; 70/30; 80/20 atau dengan pertimbangan tertentu).
2. Membangun model menggunakan data training.
3. Menguji model menggunakan data testing.

5. Berikan pendapat anda hal yang perlu diperhatikan pada permasalahan supervised learning jika ingin menerapkan teknik holdout pada tahapan validasi model!

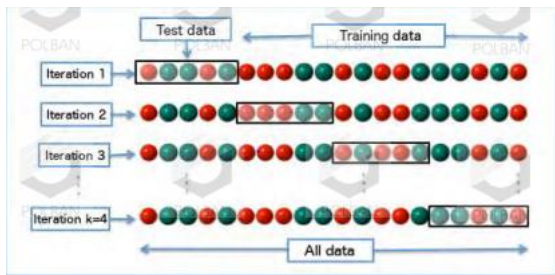
Berdasarkan yang saya baca, saat proses pengacakan data untuk dibagi sebagai data training dan testing, sangat memungkinkan terjadinya overrepresented pada salah satu atau lebih klasifikasi. Klasifikasi tersebut dominan dibandingkan klasifikasi lainnya, sehingga data training dan testing menjadi tidak representatif. Maka dari itu diperlukan prosedur stratification holdout, dimana prosedur ini menjamin bahwa setiap klasifikasi dapat terwakili pada data training dan testing yang tercipta secara proporsional. Hasil proses holdout proporsinya harus sedekat mungkin dengan proporsi aslinya.

Maka menurut saya diperlukannya sebuah perulangan terhadap seluruh proses training dan testing beberapa kali dengan data training dan testing yang teracak. Kemudian diambil nilai rata-ratanya dan prosedur ini dapat dikatakan sebagai repeated holdout.

6. Jelaskan perbedaan metode k-fold cross validation, stratified k-fold cross validation dan Leave One Out Cross Validation serta berikan contohnya!

a. K-fold cross validation

Salah satu bentuk dari cross validation adalah k-fold cross validation. Dalam metode k-fold cross validation, data dibagi ke dalam beberapa partisi yang disebut dengan fold. Masing-masing fold memiliki jumlah data dengan ukuran yang sama atau mendekati sama. Selama k iterasi, dipilih salah satu fold sebagai data uji, sedangkan sisa k-1 fold dijadikan data latih (Refaeilzadeh, et al., 2009).



Seluruh data dibagi menjadi 4 fold dengan setiap fold berisi 5 data. Dalam setiap iterasi, dipilih salah satu fold sebagai data uji dan sisanya menjadi data latih. Setiap data hanya boleh sekali menjadi data uji. Perhitungan akurasi penilaian terhadap data uji dilakukan di setiap iterasi.

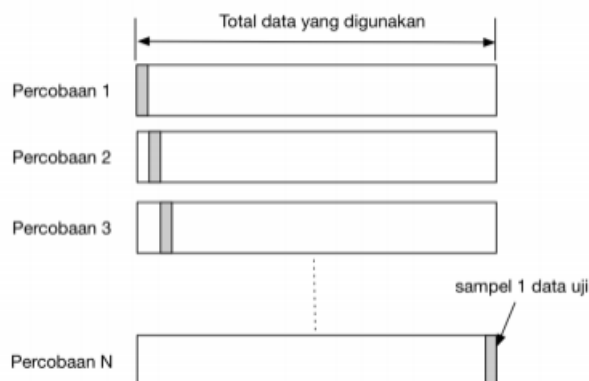
b. Stratified k-fold cross validation

Stratified merupakan teknik dalam cross validation untuk memastikan bahwa dalam data latih dan data uji harus ada perwakilan dari seluruh kelas yang ada dengan persentase yang sama. dilakukan untuk memastikan bahwa dalam setiap fold merupakan representasi data yang baik (Refaeilzadeh, et al., 2009).

Contoh penerapan stratified yaitu pada penilaian esai dengan 2 kategori yaitu A dan B, dengan persentase jumlah data yaitu 40% nilai A dan 60% nilai B. Oleh karena itu, pada data latih dan data uji masing-masing harus memiliki komposisi data 40% nilai A dan 60% nilai B.

c. Leave One Out Cross Validation

Leave one-out Cross Validation merupakan turunan dari metode K-Fold Cross validation, di mana K yang dipilih adalah sebesar jumlah data. Misalkan dataset yang memiliki N sample, maka percobaan dilakukan sebanyak N kali. Setiap percobaan menggunakan sebanyak N – 1 sample untuk data training dan sisanya untuk testing.



7. Bagaimana cara menentukan ukuran performansi (accuracy ataupun recall) ketika menggunakan k-fold cross validation dan berikan contohnya!

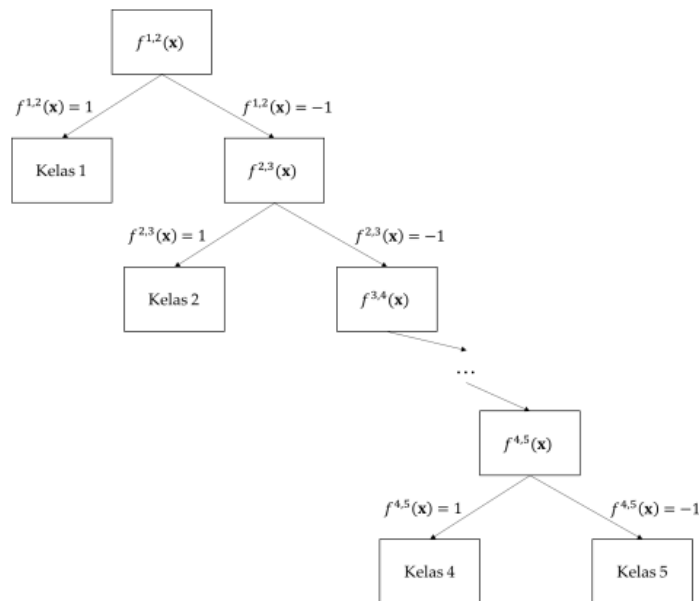
Ketika menggunakan k-fold cross validation, dimana data asli dibagi menjadi k subbagian dan pada tahapannya dilakukan sebanyak k iterasi, misalnya pada iterasi pertama dipilih suatu subbagian dari k sebagai data test dan subbagian k-1 lainnya dijadikan sebagai data train. Sehingga akan diperoleh k model untuk setiap iterasi kemudian hitung rata-rata dari setiap performansi untuk memperoleh performa accuracy dan recall dari keseluruhan model.

Contohnya :

- Akurasi yang digunakan untuk k-fold cross validation merupakan akurasi rata-rata dari setiap lipatan sebanyak K lipatan.
- Setiap lipatan dicari akurasi masing-masing kemudian di rata-rata sehingga menjadi nilai akurasi model oleh metode ini.
- Misal banyak data (n) sebanyak 12 berisi kisi-kisi untuk ujian matematika dengan  $K=3$ .
- Model I (berada di lipatan I) menjelaskan siswa belajar kisi-kisi 1-8 yang di tes 9-12 dan mendapat nilai X.
- Model II (berada di lipatan II), siswa di tes 5-8 sisanya merupakan kisi-kisi, nilai yang didapatkan adalah Y.
- Model III (berada di lipatan III) siswa belajar dari kisi-kisi 5-12 yang di tes 1-4 sehingga mendapat nilai Z.
- Kemudian akurasi model merupakan rata-rata dari akurasi setiap lipatan. Sehingga didapatkan lah akurasi dari metode K-fold Cross Validation.

8. Jelaskan teknik one vs all dan all vs rest ketika menghadapi kasus multiclass dalam proses perhitungan confusion matrix!

*One vs All (OvA)* atau *One vs Rest (OvR)* merupakan salah satu pendekatan untuk *multiclass classification*. Pendekatan ini hanya akan membangun *classifier* sesuai jumlah kelas yang akan diklasifikasikan. Pada dasarnya kita membuat sejumlah model akan tetapi kombinasinya tidak sebanyak *One vs One*. Model pertama mengklasifikasikan *input* sebagai kelas pertama atau bukan kelas pertama. Setelah itu, dilanjutkan ke model kedua mengklasifikasikan *input* sebagai kelas kedua atau bukan kelas kedua dan seterusnya. Berikut ilustrasinya terdapat lima kelas :



9. Berikanlah interpretasi jika anda menghadapi performansi model pada kasus binary classification menghasilkan detail sebagai berikut:

- a. Accuracy : 50%
- b. Precision : 50 %
- c. Recall : 50%

- a. Accuracy : 50%

Akurasi pada hasil diatas menggambarkan bahwa sebesar 50% model dapat memprediksi dan mengklasifikasi secara benar dari keseluruhan data yang ada. Namun, pada binary classification, angka 50% atau 0.5 tidak memiliki makna yang berarti penting dan hanya sebuah informasi saja. Artinya, model ini tidak cukup kuat dalam memprediksi atau mengklasifikasi secara benar data yang ada. Kemungkinan ada beberapa keasalahan yang terjadi, yaitu pada data yang diperoleh maupun pada preprocessing-nya.

- b. Precision : 50 %

Precision dengan hasil diatas menunjukkan bahwa sebesar 50% data diprediksi secara benar positif dari keseluruhan data yang sebenarnya positif. Ini berarti tingkat akurasi model dalam memprediksi data yang benar belum cukup baik, karena masih banyak data yang diprediksi dengan benar namun tidak diprediksi benar oleh model, maka diperlukan improvisasi atau perbaikan.

- c. Recall : 50%

Dari data diatas, recall dengan hasil 50% menunjukkan bahwa hanya 50% model dapat memprediksi data yang benar diprediksi secara benar. 50% sisanya model cukup lemah dalam memprediksi data yang sebenarnya benar untuk diprediksi benar. Maka, model belum berhasil dan sangat jauh dari optimal dalam memprediksi atau mengklasifikasi data

Kita juga dapat menyeimbangkan bobot Presisi dan Recall ke dalam  $F_1Score$

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} = 2 * \frac{0.5 * 0.5}{0.5 + 0.5} = 0.5$$

Berdasarkan nilai  $F_1Score$  yaitu 0.5, maka dalam kasus binary class model ini tidak berguna dalam mengklasifikasi sesuatu karena kemungkinan atau keyakinan antara positif dan negatifnya 50:50.



10. Bagaimana cara untuk menghitung presisi pada model klasifikasi jika permasalahan yang dihadapi berbentuk multiclass clasification dan berikan contohnya!

contoh kasus pada 3 kelas dengan nama yang berbeda yaitu jeruk, pepaya dan anggur.

| Confusion Matrix 3x3 |          |        |        |
|----------------------|----------|--------|--------|
| Aktual               | Prediksi |        |        |
|                      | Jeruk    | Pepaya | Anggur |
| Jeruk                | 7        | 8      | 9      |
| Pepaya               | 1        | 2      | 3      |
| Anggur               | 3        | 2      | 1      |

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                  |
| Precision | $7 / (7 + 17) = 0,29$ | $2 / (2 + 4) = 0,33$ | $1 / (1 + 5) = 0,16$ |

Tabel diatas adalah hasil perolehan nilai precision dari masing-masing tiap kelasnya. Kemudian cari nilai rata-rata dari semua presisi untuk dijadikan satu nilai presisi.

$$Presisi = \frac{Presisi A + B + C}{Jumlah Kelas} = \frac{0.29 + 0.33 + 0.16}{3} = 0.67$$

Ini berarti 29% jeruk diprediksi secara benar dari keseluruhan data yang diprediksi sebagai jeruk, 33% papaya yang diprediksi secara benar dari keseluruhan data yang diprediksi sebagai papaya, 16% anggur yang diprediksi secara benar dari keseluruhan data yang diprediksi sebagai anggur. Secara keseluruhan presisi dari model ini sebesar 67%.

11. Jelaskan pada kasus seperti apa sebaiknya metode kecerdasan buatan pada permasalahan klasifikasi menggunakan ukuran performansi  $F_\beta$  Score dibandingkan dengan menggunakan accuracy!

Terdapat keunggulan dan kelemahan  $F_\beta$  Score dan Accuracy

Keunggulan accuracy : mudah untuk ditafsirkan. Jika kita mengatakan bahwa suatu model 90% akurat, kita tahu bahwa model tersebut mengklasifikasikan 90% pengamatan dengan benar.

Kelemahan accuracy : tidak memperhitungkan bagaimana data didistribusikan. Misalnya, 90% dari semua pemain tidak direkrut ke NBA. Jika model hanya memprediksi setiap pemain untuk tidak direkrut, model tersebut akan memprediksi hasil dengan benar untuk 90% pemain. Nilai ini tampaknya tinggi, tetapi model tersebut sebenarnya tidak dapat memprediksi dengan benar pemain mana pun yang direkrut.

Keunggulan  $F_\beta$  Score : mempertimbangkan bagaimana data didistribusikan. Misalnya, jika datanya sangat tidak seimbang (misalnya 90% dari semua pemain tidak mengikuti wajib militer dan 10% mendapatkan wajib militer) maka  $F_\beta$  Score akan memberikan penilaian kinerja model yang lebih baik.

Kelemahan  $F_\beta$  Score : lebih sulit untuk ditafsirkan.  $F_\beta$  Score adalah perpaduan antara presisi dan recall model, yang membuatnya sedikit lebih sulit untuk ditafsirkan.

**Dalam praktiknya,**

Gunakan akurasi pada *balanced problem* dan tidak ada kerugian besar untuk memprediksi false negative.

Gunakan  $F_\beta$  Score pada *imbalanced problem* dan ada kerugian serius untuk memprediksi false negative.

Misalnya, jika kita menggunakan model regresi logistik untuk memprediksi apakah seseorang menderita kanker atau tidak, false negative sangat berbahaya (misalnya memprediksi bahwa seseorang tidak menderita kanker ketika mereka sebenarnya mengidap kanker) sehingga  $F_\beta$  Score akan menghitung lebih baik model yang memiliki terlalu banyak false negative dibandingkan accuracy.

12. Berikan contoh kasus yang dapat menggambarkan recall lebih baik diperhatikan dibandingkan hanya menggunakan accuracy saja!

Tidak tepat menggunakan akurasi pada imbalanced class antara False Negative dan False Positive karena mudah untuk mendapatkan skor akurasi yang tinggi hanya dengan mengklasifikasikan semua observasi sebagai kelas mayoritas.

Pilih model dengan recall tinggi jika false positive jauh lebih baik daripada false negative, dengan kata lain, jika terjadinya false negative tidak dapat diterima/tidak dapat ditoleransi. Seperti dalam contoh lebih baik algoritma memprediksi imigran positif teroris tetapi sebenarnya bukan teroris daripada algoritma salah memprediksi bahwa imigran diprediksi bukan teroris padahal sebenarnya dia teroris karena jika imigran teroris tersebut lolos dalam prediksi kemudian berbaur di lingkungan masyarakat akan mengakibatkan kondisi masyarakat dalam bahaya.

13. Berikan contoh kasus dimana menggunakan accuracy saja pada ukuran performansi model tidak cukup/atau dapat menyesatkan!

model untuk mengidentifikasi teroris yang mencoba naik pesawat. Terdapat 800 juta penumpang rata-rata pada penerbangan AS per tahun dan 19 teroris (yang dikonfirmasi) yang naik penerbangan AS dari tahun 2000–2017 dengan akurasi lebih dari 99 persen, walaupun model ini memiliki akurasi yang hampir sempurna, masalah ini adalah salah satu di mana akurasi bukanlah ukuran yang baik untuk menilai kinerja model.

Dalam kasus ini, menyatakan semua penumpang sebagai negatif (bukan teroris) dalam masalah deteksi teroris tidak membantu, dan, sebaliknya, kita harus fokus pada mengidentifikasi kasus positif. Maka kita harus memaksimalkan recall, dimana recall adalah perbandingan jumlah true positive dengan jumlah true positive dan false negative.

Jika kita melabeli semua penumpang sebagai teroris, maka recall menjadi 1.0. Namun ini masih bukan pengklasifikasian yang sempurna, seperti kebanyakan konsep dalam ilmu data, ada trade-off dalam metrik yang dipilih untuk dimaksimalkan. Dalam hal recall dan presisi, ketika meningkatkan recall, maka akan mengurangi presisi, dan sebaliknya. Dimana presisi adalah perbandingan jumlah true positive dengan jumlah true positive dan false positive.

Dalam kasus di mana kita ingin memaksimalkan antara recall atau presisi, kita dapat menggabungkan dua metrik tersebut menggunakan  $F_1 \text{ Score}$ . Dimana  $F_1 \text{ Score}$  adalah rata-rata harmonik dengan mempertimbangkan presisi dan recall dalam persamaan berikut:

$$F_1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

$F_1 \text{ Score}$  memberikan keseimbangan untuk kedua ukuran dan merupakan contoh dari  $F_\beta \text{ Score}$  secara umum di mana besaran  $\beta$  dapat disesuaikan untuk memberikan bobot lebih baik untuk recall dan presisi.

14. Jelaskan perbedaan *F!Score* untuk nilai  $\beta = \{0.5, 1\}$ !

- F-0.5 ( $\beta = 0.5$ ), untuk penggunaan beta bernilai 0.5 berarti hasil yang diinginkan lebih fokus pada mengurangi false positive atau bobot *precision* lebih besar dibanding *recall*.

$$F - 0.5 = \frac{1.25 * Precision * Recall}{Precision + Recall}$$

- F-1 ( $\beta = 1$ ), penggunaan  $\beta = 1$  merupakan default dimana hasilnya berbobot sama antara *Precision* dan *Recall*.

$$F - 1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

15. Jelaskan perbedaan Micro, Macro dan Weigthed Average pada  $F_\beta$ Score!

| Metrik              | Deskripsi   |
|---------------------|---|
| F1-micro            | <p>Melihat hasil prediksi secara keseluruhan tanpa pemisahan kinerja untuk masing-masing kelas. F1-micro = presisi = recall. Nilai F1-micro sama dengan akurasi pada <i>multi-class classification</i>.</p> $\text{micro Precision} = \frac{\sum_0^n TP_n}{\sum_0^n TP_n + \sum_0^n FP_n}$ $\text{micro Recall} = \frac{\sum_0^n TP_n}{\sum_0^n TP_n + \sum_0^n FN_n}$  |
| F1-macro            | <p>Memisahkan kinerja untuk masing-masing kelas, kemudia merata-ratakan nilai kinerja dengan kata lain <i>Menghitung F-1 score</i> berdasarkan hitungan perbaris 'kenyataan'(yang bukan dugaan).</p> $\text{macro } F - 1 = \frac{2 * \text{macro Precision} * \text{macro Recall}}{\text{macro Precision} + \text{macro Recall}}$ $\text{macro Precision} = \frac{P_0 + P_1 + \dots + P_n}{n}, n = 0,1,2, \dots, n$ $P_n = \frac{TP_n}{TP_n + FP_n}$ $\text{macro Recall} = \frac{R_0 + R_1 + \dots + R_n}{n}, n = 0,1,2, \dots, n$ $R_n = \frac{TP_n}{TP_n + FN_n}$ |
| F1-weigthed average | <p>Memisahkan kinerja untuk masing-masing kelas, kemudian merata-ratakan nilai kinerja dimana setiap kelas memiliki bobot berdasarkan banyaknya <i>desired output</i> untuk setiap kelas. Dihitung dengan mengambil rata-rata semua skor F1 per kelas sambil mempertimbangkan dukungan masing-masing kelas.</p>   |

16. Perhatikan deskripsi dibawah ini:

Misalkan kita ingin mengukur kinerja sebuah mesin prediksi interaksi protein HIV dengan manusia. Mesin tersebut bertugas mencari dan kemudian memprediksi interaksi protein tersebut dengan dua kategori, berinteraksi atau tidak berinteraksi. Kita menguji mesin tersebut dengan memberikan 100 protein dengan kategori berinteraksi dan 900 protein dengan kategori tidak berinteraksi. Hasilnya mesin tersebut mampu memisahkan 110 protein yang terdeteksi sebagai protein berinteraksi. Kemudian, 110 protein hasil prediksi yang dilakukan oleh mesin tersebut diperiksa kembali. Dari 110 protein yang terdeteksi sebagai protein yang berinteraksi oleh mesin tersebut, ternyata hanya terdapat 90 protein yang terprediksi dengan benar, sedangkan 20 lainnya adalah protein yang tidak berinteraksi.

Buatlah confusion matrix dari permasalahan diatas, kemudian hitung accuracy, precision, recall dan specificity serta jelaskan masing-masing skor yang diperoleh!

|                       | Benar berinteraksi | Tidak berinteraksi |
|-----------------------|--------------------|--------------------|
| Prediksi berinteraksi | 90                 | 20                 |
| Tidak berinteraksi    | 10                 | 880                |

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} = \frac{90 + 880}{90 + 20 + 880 + 10} = 0.97$$

artinya 97% data diprediksi secara benar baik prediksi berinteraksi maupun prediksi tidak berinteraksi dari keseluruhan data.

$$Precision = \frac{TP}{TP + FP} = \frac{90}{90 + 20} = 0.81$$

artinya 81% model mampu memprediksi secara benar protein yang berinteraksi dari keseluruhan protein yang diprediksi berinteraksi.

$$Recall = \frac{TP}{TP + FN} = \frac{90}{90 + 10} = 0.90$$

artinya 90% model mampu memprediksi secara benar protein yang berinteraksi dari keseluruhan protein yang sebenarnya berinteraksi.

$$Specificity = \frac{TN}{TN + FP} = \frac{880}{880 + 20} = 0.97$$

artinya 97% model mampu memprediksi secara benar protein yang tidak berinteraksi

17. Perhatikan gambar dibawah ini

| CASE 1: Evenly distributed errors |       |       |         | CASE 2: Small variance in errors |       |       |         | CASE 3: Large error outlier |       |       |         |
|-----------------------------------|-------|-------|---------|----------------------------------|-------|-------|---------|-----------------------------|-------|-------|---------|
| ID                                | Error | Error | Error^2 | ID                               | Error | Error | Error^2 | ID                          | Error | Error | Error^2 |
| 1                                 | 2     | 2     | 4       | 1                                | 1     | 1     | 1       | 1                           | 0     | 0     | 0       |
| 2                                 | 2     | 2     | 4       | 2                                | 1     | 1     | 1       | 2                           | 0     | 0     | 0       |
| 3                                 | 2     | 2     | 4       | 3                                | 1     | 1     | 1       | 3                           | 0     | 0     | 0       |
| 4                                 | 2     | 2     | 4       | 4                                | 1     | 1     | 1       | 4                           | 0     | 0     | 0       |
| 5                                 | 2     | 2     | 4       | 5                                | 1     | 1     | 1       | 5                           | 0     | 0     | 0       |
| 6                                 | 2     | 2     | 4       | 6                                | 3     | 3     | 9       | 6                           | 0     | 0     | 0       |
| 7                                 | 2     | 2     | 4       | 7                                | 3     | 3     | 9       | 7                           | 0     | 0     | 0       |
| 8                                 | 2     | 2     | 4       | 8                                | 3     | 3     | 9       | 8                           | 0     | 0     | 0       |
| 9                                 | 2     | 2     | 4       | 9                                | 3     | 3     | 9       | 9                           | 0     | 0     | 0       |
| 10                                | 2     | 2     | 4       | 10                               | 3     | 3     | 9       | 10                          | 20    | 20    | 400     |

| CASE 4: Errors = 0 or 5 |       |       |         | CASE 5: Errors = 3 or 4 |       |       |         |
|-------------------------|-------|-------|---------|-------------------------|-------|-------|---------|
| ID                      | Error | Error | Error^2 | ID                      | Error | Error | Error^2 |
| 1                       | 5     | 5     | 25      | 1                       | 3     | 3     | 9       |
| 2                       | 0     | 0     | 0       | 2                       | 4     | 4     | 16      |
| 3                       | 5     | 5     | 25      | 3                       | 3     | 3     | 9       |
| 4                       | 0     | 0     | 0       | 4                       | 4     | 4     | 16      |
| 5                       | 5     | 5     | 25      | 5                       | 3     | 3     | 9       |
| 6                       | 0     | 0     | 0       | 6                       | 4     | 4     | 16      |
| 7                       | 5     | 5     | 25      | 7                       | 3     | 3     | 9       |
| 8                       | 0     | 0     | 0       | 8                       | 4     | 4     | 16      |
| 9                       | 5     | 5     | 25      | 9                       | 3     | 3     | 9       |
| 10                      | 0     | 0     | 0       | 10                      | 4     | 4     | 16      |

Hitunglah MSE, MAE dan RMSE dari kelima kasus diatas kemudian analisis lah hasil yang anda peroleh!

Jawab :

a. Case 1: Evenly Distributed Error

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \frac{4 + 4 + 4 + 4 + 4 + 4 + 4 + 4 + 4 + 4}{10} = \frac{40}{10} = 4$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| = \frac{2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2}{10} = \frac{20}{10} = 2$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{Y}_i - Y_i)^2}{n}} = \sqrt{\sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{n}} = \sqrt{4} = 2$$

b. Case 2: Small variance in errors

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \frac{1 + 1 + 1 + 1 + 1 + 9 + 9 + 9 + 9 + 9}{10} = \frac{50}{10} = 5$$



$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| = \frac{1 + 1 + 1 + 1 + 1 + 3 + 3 + 3 + 3 + 3}{10} = \frac{20}{10} = 2$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{Y}_i - Y_i)^2}{n}} = \sqrt{\sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{n}} = \sqrt{5} = 2.2360679775$$

c. Case 3: Large Error Outlier

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \frac{0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 400}{10} = \frac{400}{10} = 40$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| = \frac{0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 20}{10} = \frac{20}{10} = 2$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{Y}_i - Y_i)^2}{n}} = \sqrt{\sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{n}} = \sqrt{40} = 6.3245553203$$

d. Case 4: Errors = 0 or 5

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \frac{25 + 0 + 25 + 0 + 25 + 0 + 25 + 0 + 25 + 0}{10} = \frac{125}{10} = 12.5$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| = \frac{5 + 0 + 5 + 0 + 5 + 0 + 5 + 0 + 5 + 0}{10} = \frac{25}{10} = 2.5$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{Y}_i - Y_i)^2}{n}} = \sqrt{\sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{n}} = \sqrt{12.5} = 3.5355339059$$

e. Case 5: Errors = 3 or 4

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \frac{9 + 16 + 9 + 16 + 9 + 16 + 9 + 16 + 9 + 16}{10} = \frac{125}{10} = 12.5$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| = \frac{3 + 4 + 3 + 4 + 3 + 4 + 3 + 4 + 3 + 4}{10} = \frac{35}{10} = 3.5$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{Y}_i - Y_i)^2}{n}} = \sqrt{\sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{n}} = \sqrt{12.5} = 3.5355339059$$