Machine Learning

SUPERVISED LEARNING

Feature = Independet Variable = Variable Bebas = x = Kolom selain Label = materi

Rumah: kamar, kamar mandi, mobil, luas tanah

Label = Dependent Variable = Variable Terikat = y = Jawaban Ujian

Rumah: Harga Rumah

* Regresi
  + Untuk memprediksi sebuah continuous value
  + Continous value itu contoh
    - Harga Rumah: 1,250,000 ; 550,000 ; 435,000
    - 1,000,000 – 2,000,000: 1,535,254.0987865999999999999999999
    - 1-2: 1.5654728999999999
  + Model Machine Learning
    - LinearRegresi (y = bx + a)
      * Model.fit(df[[‘penumpang’]], df[‘harga’])
      * Df[‘prediction’] = Model.predict(df[[‘penumpang’]])
      * MSE = Mean\_squared\_error(df[‘harga’], df[‘prediction’])
      * MSE = 3
      * Prediksi ala LinearRegresi / Jawaban
        + MAE (lr)
        + MSE (lr)
        + RMSE (lr)
        + R2 Score (lr)
    - Ridge (SSE.OLS + lambda (alpha) x sum(b^2))
      * Prediksi ala Ridge / Jawaban
      * MAE (rid)
      * MSE (rid)
      * RMSE (rid)
      * R2 Score (rid)
    - Lasso (SSE.OLS + lambda (alpha) x sum(|b|))
      * MAE (las)
      * MSE (las)
      * RMSE (las)
      * R2 Score (las)
    - ElasticNet
    - Polynomial Regression
    - DecisionTreeRegressor
    - RandomForestRegressor
    - GradientBoostRegressor
  + Evaluation Metrics (Regresi):
    - MAE (Mean Absolute Error) < bagus
    - MSE (Mean Squared Error) < bagus
    - RMSE (Root Mean Squared Error) < bagus
    - R2 Score > bagus
* Klasifikasi

SUPERVISED LEARNING

* + Feature = Independent Variable = x
  + Label = Dependent Variable = y
  + Klasifikasi memprediksi: [angka, text, class]:
    - Notes: kalau dia memprediksi angka, dia ngga lagi sebagai continuos
    - Notes Discrete Value
    - 1 , 2 , 3 , 4 , 5
    - 0, 1
  + Evaluation Metrics (Klasifikasi):
    - Confusion Matrix

Relatively Balanced

* + - 50 : 50
    - 60 : 40
    - 70 : 30
    - Accuracy
      * (TP + TN) / (TP + FP + FN + TN)
      * Relatively Balanced
        + 1 : 0 = 50 : 50
        + 1 : 0 = 60 : 40 / 40 : 60
        + 1 : 0 = 70 : 30 / 30 : 70

Imbalanced Data

* + - 80 : 20
    - 90 : 10
    - Type I Error (False Positive Rate)
      * (FP) / (FP + TN)
    - Type II Error (False Negative Rate)
      * (FN) / (TP + FN)
    - Recall
      * (TP) / (TP + FN)
    - Precision
      * (TP) / (TP + FP)

Cancer Detector

1 = Cancer

0 = No Cancer

Klasifikasi:

* Actual 1 > Prediksi 0 kita ingin memperkecil ratio dari FN
* Recall
  + (TP) / (TP + FN)

Spam Detector

1 = SPAM

0 = HAM

Klasifikasi:

* Actual 0 > Prediksi 1 memperkecil ratio dari FP
* Precision
  + (TP) / (TP + FP)

Andi (cara belajar A)

* Nilai Modul
  + Modul 1 + Modul 2 + Modul 3 / 3 \*40%
* Nilai Final Project
  + Final Project \* 50%
* Nilai Logic Test
  + Logic Test \* 10%
* Nilai Akhir > 80
  + Nilai Modul + Nilai Final Project + Nilai Logic Test

Andi prediksi/jawaban – kunci jawaban untuk melakukan penilaian

Budi (cara belajar B)

* Nilai Modul
  + Modul 1 + Modul 2 + Modul 3 / 3 \*40%
* Nilai Final Project
  + Final Project \* 50%
* Nilai Logic Test
  + Logic Test \* 10%
* Nilai Akhir > 80
  + Nilai Modul + Nilai Final Project + Nilai Logic Test

Budi prediksi/jawaban – Kunci jawaban untuk melakukan penilaian

Korelasi

Threshold

Correlation Poin >= +0.75

Correlation Poin <= -0.75

Correlation Poin itu

+/- itu menunjukkan apakah korelasinya positif atau negative

Angka itu menunjukkan seberapa signifikan

Machine Learning