

Simulasi Model Multidimensional Performance Based Clustering Untuk Mengenal Pola Pasar Pada Perusahaan Berskala Nasional

Tim Pencipta:

Kadek Cahya Dewi

Putu Indah Ciptayani

Ni Wayan Dewinta Ayuni

Ida Bagus Putu Sandhi Yudistira

OKTOBER

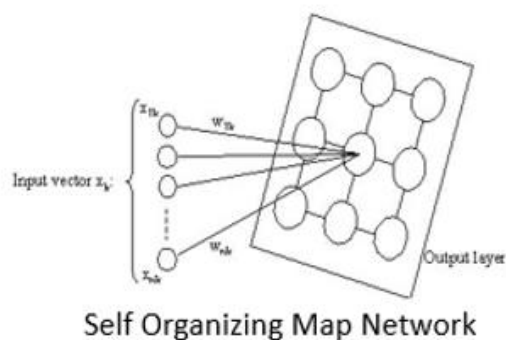
2022

Deskripsi dan Spesifikasi Produk Simulasi Model Multidimensional Performance Based Clustering Untuk Mengenali Pola Pasar pada Perusahaan Berskala Nasional

A. Deskripsi Model

PT XYZ (Persero) adalah badan usaha berskala nasional yang bergerak di bidang jasa di Indonesia. Secara teritori, PT. XYZ memiliki 7 Regional yang membawahi 61 wilayah. Setiap wilayah memiliki unit penjualan yang menyasar teritori dan segmen pasar berbeda. Fitur pelaporan performa yang saat ini dimiliki PT. XYZ belum bisa menampilkan secara detail sesuai kebutuhan level operasional. Sehingga harus dilakukan proses melalui Ms. Excel untuk mendapatkan informasi yang dibutuhkan. Unit Business Service PT. XYZ telah mengembangkan sistem internal berbasis wordpress namun sistem tersebut belum mampu melakukan pengenalan pola. Pengenalan pola dapat digunakan sebagai dasar dalam menentukan strategi penjualan. Dengan melihat pola yang tercipta maka dapat memudahkan dalam memahami perkembangan pasar. Analisa trend dan prognosa realisasi sales pun dapat dilakukan. Melihat pentingnya pengenalan pola pada performa penjualan di PT. XYZ, sangatlah penting untuk membuat sebuah produk yang dapat membantu dalam mengenali pola pasar melalui multidimensional performance based clustering. Hasil dari cluster analysis tersebut diharapkan dapat ditampilkan ke dashboard dan memberikan early warning.

Simulasi yang diciptakan bertujuan sebagai purwarupa model multidimensional performance based clustering. Melalui simulasi yang tercipta maka dapat diketahui dimensi yang paling optimal dalam multidimensional performance based clustering. Setelah melewati 23 kali uji coba, maka dimensi terbaik pada multidimensional performance based clustering adalah dimensi produk, jenis transaksi, dan bulanan. Model clustering menggunakan Artificial Neural Network Self Organizing Map dengan topologi seperti pada Gambar 1. Selain dimensi terbaik, pada model juga ditemukan size map SOM terbaik yaitu 19x10 dengan learning rate sebesar 0,9.



Gambar 1. Topologi Self Organizing Map

B. Print Out Program

Program simulasi clustering dibuat menggunakan bahasa pemrograman Python. Beberapa modul pada python yang digunakan adalah modul numpy, pandas, matplotlib, MiniSom.

```
from minisom import MiniSom
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

plt.style.use('fivethirtyeight')
df = pd.read_csv("dataset-produk-trans-bulanan.csv") #ujicoba berbagai dataset
data = df.values

som_shape = (19, 10)

som = MiniSom(som_shape[0], som_shape[1], data.shape[1], sigma=0.5,
learning_rate=0.7)

max_iter = 1000
q_error = []
t_error = []

for i in range(max_iter):
    rand_i = np.random.randint(len(data))
    som.update(data[rand_i], som.winner(data[rand_i]), i, max_iter)
    q_error.append(som.quantization_error(data))
    t_error.append(som.topographic_error(data))

plt.plot(np.arange(max_iter), q_error, label='quantization error')
plt.plot(np.arange(max_iter), t_error, label='topographic error')
plt.ylabel('Quantization error')
```

```

plt.xlabel('Iteration index')
plt.legend()
print("q_error:",q_error[max_iter-1])
print("t_error:",t_error[max_iter-1])
plt.show()

# each neuron represents a cluster
winner_coordinates = np.array([som.winner(x) for x in data]).T
print("done winner")
# with np.ravel_multi_index we convert the bidimensional
# coordinates to a monodimensional index
cluster_index = np.ravel_multi_index(winner_coordinates, som_shape)
print("done cluster", cluster_index)
# Plotting the clusters
plt.figure(figsize=(10,8))

for c in np.unique(cluster_index):
    plt.scatter(data[cluster_index == c, 0],
               data[cluster_index == c, 1], label='cluster='+str(c), alpha=.7)

# Plotting centroids
print("som.get_weights():",som.get_weights())
for centroid in som.get_weights():
    plt.scatter(centroid[:, 0], centroid[:, 1], marker='x',
               s=10, linewidths=20, color='k') # label='centroid'

plt.title("Clusters of Performance")
plt.legend();
plt.show()

```

C. Uji Coba Program

Program diuji dengan dataset dari data sales Unit Business Service PT XYZ (Persero) tahun 2019 – 2021. Uji coba program clustering dilakukan dengan 2 model dataset. Bentuk dataset yang pertama memiliki jumlah kolom sebanyak 42 yang merupakan representasi dari AM, jumlah transaksi untuk product 1 s/d jumlah transaksi untuk product 27, jumlah transaksi untuk jenis transaksi 1 s/d jumlah transaksi untuk jenis transaksi 9, serta jumlah transaksi untuk status 1 s/d jumlah transaksi untuk status 5. Sedangkan bentuk dataset yang kedua memiliki jumlah kolom sebanyak 5 yang merupakan representasi sales person, produk, jenis transaksi, status pemasangan, dan bulanan.

Untuk mengevaluasi kualitas peta fitur (*feature map*) yang dihasilkan oleh SOM, dapat dengan menggunakan dua indikator yaitu kualitas pembelajaran dan kualitas proyeksi. Indikator kualitas pembelajaran ditentukan melalui pengukuran quantization error (QE). Indikator kualitas proyeksi ditentukan melalui pengukuran topographical error (TE). Jika nilai QE dan TE kecil, maka peta fitur akan dinilai dengan kualitas yang baik

Tabel 1 menunjukkan hasil evaluasi pada uji coba pertama. Tabel 2 menunjukkan hasil evaluasi pada uji coba kedua. Penentuan size map pada SOM adalah sesuai dengan dokumentasi pada library MiniSom, yaitu untuk mengatur ukuran grid dalam pengurangan dimensi adalah grid tersebut harus berisi neuron $5*\sqrt{N}$ di mana N adalah jumlah sampel dalam kumpulan data yang akan dianalisis.

Tabel 1. Hasil Uji Coba Pertama (Size Map 5x5, learning rate 0,5)

Uji Coba	Sample	Dimensi	Field	q_error	t_error	jumlah centroid
1.1	27	42	AM, product, jenis transaksi, status	3.429490257941563	0.2692307692307692	12
1.2	27	36	AM, product, jenis transaksi	2.2963316931503885	0.23076923076923078	13
1.3	27	27	AM, product	0.4819327257840914	0.15384615384615385	14
1.4	27	32	AM, product, status	0.8639629013009049	0.4230769230769231	15
1.5	27	14	AM, jenis transaksi, status	2.544069310100512	0.38461538461538464	14
1.6	27	9	AM, jenis transaksi	0.8341135979233983	0.037037037037037035	13
1.7	27	5	AM, status	0.4317648017304434	0.34615384615384615	12

Tabel 2. Hasil Uji Coba Kedua

Uji Coba	Sample	Size Map	Learning rate	Field & Dimensi	q_error	t_error	jumlah centroid
2.1	1734	20x10	0,5	AM, product, jenis transaksi, status, bulanan (5)	2.5219058877527423	0.22606689734717417	17
2.2	1734	14x15	0,5	AM, product, jenis transaksi, status, bulanan (5)	2.2762854389591074	0.34256055363321797	12
2.3	1734	17x12	0,5	AM, product, jenis transaksi, status, bulanan (5)	2.661172915719404	0.42675893886966554	16
2.4	1517	15x13	0,5	AM, product, jenis, bulanan (4)	2.2971523252776733	0.4831905075807515	18
2.5	1517	19x10	0,5	AM, product, jenis, bulanan (4)	2.2972566027702337	0.12788398154251812	19
2.6	1510	15x13	0,5	AM, product, bulanan (3)	1.7386179761579488	0.23973509933774834	17
2.7	1510	19x10	0,5	AM, product, bulanan (3)	1.5850051615413716	0.27483443708609273	18
2.8	1734	20x10	0,7	AM, product, jenis transaksi, status, bulanan (5)	2.332054038763102	0.35294117647058826	18
2.9	1734	14x15	0,7	AM, product, jenis transaksi, status, bulanan (5)	1.9778355774944654	0.3742791234140715	19
2.10	1734	17x12	0,7	AM, product, jenis transaksi, status, bulanan (5)	2.1878320938851883	0.23241061130334487	18
2.11	1517	15x13	0,7	AM, product, jenis transaksi, bulanan (4)	2.8539734283288487	0.1034937376400791	13
2.12	1517	19x10	0,7	AM, product, jenis transaksi, bulanan (4)	1.9895143659963321	0.3197099538562953	22
2.13	1510	15x13	0,7	AM, product, bulanan (3)	1.393715804022236	0.4066225165562914	23
2.14	1510	19x10	0,7	AM, product, bulanan (3)	1.295846498239518	0.27880794701986755	22
2.15	1517	15x13	0,7	product, jenis transaksi, bulanan (3)	0.6620095117198568	0.2623599208965063	16
2.16	1517	19x10	0,7	product, jenis transaksi, bulanan (3)	0.9494841021101258	0.13052076466710613	12

Berdasarkan hasil pada tabel 1 dan 2, ada empat kombinasi dimensi yang memiliki nilai QE dan TE terbaik yaitu dimensi (1) sales person dan produk, (2) sales person dan jenis transaksi, (3) sales person, produk, jenis transaksi dan bulanan, serta (4) produk, jenis transaksi dan bulanan.

Focus group discussion dilakukan dalam rangka menentukan dimensi terbaik yang digunakan dalam clustering berdasarkan 4 kandidat dimensi dari hasil uji coba pertama dan kedua. Dengan pertimbangan bentuk dataset, kebutuhan data serta berdasarkan hasil qe dan te maka dimensi terbaik yang digunakan dalam multidimensional performance based clustering adalah produk, jenis transaksi dan bulanan dengan size map 19x10 dan besar learning rate 0,7. Dimensi – dimensi tersebut dinilai mampu merepresentasikan performa sales sehingga dapat mengenali pola pasar dan menjadi bahan dalam analisa trend dan prognosa realisasi sales.

Hasil Clustering dari dimensi terpilih adalah:

