

Seleksi Asisten Lab AI '23

Supervised Learning: KNN

1. Cara Kerja Algoritma K-Nearest Neighbors.

Konsep inti: K-Nearest Neighbors (KNN) adalah algoritma supervised learning yang fundamentalnya didasarkan pada prinsip intuisi: sebuah titik data kemungkinan besar termasuk dalam kelas yang sama dengan titik-titik data lain di sekitarnya.

Pseudocode untuk Prediksi KNN:

Algorithm 1

```

1: function PREDICT( $X_{train}, y_{train}, x_{new}, k$ )
2:    $distances \leftarrow []$ 
3:   for each point  $p$  in  $X_{train}$  do
4:      $dist \leftarrow \text{calculate\_distance}(x_{new}, p)$ 
5:     add ( $dist$ , corresponding_label) to  $distances$ 
6:   end for
7:
8:   sort  $distances$  in ascending order
9:    $k\_nearest\_neighbors \leftarrow$  first  $k$  elements from sorted  $distances$ 
10:
11:   $prediction \leftarrow$  modus label in  $k\_nearest\_neighbors$ 
12:
13:  return  $prediction$ 
14: end function

```

2. Analisis Perbandingan Eksperimen.

Table 1: Perbandingan kinerja

Metric	From Scratch	Scikit-learn
Accuration	0.7630	0.7630
Precision	0.55	0.55
Time (s)	56.7	0.018

Dapat dilihat bahwa implementasi from scratch dengan implementasi menggunakan Scikit-learn memiliki hasil metrik evaluasi yang sama selain waktu prediksi.

3. *Improvement* yang bisa dilakukan untuk mencapai hasil yang lebih baik.

Beberapa bagian yang dapat ditingkatkan adalah

- *Optimasi Parameter k .*
Nilai k yang dipilih belum tentu optimal. Jika terlalu kecil bisa terlalu sensitif, sedangkan jika terlalu besar bisa terlalu kabur atau bias. Sehingga perlu dilakukan nilai k yang optimal.
- *Imbalanced Dataset*
Pada dataset, jumlah pelanggan yang churn itu jauh lebih sedikit daripada yang tidak. Sehingga mungkin perlu ada cara untuk membuat sampel sintetis barunya.
- *Feature Selection*
Dataset memiliki 30 fitur yang mungkin beberapa tidak relevan dan akan menjadi noise yang mengganggu model, sehingga mungkin perlu ada cara untuk mencari yang penting saja.