

Coding & Machine Learning – Programming

Assignment

University of Technology Yogyakarta, academic year 2022/23.

Handed out: June 6, 2021. Due: June 11, 2021 at 23.59.

Sebuah perusahaan e-commerce bernama "SuperMart" ingin meningkatkan efisiensi pengiriman paket ke pelanggan mereka. SuperMart ingin memahami pola pembelian pelanggan untuk mengelompokkan wilayah pengiriman yang serupa. Setelah melakukan survei, mereka mendapatkan beberapa atribut pelanggan seperti usia, jenis kelamin, dan pendapatan. SuperMart juga mendapatkan data historis pembelian pelanggan di berbagai kategori produk.

Mereka memutuskan untuk menggunakan pendekatan kombinasi k-means, kolaborasi Naive Bayes, dan pohon keputusan untuk mencapai tujuan ini. Langkah-langkah yang mereka ambil adalah sebagai berikut:

1. K-means Clustering:

SuperMart menggunakan algoritma k-means untuk mengelompokkan pelanggan berdasarkan atribut usia, jenis kelamin, dan pendapatan. Mereka memutuskan untuk membentuk tiga kelompok pelanggan. Berikut adalah hasil dari k-means clustering:

- Kelompok 1: Pelanggan dengan usia muda, jenis kelamin laki-laki, dan pendapatan rendah.

- Kelompok 2: Pelanggan dengan usia sedang, jenis kelamin perempuan, dan pendapatan menengah.
- Kelompok 3: Pelanggan dengan usia tua, jenis kelamin laki-laki, dan pendapatan tinggi.

2. Kolaborasi Naive Bayes:

Setelah kelompok pelanggan terbentuk, SuperMart ingin menganalisis pola pembelian pelanggan di setiap kelompok. Mereka menggunakan algoritma kolaborasi Naive Bayes untuk menganalisis pola pembelian produk di setiap kelompok. Berikut adalah hasil analisis mereka:

- Kelompok 1: Pelanggan pada kelompok ini cenderung membeli produk-produk elektronik dan pakaian.
- Kelompok 2: Pelanggan pada kelompok ini cenderung membeli produk-produk makanan dan minuman.
- Kelompok 3: Pelanggan pada kelompok ini cenderung membeli produk-produk perawatan kesehatan dan kecantikan.

3. Pohon Keputusan:

Berdasarkan hasil analisis dari kolaborasi Naive Bayes, SuperMart ingin membangun pohon keputusan untuk mengambil keputusan pengiriman paket. Mereka menggunakan atribut kelompok pelanggan dan kategori produk yang dibeli sebagai

atribut untuk membangun pohon keputusan. Berikut adalah pohon keputusan yang terbentuk:

- Jika pelanggan termasuk dalam Kelompok 1 dan membeli produk elektronik atau pakaian, paket dikirim melalui kurir ekspres untuk pengiriman cepat.
- Jika pelanggan termasuk dalam Kelompok 2 dan membeli produk makanan atau minuman, paket dikirim melalui layanan pengiriman reguler.
- Jika pelanggan termasuk dalam Kelompok 3 dan membeli produk perawatan kesehatan atau kecantikan, paket dikirim melalui kurir khusus untuk keamanan dan perlindungan produk.

SOAL :

1. Identifikasi dan jelaskan langkah-langkah yang dilakukan SuperMart dalam mengoptimalkan pengiriman paket mereka menggunakan kombinasi k-means, kolaborasi Naive Bayes, dan pohon keputusan.
2. Berdasarkan cerita di atas, apakah Anda dapat memberikan contoh penggunaan k-means clustering, kolaborasi Naive Bayes, dan pohon keputusan dalam situasi nyata lainnya?

LANGKAH LANGKAH Pengerjaan Python :

1. Definisikan dataset pelanggan dan atributnya, termasuk kategori pembelian.
2. Lakukan pengelompokan dengan menggunakan algoritma k-means clustering pada atribut numerik yang relevan (misalnya usia dan pendapatan).
 - Lakukan pemrosesan data untuk mengubah dataset menjadi format yang sesuai dengan k-means clustering.
 - Buat objek `KMeans` dan tentukan jumlah kluster yang diinginkan.
 - Latih model menggunakan metode `.fit()` pada dataset.
 - Dapatkan hasil kelompok untuk setiap pelanggan menggunakan atribut `.labels_` pada objek k-means.
3. Lakukan analisis pembelian dengan menggunakan kolaborasi Naive Bayes pada atribut kategori pembelian.
 - Lakukan pemrosesan data untuk mengubah dataset menjadi format yang sesuai dengan kolaborasi Naive Bayes.
 - Buat objek `DictVectorizer` untuk mengubah dataset menjadi representasi vektor.
 - Buat objek `MultinomialNB` dan latih model menggunakan metode `.fit()` pada dataset yang telah diubah.
 - Dapatkan hasil analisis pembelian untuk setiap kelompok dengan menyimpan kategori pembelian setiap pelanggan pada kelompok yang sesuai.
4. Bangun pohon keputusan berdasarkan hasil analisis.
 - Lakukan pemrosesan data untuk mengubah dataset menjadi format yang sesuai dengan pohon keputusan.

- Buat objek `'DecisionTreeClassifier'`.
- Latih model menggunakan metode `'.fit()'` pada dataset.

5. Lakukan prediksi pengiriman paket berdasarkan pohon keputusan.

- Buat fungsi untuk melakukan prediksi pengiriman paket dengan memasukkan data pelanggan dan kategori pembelian.
- Gunakan metode `'.predict()'` pada objek pohon keputusan untuk mendapatkan hasil prediksi.
- Tampilkan hasil prediksi.

Berikut adalah contoh langkah-langkah pengerjaan dalam kode Python:

```
# Langkah 1: Definisikan dataset pelanggan dan atributnya
# Langkah 2: Lakukan pengelompokan dengan k-means clustering
# Langkah 3: Lakukan analisis pembelian dengan kolaborasi Naive Bayes
# Langkah 4: Bangun pohon keputusan
# Langkah 5: Lakukan prediksi pengiriman paket
'''
```

Anda dapat mengisi langkah-langkah tersebut dengan implementasi kode yang telah dijelaskan sebelumnya. Pastikan untuk memasukkan kode yang relevan untuk masing-masing langkah dalam urutan yang tepat.

Harapannya, langkah-langkah pengerjaan ini membantu Anda memahami bagaimana membangun dan menggunakan model k-means clustering, kolaborasi Naive Bayes, dan pohon keputusan dalam konteks yang diberikan. Jika Anda memiliki pertanyaan lebih lanjut, jangan ragu untuk bertanya.

LANGKAH-LANGKAH Pengerjaan Java:

1. Definisikan dataset pelanggan dan atributnya, termasuk kategori pembelian.
2. Lakukan pengelompokan dengan menggunakan algoritma k-means clustering pada atribut numerik yang relevan (misalnya usia dan pendapatan).
 - Lakukan pemrosesan data untuk mengubah dataset menjadi format yang sesuai dengan k-means clustering.
 - Buat objek `KMeans` dan tentukan jumlah kluster yang diinginkan.
 - Latih model menggunakan metode `.fit()` pada dataset.
 - Dapatkan hasil kelompok untuk setiap pelanggan menggunakan atribut `.predict()` pada objek k-means.
3. Lakukan analisis pembelian dengan menggunakan kolaborasi Naive Bayes pada atribut kategori pembelian.
 - Lakukan pemrosesan data untuk mengubah dataset menjadi format yang sesuai dengan kolaborasi Naive Bayes.
 - Buat objek `NaiveBayes` (sesuai dengan library yang digunakan) dan latih model menggunakan metode `.train()` pada dataset.
 - Dapatkan hasil analisis pembelian untuk setiap kelompok dengan menyimpan kategori pembelian setiap pelanggan pada kelompok yang sesuai.
4. Bangun pohon keputusan berdasarkan hasil analisis.
 - Lakukan pemrosesan data untuk mengubah dataset menjadi format yang sesuai dengan pohon keputusan.

- Buat objek `'DecisionTree'` (sesuai dengan library yang digunakan) dan latih model menggunakan metode `'.fit()'` pada dataset.

5. Lakukan prediksi pengiriman paket berdasarkan pohon keputusan.

- Buat metode atau fungsi untuk melakukan prediksi pengiriman paket dengan memasukkan data pelanggan dan kategori pembelian.

- Gunakan metode `'.predict()'` pada objek pohon keputusan untuk mendapatkan hasil prediksi.

- Tampilkan hasil prediksi.

Berikut adalah contoh langkah-langkah pengerjaan dalam kode Java:

// Langkah 1: Definisikan dataset pelanggan dan atributnya

// Langkah 2: Lakukan pengelompokan dengan k-means clustering

// Langkah 3: Lakukan analisis pembelian dengan kolaborasi Naive Bayes

// Langkah 4: Bangun pohon keputusan

// Langkah 5: Lakukan prediksi pengiriman paket

Anda dapat mengisi langkah-langkah tersebut dengan implementasi kode yang relevan untuk masing-masing langkah dalam urutan yang tepat. Pastikan untuk menggunakan library atau framework yang sesuai untuk implementasi algoritma k-means, kolaborasi Naive Bayes, dan pohon keputusan di Java.

Harapannya, langkah-langkah pengerjaan ini membantu Anda memahami bagaimana membangun dan menggunakan model k-means clustering, kolaborasi Naive Bayes, dan pohon keputusan dalam bahasa pemrograman Java. Jika Anda memiliki pertanyaan lebih lanjut, jangan ragu untuk bertanya.

TEKNIS PENGUMPULAN :

1. Simpan kode dengan format (.py) atau (.ipynb)
2. Apabila menggunakan data eksternal (csv, xlsx, txt, data) sertakan pada folder
3. Simpan semua menggunakan format (.ZIP) atau (.RAR)
4. Format pengumpulan NIM_NAMA_MACHINE LEARNING-PA-4.ZIP
5. Contoh : **5210411111_SUPARDI_MACHINE LEARNING-PA-4.ZIP**
6. **Kumpulkan pada : bit.ly/ML-PA4-23-IF**



TEKNIS PENILAIAN :

1. Fungsionalitas (50%):
2. Kebershan Kode (20%)
3. Konsep Pemrograman (20%)
4. Kreativitas (10%)
5. Melebihi batas pengumpulan -5 poin (setiap 1 menit keterlambatan)
6. ERROR CODE mendapatkan nilai minimal
7. Data lebih beragam membantu hasil akhir