untuk membandingkan pengaruh normalisasi min-max scaling, z-score dan decimal scaling dalam KNN, hasil evalusai model (akurasi, precision, recall, f1-score) itu sudah cukup atau enggak? bagaimana dengan P-Value, Specificity, dan ROC AUC?

Untuk membandingkan pengaruh normalisasi **Min-Max Scaling**, **Z-Score**, dan **Decimal Scaling** dalam model **KNN**, **evaluasi model** dengan menggunakan metrik seperti **akurasi**, **precision**, **recall**, dan **F1-score** adalah langkah yang **sudah cukup baik** dan dapat memberikan gambaran yang cukup lengkap tentang kinerja model. Namun, jika Anda ingin mendapatkan **analisis yang lebih mendalam** dan lebih **komprehensif**, ada beberapa metrik tambahan yang bisa dipertimbangkan, seperti **P-Value**, **Specificity**, dan **ROC AUC**. Berikut adalah penjelasan lebih lanjut mengenai metrik-metrik ini dan apakah Anda perlu menggunakannya.

**1. Akurasi, Precision, Recall, dan F1-Score**

* **Akurasi**: Mengukur seberapa sering model membuat prediksi yang benar, tetapi bisa **menyesatkan** jika dataset tidak seimbang (misalnya, kelas dominan jauh lebih banyak daripada kelas minoritas).
* **Precision**: Menunjukkan seberapa tepat prediksi positif model. Ini sangat berguna ketika **false positives** memiliki **biaya tinggi** (misalnya, memprediksi seseorang memiliki penyakit padahal tidak).
* **Recall**: Mengukur seberapa baik model dalam **mendeteksi kelas positif**. Ini sangat penting ketika **false negatives** lebih bermasalah (misalnya, melewatkan kasus penyakit yang sebenarnya ada).
* **F1-Score**: Merupakan rata-rata harmonis dari **precision** dan **recall**, memberikan keseimbangan antara keduanya, dan sering digunakan ketika Anda ingin menghindari masalah **ketidakseimbangan kelas**.

**Kapan Anda sudah cukup dengan metrik ini?**

* Jika dataset Anda **terseimbang** dan Anda hanya ingin mengevaluasi kemampuan model KNN dalam **memprediksi dengan tepat** (baik untuk kelas positif maupun negatif), maka **akurasi**, **precision**, **recall**, dan **F1-score** sudah cukup.

Namun, jika dataset Anda **tidak seimbang** (misalnya, satu kelas jauh lebih banyak daripada kelas lainnya), maka Anda perlu **menambahkan metrik lainnya** untuk menghindari bias yang mungkin terjadi dengan hanya mengandalkan akurasi.

**2. P-Value**

* **P-Value** adalah ukuran **signifikansi statistik** yang digunakan untuk menguji hipotesis. Dalam konteks evaluasi model, **P-Value** bisa digunakan untuk mengevaluasi apakah perbedaan dalam kinerja model (misalnya, perbedaan antara normalisasi yang berbeda) adalah **signifikan** secara statistik.

**Apakah Anda perlu P-Value?**

* Jika Anda ingin **menguji signifikansi statistik** dari perbedaan kinerja model yang menggunakan berbagai teknik normalisasi (misalnya, perbedaan dalam akurasi atau F1-score antara Min-Max Scaling, Z-Score, dan Decimal Scaling), maka **P-Value** bisa sangat berguna.
* Untuk mendapatkan **P-Value**, Anda bisa menggunakan **uji statistik** seperti **uji t (t-test)** atau **ANOVA** jika Anda membandingkan lebih dari dua model. Ini akan memberi tahu Anda apakah perbedaan kinerja model tersebut disebabkan oleh **faktor yang tidak kebetulan**.

**3. Specificity (True Negative Rate)**

* **Specificity** mengukur **kemampuan model untuk benar mengklasifikasikan negatif**. Ini sangat penting dalam konteks di mana **false positives** memiliki konsekuensi besar. Specificity dihitung sebagai: Specificity=TNTN+FP\text{Specificity} = \frac{TN}{TN + FP}Specificity=TN+FPTN​ di mana:
  + **TN (True Negative)** adalah jumlah prediksi yang benar untuk kelas negatif.
  + **FP (False Positive)** adalah jumlah prediksi yang salah untuk kelas negatif.

**Apakah Anda perlu Specificity?**

* **Ya**, terutama jika masalah Anda **berkaitan dengan deteksi penyakit** atau klasifikasi di mana penting untuk **menghindari false positives**. Metrik ini akan membantu Anda **menilai kemampuan model untuk tidak memprediksi kelas positif secara keliru**.

**4. ROC AUC (Area Under the Curve)**

* **ROC AUC** mengukur **kemampuan model dalam membedakan antara kelas positif dan negatif**, tanpa terpengaruh oleh ambang batas (threshold) klasifikasi. Ini menunjukkan **trade-off antara true positive rate (recall)** dan **false positive rate (1 - specificity)**.
  + **AUC = 0.5** berarti model tidak lebih baik dari **tebakan acak**.
  + **AUC = 1.0** berarti model **sempurna** dalam membedakan kelas positif dan negatif.

**Apakah Anda perlu ROC AUC?**

* **Ya**, jika Anda ingin memahami **kinerja model secara keseluruhan** dalam hal **kemampuannya membedakan antara kelas**, terutama pada dataset yang **tidak seimbang**. ROC AUC tidak terpengaruh oleh distribusi kelas atau ambang batas klasifikasi, sehingga sangat berguna dalam **kasus klasifikasi tidak seimbang**.

**5. Kesimpulan:**

Untuk membandingkan pengaruh normalisasi **Min-Max Scaling**, **Z-Score**, dan **Decimal Scaling** dalam model **KNN**, berikut adalah saran evaluasi yang tepat:

* **Akurasi**, **Precision**, **Recall**, dan **F1-Score** sudah **cukup baik** jika dataset Anda **seimbang** dan Anda ingin mengukur **performa dasar** model.
* **Specificity** dan **ROC AUC** sangat penting jika dataset Anda **tidak seimbang** atau jika Anda ingin mengukur kinerja model dalam hal **membedakan kelas positif dan negatif**.
* **P-Value** berguna untuk menguji apakah perbedaan kinerja model yang Anda temui secara statistik **signifikan**.

**Rekomendasi:**

* Gunakan **Akurasi**, **Precision**, **Recall**, dan **F1-score** untuk **perbandingan dasar**.
* Jika dataset Anda **tidak seimbang** atau Anda ingin **mendalami lebih lanjut** kinerja model, maka tambahkan **Specificity** dan **ROC AUC**.
* Jika Anda ingin **menguji signifikansi** dari perbedaan performa model, gunakan **P-Value**.

Dengan kombinasi metrik ini, Anda akan mendapatkan gambaran yang lebih lengkap tentang bagaimana normalisasi memengaruhi model KNN dan kinerja model secara keseluruhan.