

**IMPLEMENTASI ALGORITMA *BAYESIAN*
NETWORK DALAM MEMPREDIKSI INDEKS
KUALITAS UDARA**



OLEH:

**ADITIYA MUAFFAN
09011381520071**

**JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2020**

LEMBAR PENGESAHAN

IMPLEMENTASI ALGORITMA BAYESIAN NETWORK DALAM MEMPREDIKSI INDEKS KUALITAS UDARA

TUGAS AKHIR



Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

ADITIYA MUAFFAN
09011381520071

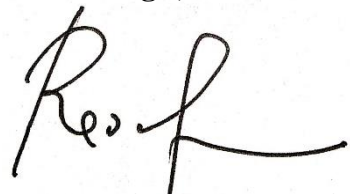
Palembang, Desember 2020

Mengetahui,
Ketua Jurusan Sistem Komputer



Dr. Ir. H. Sukemi, M.T.
NIP. 196612032006041001

Pembimbing I,



Dr. Reza Firsandaya Malik, M.T.
NIP. 197604252010121001

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada:


Hari : Kamis
Tanggal : 31 Desember 2020

Tim Penguji :

1. Ketua : Huda Ubaya, M.T.
2. Sekertaris : Aditya Putra Perdana P, M.T.
3. Pembimbing : Dr. Reza Firsandaya Malik, M.T.
4. Penguji : Kemahyanto Exaudi, M.T.



Mengetahui,
Ketua Jurusan Sistem Komputer


Dr. Ir. H. Sukemi, M.T.
NIP. 196612032006041001

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Aditiya Muaffan
NIM : 09011381520071
Judul : Implementasi Algoritma Bayesian Network Dalam
Memprediksi Indeks Kualitas Udara
Hasil Pengecekan *Software*
(*iThenticated/Turnitin*) : 11%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan / *plagiat*. Apabila ditemukan unsur penjiplakan / *plagiat* dalam laporan tugas akhir ini, maka saya bersedia menerima saksi akademik dari universitas sriwijaya.

Demikian, pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan.



Palembang, Desember 2020



Aditiya Muaffan
NIM. 09011381520071

HALAMAN PERSEMBAHAN

Everything in life happens according to our time, our clock.

You may look at some of your friends and think that they're ahead of you, maybe some of them you feel are behind, but everything happens at their own pace.

They have their own time and clock and so do you.

Be patient.

Don't let anyone rush you with their timelines. Because as Einstein Said,

*“Not Everything that counts can be counted
and not everything that's counted truly counts”*

And the most important thing is to create meaningful, purposeful fullfilling lives for yourselves and learn how to use that to make an impact and difference in the lives of others.

That will be true success.

(Everyone's Timing Different – Jay Shetty)

Tugas Akhir ini saya persembahkan kepada :

- **Kedua orang tua saya.**
- **Keluarga kecil dan besar saya.**
- **Teman dan sahabat yang telah memberikan *support*.**
- **Keluarga besar Sistem Komputer 2015.**
- **Fakultas Ilmu Komputer**
- **Universitas Sriwijaya**

KATA PENGANTAR

Puji syukur penulis penjatkan kepada Allah SWT atas limpahan berkah, rahmat, taufik, dan hidayah-nya sehingga penulis dapat menyelesaikan tugas akhir ini dan menyusun laporan tugas akhir yang berjudul “**Implementasi Algoritma Bayesian Network Dalam Memprediksi Indeks Kualitas Udara**”, dibuat dalam rangka untuk memenuhi persyaratan menyelesaikan pendidikan di jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.

Dalam penulisan Tugas Akhir ini penulis menyadari bahwa penulis banyak sekali mendapat dukungan dari berbagai pihak. Oleh karena itu dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Allah Subhanahu wa Ta'ala karena berkat rahmat-nya penulis dapat menyelesaikan tugas akhir ini.
2. Nabi Muhammad Shallallahu Alaihi Wassalam serta seluruh pengikutnya hingga pada akhir jaman.
3. Terima kasih kepada orang tua, adik-adik, keluarga besar atas doa dan dukungan yang tiada henti kepada penulis, serta memberikan pertolongan baik moril maupun materil.
4. Terima kasih banyak kepada dosen pembimbing saya yaitu, bapak Dr. Reza Firsandaya Malik, M.T. yang telah memberikan bimbingan dari awal hingga akhir dalam menyelesaikan tugas akhir ini.
5. Terima kasih kepada bapak Huda Ubaya, M.T. selaku ketua penguji dalam pengujian tugas akhir penulis, telah menyempatkan waktunya.
6. Terima kasih kepada bapak Aditya Putra Perdana P, M.T. dan bapak Kemahyanto Exaudi, M.T. atas bantuan dukungan dan dorongan untuk penulis dapat menyelesaikan tugas akhir ini.
7. Terima kasih untuk teman-teman keluarga besar Sistem Komputer, terutama untuk teman sekelas angkatan 2015 dan adik-adik kelas angkatan 2016 dan 2017 yang telah membantu penulis dalam penyelesaian tugas akhir ini.

8. Terima kasih kepada Mbak Renny Virgasari selaku admin jurusan Sistem Komputer Unggulan yang telah memberikan bantuan dalam pengurusan berkas-berkas selama masa kuliah.

Penulis menyadari bahwa baik isi maupun penyajian laporan ini masih jauh dari kesempurnaan, oleh karena itu penulis mengharapkan saran dan kritik yang sifatnya membangun untuk perbaikan laporan ini. Penulis berharap semoga laporan tugas akhir ini dapat menambah pengetahuan serta dapat menunjang perkembangan ilmu pengetahuan, khususnya bagi penulis maupun pembaca.

Palembang, Desember 2020

Penulis

BAYESIAN NETWORK ALGORITHM IMPLEMENTATION IN PREDICTION AIR QUALITY INDEX

Aditiya Muaffan (09011381520071)

Department of Computer Engineering, Faculty of Computer Science
University Sriwijaya
Email: aditiya.m05@gmail.com

Abstract

This research was conducted to be able to predict the level of air pollution by using parameters of air pollution carbon monoxide (CO), sulfur dioxide (SO₂), nitrogen dioxide (NO₂), ozone (O₃), particulate matter (PM_{2.5}) by implementing the Bayesian Network algorithm as algorithm in classifying the independent of each air pollution variable. In addition, this study also calculates the level of accuracy of the prediction results by comparing actual data with testing data. The data used in this study is data downloaded from the United States Environmental Protection Agency (EPA) website, with 365 records of data in the form of an index of the value of air pollution quality in the city of California in January-December 2019 which is used for the training process and testing process in the research. this. The distribution for testing data is 304 data records and training data is 61 data records. Tests were carried out using Jupyter Notebook software and WEKA software ver.3.8.4. The results of prediction testing with the Bayesian Network algorithm obtained in the Jupyter Notebook software are 90.14 and the results on WEKA software ver.3.8.4 are 98.63%, 99.08%, 98.63, and 100%.

Kata Kunci: Wireless Sensor Network, Bayesian Network, Prediction, Air Pollution, Air Pollution Standard Index, Accuracy

IMPLEMENTASI ALGORITMA BAYESIAN NETWORK DALAM MEMPREDIKSI INDEKS KUALITAS UDARA

Aditiya Muaffan (09011381520071)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer

Universitas Sriwijaya

Email: aditiya.m05@gmail.com

Abstrak

Penelitian ini dilakukan untuk dapat memprediksi tingkat polusi udara dengan menggunakan parameter polusi udara Karbon Monoksida (CO), Sulfur Dioksida (SO₂), Nitrogen Dioksida (NO₂), Ozon (O³), Partikulat Material (PM_{2.5}) dengan mengimplementasikan algoritma *Bayesian Network* sebagai algoritma dalam mengklasifikasi independen setiap variabel polusi udara. Selain itu penelitian ini juga menghitung tingkat akurasi hasil prediksi dengan membandingkan data aktual dengan data *testing*. Data yang digunakan dalam penelitian ini adalah data yang diunduh dari situs *United States Environmental Protection Agency* (EPA), sebanyak 365 *record* data berupa indeks nilai kualitas polusi udara kota California pada bulan Januari-Desember tahun 2019 yang digunakan untuk proses *training* dan proses *testing* pada penelitian ini. Pembagian untuk data *testing* adalah 304 *record* data dan data *training* adalah 61 *record* data. Pengujian dilakukan menggunakan perangkat lunak Jupyter Notebook dan perangkat lunak WEKA ver.3.8.4. Hasil pengujian prediksi dengan algoritma *Bayesian Network* yang didapat pada perangkat lunak Jupyter Notebook adalah 90.14 dan hasil pada perangkat lunak WEKA ver.3.8.4 adalah 98,63%, 99,08%, 98,63, dan 100%.

Kata Kunci: Jaringan Sensor Nirkabel, *Bayesian Network*, Prediksi, Polusi Udara, Indeks Standar Pencemaran Udara, Akurasi

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PERNYATAAN	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
ABSTRACT	viii
ABSTRAK	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN	xv
BAB I. PENDAHULUAN	
1.1. Latar Belakang	1
1.2. Tujuan dan Manfaat	2
1.3. Rumusan Masalah	2
1.4. Batasan Masalah	2
1.5. Metodologi Penelitian	3
1.6. Sistematika Penulisan	4
BAB II. TINJAUAN PUSTAKA	
2.1. Pencemaran Udara	5
2.2. Indeks Standar Pencemaran Udara (ISPU)	6
2.3. <i>United Environment Protection Agency</i> (EPA)	8
2.4. Data Mining	10
2.4.1. Fungsi Data Mining	11
2.4.2. Proses Data Mining	11

2.4.3. Tahapan Data Mining	13
2.5. Klasifikasi	14
2.6. Pengukuran Kinerja Klasifikasi	15
2.7. Jaringan Sensor Network (JSN)	16
2.8. Teori Bayes	17
2.9. <i>Bayesian Networks</i>	18
2.10. <i>Multinomial Bayesian Networks</i>	18
2.11. <i>Gaussian Bayesian Networks</i>	19
2.12. Struktur <i>Bayesian Networks</i>	20
BAB III. METODOLOGI PENELITIAN	
3.1. Pendahuluan	22
3.2. Kerangka Kerja	22
3.3. Studi Pustaka dan Literatur	24
3.4. Pengambilan Data	24
3.5. Penentuan Parameter Pengujian	24
3.6. Pengolahan Data Awal	24
3.7. Penerapan Algoritma <i>Bayesian Networks</i>	27
3.7.1. Praproses Data	27
3.7.2. Struktur <i>Bayesian Network</i>	28
3.7.3. Inferensi <i>Bayesian Network</i>	28
BAB IV. PENGUJIAN DAN ANALISA	
4.1. Pendahuluan	30
4.2. <i>Pre-Processing Data</i>	30
4.3. Proses <i>Learning Data</i>	32
4.4. Proses <i>Testing Data</i>	33
4.5. Visualisasi DAG	39
4.6. Pengujian Lain	40
BAB V. KESIMPULAN DAN SARAN	
5.1. Kesimpulan	44

5.2. Saran	44
DAFTAR PUSTAKA	45
LAMPIRAN	

DAFTAR GAMBAR

	Halaman
Gambar 2.1. Siklus pencemaran udara	6
Gambar 2.2. Proses pengiriman data hasil pemantauan	9
Gambar 2.3. Tahapan data mining	13
Gambar 2.4. Topologi JSN	17
Gambar 2.5. <i>Directed Acyclic Graph</i> (DAG)	21
Gambar 3.1. <i>Flowchart</i> penelitian	23
Gambar 3.2. Data mentah atau data awal	25
Gambar 3.3. <i>Missing value</i> pada data	26
Gambar 3.4. Dataset	26
Gambar 3.5. <i>Flowchart</i> metode <i>Bayesian Network</i>	27
Gambar 3.6. Inferensi <i>Bayesian Network</i>	28
Gambar 4.1. Contoh <i>missing value</i> pada data	31
Gambar 4.2. Cek info dataset	31
Gambar 4.3. Pembagian data <i>training</i> dan data <i>testing</i>	32
Gambar 4.4. Proses <i>learning</i> data	32
Gambar 4.5. Hasil Prediksi	37
Gambar 4.6. Diagram DAG	39
Gambar 4.7. Hasil pengujian dengan percentage split 60%	40
Gambar 4.8. Hasil pengujian dengan percentage split 70%	41
Gambar 4.9. Hasil pengujian dengan percentage split 80%	42
Gambar 4.10. Hasil pengujian dengan percentage split 90%	43

DAFTAR TABEL

	Halaman
Tabel 2.1. Rentang Indeks Standar Pencemaran Udara	7
Tabel 2.2. Batas Indeks Standar Pencemaran Udara Dalam Satuan SI	8
Tabel 2.3. Matriks Konfusi Untuk Klasifikasi dua Kelas	16
Tabel 4.1. Data <i>Testing</i>	33
Tabel 4.2. Dataset Proses <i>Training</i>	35
Tabel 4.3. Data Perbandingan Hasil Prediksi	37
Tabel 4.4. Hasil Kinerja Algoritma <i>Bayesian Network</i>	43

DAFTAR LAMPIRAN

Lampiran 1. Data ISPU Bulan Januari-Desember 2019

Lampiran 2. Data *Training*

Lampiran 3. Data *Testing*

Lampiran 4. Berkas-Berkas Persyaratan

BAB I

PENDAHULUAN

1.1. Latar Belakang

Udara merupakan komponen penting yang dibutuhkan manusia dalam proses transpirasi. Meningkatnya pembangunan fisik kota dan pusat-pusat industri, komponen udara tersebut telah mengalami perubahan. Perubahan komponen udara tersebut berpengaruh pada perubahan kualitas udara dan berakibat pada pencemaran udara. Penurunan kualitas udara ini dapat mengganggu kesehatan masyarakat disekitarnya [1].

Beberapa penyebab polusi antara lain reaksi kimia limbah, sisa pembuangan kendaraan bermotor, kebakaran hutan, gunung meletus, sisa pembakaran dari pabrik dan mesin-mesin industri.

Di Indonesia ada lima parameter utama yang menjadi standar gas-gas penyebab pencemaran udara berupa Karbon Monoksida (CO), Sulfur Dioksida (SO₂), Nitrogen Dioksida (NO₂), Ozon (O³), Tingkat Partikulat (PM₁₀). Kelima parameter ini dituangkan sebagai Indeks Standar Pencemar Udara yang selanjutnya disebut ISPU [2]. Selanjutnya informasi tentang kualitas udara ini disampaikan kepada masyarakat dalam bentuk grafik warna ISPU.

Gas-gas polutan adalah gas yang tidak berbau serta tidak bisa dilihat, sehingga dibutuhkan suatu perangkat untuk mengukur dan memprediksi indeks kualitas udara sebagai pencegahan dini terhadap dampak berbahaya dari gas-gas polutan yang dapat terhirup dan membahayakan kesehatan [3].

Corani G & Scanagatta M [4] dalam “*Air Pollution Prediction via Multi-Label Classification*” mencoba menerapkan algoritma *Bayesian Network Classifier* untuk dapat mengklasifikasikan independen variabel polusi udara, kemudian dengan pemodelan dependensi dapat melakukan prakiraan/prediksi variabel polusi udara tersebut. Dalam paper tersebut variabel polusi udara yang digunakan dalam

percobaan adalah polusi udara PM-2.5 dengan menggunakan dataset dari Shanghai dan variabel polusi udara Ozon (O_3) dengan menggunakan dataset Berlin.

Merujuk pada penelitian Corani G & Scanagatta M [4], akan dilakukan analisis prediksi kualitas udara dari data pengukuran di Jaringan Sensor Nirkabel (JSN) dengan menggunakan dataset negara lain dan variabel yang di prediksi adalah Indeks Kualitas Udara.

1.2. Tujuan dan Manfaat

Tujuan dilakukannya penelitian ini adalah sebagai berikut :

1. Menerapkan metode *Bayesian Network* untuk dapat memprediksi indeks kualitas udara di hari berikutnya.
2. Menghitung tingkat akurasi metode *Bayesian Network* dengan data actual.

Adapun manfaat dari dilakukannya penelitian ini, sebagai berikut :

1. Membantu untuk dapat memprediksi indeks kualitas udara di hari berikutnya.
2. Menjadi referensi dalam pengembangan teknologi WSN selanjutnya.

1.3. Rumusan Masalah

Perancangan dan Analisa system prediksi kualitas udara berbasis *Bayesian Network* ini berdasarkan permasalahan sebagai berikut :

1. Bagaimana mengimplementasikan algoritma *Bayesian Network* dalam memprediksi kualitas udara ?
2. Berapa besar akurasi penerapan algoritma *Bayesian Network* dalam memprediksi kualitas udara ?

1.4. Batasan Masalah

Agar tujuan dari penelitian ini dapat tercapai, maka permasalahan tugas akhir ini terbatas pada ruang lingkup sebagai berikut :

1. Parameter yang akan digunakan dalam penelitian adalah CO, SO₂, NO₂, O₃, PM_{2.5}, dan ISPU.

2. Data yang digunakan dalam penelitian adalah data publik yang diunduh melalui situs <https://www.epa.gov/outdoor-air-quality-data>, data yang digunakan adalah bulan januari-desember 2019.

1.5. Metodologi Penelitian

Adapun metodologi yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Tahap Pertama (Studi Pustaka/Literatur dan Konsultasi)
Studi literatur merupakan dasar dari sebuah penelitian, yaitu dengan cara mencari, mengumpulkan, dan mempelajari informasi dari berbagai literatur atau sumber sebagai penunjang tugas akhir. Selain itu, untuk menunjang tugas akhir ini dilakukan konsultasi dengan orang-orang yang berkompetensi di bidang WSN dan pemograman dalam pengolahan data.
2. Tahap Kedua (Perancangan Sistem)
Pada tahap ini dilakukan perancangan algoritma *Bayesian Network* pada simulasi prediksi kualitas udara, dan menentukan parameter yang akan digunakan dalam mengukur prediksi kualitas udara.
3. Tahap Ketiga (Pengujian Eksperimen)
Pada tahap ini dilakukan pembuatan rangkaian system prediksi kualitas udara dengan parameter CO, SO₂, NO₂, O₃, PM_{2.5}, dan ISPU. Kemudian dilakukan pengujian untuk mengetahui bagaimana algoritma *Bayesian Network* dapat memprediksi indeks kualitas udara di hari berikutnya.
4. Tahap Keempat (Analisis Sistem)
Pada tahap ini merupakan hasil dari pengujian pada tahap-tahap sebelumnya yang kemudian dianalisis dengan tujuan mengetahui kekurangan dari hasil perancangan dan menentukan factor penyebabnya sehingga dapat dilakukan perbaikan pada penelitian selanjutnya.
5. Tahap Kelima (Kesimpulan)
6. Pada tahap ini menarik kesimpulan setelah proses analisa dan pengolahan data dilakukan.

1.6. Sistematika Penulisan

Agar lebih memudahkan dalam menyusun tugas akhir dan memperjelas isi dari setiap bab yang terdapat pada laporan ini, maka dibuatlah sistematika penulisan sebagai berikut ini.

BAB I PENDAHULUAN

Bab ini berisi penjabaran secara sistematis topik yang di ambil berupa latar belakang masalah, tujuan, manfaat, rumusan dan batasan masalah.

BAB II TINJAUAN PUSTAKA

Bab ini berisi kerangka teori dan kerangka berpikir terkait perancangan sistem dan metode yang digunakan dalam menyelesaikan tugas akhir.

BAB III METODOLOGI

Bab ini menjelaskan secara bertahap dan terperinci tentang langkah-langkah yang digunakan untuk mencari, mengumpulkan dan menganalisis tema dalam penulisan akhir.

BAB IV HASIL DAN ANALISA

Bab ini menjelaskan mengenai hasil pengujian yang telah dilakukan dan analisis terhadap hasil perancangan yang telah dibuat.

BAB V KESIMPULAN

Bab ini berisi kesimpulan tentang apa yang akan di peroleh oleh penulis serta merupakan jawaban dari setiap tujuan yang ingin dicapai pada bab I (Pendahuluan).

DAFTAR PUTSAKA

- [1] Santi, P, “**Analisis Kualitas Udara Stasiun Global Atmosphere Watch (GAW) Bukit Kototabang Kabupaten Agam Sumatera Barat**”, Skripsi. Universitas Gajah Mada. 2012.
- [2] Kementerian Lingkungan Hidup Republik Indonesia, "**Keputusan Menteri Negara Lingkungan Hidup No.45 Tahun 1997 Tentang Indeks Standar Pencemaran Udara**", no. 45. 1997.
- [3] United States Environmental Protection Agency , “**Air Quality Index**”,p.1-11. 2014.
- [4] Corani G, Scanagatta M, “**Air Pollution Prediction via Multi-Label Classification**”, Environmental Modelling & Software, 2016. 80: 259-264.
- [5] Turban, E, “**Decision Support Systems and Intelligent Systems Edisi Bahasa Indonesia Jilid 1**”. Andi: Yogyakarta. 2005.
- [6] Larose, Daniel T, “**Discovering Knowledge in Data : An Introduction to Data Mining**”. John Willey & Sons, Inc. 2005.
- [7] Ayyad, Usama, “**Advances in Knowledge Discovery and Data Mining**”, MIT Press. 1996.
- [8] A. Y. T. Leung, H. Y. Fan, J. C. K. Wong, and W.Z. Lu, “**Analysis of Pollutant Levels in Central Hongkong Applying Neural Network with Particle Swarm Optimization**”, Environ. Monit. Assess. 79, vol. 79, pp. 217–230, 2001.
- [9] G. Swagarya, R. S. Sinde, and S. Kaijage, “**a Survey on Wireless Sensor Networks Application for Air Pollution Monitoring**”, Int. J. Eng. Comput. Sci., vol. 3, no. 5 May, pp. 5975–5979, 2014.

- [10] Yahdin, Sugandi, dkk, “**Aplikasi Pengambilan Keputusan Pada Perencanaan Produksi Berdasarkan Teorema Bayes**”, Media Informatika, No.1, Hal. 25-38, Juni 2008, Palembang.
- [11] Roselina, Yunita Nancy, dkk, “**Aplikasi Diagnosa Penyakit Asma Menggunakan Bayesian Network Berbasis Web**”, Jurnal Teknik Informatika, Vol. 1, September 2012, Riau.
- [12] Cooper GF, Herskovits E, “**A Bayesian Method for the Induction of Probabilistic Networks From Data**” Machine Learning Journal 9:308-347. 1992.
- [13] Cano R, Sordo C, Guitierrez JM, “**Application of Bayesian Network in Meteorology**” Springer 309-327. 2004.
- [14] Han, J., Kamber, M. & Pei, J., “**Data Mining : Concept and Techniques**” 3rd ed. USA: ELSEVIER. 2012.
- [15] Prasetyo, E., “**Data Mining : Konsep dan Aplikasi Menggunakan Matlab**”, Yogyakarta: C.V Andi Offset. 2012.
- [16] Yang R, Yan F, Zhao N, “**Urban Air Quality Based on Bayesian Network**”, International Conference on Communication Software and Networks (ICCSN), pp. 1003-1006. 2017.