

# ANALISIS FITUR REDUKSI *PRINCIPAL COMPONENT ANALYSIS* (PCA) PADA AKURASI KLASIFIKASI OZON

Anang Fajrial Prathama<sup>1</sup>, Guruh Fajar Shidik<sup>2</sup>

Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

Jl. Nakula I No. 5-11, Semarang, Jawa Tengah 50131 – (024)3517261

E-mail : anang.fajrial10@gmail.com<sup>1</sup>, guruh.shidik@dsn.dinus.ac.id<sup>2</sup>

---

## **Abstrak**

Ozon adalah gas yang secara alami terdapat di atmosfer bumi. Tersusun oleh tiga molekul oksigen atau biasa dilambangkan  $O_3$ . Ozon berfungsi untuk melindungi bumi dari radiasi sinar ultraviolet yang dipancarkan matahari. Banyak sekali hal yang mempengaruhi lapisan ozon, diantaranya yaitu suhu, kecepatan angin dan sebagainya. Dalam hal ini, banyak sekali data yang di dapat dari pengukuran variabel yang mempengaruhi lapisan ozon tersebut. Data mining klasifikasi bisa digunakan untuk solusi dimana bisa mengklasifikasikan lapisan ozon dalam keadaan berbahaya (ozon day) atau normal (normal day). Metode reduksi *Principal Component Analysis* (PCA) yang berfungsi untuk mereduksi atribut yang terlampaui banyak tanpa mengurangi tingkat informasi dari data aslinya. *Neural Network* digunakan untuk mengklasifikasikan data ozon untuk menentukan lapisan ozon dalam keadaan berbahaya (ozon day) atau normal (normal day). *Neural Network* juga diperlukan untuk mengetahui tingkat akurasi pada prediksi ozon. Hasil pengujian menunjukkan bahwa fitur reduksi *Principal Component Analysis* (PCA) yang dikombinasikan dengan *Neural Network* mendapatkan tingkat keakuratan pada prediksi sebesar 95%.

**Kata Kunci:** Ozon, Data Mining, PCA, Neural Network.

## **Abstract**

Ozone is a gas that is present in the earth's atmosphere naturally. It is composed of three oxygen molecules or commonly denoted  $O_3$ . The function of ozone is to protect the earth from ultraviolet radiation emitted by the sun. There are so many things that affect the ozone layer, such as temperature, wind speed and so on. In this case, a lot of data obtained from the measurement of the variables that affect of ozone layer. Data mining can be used for the solutions which can classify the ozone layer in a dangerous situation (ozone day) or normal (normal day). Feature of *Principal Component Analysis* (PCA) reduction, purpose to reduce attribute too much without reducing the level information of the original data. *Neural Network* are used to classify the data to determine ozone layer in a state of dangerous ozone (ozone day) or normal (normal day). *Neural Network* are also needed to determine the level of accuracy in the ozone prediction. Result of test indicate that the reduction features of *Principal Component Analysis* (PCA) combined with *Neural Network* to get the level of accuracy in the prediction of 95%.

**Keywords:** Ozone, Data Mining, PCA, Neural Network.

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Ozon atau yang dalam simbol kimia dinyatakan dengan  $O_3$  merupakan salah satu gas yang membentuk atmosfer bumi. Ozon berada di bawah lapisan stratosfer antara 15 dan 30 km di atas permukaan bumi yang dikenal sebagai lapisan ozon. Ozon di lapisan stratosfer sangat berguna untuk kehidupan makhluk hidup di bumi. Fungsi dari ozon di lapisan stratosfer yaitu menyerap sebagian besar radiasi sinar ultraviolet matahari yang sangat berbahaya bagi kehidupan di bumi apabila secara langsung terkena sinar ini. Ozon juga melindungi bumi dari batuan luar angkasa yang tertarik ke pusat bumi, menstabilkan suhu permukaan bumi sehingga mengurangi pemanasan global. Tingkat dari ozon, tergantung pada beberapa faktor yang terjadi di dalam bumi. Tingkat ozon dapat berbahaya bagi kesehatan manusia dan mempengaruhi kegiatan sehari - hari seperti dalam bidang pertanian, pariwisata dan lain sebagainya [1].

Pada awal penemuan ozon oleh seorang ilmuwan berkebangsaan Jerman yang bernama Christian Freiderich Schoubin, hanya terdapat pada lapisan stratosfer. Seiring dengan perkembangan ilmu pengetahuan dan teknologi, molekul – molekul ozon tidak hanya terdapat di lapisan stratosfer saja, tetapi telah memasuki lapisan troposfer. Masuknya molekul ozon atau  $O_3$  ke dalam lapisan troposfer dikarenakan adanya polusi udara (zat pencemar) yang berinteraksi dengan molekul  $O_3$  sehingga molekul  $O_3$  yang dibawa oleh dinamika atmosfer memasuki lapisan troposfer. Polusi udara (zat pencemar) berasal dari proses alami bumi sendiri maupun disengaja oleh perbuatan manusia yang secara sengaja membangun pabrik – pabrik dengan tidak memperhitungkan limbah

gas yang dihasilkannya [2].

Pada kondisi ozon yang tidak menentu akibat dari pengaruh polusi udara yang terjadi secara alami maupun hasil dari perbuatan manusia, maka diperlukan sistem peramalan peringkat ozon untuk menghasilkan ramalan peringkat ozon berbahaya atau normal kepada publik sebelum mencapai tingkat ozon yang membahayakan bagi manusia khususnya [1].

Pada akhir tahun 1980-an pada saat istilah *Data Mining* mulai digunakan, paling tidak dalam kalangan komunitas riset. Dalam arti luas *data mining* dapat didefinisikan sebagai sekumpulan mekanisme dan teknik yang direalisasikan dalam perangkat lunak untuk mengekstrak informasi tersembunyi dalam kumpulan data. Sebelum tahun 1990-an *data mining* umumnya dikenal sebagai sub proses dalam lingkup lebih besar yang disebut *Knowledge Discovery in Database* (KDD). Meskipun dalam konteks modern dari *data mining* KDD akan lebih sesuai, karena sumber pengetahuan bukan lagi terbatas pada *database*. Data yang digunakan dalam proses *data mining* pada awalnya hanya untuk data dalam bentuk tabel (relasional) mengingat keterbatasan kemampuan komputasi saat itu. Dengan peningkatan kemampuan komputasi, maka waktu komputasi tidak lagi menjadi persoalan utama dan digantikan dengan tujuan lain yakni akurasi dan keinginan untuk menambang data yang jauh lebih besar [14].

Dataset yang digunakan yaitu tentang peramalan “*ozon day*” atau hari ozon, maka didapatkan banyak sekali variabel yang mempengaruhi hari ozon. Variabel tersebut nantinya akan dijadikan label dua kelas, yaitu “*ozon day*” dan “*normal day*”. Dalam kondisi “*ozon day*” artinya, ozon dalam keadaan sedang dalam ambang berbahaya, karena lapisan ozon mulai menipis. Sedangkan “*normal day*”

artinya ozon dalam ambang batas yang tidak berbahaya atau normal [1].

*Principal Component Analysis* (PCA) merupakan teknik statistik yang sudah digunakan secara luas baik dalam hal pengolahan data, pembelajaran mesin, maupun pengolahan citra atau pemrosesan signal. Pada tahun 1901, Karl Pearson menemukan metode *Principal Component Analysis* (PCA) dan memakai metode ini pada bidang biologi. *Principal Component Analysis* (PCA) digunakan untuk mereduksi dimensi yang bertujuan meminimalkan kerugian dalam varian dalam data asli [3].

*Neural Network* (NN) merupakan model komputasi yang terinspirasi oleh prinsip-prinsip mengenai bagaimana cara otak manusia bekerja. *Neural network* dapat mempelajari dari data, mampu men-generalisasi dengan baik, dan tahan dengan kebisingan. Secara umum *Neural Network* (NN) adalah jaringan dari sekelompok unit pemroses kecil yang dimodelkan berdasarkan jaringan syaraf manusia. *Neural Network* merupakan sistem adaptif yang dapat merubah struktur untuk memecahkan masalah berdasarkan informasi eksternal maupun internal yang mengalir melalui jaringan tersebut [4].

## 1.2 Rumusan Masalah

Ozon mempunyai batas ambang yang berbahaya bagi manusia dan juga makhluk hidup yang ada di bumi [1]. Sehingga diperlukan metode yang dapat memprediksi kondisi ozon secara akurat.

## 1.3 Pertanyaan Penelitian

1. Apakah fitur reduksi *Principal Component Analysis* (PCA) berpengaruh pada akurasi dan waktu proses klasifikasi kondisi ozon.
2. Apakah metode PCA dengan *neural network* untuk prediksi

kondisi ozon memiliki akurasi terbaik dibandingkan dengan metode SVM, SVM – PCA, Naive Bayes, Naive Bayes – PCA, Decision tree, PCA - Decision tree

## 1.4 Tujuan Penelitian

1. Mengetahui pengaruh metode fitur reduksi *Principal Component Analysis* (PCA) dan metode klasifikasi *Neural Network* untuk mengklasifikasi kondisi ozon.
2. Mengevaluasi akurasi prediksi ozon menggunakan fitur reduksi *Principal Component Analysis* (PCA) dan metode klasifikasi *Neural Network*.

## 2. METODE PENELITIAN

Jenis penelitian dari proposal ini adalah kualitatif, yaitu penelitian dengan menggunakan data-set yang sudah ada, namun menggunakan metode yang berbeda dengan hasilnya yaitu akurasi dari prediksi ozon. Penelitian ini bertujuan untuk mengukur akurasi dari prediksi ozon dimana hasilnya berupa “ozon day” atau “normal day”. Secara umum proses penelitian ditunjukkan pada gambar sebagai berikut :



Gambar 1. Desain Penelitian

## 2.1 Pengumpulan Data

Data yang digunakan untuk analisis akurasi ozon berasal dari jurnal yang berjudul “*Forecasting Skewed Biased*

*Stochastic Ozone Days: Analyses and Solutions*” tentang peramalan hari ozon.

Label dalam data ini, dibagi menjadi dua kelas yaitu 0 (normal day) dan 1 (ozon day). Normal day di sini yaitu kondisi dimana ozon dalam keadaan normal, tidak berbahaya apabila beraktivitas di luar ruangan yang terkena sinar matahari secara langsung. Sedangkan 1 (ozon day) yaitu kondisi dimana ozon menjadi menipis dan mempunyai resiko bahaya apabila terkena sinar matahari secara langsung karena ozon menipis dan sinar UV tidak bisa tersaring dengan baik dan bisa membahayakan makhluk hidup khususnya manusia yang terpapar sinar matahari secara langsung. Metode analisis yang digunakan yaitu menggunakan reduksi *Principal Component Analysis* (PCA) dan menggunakan algoritma *neural network*. PCA di sini berfungsi untuk mereduksi jumlah variabel tanpa mengurangi tingkat informasinya. Taraf kesalahan indentifikasi menjadi lebih kecil dan periode latihan yang lebih pendek karena variabel yang digunakan sudah di reduksi atau sudah berkurang.

## 2.2 Preprocessing

Penarikan contoh (*sample*) yang dilakukan dalam data-set, masih terdapat *missing value* atau hilangnya nilai pada sebuah atribut yang ada. Penanganan *missing value* pada penelitian ini dilakukan dengan menggunakan *mean imputation*. *Mean imputation* adalah metode yang mengganti *missing value* pada atribut dengan nilai rata-rata yang diperoleh dari seluruh atribut yang diketahui nilainya [15].

Berikut merupakan rumus yang digunakan untuk *mean imputation* :

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad (2.1)$$

## 2.3 Fitur Reduksi

*Principal Component Analysis*

(PCA) merupakan teknik statistik yang sudah digunakan secara luas baik dalam hal pengolahan data, pembelajaran mesin, maupun pengolahan citra atau pemrosesan signal yang berguna untuk mereduksi jumlah variabel masukan tanpa mengurangi tanpa mengurangi tingkat informasinya. Pada tahun 1901, Karl Pearson menemukan metode *Principal Component Analysis* (PCA) dan memakai metode ini pada bidang biologi. Pada tahun 1947 teori ini ditemukan kembali oleh Karhunen, dan kemudian dikembangkan oleh Loeve pada tahun 1963, sehingga teori ini juga dinamakan Karhunen – Loeve transform pada bidang ilmu telekomunikasi. *Principal Component Analysis* (PCA) digunakan untuk mereduksi dimensi yang bertujuan meminimalkan kerugian dalam varian dalam data asli [5].

*Principal Component Analysis* (PCA) ini digunakan untuk mereduksi atribut tanpa mengurangi nilai informasi di dalamnya. Pada metode PCA hal yang pertama dilakukan pada data-set yaitu standarisasi / normalisasi data dengan mengurangkan masing-masing data dengan mean. Hasil dari normalisasi disebut dengan *DataAdjust*.

Persamaan yang digunakan untuk menghitung kovarian untuk suatu sampel berukuran n yang diambil dari suatu populasi berukuran n adalah :

$$\text{Cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)} \quad (2.2)$$

*Eigenvalue* didefinisikan sebagai nilai karakteristik suatu matrik. *Eigenvalue* didapatkan dari matrik kovarian dengan menggunakan rumus berikut :

$$\det(A - \lambda I) = 0 \quad (2.3)$$

*Eigenvalue* digunakan untuk mendapatkan *Eigenvector* dengan menggunakan rumus berikut :

$$\det(A - \lambda I) x = 0 \quad (2.4)$$

## 2.4 Klasifikasi

*Neural Network* (NN) merupakan model komputasi yang terinspirasi oleh prinsip-prinsip mengenai bagaimana

cara otak manusia bekerja. *Neural network* dapat mempelajari dari data, mampu men-generalisasi dengan baik, dan tahan dengan kebisingan [5]. Secara umum *Neural Network* (NN) adalah jaringan dari sekelompok unit pemroses kecil yang dimodelkan berdasarkan jaringan syaraf manusia. *Neural Network* merupakan sistem adaptif yang dapat merubah struktur untuk memecahkan masalah berdasarkan informasi eksternal maupun internal yang mengalir melalui jaringan tersebut.

Neural network merupakan salah satu metode klasifikasi yang bisa digunakan untuk menyelesaikan klasifikasi dari suatu dataset. Model yang digunakan dalam penelitian ini yaitu model jaringan *backpropagation*. Algoritma pelatihan *Backpropagation* yaitu sebagai berikut :

1. Inisialisasi bobot (ambil dengan nilai random yang cukup kecil).
2. Selama kondisi berhenti bernilai salah, maka di kerjakan :

Tahap perambatan maju (*forward propagation*)

Setiap unit input ( $X_i$ ,  $i = 1,2,3,\dots,n$ ) menerima sinyal  $x_i$  dan meneruskan sinyal tersebut ke semua unit pada lapisan tersembunyi.

Setiap unit tersembunyi ( $Z_j$ ,  $j = 1,2,3,\dots,p$ ) menjumlahkan bobot sinyal input dengan persamaan berikut :

$$z_{inj} = v_{0j} + \sum_{i=1}^n x_i v_{ij} \quad (2.5)$$

Dan menerapkan fungsi aktivasi untuk menghitung sinyal output-nya :

$$z_j = f(z_{inj}) \quad (2.6)$$

Tahap perambatan Balik (*Backpropagation*)

Setiap unit output ( $Y_k$ ,  $k = 1,2,3,\dots,m$ ) menerima pola target yang sesuai dengan pola input pelatihan, kemudian hitung error dengan persamaan berikut :

$$\delta_k = (t_k - y_k) f'(y_{ink}) \quad (2.7)$$

$f'$  adalah turunan dari fungsi aktivasi kemudian hitung koreksi bobot dengan persamaan berikut :

$$\Delta w_{jk} = \alpha \delta_k z_j \quad (2.8)$$

Dan menghitung koreksi bias dengan persamaan berikut :

$$\Delta w_{0k} = \alpha \delta_k \quad (2.9)$$

Sekaligus mengirimkan  $\delta_k$  ke unit-unit yang ada di lapisan paling kanan.

Setiap unit tersembunyi ( $Z_j$ ,  $j = 1,2,3,\dots,p$ ) menjumlahkan delta inputnya (dari-unit-unit yang berada pada lapisan di kanannya) :

$$\delta in_j = \sum_{k=1}^m \delta_k w_{jk} \quad (2.10)$$

Untuk menghitung informasi error, kalikan nilai ini dengan turunan dari fungsi aktivasinya :

$$\delta_j = \delta in_j f'(z_{int_j}) \quad (2.11)$$

Kemudian hitung koreksi bobot dengan persamaan berikut :

$$\Delta v_{jk} = \alpha \delta_j x_i \quad (2.12)$$

Setelah itu, hitung juga koreksi bias dengan persamaan berikut :

$$\Delta v_{0j} = \alpha \delta_j \quad (2.13)$$

Tahap Perubahan Bobot dan Bias

Setiap unit output ( $Y_k$ ,  $k = 1,2,3,\dots,m$ ) dilakukan perubahan bobot dan bias ( $j = 0,1,2,\dots,p$ ) dengan persamaan berikut

$$w_{jk}(\text{baru}) = w_{jk}(\text{lama}) + \Delta w_{jk} \quad (2.14)$$

Setiap unit tersembunyi ( $z_j$ ,  $j = 1,2,3,\dots,p$ ) dilakukan perubahan bobot dan bias

$$v_{ij} = v_{ij}(\text{lama}) + \Delta v_{ij} \quad (2.15)$$

## 2.5 Pengujian

Pengujian data tanpa menggunakan PCA (non-PCA), adalah dengan :

Data dengan 72 atribut di olah dengan menggunakan metode *Neural Network*, SVM, dan Naive Bayes.

Pengujian data menggunakan fitur reduksi PCA, adalah sebagai berikut :

Data dengan 72 atribut, di reduksi menggunakan fitur reduksi PCA menjadi 5, 10, dan 15 atribut.

Data yang sudah di reduksi dengan fitur reduksi PCA, diolah dengan metode klasifikasi *Neural Network*, SVM, dan Naive Bayes.

## 2.6 Evaluasi

*Confusion matrix* adalah parameter baik atau buruknya sebuah

pengklasifikasian atas record dalam kelas yang berbeda. Hasil dari *confusion matrix* yaitu *accuracy*. *Accuracy* adalah presentase ketepatan dalam hal pengklasifikasian record. Berikut rumus perhitungan dari bagian-bagian *confusion matrix* [15]:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (2.1)$$

### 3. HASIL DAN PEMBAHASAN

Dengan menggunakan rapidminer, maka di dapatkan akurasi prediksi ozon sebagai berikut :

**Tabel 4.11.** Akurasi, Waktu Proses dengan PCA 5 atribut

Metode	Akurasi	Waktu
Neural Network	95,03%	9 detik
SVM	95,03%	2 detik
Naive Bayes	94,75%	1 detik
Decision Tree	95,03%	1 detik

**Tabel 4.11.** Akurasi, Waktu Proses dengan PCA 10 atribut

Metode	Akurasi	Waktu
Neural Network	93,51%	17 detik
SVM	95,03%	2 detik
Naive Bayes	91,85%	1 detik
Decision Tree	94,89%	1 detik

**Tabel 4.11.** Akurasi, Waktu Proses dengan PCA 15 atribut

Metode	Akurasi	Waktu
Neural Network	94,06%	24 detik
SVM	95,03%	2 detik
Naive Bayes	90,47%	1 detik
Decision Tree	94,75%	1 detik

**Tabel 4.11.** Akurasi, Waktu Proses dengan PCA 10 atribut

Metode	Akurasi	Waktu
Neural Network	93,51%	17 detik
SVM	95,03%	2 detik
Naive Bayes	91,85%	1 detik
Decision Tree	94,89%	1 detik

**Tabel 4.11.** Akurasi, Waktu Proses non-PCA

Metode	Akurasi	Waktu
Neural Network	94,61%	4 menit 2 detik
SVM	95,03%	3 detik
Naive Bayes	74,45%	1 detik
Decision Tree	93,09%	8 detik

## 4. KESIMPULAN DAN SARAN

### 4.1 Kesimpulan

1. Fitur reduksi *Principal Component Analysis* (PCA) pada proses klasifikasi dengan metode *Neural Network* berpengaruh dan memperbaiki akurasi maupun waktu proses.
2. Akurasi klasifikasi menggunakan fitur reduksi *Principal Component Analysis* (PCA) dengan metode *Neural Network* mencapai 95,03% yang sama baiknya dengan metode SVM dan Decision Tree.

### 4.2 Saran

1. Penelitian dapat dilanjutkan dengan menggunakan fitur reduksi yang lain, selain *Principal Component Analysis* (PCA).
2. Penelitian dapat dilanjutkan menggunakan metode klasifikasi lain.

## 5. DAFTAR PUSTAKA

- [1] K. Zhang, W. Fan, X. Yuan, I. Davidson and X. Li, "Forecasting Skewed Biased Stochastic Ozone Days: Analyses and Solutions," vol. 14, no. 3, 2008.
- [2] S. Lilik, "Pemanfaatan Potensi Ozon di Indonesia," 2002.
- [3] Y. Puspitaningrum and S. Syidada, "Reduksi Fitur untuk Kategorisasi Text dengan Klasifikasi Menggunakan Neural Network," 2009.

- [4] D. Aprilla, A. Baskoro and e. al, "Belajar Data Mining dengan Rapid Miner," 2013.
- [5] M. Rivai, "Pengaruh Principle Component Analysis Terhadap Tingkat Identifikasi Neural Network pada Sistem Sensor Gas," pp. 159-167, 2007.
- [6] A. Wahab and A. Alawi, "Assessment and prediction of tropospheric ozone concentration," pp. 219-228, 2002.
- [7] S. Al-Alawi, S. Abdul-Wahab and e. al, "Combining principal component regression and artificial neural," pp. 396-403, 2008.
- [8] H. Zhang, A. Palazoglu and e. al, "Prediction of surface ozone exceedance days using PCA with a," pp. 42-48, 2014.
- [9] F. Harrou, M. Nounou and e. al, "Detecting Abnormal Ozone Levels using," 2013.
- [10] H.-C. Lu and e. al, "Prediction of daily maximum ozone concentrations," pp. 124-139, 2006.
- [11] A. Luna, Paredes and e. al, "Prediction of ozone concentration in tropospheric levels using," pp. 98-104, 2014.
- [12] S. Sousa, Martins and e. al, "Multiple linear regression and artificial neural networks based," pp. 97-103, 2007.
- [13] C.-j. Tsai, L.-c. Chang and e. al, "Forecasting of ozone episode days by cost-sensitive neural," 2009.
- [14] A. Hermawati, Data Mining, Yogyakarta: CV ANDI OFFSET, 2009, pp. 2124-2135.
- [15] R. Purba, "Data Mining : Masa Lalu, Sekarang dan Masa Mendatang," vol. XIII, pp. 98-104, 2012.
- [16] T. Sutojo, E. Mulyanto and V. Suhartono, Kecerdasan Buatan, Yogyakarta: CV ANDI OFFSET, 2011.