

Penerapan *Principal Component Analysis* (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro

Dyah Hedyati¹, I Made Suartana²

^{1,2} Teknik Informatika, Universitas Negeri Surabaya
dyahhedyati16051204013@mhs.unesa.ac.id

Abstrak—Dalam penelitian, data harus melalui proses pengolahan agar dapat digunakan dalam penelitian tersebut. Data yang digunakan haruslah valid untuk dapat menghasilkan solusi yang tepat guna. Pengolahan data dalam jumlah besar secara manual berpeluang menghasilkan banyak kesalahan. Untuk itu diperlukan pendekatan teknologi untuk dapat meminimalisir kesalahan yang dapat terjadi. Data mining merupakan suatu proses pengekstrakan informasi dari kumpulan data yang besar. Proses ini bertujuan untuk mendapatkan intisari dari kumpulan data tersebut. Proses data mining dapat menghasilkan informasi penting berupa klasifikasi (*classification*), pengelompokan (*clustering*), bahkan prediksi (*prediction*). Clustering merupakan suatu proses analisis data untuk membentuk sekelompok objek berdasarkan sifat dan cirinya sehingga terbentuk suatu kelompok yang bersifat homogen antar anggota pada kelompok yang sama. Namun, beberapa algoritma clustering menemui masalah ketika dihadapkan pada data dengan dimensi tinggi, termasuk juga K-Means. Reduksi dimensi dapat dijadikan sebagai salah satu langkah optimasi algoritma clustering. Proses reduksi dimensi yang umumnya diterapkan pada tahap pre-processing data bertujuan untuk mengurangi jumlah fitur (dimensi) tanpa menghilangkan informasi penting dari suatu data. Metode PCA akan membentuk sekumpulan dimensi baru yang kemudian di ranking berdasarkan varian datanya, sehingga tercipta kumpulan data dengan fitur yang lebih sederhana. Penelitian ini akan menguji kinerja PCA sebagai salah satu metode optimasi algoritma clustering K-Means yang diterapkan pada data pertanian Kab. Bojonegoro pada tahun 2017 hingga 2020. Dataset hasil clustering yang didapatkan dari situs BPS akan dibandingkan dengan dataset dari sumber yang sama namun telah mengalami proses reduksi dimensi menjadi 1 PC, 2 PC, dan 3 PC. Evaluasi data hasil clustering menggunakan nilai DB Index menunjukkan nilai paling optimal pada dataset yang direduksi menjadi 1 PC dan dibentuk menjadi 3 klaster, yaitu 0.4072. Sedangkan dengan jumlah klaster yang sama, dataset dengan 2PC menghasilkan nilai DB Index 0.6168, dataset dengan 3 PC menghasilkan nilai 0.6598, dan dataset tanpa proses reduksi dimensi menghasilkan nilai DB Index 0.4598.

Kata Kunci— Data Mining, Clustering, Reduksi Dimensi, PCA

I. PENDAHULUAN

Memiliki hamparan tanah subur yang luas, sebagian besar rakyat Indonesia memilih bercocok tanam sebagai pencahariannya, hingga Indonesia dikenal sebagai Negara agraris[1]. Pertanian memegang peran penting bagi perekonomian negara, karenanya banyak studi yang dilakukan untuk menyokong kemajuan pertanian di Indonesia. Studi yang dilakukan membutuhkan data valid sebagai bahan kajian agar penelitian tersebut dapat menghasilkan sebuah solusi yang tepat guna.

Kehadiran BPS (Badan Pusat Statistik) sebagai Lembaga Pemerintahan non Kementrian yang menyediakan kebutuhan data bagi pemerintah maupun masyarakat menjadi salah satu keuntungan bagi akademisi dalam menjalankan penelitiannya. Data-data yang disediakan oleh BPS telah melalui proses pengolahan data menggunakan teknologi komputer, sehingga data ditampilkan berupa statistik yang mudah dipahami oleh masyarakat dan dapat diakses secara mudah oleh siapa saja melalui *website* BPS[2].

Tidak jarang dalam suatu penelitian dibutuhkan data yang sangat banyak dan beragam. Pemrosesan data secara manual memungkinkan terjadi banyak kesalahan. Untuk itu, penggunaan teknologi untuk mengotomatisasi proses pengolahan data yang besar sangat diperlukan agar informasi yang didapatkan sesuai dengan data yang disajikan. Teknik *data mining* bertujuan untuk mendapatkan intisari dari sebuah data yang besar. *Data mining* menjadi topik menarik bagi peneliti karena kegunaannya yang beragam. Teknik data mining digunakan untuk mengekstrak informasi dari *dataset* hingga dapat menghasilkan informasi penting berupa klasifikasi (*classification*), pengelompokan (*clustering*), bahkan prediksi (*prediction*). Penerapan data mining telah banyak dilakukan untuk mengolah data pertanian, seperti deteksi dan klasifikasi serta prediksi penyakit tanaman, prediksi hasil, manajemen input (perencanaan irigasi dan pestisida), saran pupuk, prediksi kelembaban tanah secara waktu langsung, dll[3].

Salah satu teknik *data mining* adalah *clustering*. *Clustering* merupakan suatu proses analisis data untuk membentuk sekelompok objek berdasarkan sifat dan cirinya sehingga terbentuk suatu kelompok yang bersifat homogen antar anggota pada kelompok yang sama[4]. Pendekatan *K-Means* banyak digunakan sebagai metode *clustering*, karena algoritma *K-Means* merupakan algoritma yang sederhana dan mudah diimplementasikan. algoritma ini dapat bekerja pada jumlah data yang besar dengan waktu komputasi relatif cepat dan efisien [5]. Penggunaan algoritma *K-Means* untuk

mengelompokkan data pertanian telah dilakukan oleh Putrama Alkhairi dan Agus Perdana Windarto. Pada [6], menganalisa kinerja *K-Means* dalam mengelompokkan daerah potensi karet produktif di Provinsi Sumatera Utara menjadi 3 *cluster*. Kinerja algoritma dilihat dari hasil *cluster* dengan menghitung nilai kemurnian (*purity measure*) masing – masing *cluster*/kelompok yang di hasilkan. Hasilnya, jumlah *cluster* Kelompok tertinggi 1 item, *cluster* Kelompok menengah 6 item dan, *cluster* Terendah 19 item dengan total jumlah data sebanyak 26 data.

Namun, beberapa algoritma *clustering* menemui masalah ketika dihadapkan pada data dengan dimensi tinggi, termasuk juga *K-Means*. Masalah yang mungkin muncul adalah menurunnya akurasi klasifikasi, kualitas *cluster* yang buruk, hingga waktu komputasi yang lama. Untuk itu diperlukan proses optimasi agar kinerja algoritma *K-Means* tetap stabil ketika dihadapkan pada data dengan fitur yang tinggi. Reduksi dimensi merupakan salah satu proses yang dapat mengoptimalkan kinerja algoritma *K-Means*. Reduksi dimensi adalah proses pengurangan jumlah fitur (dimensi) tanpa menghilangkan informasi penting dari suatu data. Reduksi dimensi umum diterapkan sebagai metode *pre-processing* untuk menciptakan model klasifikasi yang lebih baik[7][8].

Keuntungan lain dari diterapkannya reduksi dimensi terkait dengan penggunaan memori penyimpanan data yang lebih sedikit, mengurangi waktu komputasi, kualitas data meningkat karena reduksi dan data yang tidak relevan dapat dihilangkan. Selain itu, penerapan reduksi dimensi membuat kinerja beberapa algoritma klasifikasi menjadi lebih baik dan efisien karena tidak perlu memproses terlalu banyak dimensi[8].

Melalui eksperimen perbandingan kinerja dua metode reduksi dimensi yang paling banyak digunakan, menghasilkan PCA yang unggul dari LDA di semua pengukuran. Pengujian lebih lanjut dilakukan pada *dataset Diabetic Retinopathy* dan *Intrusion Detection System*. Hasil akhir dari percobaan menunjukkan bahwa kinerja algoritma *Machine Learning* dengan PCA pada tahap *pre-processing* akan lebih baik ketika mengolah kumpulan data dengan dimensi yang tinggi[9].

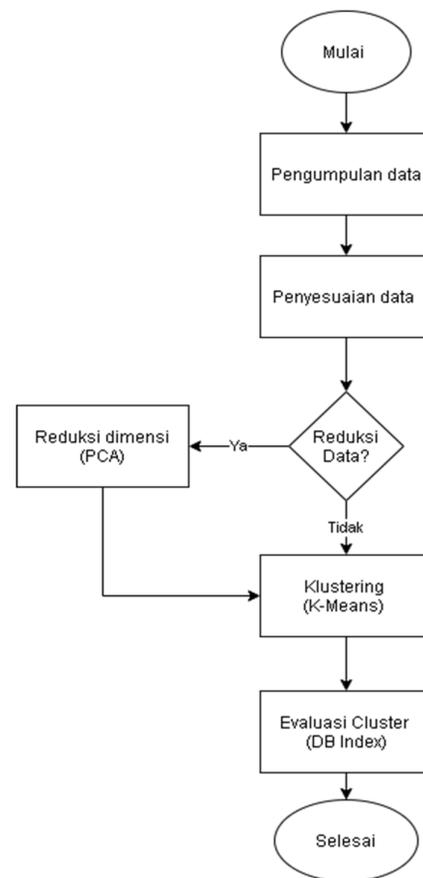
Penelitian [10] mengusulkan metode baru untuk memilih subset dari fungsi optimal untuk meningkatkan akurasi penggunaan klasifikasi patologis dari CTG. Penerapan PCA pada tahap *pre-processing* telah terbukti efektif dalam meningkatkan waktu komputasi dan akurasi. Penelitian ini membantu staf medis dalam membuat keputusan medis yang lebih efisien dan cepat dengan melakukan penafsiran pembacaan hasil CTG.

Pada [5], Metode PCA digunakan untuk mereduksi dimensi *dataset* kinerja dosen Program Studi Teknik Elektro di Universitas Panca Marga Probolinggo sebelum dikelompokkan menggunakan *K-Means*. Hasilnya, reduksi dimensi *dataset* menggunakan metode PCA terbukti dapat meningkatkan kualitas *cluster*. Terdapat peningkatan kualitas model pada penerapan *PCA-K-Means* dibandingkan dengan *K-Means* tradisional. Peningkatan kualitas model terlihat dari meningkatnya akurasi *clustering* dan hasil pengukuran validitas *cluster* dengan *Davies-Bouldin Index* menunjukkan *PCA-K-Means* menghasilkan nilai *DB Index* paling kecil dibandingkan dengan *K-Means* tradisional.

Penelitian ini berfokus pada penerapan PCA untuk mereduksi dimensi data produksi pertanian di Kabupaten Bojonegoro sebelum dikelompokkan menggunakan *K-Means* kedalam 3 *cluster*, baik, buruk, sedang. Data produksi yang digunakan pada penelitian ini adalah data produksi pertanian tahun 2017 hingga 2020 yang disediakan oleh BPS. Hasil penerapan PCA pada proses pengelompokan menggunakan *K-Means* akan dibandingkan dengan hasil *clustering* menggunakan *K-Means* tradisional berdasarkan kualitas *cluster* yang diukur menggunakan *Davies-Bouldin Index*.

II. METODOLOGI PENELITIAN

Pada penelitian ini, data yang digunakan merupakan data produksi pertanian dari tahun 2017 hingga 2020. Data yang didapat kemudian diolah menggunakan PCA sebelum diklasifikasikan menggunakan algoritma *K-Means*. Model yang dihasilkan oleh data latih kemudian digunakan untuk mengklasifikasikan wilayah produksi pertanian kedalam *cluster* baik, buruk, maupun. Rancangan penelitian dapat dilihat pada Gambar 1



Gambar 1 Rancangan Penelitian

A. Pengumpulan data

Data yang diperlukan adalah kumpulan data hasil pertanian tahunan tiap-tiap kecamatan di Kabupaten Bojonegoro yang dikumpulkan dari situs BPS Kabupaten Bojonegoro. Berkas yang dikumpulkan berbentuk file PDF dan berisi beragam data dari berbagai kategori yang disediakan oleh BPS. Sehingga untuk memenuhi kebutuhan penelitian, perlu dilakukan pengumpulan data khusus kategori pertanian dari tiap-tiap kecamatan se-Kabupaten Bojonegoro.

B. Penyesuaian Data

Data yang akan digunakan akan disesuaikan untuk memudahkan dalam penggunaannya. Penyesuaian meliputi penyesuaian format file agar bisa diproses dan pemberian atribut agar data dapat dikenali. Data yang diperlukan pada penelitian ini antara lain: nama desa, luas tanam, produksi, produktivitas dan tahun. Data disimpan dalam format *comma separated value*(csv).

C. Reduksi Dimensi

Dimensi dan kompleksitas data yang tinggi dapat mempengaruhi hasil klasifikasi. Untuk itu, perlu adanya proses reduksi dimensi dan kompleksitas dari data yang digunakan untuk mengurangi kesalahan pada proses klasifikasi. Reduksi dimensi ini dapat dilakukan menggunakan algoritma *Principal Component Analysis (PCA)*[11]. PCA akan membentuk sekumpulan dimensi baru yang kemudian di ranking berdasarkan varian datanya. PCA akan menghasilkan *Principal component* yang didapat dari dekomposisi *eigen value* dan *eigen vector* dari matriks kovariansi[12]. Langkah dari algoritma PCA adalah sebagai berikut[11]:

- 1) Menghitung *mean* (\bar{X}) dari data pada tiap dimensi menggunakan persamaan (1):

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

Dengan:

n = jumlah data sampel
 X_i = data sampel

- 2) Menghitung *covariance matrix* (C_X) menggunakan persamaan (2):

$$C_X = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T \quad (2)$$

Dengan:

n = jumlah data sampel
 X_i = data sampel
 \bar{X} = *mean*

- 3) Menghitung *eigenvector* (v_m) dan *eigenvalue* (λ_m) dari *covariance matrix* menggunakan persamaan (3):

$$C_X v_m = \lambda_m v_m \quad (3)$$

- 4) Urutkan *eigenvalue* secara *descending*. *Principal Component* (PC) adalah deretan *eigenvector* sesuai dengan urutan *eigenvalue* pada tahap 3.
- 5) Menghasilkan *dataset* baru.

D. K-Means

Clustering merupakan suatu teknik untuk membagi objek kedalam kelompok-kelompok. *Cluster* adalah kumpulan dari objek yang memiliki kemiripan sifat yang sama dan berbeda dengan objek anggota *cluster* yang lain. Pada *unsupervised learning* dikenal satu algoritma paling sederhana, yaitu *K-Means*. Algoritma K-Means ini bekerja dengan membagi n himpunan data menjadi k cluster sehingga menghasilkan kesamaan *intracluster* yang tinggi namun kesamaan *intercluster* tetap rendah[13]. Tahapan pada algoritma *K-Means*[5] adalah sebagai berikut:

- 1) Membuat partisi sebanyak k cluster yang berisi sampel acak kemudian perhitungkan *centroid cluster* dengan cara memilih salah satu k -point secara acak sebagai *center*.
- 2) Hitung jarak antar objek dengan *center* menggunakan persamaan *Euclidean Distance*. *Euclidean distance* adalah perhitungan untuk menentukan jarak dari 2 buah titik[14]. Persamaan *Euclidean Distance* (4):

$$d_{ij} = \sqrt{\sum_{k=1}^p \{x_{ik} - x_{jk}\}^2} \quad (4)$$

Dengan:

d_{ij} = jarak antar objek i dengan j

p = dimensi data

X_{ik} = koordinat objek i pada dimensi k

X_{jk} = koordinat objek j pada dimensi k

- 3) Masukkan tiap objek ke segmen terdekat.
- 4) Lakukan pengukuran kembali tiap kali terdapat perubahan segmen.
- 5) Ulangi langkah hingga objek dalam tiap segmen tidak mengalami perubahan.

E. Davies-Bouldin Index

Davies boulden index atau DB Index merupakan suatu algoritma yang digunakan untuk mengevaluasi suatu *cluster*. Cara kerja atau skema dari evaluasi tersebut ialah dengan cara memvalidasi seberapa baik cluster yang ada menggunakan jumlah data dan properti yang tersedia pada dataset atau sumber data yang digunakan, sehingga mendapat 3 nilai yaitu jarak antara data dengan centroid, jarak antar cluster dan rasio yang dimiliki masing masing cluster.

Tahapan tahapan dalam perhitungan DB Index antara lain :

- 1) *Sum of Square within Cluster* (SSW) merupakan cara yang dipakai untuk mendapatkan nilai ketertarikan matriks pada sebuah cluster ke- i seperti persamaan (5) :

$$SSWi = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad (5)$$

dengan :

m_i = total dari data dalam cluster ke-i

c_i = centroid dari cluster ke-n

$d(x_j, c_i)$ = jarak dari setiap data terhadap centroid

- Menghitung *Sum of Square Between Cluster* (SSB) merupakan persamaan yang digunakan untuk menghitung nilai pemisahan / jarak antar cluster seperti persamaan (6) :

$$SSB_{i,j} = d(c_i, c_j) \quad (6)$$

- Menghitung total rasio, sesudah nilai dari SSW dan SSB sudah didapatkan langkah berikutnya menemukan nilai evaluasi rasio ($R_{i,j}$) untuk memperoleh nilai perbandingan antara cluster ke-i dan ke-j, rumus perhitungan sebagai berikut :

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSW_{ij}} \quad (7)$$

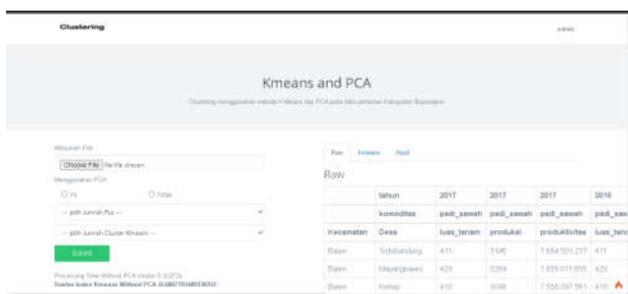
- Langkah terakhir adalah menghitung nilai DB Index dari hasil R_{ij} menggunakan persamaan (8) berikut :

$$DBIndex = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (8)$$

k = jumlah cluster yang telah ditentukan

III. HASIL DAN PEMBAHASAN

Sistem yang dirancang pada penelitian ini berhasil berjalan dan memberikan hasil sesuai yang diharapkan. Sistem mampu melakukan *clustering* menggunakan algoritma *K-Means* dengan PCA. Gambar 3 adalah hasil tangkapan layar tampilan sistem.



Gambar 2Tangkapan Layar Sistem

Sistem akan melakukan *clustering* pada data masukan pengguna, dengan pilihan untuk menggunakan PCA atau tidak. Setelah pengguna melakukan pengaturan sesuai dengan kebutuhan, system kemudian akan melakukan proses clustering

dan menampilkan data asli, data *clustering*, dan grafik sebaran data hasil *clustering*.

A. Dataset

Dataset yang digunakan pada penelitian ini merupakan data hasil pertanian Kabupaten Bojonegoro. Komoditas yang digunakan adalah padi sawah karena hampir semua daerah memiliki komoditas tersebut. Data yang terdiri atas 12 kolom dan 430 baris didapatkan dari situs resmi BPS dan disimpan dalam format .csv (*comma separated value*). Atribut data yang digunakan dalam penelitian ditampilkan pada tabel I.

TABEL I
 ATRIBUT DATA YANG DIGUNAKAN DALAM PENELITIAN

Nomor	Atribut Data
1	Tahun Produksi
2	Luas Tanam (Ha)
3	Jumlah Produksi (Ton)
4	Produktivitas

Contoh data yang disimpan untuk penelitian ini disajikan dalam gambar 2.

tahun	2017	2017	2017	2018	2018	2018	2019	2019	2019	2020	2020	2020
Desa	luas_tanam	produksi	produktivitas									
1 Subdandang	411	3146	7,654501217	411	3146	7,654501217	411	3146	7,654501217	411	3146	7,654501217
2 Masyrakawati	429	3284	7,655011855	429	3284	7,655011855	429	3284	7,655011855	429	3284	7,655011855
3 Kriep	410	3098	7,556097541	410	3098	7,556097541	410	3098	7,556097541	410	3098	7,556097541
4 Pangago	217	1658	7,640523985	217	1658	7,640523985	217	1658	7,640523985	217	1658	7,640523985
...
428 Gugangan	270	1647	6,1	270	1647	6,1	285	1647	5,77047368	285	1938	6,834736842
429 Sramak	188	1116	6,2	188	1152	6,4	189	1132	6,09528095	188	1124	6,8
430 Bangsari	211	1287	6,09528095	185	704	6,704762385	195	704	4,541935484	174	1148	6,597701148

Gambar 2 Data hasil pertanian Kab. Bojonegoro dari situs BPS

B. Reduksi Dimensi (PCA)

Dilakukan proses reduksi dimensi pada *dataset* yang telah disusun. Proses ini bertujuan untuk mereduksi dimensi/fitur dari dataset. Proses PCA mampu menghasilkan *principal component* (PC) 1 hingga jumlah kolom maksimal dari data yang digunakan, sehingga pada penelitian ini proses PCA menghasilkan 12 *principal component*, karena fungsi dari PCA adalah untuk mereduksi dimensi maka hanya disimpan sebanyak 3 PC.

Dataset pada penelitian ini diproses menggunakan PCA dengan perhitungan *eigen value* dan *eigen vector*, kemudian disimpan kedalam 3 kolom dengan jumlah baris yang sama untuk tiap kolom, yaitu 430 baris. Data ini yang kemudian digunakan pada proses *clustering*. Contoh baris data hasil proses PCA terdapat pada gambar 4.

PCA

No	Nama	1	2	3
1	Sidobandung	2.073266104026805	-0.8019421745907553	-0.010118513239955487
2	Mayangkawis	2.27708068197293	-0.741422517397017	-0.01299443600697195
3	Kenep	2.0014639033256634	-0.7240996311924432	-0.012837502970119414
4	Pohbogo	-0.12628930434513216	-1.4470205466273476	0.02056186701790719
5	Penganten	7.934684551734473	1.3325190007542527	-0.10401421535346266
6	Bulaklo	1.7285832929135536	-0.6499407100628007	-0.0149055665701182748
7	Bulu	3.1071079588150754	-0.33368556621909884	-0.0304732776323258
8	Kemamang	2.2978246772908117	-4.720222969875995	0.1466480439220598
9	Ngadiluhur	0.07816981538270601	-1.1794051670711927	0.009228593707936043

Gambar 3 Hasil Proses Reduksi Dimensi

C. Clustering (K-Means)

Proses clustering dilakukan terhadap dua jenis data berbeda, yaitu dataset asli dan dataset hasil proses PCA yang menghasilkan data dengan jumlah 1 PC, 2 PC, dan 3 PC. Clustering dilakukan menggunakan algoritma K-Means dengan inialisasi centroid awal secara acak.

Pada proses clustering, dari total keseluruhan jumlah desa, sebanyak 430 desa akan dibentuk menjadi kelompok-kelompok data dengan 4 jumlah klaster berbeda, yaitu berjumlah 2,3,4, dan 5 klaster. Hal ini dilakukan untuk mengetahui jumlah klaster paling optimal untuk data yang digunakan.

Tabel II merupakan jumlah anggota dari tiap-tiap cluster yang didapatkan melalui proses clustering.

TABEL II
 JUMLAH ANGGOTATIAP-TIAP CLUSTER

Jumlah Cluster	Cluster	Tanpa PCA	PC 1	PC 2	PC 3
2	1	100	104	99	99
	2	330	326	331	331
3	1	1	1	43	43
	2	103	119	83	82
	3	326	310	304	305
4	1	1	1	1	1
	2	30	31	43	43
	3	143	151	96	96
	4	256	247	290	290
5	1	1	1	1	1
	2	1	31	29	29
	3	30	100	43	43
	4	144	103	108	107
	5	254	195	249	250

D. Pengujian (DB Index)

Data hasil clustering kemudian dievaluasi menggunakan nilai DB Index. Perhitungan nilai DB Index dilakukan pada dataset tanpa proses reduksi dimensi dan dataset dengan proses reduksi dimensi. Hasil perhitungan nilai DB Index untuk

dataset tanpa proses reduksi dimensi dapat dilihat pada tabel III

TABEL III
 NILAI DB INDEX TANPA PROSES REDUKSI DIMENSI

Uji	Jumlah Cluster	Nilai DB Index
1	2	0,6812
2	3	0,4598
3	4	0,5341
4	5	0,4126

Perhitungan nilai DB Index pada dataset clustering yang melalui tahap reduksi dimensi menunjukkan hasil yang berbeda. Percobaan juga telah dilakukan sebanyak 4 kali uji. Nilai DB Index yang telah dihitung dirincikan pada tabel IV.

TABEL IV
 NILAI DB INDEX DENGAN PROSES REDUKSI DIMENSI

Jumlah PC	Jumlah Cluster			
	2	3	4	5
PC 1	0,6079	0,4072	0,4335	0,4814
PC 2	0,8374	0,6168	0,4678	0,5268
PC 3	0,8828	0,6598	0,5018	0,5988

Kedua data di atas kemudian dibandingkan untuk mengetahui pengaruh tahap reduksi dimensi dataset sebelum data tersebut di-clustering menggunakan algoritma K-means. Keseluruhan nilai DB Index dari kedua jenis dataset disajikan pada tabel V.

TABEL V
 KESELURUHAN NILAI DB INDEX DATASET

Clustering	Jumlah Cluster			
	2	3	4	5
K-Means	0,6812	0,4598	0,5341	0,4126
PCA-K-Means (1 dimensi)	0,6079	0,4072	0,4335	0,4814
PCA-K-Means (2 dimensi)	0,8374	0,6168	0,4678	0,5268
PCA-K-Means (3 dimensi)	0,8828	0,6598	0,5018	0,5988

Melalui percobaan yang telah dilakukan, dicari nilai DB Index paling optimal. Nilai DB Index paling optimal merupakan nilai DB Index terendah yang berarti perbedaan antar cluster/jarak intracuster paling kecil.

Dari data di atas dapat ditarik kesimpulan bahwa nilai DB Index paling optimal terdapat pada dataset yang direduksi menjadi 1 PC dan dibentuk menjadi 3 klaster. Penerapan tahap reduksi dimensi membuat adanya selisih pada nilai DB Index antar dataset dengan jumlah data dan cluster yang sama. Perbedaan nilai DB Index pada dataset dipengaruhi oleh jumlah

fitur yang terdapat pada *dataset* tersebut. *Dataset* yang melalui tahap reduksi dimensi akan mengalami peningkatan nilai *DB Index* seiring meningkatnya jumlah fitur yang ada

IV. KESIMPULAN

Melalui penelitian ini dapat disimpulkan bahwa, sistem yang dirancang telah berhasil menggunakan PCA untuk mereduksi fitur dari *dataset* yang didapatkan dari data hasil produksi pertanian Kab. Bojonegoro pada tahun 2017-2020 yang semula terdiri atas 12 kolom dan 430 baris, menjadi *dataset* dengan 1 PC, 2 PC, dan 3 PC.

Sistem mampu melakukan *clustering* menggunakan algoritma *K-Means* dengan jumlah kluster berbeda pada dua jenis *dataset* berbeda, yaitu *dataset* asli dan *dataset* yang telah mengalami proses reduksi dimensi. Proses *clustering* menghasilkan data dengan jumlah kluster 2, 3, 4 dan 5 dari tiap-tiap *dataset* yang digunakan.

Evaluasi data hasil *clustering* menggunakan nilai *DB Index* menunjukkan nilai paling optimal pada *dataset* yang direduksi menjadi 1 PC dan dibentuk menjadi 3 kluster, yaitu 0.4072. Sedangkan dengan jumlah kluster yang sama, *dataset* dengan PC 2 menghasilkan nilai *DB Index* 0.6168, *dataset* dengan PC 3 menghasilkan nilai 0.6598, dan *dataset* tanpa proses reduksi dimensi menghasilkan nilai *DB Index* 0.4598.

Untuk penelitian yang akan datang, disarankan untuk menguji efisiensi algoritma *PCA K-Means* menggunakan data dengan tingkat kompleksitas tinggi dan jumlah data yang lebih banyak. Selain itu, perlu mempertimbangkan penggunaan algoritma *clustering* dan reduksi dimensi lainnya beserta parameter-parameter uji lain untuk menilai kinerja algoritma, sehingga tingkat efisiensi dan keberhasilan algoritma dalam melakukan *clustering* lebih kredibel karena mempertimbangkan banyak aspek.

UCAPAN TERIMA KASIH

Terima kasih kepada Tuhan YME atah rahmat-Nya. Terima kasih pula kepada BPS Kabupaten Bojonegoro atas dukungan data untuk penelitian ini. Tak lupa orang tua dan keluarga yang selalu memberikan kasih sayang dan kesabaran serta semua pihak yang telah memberikan bantuan dan dukungan sehingga penelitian ini dapat dilakukan hingga tuntas.

REFERENSI

- [1] Kompas, "Kompas.com," 19 Februari 2017. [Online]. Available: <https://money.kompas.com/read/2017/02/19/163912926/negara.agraris.mengapa.harga.pangan.di.indonesia.raw.an.bergejolak.?page=all#:~:text=Indonesia%20juga%20dikenal%20sebagai%20negara,bidang%20pertanian%20atau%20bercocok%20tanam.&text=Selain%20itu%2C%20Indon.> [Diakses 23 November 2020].
- [2] Badan Pusat Statistik, "BPS - Statistics Indonesia," 2020. [Online]. Available: [https://www.bps.go.id/menu/1/pengolahan-](https://www.bps.go.id/menu/1/pengolahan-data.html#masterMenuTab6)
- [3] H. A. Issad, R. Aoudjit dan J. J. Rodrigues, "A comprehensive review of Data Mining techniques in smart agriculture," *Engineering in Agriculture, Environment and Food*, 2019.
- [4] B. Poerwanto dan R. Y. Fa'rifah, "Algoritma K-Means Dalam Mengelompokkan Kecamatan Di Tana Luwu Berdasarkan Produktifitas Hasil Pertanian," *Jurnal Ilmiah d'Computare*, vol. 9, 2019.
- [5] A. Izzuddin, "Optimasi Cluster pada Algoritma K-Means dengan Reduksi Dimensi Dataset Menggunakan Principal Component Analysis untuk Pemetaan Kinerja Dosen," vol. 5, no. [2], 2015.
- [6] P. Alkhairi dan A. P. Windarto, "Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," dalam *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 2019.
- [7] C.-H. Yu, F. Gao, S. Lin dan J. Wang, "Quantum data compression by principal component," *Quantum Information Processing*, 2019.
- [8] S.Velliangiri, S.Alagumuthukrishnan dan S. I. T. Joseph, "A Review of Dimensionality Reduction Techniques for Efficient Computation," dalam *International Conference on Recent Trends in Advanced Computing*, 2019.
- [9] G. T. Reddy, P. K. Reddy, K. Lakshmana, R. Kaluri dan D. S. Rajput, "Analysis of Dimensionality Reduction Techniques on Big Data," *IEEE Access*, 2020.
- [10] Y. Zhang dan Z. Zhao, "Fetal State Assessment based on Cardiotocography Parameters Using PCA and AdaBoost," dalam *10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, 2017.
- [11] Adiwijaya, U. N. Wisesty, E. Lisnawati, A. Aditsania dan D. S. Kusumo, "Dimensionality Reduction using Principal Component Analysis for Cancer Detection based on Microarray Data Classification," *Jurnal of Computer Science*, vol. 14 (11), pp. 1521-1530, 2018.
- [12] A. Jamala, A. Handayana, A. A. Septiandria, E. Ripmiatina dan Y. Effendi, "Dimensionality Reduction using PCA and K-Means Clustering for Breast Cancer Prediction," *Lontar Komputer*, vol. 9 (3), pp. 192-201, 2018.
- [13] P.Surya dan I. Aroquiaraj, "Performance Analysis Of K-Means And K-Medoid Clustering Algorithms Using Agriculture Dataset," *Journal of Emerging Technologies and Innovative Research*, vol. 6 (1), pp. 539-545, 2019.
- [14] J. Pardede, M. Gustiana dan M. Nurhasan, "Implementasi K-Means Clustering Pada Aplikasi Gis (Geographic Information System) Studi Kasus Pertanian Padi," dalam *Seminar Nasional APTIKOM (SEMNASITIKOM)*, Mataram, 2016.