Analisa Implementasi Naïve Bayes Classifier menggunakan Python

Machine Learning A

Vincent Darmawan / c14190162

# Implementasi menggunakan Library Scikit-learn

Bantuan Library:

* Pandas
* Sklearn.model\_selection
* Sklearn.metrics
* Sklearn.naive\_bayes

Hasil output program scikit\_nb.py:

precision recall f1-score support

1 0.17 1.00 0.30 4

2 1.00 0.30 0.46 10

3 0.75 0.14 0.24 21

4 0.67 1.00 0.80 10

accuracy 0.44 45

macro avg 0.65 0.61 0.45 45

weighted avg 0.74 0.44 0.42 45

Program dibuat menggunakan function GaussianNB yang disediakan oleh library scikit-learn. Dataset dipisah menjadi 0.8 training set dan 0.2 testing set. Random state dari function train\_test\_split() milik scikit-learn di set menjadi 0.

# Implementasi from scratch

Bantuan Library:

* Math
* Scipy.stats
* Numpy
* Pandas
* Sklearn.model\_selection
* Sklearn.metrics

Hasil output program scratch\_nb.py:

precision recall f1-score support

1 0.43 0.75 0.55 4

2 0.86 0.60 0.71 10

3 0.94 0.76 0.84 21

4 0.71 1.00 0.83 10

accuracy 0.78 45

macro avg 0.74 0.78 0.73 45

weighted avg 0.83 0.78 0.78 45

Program dibuat dari scratch untuk algoritma utama naïve bayes. Implementasi algoritma mendapat modifikasi yaitu daripada menggunakan product of probabilities menjadi sum of log probabilites untuk menghindari masalah numerical underflow.

P(yi | x1, x2, …, xn) = log(P(x1|y1)) + log(P(x2|y1)) + … log(P(xn|y1)) + log(P(yi))

Dataset tetap sama yaitu dipisah menjadi 0.8 training set dan 0.2 testing set. Random state dari function train\_test\_split() milik scikit-learn di set menjadi 0.

# Analisa

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Category | Precision | Recall | F1-score | Support |
| 1 | 0.17 | 1.00 | 0.30 | 4 |
| 2 | 1.00 | 0.30 | 0.46 | 10 |
| 3 | 0.75 | 0.14 | 0.24 | 21 |
| 4 | 0.67 | 1.00 | 0.80 | 10 |
| Accuracy |  |  | 0.44 | 45 |

Table 1 Classification Report scikit\_nb.py

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Category | Precision | Recall | F1-score | Support |
| 1 | 0.43 | 0.75 | 0.55 | 4 |
| 2 | 0.86 | 0.60 | 0.71 | 10 |
| 3 | 0.94 | 0.76 | 0.84 | 21 |
| 4 | 0.71 | 1.00 | 0.83 | 10 |
| Accuracy |  |  | 0.78 | 45 |

Table 2 Classification Report scratch\_nb.py

Dari hasil classification report diatas, program scratch\_nb.py memiliki akurasi yang lebih tinggi daripada scikit\_nb.py dengan akurasi sebesar 0.78 dibandingkan 0.44. scratch\_nb.py memiliki presisi untuk kategori 1, 3, dan 4 lebih tinggi sedangkan scikit\_nb.py unggul di presisi kategori 2. Scratch\_nb.py juga unggul di bagian recall pada kategori 2 dan 3 dengan kategori 4 memiliki skor yang sama dan kalah pada recall kategori 1 dengan scikit\_nb.py. F1 score dari scratch\_nb.py unggul di semua kategori dibandingkan dengan scikit\_nb.py

Program scratch\_nb.py meskipun tidak dituliskan pada laporan ini dapat berhasil menebak benar sebanyak 25 kali dari total 45 data di test set.

Program scikit\_nb.py menebak benar sebanyak 20 kali dari 45 data di test set.

# Referensi

* <https://scikit-learn.org/stable/modules/naive_bayes.html>
* <https://machinelearningmastery.com/classification-as-conditional-probability-and-the-naive-bayes-algorithm/>
* <https://machinelearningmastery.com/naive-bayes-classifier-scratch-python/>
* <https://www.youtube.com/watch?v=H3EjCKtlVog&t=222s>