

Data Mining Pada Faktor-Faktor Potensi Daerah di Kabupaten Sidoarjo Provinsi Jawa Timur

Trianingsih Eni Lestari

Jurusan Matematika, FMIPA – Universitas Negeri Malang
e-mail: trianingsih.eni.fmipa@um.ac.id

Hendro Permadi

Jurusan Matematika, FMIPA – Universitas Negeri Malang
e-mail: hendro.permadi.fmipa@um.ac.id

Sri Susilowati

Jurusan Matematika, FMIPA – Universitas Negeri Malang
e-mail: shusilw@gmail.com

Abstract: *Sidoarjo is one of the districts located in East Java that has developed rapidly. The remarkable progress can be achieved due to several potentials had by its people, for instance, industries, trades, small and medium businesses. Therefore, this research aims to find out the information regarding dominating factors had by the Sidoarjo using data mining. The result shows that Keboansikep, Sawotratap, Tebel, Keboananom, Gedangan, Keboguyang, Ketajen, Sidomulyo, Terik, Ponokawan, Sedengan Mijen, and Barengkrajan villages are the most potential villages in Sidoarjo. Based on the classification method, it is found that the villages of Keboansikep, Sawotratap, Tebel, Keboananom, Gedangan, and Ketajen (Gedangan District) have local potential in the form of agricultural factors such as rice and secondary crops. All residences in Keboguyang Village (Jabon District) already have an IMB. Meanwhile, the villages of Sidomulyo, Terik, Ponokawan, Sedengan Mijen, and Barengkrajan (Krian District) have high early childhood education factors such as kindergarten students, kindergarten teachers, and kindergarten schools*

Keywords: *Biplot Analysis, Data Mining, Local Potential*

Abstrak: *Kabupaten Sidoarjo merupakan salah satu kabupaten yang terletak di provinsi Jawa Timur dan merupakan daerah yang mengalami perkembangan cukup pesat. Keberhasilan ini dicapai karena berbagai potensi yang ada di wilayahnya. Artikel ini bertujuan untuk mengetahui faktor-faktor yang mendominasi potensi daerah Kabupaten Sidoarjo dengan menggunakan data mining. Hasil penelitian menunjukkan bahwa Desa Keboansikep, Sawotratap, Tebel, Keboananom, Gedangan, Keboguyang, Ketajen, Sidomulyo, Terik, Ponokawan, Sedengan Mijen dan Barengkrajan merupakan desa yang memiliki potensi daerah terbaik di Kabupaten Sidoarjo. Berdasarkan metode klasifikasi diperoleh Desa Keboansikep, Sawotratap, Tebel, Keboananom, Gedangan, dan Ketajen (Kecamatan Gedangan) memiliki potensi daerah berupa faktor pertanian yakni hasil panen padi dan palawija. Seluruh hunian di Desa Keboguyang (Kecamatan Jabon) telah memiliki IMB. Sedangkan Desa Sidomulyo, Terik, Ponokawan, Sedengan Mijen*

dan Barengkrajan (Kecamatan Krian) mempunyai faktor pendidikan usia dini yang tinggi yakni murid TK, Guru TK dan Sekolah TK.

Kata Kunci: Analisis Biplot, Data Mining, Potensi Daerah

1. Pendahuluan

Pertumbuhan *Big Data* bertambah pesat di era modern ini. *Big Data* semakin populer seiring dengan semakin majunya teknologi, informasi dan komunikasi. *Big Data* terdiri dari himpunan data dengan jumlah besar dan tidak terstruktur sehingga sukar untuk ditangani secara manual. *Big Data* dapat menghasilkan informasi yang lebih cepat sehingga pengumpulan data seperti *survey* sudah mulai ditinggalkan (Maryanto, 2017).

Data mining merupakan proses untuk mendapatkan informasi berdasarkan himpunan data dengan jumlah besar atau *Big Data* yang dapat digunakan untuk mengambil sebuah keputusan (Sulianta & Juju, 2010). *Data mining* dapat dikatakan sebagai inovasi dalam ilmu statistika. Perkembangan *data mining* tidak lepas dari perkembangan teknologi (Santosa & Umam, 2018). Teknologi mempunyai peran dalam merubah pola pikir individu, sehingga dapat digunakan untuk menentukan suatu kebijakan pemerintah dalam jangka panjang. *data mining* memiliki beberapa fungsi yaitu klasifikasi, *clustering*, asosiasi, dan estimasi (Kusrini & Luthfi, 2009). Klasifikasi digunakan untuk menemukan model untuk kepentingan tertentu. *Clustering* digunakan untuk mengelompokkan data berdasarkan kemiripan. Tingkat akurasi pada masing-masing teknik memiliki perbedaan dari setiap model yang dihasilkan. Asosiasi digunakan untuk mengidentifikasi suatu kombinasi item yang dapat muncul secara bersamaan. Estimasi digunakan untuk mengestimasi suatu nilai yang belum diketahui. Penelitian terdahulu tentang *data mining* pernah dibahas oleh (Jain & Singh, 2013) yang menerapkan dalam sektor primer, sekunder dan tersier pada Produk Domestik Negara di India. Selain itu, *data mining* juga diterapkan pada Produk Domestik Bruto di masa yang akan datang berdasarkan pertanian, jasa dan hasil produksi (Kunda & Chihana, 2017).

Kabupaten Sidoarjo adalah salah satu daerah yang mengalami perkembangan cukup pesat di Jawa Timur. Potensi daerah seperti pariwisata, perdagangan, industri, serta usaha kecil yang dapat dikemas secara terarah menjadi faktor keberhasilan bagi Kabupaten Sidoarjo (Sungai et al., 1994). Dukungan sumber daya manusia yang memadai, menjadikan daerah yang strategis bagi pengembalian perekonomian regional (Supriyati et al., 2017). Jumlah penduduk yang meningkat setiap tahun dan seimbang dengan lapangan pekerjaan membuat Kabupaten Sidoarjo menjadi daerah yang sejahtera (Kustanto & Sholihah, 2018). Perkembangan teknologi, informasi dan komunikasi telah membuat Kabupaten Sidoarjo semakin membaik. Seluruh data akan semakin meningkat setiap tahun dan membutuhkan ruang penyimpanan data yang cukup besar (Santoso, 2010). Jika

hal ini tidak diatasi maka akan kesulitan untuk mengelola *Big Data*. Sehingga tidak dapat diketahui faktor-faktor apa saja yang mendominasi potensi daerah di Kabupaten Sidoarjo. Dengan begitu, tujuan dari penelitian ini adalah untuk menganalisis *Big Data* dengan menggunakan *data mining* sehingga dapat membantu Pemerintah Daerah Kabupaten Sidoarjo untuk membuat suatu kebijakan lebih lanjut mengenai faktor-faktor yang mendominasi potensi daerah di Kabupaten Sidoarjo.

2. Metode Penelitian

Data yang digunakan adalah data primer dan data sekunder. Adapun data primer diperoleh dengan melakukan survei di 4 kecamatan dengan 71 desa (Kecamatan Gedangan, Kecamatan Jabon, Kecamatan Krian, dan Kecamatan Tarik) di Kabupaten Sidoarjo. Sedangkan data sekunder berasal dari Badan Pusat Statistik Kabupaten Sidoarjo berdasarkan Desa tahun 2019.

Variabel yang diamati dalam penelitian ini dikelompokkan menjadi beberapa faktor berdasarkan karakteristik yang dimiliki oleh setiap variabel antara lain:

1. Faktor Pemerintahan: jumlah dusun (X_1), jumlah perangkat desa (X_2), jumlah hansip (X_3), jumlah RW (X_4), jumlah RT (X_5), jarak ke kecamatan (X_6).
2. Faktor Kependudukan: jumlah kepala keluarga (X_7), jumlah penduduk (X_8), jumlah pendatang (X_9), jumlah kepindahan (X_{10}).
3. Faktor Jumlah Guru: guru TK (X_{11}), guru SD (X_{12}), guru SMP (X_{13}), guru SMA (X_{14}).
4. Faktor Jumlah Murid: murid TK (X_{15}), murid SD (X_{16}), murid SMP (X_{17}), murid SMA (X_{18}).
5. Faktor Fasilitas Pendidikan: TK (X_{19}), SD (X_{20}), SMP (X_{21}), SMA (X_{22}).
6. Faktor Pertanian: padi (X_{23}), palawija (X_{24}), buah (X_{25}), populasi ternak (X_{26}), populasi unggas (X_{27}).
7. Faktor Kendaraan: mobil (X_{28}), bus (X_{29}), truck (X_{30}), sepeda (X_{31}), becak (X_{32}), kendaraan lain (X_{33}).
8. Faktor Pekerjaan: PNS (X_{34}), ABRI (X_{35}), petani (X_{36}), buruh tani (X_{37}), buruh swasta (X_{38}), pedagang (X_{39}), usaha industri (X_{40}), usaha angkutan (X_{41}).
9. Faktor Pemukiman dan Wilayah: non IMB (X_{42}), IMB (X_{43}), hunian (X_{44}).

Tahap analisis data yang akan dilakukan dalam penelitian ini sebagai berikut:

1. Seleksi data.

Seleksi data dilakukan berdasarkan variabel yang akan diteliti menggunakan analisis faktor. Langkah-langkah dalam seleksi data sebagai berikut:

- a. Asumsi-asumsi yang harus dipenuhi adalah:
 - Nilai Uji *Kaiser-Meyer-Olkin* (KMO) $> 0,5$.
 - Nilai *Measure of Sampling Adequacy* $> 0,5$.

- Nilai Uji *Barlett Test of Sphericity* < 0,5.
- b. Menentukan jumlah faktor yang dapat diekstrak.
- c. Menentukan jumlah faktor yang terbentuk dengan melihat nilai eigen. Faktor signifikan jika nilai eigen yang dihasilkan lebih besar dari 1.
- d. Menggunakan metode Varimax untuk rotasi faktor. Hal ini bertujuan agar faktor lebih mudah untuk diinterpretasikan.
- e. Melakukan interpretasi faktor dengan cara mengelompokkan variabel.

2. Data Mining

Proses data mining mencakup klasifikasi, *clustering*, asosiasi dan estimasi., dimana tahapannya dijelaskan sebagai berikut:

a. Klasifikasi

Tahap klasifikasi dilakukan dengan cara mengelompokkan variabel hasil dari seleksi data berdasarkan karakteristik yang dimiliki untuk menemukan model atau fungsi untuk menggambarkan konsep dari suatu data dan merupakan proses yang digunakan untuk mendeskripsikan data yang penting.

b. *Clustering*

Setelah mendapatkan variabel yang dominan, tahap selanjutnya adalah melakukan analisis *cluster* berdasarkan variabel yang dominan. Teknik analisis cluster yang digunakan adalah hierarki dengan menggunakan metode *Single Linkage, Complete Linkage, Average Linkage, Centroid dan Ward's* (Johnson & Wichern, 2007). Adapun penentuan metode terbaik menggunakan kriteria RMSSTD terkecil. Semakin kecil nilai RMSSTD maka semakin tinggi kehomogenan kelompok yang terbentuk. Adapun rumus nilai RMSSTD sebagai berikut:

$$RMSSTD = \sqrt{\frac{\sum_{u=1}^{n_a} \sum_{v=1}^{r_b} (x_{uv} - \bar{x}_u)^2}{\sum_{b=1}^{n_a} (r_b - 1)}}$$

Keterangan:

x_{uv} = nilai obyek ke-v pada kelompok ke-u

\bar{x}_u = nilai pusat kelompok ke-u (*centroid*)

n_a = jumlah kelompok yang terbentuk

r_b = banyaknya obyek yang termasuk ke dalam kelompok ke-b

c. Asosiasi

Tahap asosiasi dilakukan dengan analisis biplot berdasarkan variabel yang dominan dalam analisis faktor. Analisis biplot merupakan metode multivariat yang bertujuan untuk mendeskripsikan data dengan cara membuat pemetaan dalam tampilan plot. seperti yang pernah dilakukan oleh (Srinadi & Sumarjaya, 2016) yang memetakan posisi dan karakteristik usaha pariwisata di Provinsi Bali. Analisis biplot banyak diterapkan untuk melihat kedekatan antar objek berdasarkan variabel penyusunnya. Penelitian sebelumnya tentang penerapan analisis biplot antara lain pernah diterapkan pada indikator kesejahteraan di Provinsi Bali (Sukarsa & Gandhiadi, 2020).

Perhitungan dengan analisis biplot berdasarkan pada dekomposisi nilai singular (SVD) terhadap matriks data (Jolliffe, 2002). Matriks X dinyatakan sebagai SVD seperti berikut:

$$X = ULA^B$$

Keterangan:

- X = Matriks berukuran $n \times p$ dengan n obyek pengamatan dan p peubah.
- U = Matriks berukuran $n \times r$ dengan kolom-kolomnya disebut vektor singular kolom dan matriks U merupakan matriks orthonormal.
- L = Matriks berukuran $r \times r$ dengan unsur diagonal utama adalah nilai dari singular matriks X yang merupakan hasil dari nilai eigen matriks $X'X$.
- A = Matriks berukuran $r \times p$ dengan kolom-kolomnya disebut vektor eigen dari matriks $X'X$ Kolom pada matriks A disebut vektor singular baris dan matriks A merupakan matriks orthonormal.

Langkah-langkah dalam metode ini sebagai berikut:

1. Menentukan variabel yang akan diteliti berdasarkan variabel yang dominan dalam analisis faktor.
2. Membuat matriks U , L dan A dengan menggunakan metode *Singular Value Decomposition* (SVD).
3. Melakukan pemetaan karakteristik berdasarkan variabel yang dominan yaitu dengan cara melakukan proyeksi orthogonal pada setiap objek terhadap semua vektor yang ada di dalam tampilan analisis biplot.
4. Uji kesesuaian keragaman data dengan syarat keragaman data $\geq 70\%$.
5. Interpretasi hasil analisis biplot.
- d. Estimasi

Tahap estimasi dilakukan dengan analisis analisis korelasi berdasarkan variabel yang dominan dalam analisis faktor. Langkah-langkah dalam metode ini adalah sebagai berikut:

1. Analisis korelasi dengan variabel dominan yang diperoleh dari analisis faktor.
2. Menentukan variabel yang memiliki korelasi (r) yang tertinggi dengan variabel yang lain.
3. Menentukan hubungan antar variabel berdasarkan kategori:
 - a. Korelasi kuat secara negatif : $-1,00 \leq r \leq -0,80$
 - b. Korelasi sedang secara negatif : $-0,79 \leq r \leq -0,50$
 - c. Korelasi lemah : $-0,49 \leq r \leq 0,49$
 - d. Korelasi sedang secara positif : $0,50 \leq r \leq 0,79$
4. Korelasi kuat secara positif : $0,80 \leq r \leq 1,00$
5. Mencari distribusi terbaik dari variabel tersebut.
6. Menghitung nilai estimasi dan interpretasi hasil

3. Hasil dan Pembahasan

3.1 Seleksi Data

Tahapan seleksi data dilakukan untuk mengetahui variabel-variabel utama yang akan diproses ke dalam tahapan selanjutnya. Sebanyak 44 variabel akan dilakukan seleksi menggunakan analisis faktor. Hasil seleksi variabel pada Kabupaten Sidoarjo ditampilkan pada Tabel 1. Berdasarkan Tabel 1, diperoleh bahwa seluruh faktor telah memenuhi asumsi yakni nilai KMO > 0,5, nilai Barlett's Test < 0,5 dan nilai MSA > 0,5 serta terbentuk variabel utama pada masing-masing faktor. Setelah dilakukan analisis dengan variabel-variabel utama, nilai MSA pada variabel Petani (X36) < 0,5 sehingga variabel tersebut harus dikeluarkan dari analisis. Setelah itu, uji kembali variabel yang telah memenuhi asumsi karena masing-masing nilainya harus lebih dari 0,5

Tabel 1. Hasil Uji Seleksi *Big Data* di Kabupaten Sidoarjo Tahun 2019

No	Faktor	KMO	Barlett's	Nilai MSA	Kumulatif	Variabel Utama
1	Pemerintahan	0,634	0	Memenuhi	60%	Jumlah RT (X_5)
2	Kependudukan	0,502	0	Memenuhi	79,7%	KK (X_7)
3	Jumlah Guru	0,602	0	Memenuhi	72,9%	Guru TK (X_{11})
4	Jumlah Murid	0,511	0	Memenuhi	71%	Murid TK (X_{15})
5	Fasilitas Pendidikan	0,661	0	Memenuhi	81,3%	TK (X_{19})
6	Pertanian	0,509	0	Memenuhi	66,7%	Padi (X_{23}) Palawija (X_{24})
7	Kendaraan	0,705	0	Memenuhi	63,3%	Becak (X_{32})
9	Pemukiman dan Wilayah	0,562	0,001	Memenuhi	82,3%	Non IMB (X_{42})

Hasil lainnya adalah nilai Eigen dari variabel-variabel terbaik yang disajikan dalam Tabel 2.

Tabel 2. Nilai Eigen Variabel-variabel Terbaik

Component	Initial Eigenvalues		
	Total	% of Variance	Comulative (%)
1	4,167	46,302	46,302
2	1,625	18,053	64,356
3	1,301	14,451	78,807
4	0,681	7,569	86,376
5	0,599	6,656	93,032
6	0,295	3,280	96,312
7	0,189	2,104	98,416
8	0,082	0,916	99,332
9	0,060	0,668	100,000

Berdasarkan Tabel 2, terdapat 3 komponen yang memiliki nilai *eigen* > 1, sehingga dapat disimpulkan bahwa keragaman data dengan menggunakan 3 komponen yaitu sebesar 78,807%.

3.2 Klasifikasi

Berdasarkan hasil dari seleksi data dengan menggunakan metode analisis faktor, hasil klasifikasi dari variabel terbaik berdasarkan karakteristik yang dimiliki oleh variabel-variabel tersebut adalah sebagai berikut:

1. Pendidikan Usia Dini, terdiri dari variabel Guru TK (X_{11}), Murid TK (X_{15}) dan Sekolah TK (X_{19}).
2. Sosial, terdiri dari variabel Jumlah RT (X_5), Jumlah Kepala keluarga (X_7), Non IMB (X_{42}).
3. Pertanian, terdiri dari variabel Padi (X_{23}) dan Palawija (X_{24}).
4. Kendaraan, terdiri dari variabel Becak (X_{32}).

3.3 Clustering

Tahap *clustering* dilakukan dengan menggunakan metode *Cluster Hierarki*. Pada metode hierarki, analisis cluster dilakukan dengan metode *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Centroid* dan *Ward*. Berdasarkan lima metode tersebut, akan dipilih satu metode yang memiliki nilai RMSSTD terkecil untuk jumlah kelompok (*cluster*) yang terpilih. Nilai RMSSTD dari kelima metode disajikan pada Tabel 3. Berdasarkan Tabel 3, metode terbaik yang dapat digunakan adalah metode *Complete Linkage* karena memiliki nilai RMSSTD terkecil yaitu 3,581

Tabel 3. Nilai RMSSTD Metode Cluster Hierarki

Kelompok	Metode				
	<i>Single Linkage</i>	<i>Complete Linkage</i>	<i>Average Linkage</i>	<i>Centroid</i>	<i>Ward</i>
1	27,062	27,062	27,062	27,062	27,062
2	26,676	26,676	26,676	26,676	20,532
3	26,286	13,153	26,286	26,286	13,577
4	25,914	9,633	25,890	25,890	7,580
5	25,530	3,581	25,135	25,496	6,728

Setelah terpilih metode *Complete Linkage* yang memiliki nilai RMSSTD terkecil maka berikutnya akan disajikan hasil pengelompokan dari potensi daerah Kabupaten Sidoarjo pada Tabel 4. Hasil pengelompokan yang terbentuk menggunakan metode *Complete Linkage* adalah 71 desa membentuk 5 kelompok dimana ada satu kelompok yaitu kelompok 5 yang hanya terdiri dari satu desa saja Krian. Penjelasan kelompok mana yang termasuk kelompok dengan potensi daerah terbaik, maka digunakan peringkat dari masing-masing variabel. Hasil peringkat disajikan dalam Tabel 5 dibawah ini.

Tabel 4. Hasil Pengelompokan dengan Metode *Complete Linkage*

Kelompok	Desa
Kelompok 1	Sruni, Tarik, Kedungcangkring, Kramat Temenggung, Gempolklutuk, Singogalih, Kedinding, Gampingrowo, Balongmacekan, Miriprowo, Klantingsari, Mindugading, Sebani, Gemurung, Gamping, Kemas, Junwangi, Semambung, Tempel, Jaticalang, Tropodo, Karangbong, Punggul, Bangah, Wedi, Kragan, Terung Wetan, Permisian, Tambak-kalisogo, Jeruk Gamping, Sidomojo, Keboarahan, Pejarakan, Ganting, Kendalsewu
Kelompok 2	Keboansikep, Sawotratap, Tebel, Keboananom, Gedangan, Keboguyang, Ketajen, Sidomulyo, Terik, Ponokawan, Sedengan Mijen, Barengkrajan
Kelompok 3	Jemirahan, Pangreh, Dukuhsari, Semambung, Balongtani, Kedungpandan, Trompoasri, Kedungrejo, Kupang, Watugolong, Terung Kulon, Tambak Kemerak, Kraton, Katrungan
Kelompok 4	Mergobener, Mergosari, Banjarwungu, Kedungbocok, Kalimati, Segodobancang, Janti, Kemuning, Sidorejo
Kelompok 5	Krian

Tabel 5. Rata-rata Setiap Kelompok hasil *clustering* dari variabel-variabel Terbaik

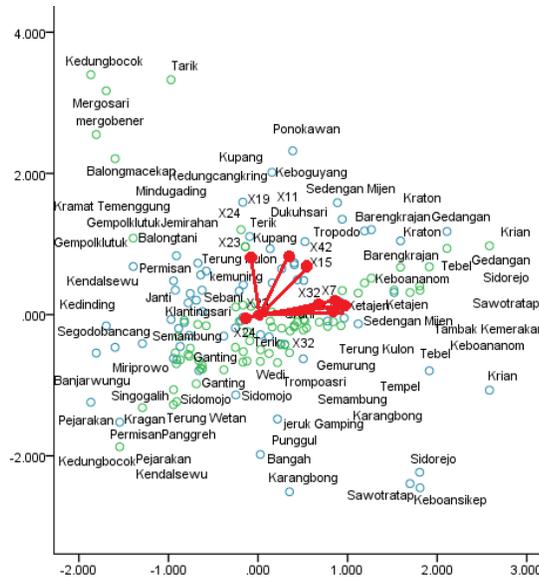
Variabel	Kelompok 1	Kelompok 2	Kelompok 3	Kelompok 4	Kelompok 5
X5	20,25	34	18	22,5	40
X7	1385,88	2796,25	1836,55	1294,3	3445
X11	7,51	15,91	12,28	5,1	15
X15	88,71	172,08	119,85	47,8	221
X19	2,4	3,08	2,78	2,3	3
X23	81,16	55,8	154,28	208,5	24
X24	45,98	3,83	1,14	207,84	0
X32	9,77	20,5	13,5	9,5	178
X42	1059,57	1676,9	1251,07	1131,67	2153

Berdasarkan Tabel 5, diperoleh bahwa kelompok 2 adalah kelompok dengan potensi daerah sangat tinggi, kelompok 5 dengan potensi daerah tinggi, kelompok 4 dengan potensi daerah sedang, kelompok 1 dengan potensi daerah rendah, dan kelompok 3 dengan potensi daerah sangat rendah.

3.4 Asosiasi

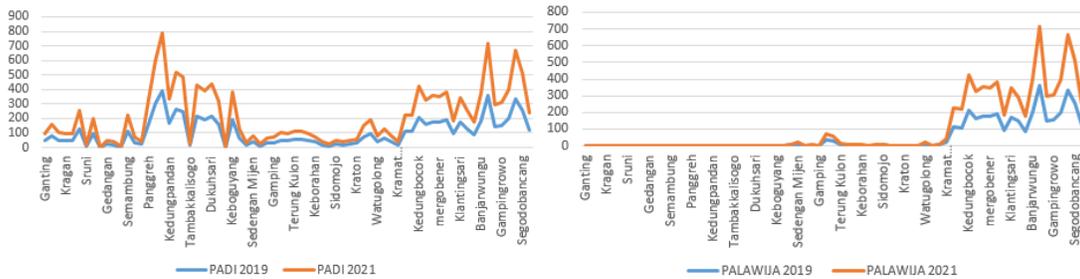
Pada tahap asosiasi dihasilkan analisis biplot pada Gambar 1. Hasil analisis biplot diperoleh keragaman data sebesar 77,3%. Berdasarkan Gambar 1, didapatkan hasil bahwa vektor yang paling panjang adalah variabel Jumlah Kepala Keluarga (X_7). Hal ini menunjukkan bahwa jumlah kepala keluarga yang ada di 71 desa di Kabupaten Sidoarjo tahun 2019 memiliki keragaman paling besar artinya jumlah kepala keluarga yang ada di 71 desa di Kabupaten Sidoarjo tahun 2019 menyebar secara heterogen diantara variabel lainnya. Vektor yang memiliki panjang terkecil adalah variabel Sekolah TK (X_{19}). Hal ini

menunjukkan bahwa jumlah Sekolah TK yang ada di 71 desa di Kabupaten Sidoarjo tahun 2019 memiliki keragaman paling kecil dan menyebar secara homogen



Gambar 1. Hasil Analisis Biplot

Hasil lainnya adalah berdasarkan sudut antar variabel. Sudut antara dua variabel didapatkan hasil bahwa sudut Murid TK (X_{15}) dengan Guru TK (X_{11}), Sekolah TK (X_{19}) dan Jumlah KK (X_7) kurang dari 90° artinya berkorelasi positif. Hal ini menunjukkan jika jumlah murid TK meningkat maka akan meningkatkan jumlah guru TK dan sekolah TK. Sebaliknya, sudut Jumlah KK (X_7) dengan Padi (X_{23}) dan Palawija (X_{24}) lebih dari 90° sehingga berkorelasi negatif.



Gambar 2. Grafik Nilai Estimasi Variabel Padi dan Palawija Tahun 2021

Hasil estimasi dari variabel padi dan palawija disajikan Gambar 2. Berdasarkan Gambar 2, diperoleh bahwa pada tahun 2021 hasil panen padi dan palawija setiap desa di Kabupaten Sidoarjo mengalami kenaikan, hal ini menunjukkan bahwa semakin banyak

hasil panen padi maka semakin banyak pula hasil panen palawija yang ada di kabupaten Sidoarjo tahun 2021.

4. Kesimpulan dan Saran

Hasil *data mining* pada kasus *Big Data* terhadap potensi daerah di Kabupaten Sidoarjo tahun 2019 pada bab sebelumnya dengan menggunakan metode *cluster* didapatkan 5 kelompok desa. Berdasarkan hasil seluruh analisis diperoleh Desa Keboansikep, Sawotratap, Tebel, Keboananom, Gedangan, Keboguyang, Ketajen, Sidomulyo, Terik, Ponokawan, Sedengan Mijen, dan Barengkrajan merupakan desa yang memiliki potensi daerah terbaik di Kabupaten Sidoarjo.

Berdasarkan metode klasifikasi diperoleh Desa Keboansikep, Sawotratap, Tebel, Keboananom, Gedangan, dan Ketajen (Kecamatan Gedangan) memiliki potensi daerah berupa faktor pertanian yakni hasil panen padi dan palawija. Seluruh hunian di Desa Keboguyang (Kecamatan Jabon) telah memiliki IMB. Sedangkan Desa Sidomulyo, Terik, Ponokawan, Sedengan Mijen dan Barengkrajan (Kecamatan Krian) mempunyai faktor pendidikan usia dini yang tinggi yakni murid TK, Guru TK dan Sekolah TK.

Hasil metode menggunakan analisis biplot berdasarkan panjang vektor diperoleh variabel Kepala Keluarg (X_7) memiliki keragaman paling besar atau paling heterogen dan variabel Sekolah TK (X_{19}) paling homogen. Sedangkan berdasarkan sudut yang terbentuk Sudut Murid TK (X_{15}) dengan Guru TK (X_{11}), Sekolah TK (X_{19}) dan Jumlah Kepala Keluarga (X_7) kurang dari 90° artinya berkorelasi positif, jadi jika jumlah murid TK meningkat maka akan meningkatkan jumlah guru TK dan sekolah TK. Sudut Jumlah Kepala Keluarga (X_7) dengan Padi (X_{23}) dan Palawija (X_{24}) lebih dari 90° artinya berkorelasi negatif, jadi jika jumlah kepala keluarga meningkat akan menurunkan hasil panen padi dan palawija karena tingkat konsumsi akan meningkat.

Hasil estimasi untuk tahun 2021 diperoleh bahwa hasil panen padi (X_{23}) dan palawija (X_{24}) setiap desa di Kabupaten Sidoarjo mengalami peningkatan artinya semakin bertambah hasil panen padi maka semakin bertambah pula hasil panen palawija di Kabupaten Sidoarjo tahun 2021.

Saran untuk penelitian berikutnya adalah *data mining* dapat diterapkan pada daerah yang lain untuk mengetahui faktor-faktor yang mendominasi potensi daerah sehingga bisa memberikan masukan dalam pengambilan kebijakan.

Ucapan Terima Kasih

Artikel ini merupakan bagian dari penelitian yang didanai PNBPU UM tahun 2020. Terimakasih kepada Universitas Negeri Malang yang telah mendukung dan mendanai penelitian ini

Daftar Pustaka

- Jain, R., & Singh, D. (2013). Data Mining and Analysis of Economic Data. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(August 2013), 683–688.
- Johnson, R., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall.
- Jolliffe, I. T. (2002). *Principal Component Analysis, Second Edition*. New York : Springer.
- Kunda, D., & Chihana, S. (2017). Analysis of value added services on GDP Growth rate using Data Mining Techniques. *Database System Journal*, 8 No 2, 29–43.
- Kusrini, & Luthfi, E. (2009). *Algoritma Data Mining*. CV.Andi Offset.
- Kustanto, M., & Sholihah, F. (2018). *Data Dinamis Kabupaten Sidoarjo triwulan 1-2018*. Bappeda.
- Maryanto, B. (2017). Big Data dan Pemanfaatannya. *Media Informatika*, 16(2), 14–19.
- Santosa, B., & Umam, A. (2018). *Data Mining dan Big Data Analytics: Teori dan Implementasi Menggunakan Python & Apache Spark Edisi 2*. Penebar Media Pustaka.
- Santoso, S. (2010). *Mahir Statistik Multivariat dengan SPSS*. PT Gramedia.
- Srinadi, I. G. A. M., & Sumarjaya, I. W. (2016). Analisis Biplot untuk Pemetaan Posisi dan Karakteristik Usaha Pariwisata di Provinsi Bali. *Jurnal Matematika*, 6(1), 34–45.
- Sukarsa, I. K., & Gandhiadi, G. (2020). *Indikator Kesejahteraan di Provinsi Bali : Suatu Pendekatan Analisis Biplot*. 10(1), 22–31.
<https://doi.org/10.24843/JMAT.2020.v10.i01.p120>
- Sulianta, F., & Juju, D. (2010). *Data Mining: Meramalkan Bisnis Perusahaan*. PT Elex Media Komputindo.
- Sungai, D. A. N. M., Mansyur, A., & Tonnek, S. (1994). *Prospek budi daya bandeng dalam karamba jaring apung laut dan muara sungai*. 129, 79–85.
- Supriyati, Meliza, & Anik. (2017). Industrialisasi Pertambakkan Kabupaten Sidoarjo Sebagai Upaya Peningkatan Kemakmuran Masyarakat. *Jurnal Pengabdian LPPM Untag Surabaya*, 02(03), 26–32.