

Pertemuan 5

Evaluasi dan Validasi

Evaluasi

- Evaluasi merupakan kegiatan yang dilakukan untuk menentukan nilai dari suatu hal, dalam hal ini yang dinilai adalah kinerja metode pembelajaran mesin
- Tujuan dilakukan evaluasi adalah untuk menganalisa hasil kinerja metode pembelajaran mesin

Kriteria Evaluasi

- Secara umum pengukuran model data mining mengacu kepada tiga kriteria: Akurasi (Accuracy), Keandalan(Reliability) dan Kegunaan (Usefulness)
- Keseimbangan diantaranya ketiganya diperlukan karena belum tentu model yang akurat adalah handal, dan yang handal atau akurat belum tentu berguna

Kriteria Evaluasi

1. **Akurasi** adalah ukuran dari seberapa baik model mengkorelasikan antara hasil dengan atribut dalam data yang telah disediakan. Terdapat berbagai model akurasi, tetapi semua model akurasi tergantung pada data yang digunakan
2. **Kehandalan** adalah ukuran di mana model data mining diterapkan pada dataset yang berbeda akan menghasilkan sebuah model data mining dapat diandalkan jika menghasilkan pola umum sama terlepas dari data testing yang disediakan
3. **Kegunaan** mencakup berbagai metrik yang mengukur apakah model tersebut memberikan informasi yang berguna.

Evaluasi Berdasarkan Metode Pembelajaran Mesin

1. Estimation:
 - **Error**: Root Mean Square Error (RMSE), MSE, MAPE, etc
2. Prediction/Forecasting (Prediksi/Peramalan):
 - **Error**: Root Mean Square Error (RMSE) , MSE, MAPE, etc
3. Classification:
 - **Confusion Matrix**: Accuracy
 - **ROC Curve**: Area Under Curve (AUC)
4. Clustering:
 - **Internal Evaluation**: Davies–Bouldin index, Dunn index,
 - **External Evaluation**: Rand measure, F-measure, Jaccard index, Fowlkes–Mallows index, Confusion matrix

Root Mean Square Error

- Root Mean Square Error adalah metode alternative untuk mengevaluasi Teknik peramalan yang digunakan untuk mengukur tingkat akurasi hasil perkiraan suatu model.
- RMSE merupakan nilai rata-rata dari jumlah kuadrat kesalahan.
- Semakin rendah RMSE semakin baik kinerja metode pembelajaran mesin

Confussion Matrix

- Confussion Matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi.
- Pada dasarnya confusion matrix mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya.

Confussion Matrix (2)

		Actual Class		
		+	-	
Predicted Class	+	105 TP	12 FP	Predicted as positive = 117
	-	3 FN	53 TN	Predicted as negative = 56
		Positive data= 108	Negative data= 65	Total = 173

- Akurasi = $\frac{TP+TN}{TP+TN+FP+FN}$
- Rasio kesalahan = 1 - akurasi

Confussion Matrix (3)

- True Positive Rate = $\frac{TP}{TP+FN}$
 - Recall
 - Sensitivity
- True Negative Rate = $\frac{TN}{TN+FP}$
 - Specificity

Evaluasi Clustering

- Evaluasi Internal

- Ketika hasil pengelompokan dievaluasi berdasarkan data yang mengelompok itu sendiri
- Metode-metode ini biasanya menetapkan skor terbaik untuk algoritma yang menghasilkan cluster dengan kesamaan tinggi dalam sebuah cluster dan kesamaan rendah antara kluster
- Evaluasi Internal dapat menggunakan Davies Bouldin Index atau Dunn Index

Evaluasi Clustering (2)

- Evaluasi Eksternal

- hasil pengelompokan dievaluasi berdasarkan data yang tidak digunakan untuk pengelompokan, seperti label kelas yang dikenal dan tolok ukur eksternal.
- Tolak ukur tersebut terdiri dari seperangkat item pra-diklasifikasikan, dan set ini sering dibuat oleh (ahli) manusia.
- Evaluasi eksternal dapat menggunakan confusion matrix

Validasi

- Validasi dalam pembelajaran mesin adalah validasi data untuk memastikan bahwa program beroperasi pada data yang benar.

Validasi

- Pembagian dataset:
 - Dua subset: **data training** dan **data testing**
 - Tiga subset: **data training**, **data validation** dan **data testing**
- Data **training** untuk pembentukan model, dan data **testing** digunakan untuk pengujian model
- Data validation untuk memvalidasi model kita valid atau tidak

Cross Validation

- Metode cross-validation digunakan untuk **menghindari overlapping** pada data testing
- Tahapan cross-validation:
 1. Bagi data menjadi **k subset** yg berukuran sama
 2. Gunakan **setiap subset untuk data testing** dan sisanya untuk data training
- Disebut juga dengan **k-fold cross-validation**
- Seringkali subset dibuat stratified (bertingkat) sebelum cross-validation dilakukan, karena **stratifikasi akan mengurangi variansi** dari estimasi

Cross Validation

- Metode evaluasi standard: **stratified 10-fold cross-validation**
- Mengapa **10**? Hasil dari berbagai percobaan yang ekstensif dan pembuktian teoritis, menunjukkan bahwa **10-fold cross-validation adalah pilihan terbaik** untuk mendapatkan hasil validasi yang akurat
- 10-fold cross-validation akan **mengulang pengujian sebanyak 10 kali** dan **hasil pengukuran adalah nilai rata-rata dari 10 kali pengujian**

Cross Validation

Testing	Dataset									
1	Orange	Light Gray	Gray	Light Gray	Gray	Light Gray	Gray	Light Gray	Gray	Light Gray
2	Light Gray	Orange	Gray	Light Gray	Gray	Light Gray	Gray	Light Gray	Gray	Light Gray
3	Light Gray	Light Gray	Orange	Light Gray	Gray	Light Gray	Gray	Light Gray	Gray	Light Gray
4	Light Gray	Light Gray	Gray	Orange	Gray	Light Gray	Gray	Light Gray	Gray	Light Gray
5	Light Gray	Light Gray	Gray	Light Gray	Orange	Light Gray	Gray	Light Gray	Gray	Light Gray
6	Light Gray	Light Gray	Gray	Light Gray	Gray	Orange	Gray	Light Gray	Gray	Light Gray
7	Light Gray	Light Gray	Gray	Light Gray	Gray	Light Gray	Orange	Light Gray	Gray	Light Gray
8	Light Gray	Light Gray	Gray	Light Gray	Gray	Light Gray	Gray	Orange	Gray	Light Gray
9	Light Gray	Light Gray	Gray	Light Gray	Gray	Light Gray	Gray	Light Gray	Orange	Light Gray
10	Light Gray	Light Gray	Gray	Light Gray	Gray	Light Gray	Gray	Light Gray	Gray	Orange