

## Penerapan *Learning Vector Quantization* (LVQ) pada Klasifikasi Indeks Pembangunan Manusia di Indonesia

Fryska Alvianita<sup>1\*)</sup>, Bowo Winarno<sup>2</sup>, & Titin Sri Martini<sup>3</sup>  
<sup>1,2,3</sup>Universitas Sebelas Maret

### INFO ARTICLES

#### Key Words:

Data mining, klasifikasi, *learning vector quantization*



This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

**Abstract:** *The Human Development Index (HDI) is an important indicator to measure success in efforts to build the quality of human life. HDI explains how residents access development results in obtaining income, health, education, and so on. This research aim to apply the Learning Vector Quantization (LVQ) algorithm to the classification of HDI data at the district/city level. In data mining, an object whose label is unknown can classify it by classification. The LVQ algorithm is useful for speeding up the calculation process effectively and efficiently. The K-Fold Cross Validation method is used to obtain maximum classification results by dividing training data and testing data randomly by K in K-Fold. The results of the classification by applying the LVQ algorithm using 10-Fold obtained the best accuracy of 92.59%.*

**Abstrak:** *Indeks Pembangunan Manusia (IPM) merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia. IPM menjelaskan bagaimana penduduk dalam mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya. Penelitian ini bertujuan untuk menerapkan algoritme Learning Vector Quantization (LVQ) pada klasifikasi data IPM di tingkat kabupaten/kota. Dalam data mining, suatu objek yang labelnya belum diketahui dapat ditentukan kelasnya dengan klasifikasi. Algoritme LVQ bermanfaat untuk mempercepat proses perhitungan secara efektif dan efisien. Metode K-Fold Cross Validation digunakan untuk memperoleh hasil klasifikasi yang maksimum dengan membagi data training dan data testing secara acak sebanyak K pada K-Fold. Hasil dari klasifikasi dengan menerapkan algoritme LVQ menggunakan 10-Fold diperoleh akurasi terbaik yaitu 92.59 %.*

**Correspondence Address:** Jl. Ir Sutami No.36, Kentingan, Kec. Jebres, Kota Surakarta, Jawa Tengah 57126, Indonesia; e-mail: [fryskaalvia27@student.uns.ac.id](mailto:fryskaalvia27@student.uns.ac.id)

**How to Cite (APA 6<sup>th</sup> Style):** Alvianita, F., Winarno, B., & Martini, T. S. (2023). Penerapan *Learning Vector Quantization* (LVQ) pada Klasifikasi Indeks Pembangunan Manusia di Indonesia. *Prosiding Diskusi Panel Nasional Pendidikan Matematika*, 149-154.

**Copyright:** Alvianita, F., Winarno, B., & Martini, T. S. (2023)

## PENDAHULUAN

Indonesia merupakan negara keempat yang memiliki jumlah penduduk terbesar di dunia. Berdasarkan Kemendagri diketahui bahwa jumlah penduduk di Indonesia mencapai 273.879.750 jiwa (Direktorat Jenderal Duckcapil, 2022). Dengan meningkatnya jumlah penduduk diperlukan peningkatan dalam pembangunan kualitas sumber daya manusia melalui Indeks Pembangunan Manusia (IPM). IPM merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia (masyarakat/penduduk)(Badan Pusat Statistika, 2022). IPM menjelaskan bagaimana penduduk dalam mengakses hasil pembangunan untuk memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya.

Peningkatan kualitas sumber daya manusia diperlukan guna mempersiapkan manusia Indonesia yang mampu bersaing dalam berbagai aspek baik dalam lingkup nasional maupun secara internasional. Adanya upaya peningkatan kualitas manusia dengan IPM, mendukung pemerataan dan pengadaan berbagai fasilitas untuk menunjang kehidupan yang layak di Indonesia. *Data mining* diperlukan dalam mendapatkan informasi yang bermanfaat dari data IPM di Indonesia. *Data mining* disebut sebagai proses penemuan, penggalian, dan ekstraksi pengetahuan dari suatu data dengan pola analisis (Raval, 2012). *Data mining* digunakan untuk mengklasifikasi, memprediksi, memperkirakan dan mendapatkan informasi lain yang bermanfaat dari suatu data (Mardi, 2017). Klasifikasi adalah proses untuk menetapkan label kelas dari dataset yang label kelasnya tidak diketahui (Nikam, 2015).

Dalam *machine learning*, klasifikasi dilakukan pada data yang sudah diberi label menggunakan pendekatan terawasi (*supervised learning*). Algoritme klasifikasi yang digunakan merupakan penerapan dari Jaringan Syaraf Tiruan yaitu *Learning Vector Quantization* (LVQ). Pada algoritme LVQ, proses yang terjadi pada setiap neuron yaitu mencari jarak terdekat antara suatu vektor masukan ke bobot yang berkaitan. Kelebihan algoritme ini selain mencari jarak terdekat, yaitu selama pembelajaran unit *output* diposisikan dengan mengatur dan memperbarui bobot melalui pembelajaran yang terawasi untuk memperkirakan keputusan klasifikasi (Kusumadewi, 2006).

Pada penelitian ini dengan menerapkan algoritme LVQ dilakukan klasifikasi IPM di Indonesia. Penelitian dengan menggunakan algoritme LVQ pernah dilakukan oleh Hamidi dkk (2017) untuk melakukan klasifikasi kualitas air sungai. K lasifikasi menggunakan algoritme LVQ pada data IPM ini dilakukan pembagian data *training* dan data *testing* menggunakan metode *K-Fold Cross Validation* sehingga mendapatkan hasil klasifikasi dengan akurasi terbaik.

## METODE

Metode penelitian dibagi menjadi dua bagian yaitu data penelitian dan langkah penelitian.

### 1. Data Penelitian

Data yang digunakan merupakan data sekunder yang dipublikasikan oleh Badan Pusat Statistika tahun 2021 yang berupa data numerik. Dalam penelitian ini terdapat lima variabel pengamatan yang terdiri dari variabel bebas dan variabel terikat. Variabel bebas yang digunakan yaitu umur harapan hidup saat lahir (UHH), rata-rata lama sekolah (RLS), harapan lama sekolah (HLS), pengeluaran per kapita disesuaikan. Sedangkan, variabel terikatnya yaitu indeks pembangunan manusia (IPM). Menurut Badan Pusat Statistika (2022) data IPM dapat dikategorikan, sehingga diperoleh kelas target dengan capaian IPM di suatu wilayah dapat dikelompokkan menjadi empat kategori, yaitu rendah ( $IPM < 60$ ), sedang ( $60 \leq IPM < 70$ ), tinggi ( $70 \leq IPM < 80$ ), dan sangat tinggi ( $IPM \geq 80$ ).

### 2. Langkah Penelitian

Berikut merupakan langkah penelitian yang dilakukan untuk mencapai tujuan penelitian.

- (1) Mempelajari literatur-literatur tentang algoritme LVQ dan metode *K-Fold Cross Validation*.
- (2) Mengumpulkan data penelitian
- (3) Melakukan normalisasi data
- (4) Membagi data menjadi data *testing* dan data *training* dengan metode *K-Fold Cross Validation*
- (5) Melakukan klasifikasi pada data *training* dengan algoritme LVQ
- (6) Menguji model pada data *testing*
- (7) Melakukan analisis dari hasil klasifikasi dan akurasinya

## HASIL

### 1. Normalisasi *Min-Max*

Normalisasi *Min-Max* merupakan metode normalisasi dengan melakukan transformasi linier terhadap data asli sehingga menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses (Hanifa, 2017). Metode ini dapat menggunakan rumus sebagai berikut.

$$Normalized(x) = \frac{minRange + (x - minValue)(maxRange - minRange)}{maxValue - minValue}$$

Dilakukan normalisasi pada data IPM menggunakan metode normalisasi *Min-Max* sehingga diperoleh data IPM pada rentang yang sama, ditunjukkan sampel data setelah dilakukan normalisasi pada Tabel 1 berikut.

**Tabel 1 Sampel Data Hasil Normalisasi *Min-Max***

UHH	RLS	HLS	Pengeluaran per kapita disesuaikan	IPM
0.65	0.70	0.75	0.28	2
0.44	0.71	0.72	0.16	1
0.54	0.64	0.75	0.24	1
0.40	0.65	0.77	0.21	1
0.57	0.72	0.73	0.20	1
0.60	0.60	0.66	0.23	1
0.60	0.74	0.75	0.34	2
0.56	0.71	0.77	0.28	2
0.64	0.78	0.78	0.28	2

### 2. *K-Fold Cross Validation*

*Cross Validation* adalah sebuah metode dari teknik *data mining* yang bertujuan untuk memperoleh hasil akurasi maksimum saat data dibagi menjadi data *training* dan data *testing*. *K-Fold Cross Validation* adalah salah satu dari jenis pengujian *Cross Validation*, yang berfungsi untuk menilai kinerja sebuah algoritme dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai *K* pada *K-Fold*. Data yang telah dibagi (data partisi) tersebut diolah sejumlah *K* kali eksperimen dengan masing-masing eksperimen menggunakan data partisi ke-*K* sebagai data *testing* dan menggunakan sisa partisi lainnya sebagai data *training* (Kurniawan, 2017).

Pada penelitian ini digunakan nilai *K=10*, yaitu data dibagi secara acak dan dikelompokkan sebanyak 10 kelompok data dengan *10-Fold*. Berikut pada Gambar 1 rancangan pembagian data *testing* dan data *training* berdasarkan 10 kelompok data dengan *10-Fold*.

Fold	1	1	1	1	1	1	1	1	1	1
	2	2	2	2	2	2	2	2	2	2
	3	3	3	3	3	3	3	3	3	3
	4	4	4	4	4	4	4	4	4	4
	5	5	5	5	5	5	5	5	5	5
	6	6	6	6	6	6	6	6	6	6
	7	7	7	7	7	7	7	7	7	7
	8	8	8	8	8	8	8	8	8	8
	9	9	9	9	9	9	9	9	9	9
	10	10	10	10	10	10	10	10	10	10

data testing
  data training

**Gambar 1. Pembagian Data dengan 10-Fold**

### 3. Learning Vector Quantization

Menurut Wibowo (2014) LVQ digunakan untuk pengelompokkan dimana jumlah kelompok sudah ditentukan arsitekturnya (target/kelas sudah ditentukan). Algoritme diusulkan oleh Kohonen pada tahun 1986 sebagai perbaikan dari *Vector Quantization*. Arsitektur LVQ terdiri dari lapisan *input* (*input layer*), lapisan kompetitif dimana terjadi kompetisi pada *input* untuk masuk ke dalam suatu kelas berdasarkan kedekatan jaraknya, dan lapisan *output* (*output layer*). Lapisan *input* dihubungkan dengan lapisan kompetitif oleh bobot. Dalam lapisan kompetitif, proses pembelajaran dilakukan secara terawasi. *Input* bersaing untuk masuk ke dalam suatu kelas.

Menurut Fausset (1994), pada arsitektur LVQ terdapat  $x$  vektor *input* ( $x_1, \dots, x_i, \dots, x_n$ ),  $w_j$  vektor bobot untuk *output* ke- $j$  ( $w_{1j}, \dots, w_{ij}, \dots, w_{nj}$ ), dan  $Y$  vektor *output* ( $y_1, \dots, y_j, \dots, y_m$ ). Ditetapkan parameter yaitu error minimum  $\min \alpha$ , fungsi pembelajaran  $dec \alpha$ , iterasi (*epoch*), iterasi maksimum (*max epoch*), target atau label kelas ( $T$ ),  $C_j$  kelas ke- $j$  dari selisih bobot terkecil,  $J$  kelas ke- $j$  untuk  $\|x_i - w_j\|$  minimum,  $\|x_i - w_j\|$  merupakan jarak *euclidean* antara vektor *input* dan vektor bobot untuk *output* ke- $j$ , sehingga algoritme pelatihan LVQ sebagai berikut.

Langkah 0. Inisialisasi bobot awal  $w_j$  dan *learning rate*  $\alpha$ .

Langkah 1. Lakukan langkah 2-6 jika  $epoch < max\ epoch$  dan  $\alpha > min\ \alpha$ . Dan kondisi berhenti saat bernilai salah.

Langkah 2. Untuk setiap vektor *input*  $x$  dari data *training*, lakukan langkah 3-4.

Langkah 3. Tentukan  $J$  sedemikian hingga  $\|x_i - w_j\|$  minimum dengan menggunakan perhitungan rumus jarak *euclidean*.

$$D(j) = \sqrt{\sum_{i=1}^n (x_i - w_{ij})^2}$$

Langkah 4. Perbaiki  $w_j$  dengan ketentuan :

(1) Jika  $T = C_j$ , maka digunakan rumus :

$$w_j(\text{baru}) = w_j(\text{lama}) + \alpha(x - w_j(\text{lama}))$$

(2) Jika  $T \neq C_j$ , maka digunakan rumus :

$$w_j(\text{baru}) = w_j(\text{lama}) - \alpha(x - w_j(\text{lama}))$$

Langkah 5. Kurangi nilai *learning rate*  $\alpha$  dengan :

$$\alpha(\text{baru}) = \alpha(\text{lama}) \times dec\ \alpha$$

Langkah 6. Tes kondisi berhenti yaitu saat iterasi mencapai *max epoch* (pada langkah 1) atau *learning rate*  $\alpha$  mencapai  $\min \alpha$ , dengan *output* berupa bobot optimal.

## PEMBAHASAN

Pengujian pada data *testing* dilakukan setelah proses klasifikasi pada data *training* menggunakan algoritme LVQ. Terdapat 10-Fold pada pembagian data *testing* dan data *training*, sehingga terdapat 10 percobaan klasifikasi dengan pembagian data yang berbeda dimana data dengan *Fold* ke-*K* digunakan sebagai data *testing*. Dari pengujian pada data *testing*, diperoleh akurasi dari tiap percobaan. Berikut ini ditunjukkan akurasi tiap *Fold* untuk klasifikasi LVQ pada data IPM pada Tabel 2.

**Tabel 2 Akurasi Hasil Klasifikasi tiap *Fold***

Fold	Akurasi	Fold	Akurasi
1	81.48148	6	92.59259
2	83.33333	7	83.33333
3	83.33333	8	92.59259
4	92.59259	9	92.59259
5	88.88889	10	90.74074

Berdasarkan Tabel 2, diketahui terdapat 10 Fold dengan akurasi masing-masing. Diperoleh akurasi maksimum yaitu sebesar 92.59 % untuk klasifikasi data IPM menggunakan algoritme LVQ dengan menerapkan metode *K-Fold Cross Validation*.

## SIMPULAN

Berdasarkan hasil dan pembahasan diperoleh simpulan bahwa algoritme LVQ menggunakan metode *K-Fold Cross Validation* untuk memperoleh hasil akurasi maksimum dari klasifikasi pada data IPM yaitu sebesar 92.59 %. Diketahui bahwa pembagian data *testing* dan data *training* mempengaruhi hasil akurasi dari klasifikasi IPM di kabupaten/kota di Indonesia dengan menerapkan algoritme LVQ.

## UCAPAN TERIMA KASIH

Pada kesempatan ini, diucapkan terima kasih kepada Universitas Sebelas Maret atas pendanaan dan dukungan untuk penelitian ini dibawah Penelitian HGR-UNS A 2023.

## DAFTAR RUJUKAN

- Badan Pusat Statistika. (2022). Indeks Pembangunan Manusia:Konsep. Retrieved from <https://www.bps.go.id/subject/26/indeks-pembangunan-manusia.html>.
- Dukcapil Kemendagri. (2022). 273 Juta Penduduk Indonesia Terupdate Versi Kemendagri. Retrieved from <https://dukcapil.kemendagri.go.id/berita/baca/1032/273-juta-penduduk-indonesiaterupdate-versi-kemendagri>.
- Fausett, L. (1994). *Fundamentals of neural networks: architectures, algorithms, and applications*. New Jersey: Prentice-Hall, Inc.
- Hanifa, T. T., S. Al-faraby. (2017) Analisis Churn Prediction pada Data Pelanggan PT. Telekomunikasi dengan Logistic Regression dan Underbagging. *F. Informatika and U. Telkom*, 4(2), 3210–3225.
- Hamidi, Rifwan, M. Tanzil Furqon, Bayu Rahayudi. (2017). Implementasi Learning Vector Quantization (LVQ) untuk Klasifikasi Kualitas Air Sungai. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(12), 1758-1763.
- Kusumadewi. (2006). *Neuro Fuzzy-Integrasi Sistem Fuzzy dan Jaringan Saraf*. *Journal of Multivariate*. Yogyakarta: Graha Ilmu.
- Kurniawan, T. (2017). Implementasi Text Mining pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan Naïve Bayes Classifier dan Support Vector Machine. *Institut Teknologi Sepuluh Nopember*.

- Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4. 5. *Jurnal Edik Informatika Penelitian Bidang Komputer Sains dan Pendidikan Informatika*, 2(2), 213-219.
- Nikam, Sagar S. (2015). A Comparative Study of Classification Techniques in Data Mining Algorithms. *Oriental Journal Of Computer Science Technology*, 8(1), 13-19.
- Raval, Kalyani M. (2012). Data Mining Techniques. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(10), 439-442.
- Wibowo, Agus W.S., Yoyon K. (2014). Verifikasi Dan Identifikasi Tandatangan Offline Menggunakan Wavelet Dan Learning Vector Qantization. *Prosiding Seminar Nasional Sains dan Pendidikan Sains IX, Fakultas Sains dan Matematika, UKSW Salatiga*, 5(3), 649-655.