

PENGARUH PERBAIKAN KUALITAS CITRA DAN POSISI AWAL CENTROID PADA ITERASI K-MEANS KLASTER

Budi Hartono¹, Veronica Lusiana²

^{1,2}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Stikubank Semarang
e-mail: ¹budihartono@edu.unisbank.ac.id, ²vero@edu.unisbank.ac.id

ABSTRAK

Penelitian ini melakukan analisa terhadap hasil perbaikan kualitas citra buah alpukat mentah dan matang menggunakan modifikasi intensitas warna dengan metode peregangan kontras dan ekualisasi histogram. Proses ekstraksi menggunakan matriks GLCM dengan fitur energi dan kontras. Proses komputasi dan visualisasi menggunakan perangkat lunak Octave. Pengamatan dilakukan pada pengaruh hasil perbaikan kualitas citra dan pemilihan posisi awal centroid pada jumlah iterasi k-means klaster. Secara umum, hasil perbaikan kualitas citra peregangan kontras membutuhkan jumlah iterasi yang lebih sedikit dibandingkan ekualisasi histogram. Melalui pemilihan nilai awal centroid yang berbeda dapat menghasilkan nilai akhir centroid yang sama, namun dengan jumlah iterasi yang berbeda.

Kata Kunci: *Peregangan kontras, Ekualisasi histogram, GLCM, k-means klaster*

1. PENDAHULUAN

Citra digital dengan kontras yang baik yaitu citra yang memiliki distribusi intensitas lebar. Pada histogram tidak terlihat puncak yang terlalu dominan. Sebaliknya, citra dengan kontras rendah yaitu citra yang memiliki distribusi intensitas sempit. Kontras pada sebuah citra adalah berhubungan dengan distribusi intensitas piksel, yaitu proses untuk memperluas jangkauan intensitas [1]. Modifikasi intensitas warna pada sebuah citra berhubungan dengan distribusi intensitas piksel. Memperbaiki kualitas citra yang sering dilakukan dengan memodifikasi intensitas warna yaitu dengan cara peregangan kontras (*contrast stretching*) dan ekualisasi histogram (*histogram equalization* atau perataan histogram) [2].

Setelah kualitas citra menjadi lebih baik, proses berikutnya adalah mengekstraksi fitur. Ekstraksi fitur citra yaitu untuk memperoleh ciri citra melalui isi yang dimiliki oleh citra, disini menggunakan matriks GLCM (*Gray level co-occurrence matrix*). Tiga jenis fitur citra yang umum digunakan pada proses ekstraksi yaitu fitur: warna, bentuk obyek, dan tekstur [3]. Melalui pengolahan data fitur bisa digunakan untuk mengenali obyek yang ada pada citra. Jarak Euclidean (*Euclidean distance*) adalah jarak garis lurus antara dua titik dalam ruang. Jarak Euclidean dapat diimplementasikan untuk menghitung jarak antara data fitur dengan centroid pada K-means clustering.

Penelitian ini akan melakukan analisa terhadap hasil perbaikan kualitas citra buah Alpukat menggunakan modifikasi intensitas warna. Proses modifikasi intensitas warna menggunakan metode peregangan kontras dan ekualisasi histogram. Citra setelah melalui kedua proses ini menjadi dalam kondisi yang lebih baik, sehingga identifikasi buah alpukat yang matang dan alpukat yang mentah menjadi lebih baik. Proses komputasi dan visualisasi menggunakan perangkat lunak Octave. Bahan penelitian citra buah alpukat yang diperoleh dari datasets [4, 5]. Pada Gambar 1 dan Gambar 2 adalah data citra percobaan, yaitu citra buah Alpukat mentah dan Alpukat matang. Pengamatan dilakukan pada pengaruh hasil perbaikan kualitas citra dan pemilihan posisi awal centroid pada jumlah iterasi k-means klaster.

2. TINJAUAN PUSTAKA

2.1. Gray level co-occurrence matrix (GLCM)

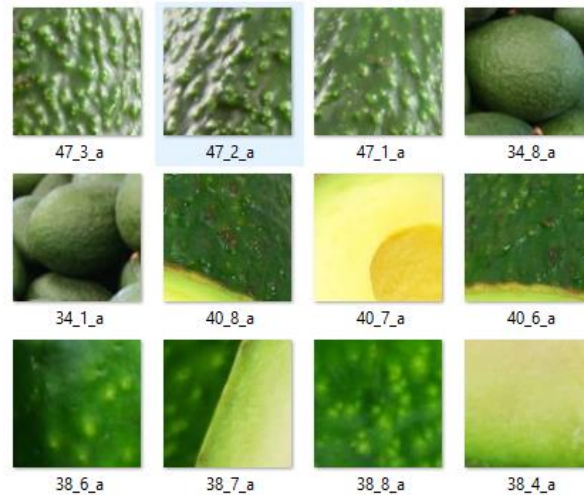
Matriks co-occurrence dapat digunakan untuk menghitung beberapa fitur tekstur antara lain fitur energi dan fitur kontras. Fitur energi digunakan untuk mengukur konsentrasi pasangan intensitas. Nilai energi membesar apabila pasangan piksel terkonsentrasi pada beberapa koordinat, dan sebaliknya akan mengecil jika letaknya menyebar. Fitur kontras digunakan untuk mengukur kekuatan perbedaan intensitas dalam citra. Nilai kontras membesar jika variasi intensitas dalam citra tinggi, dan sebaliknya akan mengecil jika variasinya rendah [6, 7].

Matriks intensitas co-occurrence adalah suatu matriks yang menggambarkan frekuensi munculnya pasangan dua piksel dengan intensitas tertentu dalam jarak dan arah tertentu di dalam citra. Menyusun matriks intensitas *co-occurrence* $P(i_1, i_2)$, didefinisikan dengan dua langkah:

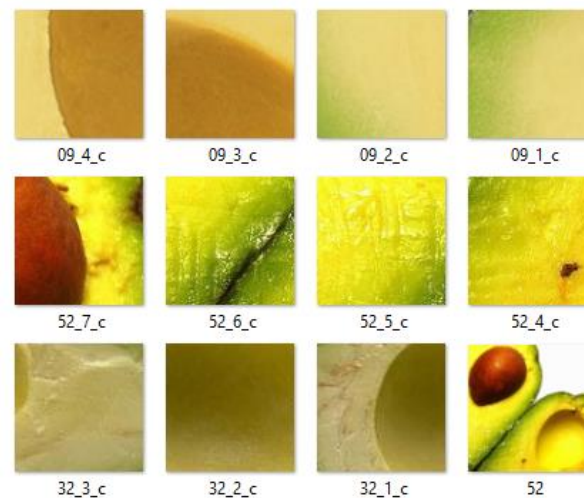
1. Menentukan jarak antara dua titik dalam arah vertikal dan horisontal, vektor $d = (dx, dy)$, dimana dx dan dy posisi piksel citra.

2. Menghitung pasangan piksel yang mempunyai nilai intensitas i_1 dan i_2 dan berjarak d piksel. Hasil perhitungan setiap pasangan nilai intensitas diletakkan pada matriks sesuai dengan koordinatnya, dimana absis untuk mencatat nilai intensitas i_1 dan ordinat untuk nilai intensitas i_2 .

Matriks GLCM berisi data *co-occurrence* tingkat keabuan piksel untuk ketetanggaan pada arah 0° , 45° , 90° , dan 135° . Arah ini diperoleh dari intensitas piksel sebagai pusat dengan piksel tetangga terdekat yang berjarak 1.



Gambar 1. Citra buah Alpukat mentah.



Gambar 2. Citra buah Alpukat matang.

2.2. Euclidean Distance

Metode pengukuran kemiripan citra dapat dilakukan dengan cara menentukan tingkat ketidaksamaan atau kesamaan (*similarity degree*) dari dua buah vektor ciri yang dimiliki oleh citra. Vektor ciri citra diperoleh melalui proses ekstraksi fitur citra. Tingkat kesamaan adalah suatu nilai yang menunjukkan apakah kedua vektor tersebut mirip atau tidak. Metode ini dapat diterapkan secara umum untuk berbagai jenis data, tidak sebatas pada data citra [2]. Hasil penghitungan jarak Euclidean, semakin kecil skor $d(p,q)$ maka semakin mirip kedua vektor fitur yang dicocokkan. Metode ini bisa digunakan untuk mendeteksi kemiripan citra dengan mengisi nilai vektor (p, q) dengan nilai fitur citra yang akan dideteksi tingkat kemiripannya. Pada Tabel 1 mencatat beberapa penelitian sejenis yang berkaitan dengan perbaikan kualitas citra, ekstraksi fitur, dan matriks GLCM.

Tabel 1. Penelitian sejenis.

No.	Peneliti	Tujuan	Obyek	Metode
1	Widodo, dkk. [8]	Perbaikan kualitas citra	Radiologi gigi	Peregangan kontras
2	Syahputra, dkk.[9]	Pengenalan	Daun, tanaman obat Indonesia	GLCM, Moment invariant, dan Morfologi bentuk
3	Sutarno dan Fauliah [10]	Klasifikasi	Buah Durian	GLCM dan LVQ
4	Andrian, dkk. [11]	Identifikasi	Kupu-kupu	GLCM dan KNN

3. METODE PENELITIAN

3.1 Kuantisasi

Setelah diperoleh citra dalam kondisi kontras yang lebih baik, langkah berikutnya adalah melakukan kuantisasi. Kuantisasi bertujuan untuk mereduksi jumlah variasi nilai piksel yang muncul pada citra. Terdapat dua macam kuantisasi, yaitu kuantisasi skalar (*scalar quantizers*) dan kuantisasi vektor (*vector quantizers*). Kuantisasi skalar dalam mengelompokkan nilai masukan menggunakan parameter tunggal, contoh melalui nilai abu-abu. Sedangkan kuantisasi vektor untuk mengelompokkan nilai masukan menggunakan dua nilai atau lebih. Kuantisasi vektor membutuhkan jumlah komputasi lebih banyak [12]. Pada penelitian ini, menggunakan kuantisasi skalar seragam dengan keluaran yaitu 3 bit, yang diperoleh dari masukan nilai abu-abu 8 bit.

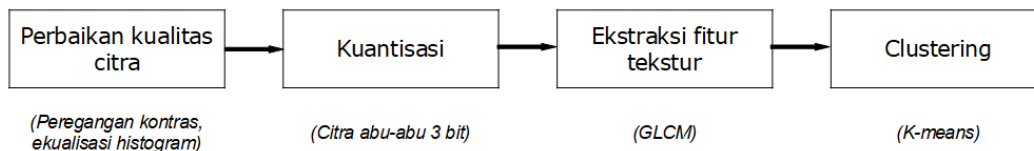
3.2 K-means cluster

K-means merupakan salah satu algoritma *clustering*. Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok, dengan masukan berupa data tanpa label kelas. Pada algoritma ini, komputer mengelompokkan data yang menjadi masukannya tanpa mengetahui target kelasnya. Pembelajaran ini termasuk dalam *unsupervised learning*. Masukan yang diterima adalah data atau objek dan k-buah kelompok (*cluster*) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam k-buah kelompok, dimana pada setiap kelompok memiliki titik pusat (*centroid*).

Terdapat dua jenis *data clustering* yang sering dipergunakan dalam proses pengelompokan data yaitu *Hierarchical* dan *non-Hierarchical*. K-Means merupakan salah satu metode data *clustering non-hierarchical* atau *Partitional Clustering* [13, 14]. Metode K-Means Clustering berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan sebaliknya mempunyai karakteristik berbeda dengan data yang ada di dalam kelompok lain.

3.3 Proses penelitian

Proses penelitian diawali dengan perbaikan kualitas citra menggunakan metode peregangan kontras dan ekualisasi histogram. Citra percobaan adalah 32 citra alpukat mentah dan 64 citra alpukat matang, masing-masing berdimensi 100x100 piksel. Data hasil perbaikan kualitas citra akan dilakukan proses kuantisasi menjadi citra abu-abu 3 bit. Proses selanjutnya adalah ekstraksi fitur tekstur menggunakan matriks GLCM dengan fitur energi dan kontras. Pengelompokkan citra alpukat mentah dan matang menggunakan algoritma *k-means cluster*. Pada Gambar 3 dapat dilihat blok diagram tahap penelitian. Pengamatan dilakukan pada pengaruh hasil perbaikan kualitas citra dan pemilihan posisi awal centroid terhadap jumlah iterasi yang diperlukan.



Gambar 3. Tahap penelitian.

4. HASIL DAN PEMBAHASAN

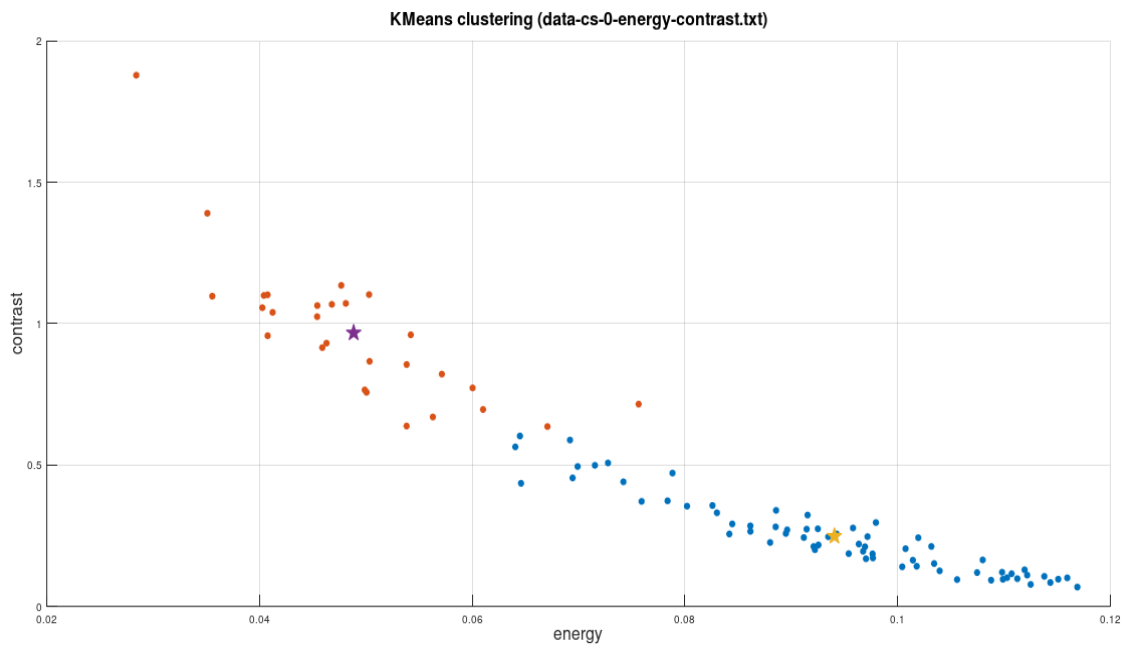
Pengamatan dilakukan pada pengaruh hasil perbaikan kualitas citra dan pemilihan posisi awal centroid terhadap jumlah iterasi yang diperlukan untuk mengelompokkan data citra Alpukat mentah dan matang menggunakan k-means klaster. Fitur yang digunakan adalah fitur energi dan fitur kontras. Pada Tabel 2 mencatat nilai awal centroid C1 dan C2 dengan jumlah iterasi. Terdapat dua macam nilai awal centroid yang dipilih yaitu:

- a. nilai awal centroid I C1(0.1, 0.6) dan C2(0.2, 0.9)
- b. nilai awal centroid II C1(0.1, 0.9) dan C2(0.3, 0.5)

Pada metode perbaikan kualitas citra peregangan kontras (pada baris 1), nilai akhir centroid adalah C1(0.094 , 0.248) dan C2(0.049 , 0.968). Nilai akhir C1 dan C2 ini bertukar tempat dengan nilai akhir C1 dan C2 pada baris 3. Pada metode perbaikan kualitas citra ekualisasi histogram (pada baris 2), nilai akhir centroid adalah C1(0.099 , 0.246) dan C2(0.052 , 0.931). Nilai akhir C1 dan C2 ini bertukar tempat dengan nilai akhir C1 dan C2 pada baris 4. Melalui percobaan ini, dengan nilai awal centroid yang berbeda akan menghasilkan nilai akhir centroid yang sama.

Tabel 2. Nilai awal centroid C1 dan C2 dengan jumlah iterasi.

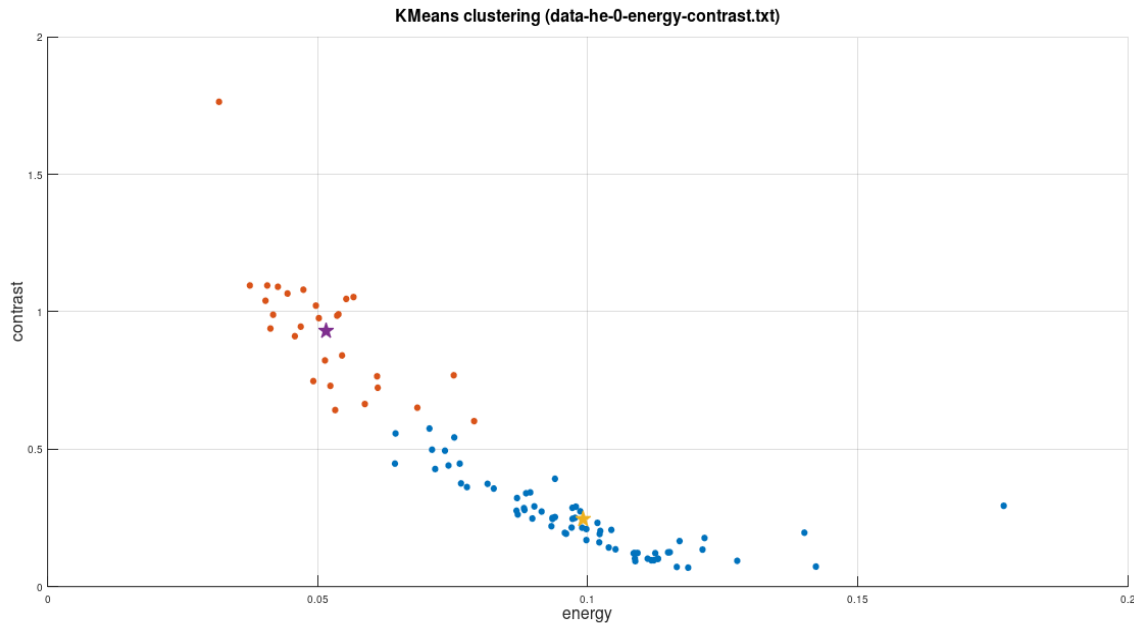
No.	Metode perbaikan kualitas citra	Nilai C1 dan C2 awal	Nilai C1 dan C2 akhir	Jumlah iterasi
1	Peregangan kontras	C1(0.1 , 0.6) C2(0.2 , 0.9)	C1(0.094 , 0.248) C2(0.049 , 0.968)	4
2	Ekualisasi histogram	C1(0.1 , 0.6) C2(0.2 , 0.9)	C1(0.099 , 0.246) C2(0.052 , 0.931)	5
3	Peregangan kontras	C1(0.1 , 0.9) C2(0.3 , 0.5)	C1(0.049 , 0.968) C2(0.094 , 0.248)	2
4	Ekualisasi histogram	C1(0.1 , 0.9) C2(0.3 , 0.5)	C1(0.052 , 0.931) C2(0.099 , 0.246)	3



Gambar 4. Perbaikan citra peregangan kontras, nilai awal C1(0.1, 0.6) C2(0.2, 0.9).

Pada nilai awal centroid I, dengan metode perbaikan kualitas citra peregangan kontras memerlukan jumlah iterasi lebih sedikit yaitu 4 kali, dibandingkan dengan metode perbaikan kualitas citra ekualisasi histogram yaitu 5. Pada nilai awal centroid II, dengan metode perbaikan kualitas citra peregangan kontras memerlukan jumlah iterasi lebih sedikit yaitu 2 kali, dibandingkan dengan metode perbaikan kualitas citra ekualisasi histogram yaitu 3. Secara umum, hasil perbaikan kualitas citra peregangan kontras membutuhkan jumlah iterasi yang lebih sedikit dibandingkan ekualisasi histogram.

Pada Gambar 4 adalah sebaran data citra Alpukat mentah dan Alpukat matang dengan pusat klaster pada posisi akhir centroid C1(0.094 , 0.248) di kanan bawah dan C2 (0.049 , 0.968) di kiri atas. Pada Gambar 5 adalah sebaran data citra Alpukat mentah dan Alpukat matang dengan pusat klaster pada posisi akhir centroid C1(0.099 , 0.246) di kanan bawah dan C2 (0.052 , 0.931) di kiri atas.



Gambar 5. Perbaikan citra ekualisasi histogram, nilai awal C1(0.1, 0.6) C2(0.2, 0.9).

5. KESIMPULAN

Penelitian ini mengamati pengaruh hasil perbaikan kualitas citra dan pemilihan posisi awal centroid terhadap jumlah iterasi yang diperlukan untuk mengelompokkan data citra Alpukat mentah dan matang menggunakan k-means klaster. Pada nilai awal centroid I, dengan metode perbaikan kualitas citra peregangan kontras memerlukan jumlah iterasi lebih sedikit yaitu 4 kali, dibandingkan dengan metode perbaikan kualitas citra ekualisasi histogram yaitu 5. Melalui nilai awal centroid II, dengan metode perbaikan kualitas citra peregangan kontras memerlukan jumlah iterasi lebih sedikit yaitu 2 kali, dibandingkan dengan metode perbaikan kualitas citra ekualisasi histogram yaitu 3. Secara umum, pada hasil perbaikan kualitas citra peregangan kontras membutuhkan jumlah iterasi yang lebih sedikit dibandingkan ekualisasi histogram. Melalui percobaan diperoleh dengan nilai awal centroid yang berbeda dapat menghasilkan nilai akhir centroid yang sama, namun dengan jumlah iterasi yang berbeda.

UCAPAN TERIMA KASIH

Terima kasih kepada Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Universitas Stikubank Semarang untuk dukungan dan kesempatan kepada kami mengerjakan Penelitian ini.

DAFTAR PUSTAKA

- [1] Gonzalez, R.C., dan Woods, R.E., 2018, *Digital Image Processing, Global Edition*, Ed.4, Pearson Education.
- [2] Putra, D., 2010, *Pengolahan Citra Digital*, Ed.1, Andi, Yogyakarta.
- [3] Nixon, M.S., dan Aguado, A.S., 2020, *Feature Extraction and Image Processing for Computer Vision*, Fourth Edition, Elsevier, Academic Press.
- [4] Oltean, M., 2018, *Fruits 360 dataset*, Mendeley Data, V1.
- [5] Škrjanec, M., 2013, Automatic fruit recognition using computer vision, *Bsc. Thesis*, (Mentor: Matej Kristan), Fakulteta za računalništvo in informatiko, Univerza v Ljubljani.
- [6] Kadir, A., dan Susanto, A., 2013, *Teori dan Aplikasi Pengolahan Citra*, Penerbit Andi, Yogyakarta.
- [7] Ahmad, U., 2005, *Pengolahan Citra Digital dan Teknik Pemrogramannya*, Ed.1, Graha Ilmu, Yogyakarta.
- [8] Widodo, H.B., Soelaiman, A., Ramadhani, Y., dan Supriyanti, R., 2016, Calculating Contrast Stretching Variables in Order to Improve Dental Radiology Image Quality, *IOP Conf. Series: Materials Science and Engineering* 105 012002, pp.1-8.

- [9] Syahputra, H., Indra, Z., Febrian, D., dan Adriani, D.P., 2019, Leaf feature extraction using glcm, moment invariant and shape morphology for indonesian medicinal plants recognition, *IOP Conf. Series: Journal of Physics: Conf. Series* 1317 012008.
- [10] Sutarno, dan Fauliah, S.P., 2019, Implementation of Learning Vector Quantization (LVQ) Algorithm for Durian Fruit Classification Using Gray Level Co-occurrence Matrix (GLCM) Parameters, *IOP Conf. Series: Journal of Physics: Conf. Series* 1196 012040.
- [11] Andrian, R., Maharani, D., Muhammad, M.A., dan Junaidi, A., 2020, Butterfly identification using gray level co-occurrence matrix (glcm) extraction feature and k-nearest neighbor (knn) classification, *Jurnal Register*, Vol. 6, No. 1, pp. 11-21.
- [12] Symes, P.D., 2000, *Video Compression Demystified*, McGraw-Hill Professional Publishing, New York, NY.
- [13] Han, J., Kamber, M., dan Pei, J., 2011, *Data Mining: Concepts and Techniques*, 3rd Edition, Morgan Kaufmann.
- [14] Ian H.W., Eibe F., Mark A.H., dan Christopher J.P., 2017, *Data Mining, Practical Machine Learning Tools and Techniques*, Fourth Ed., Morgan Kaufmann, Elsevier.