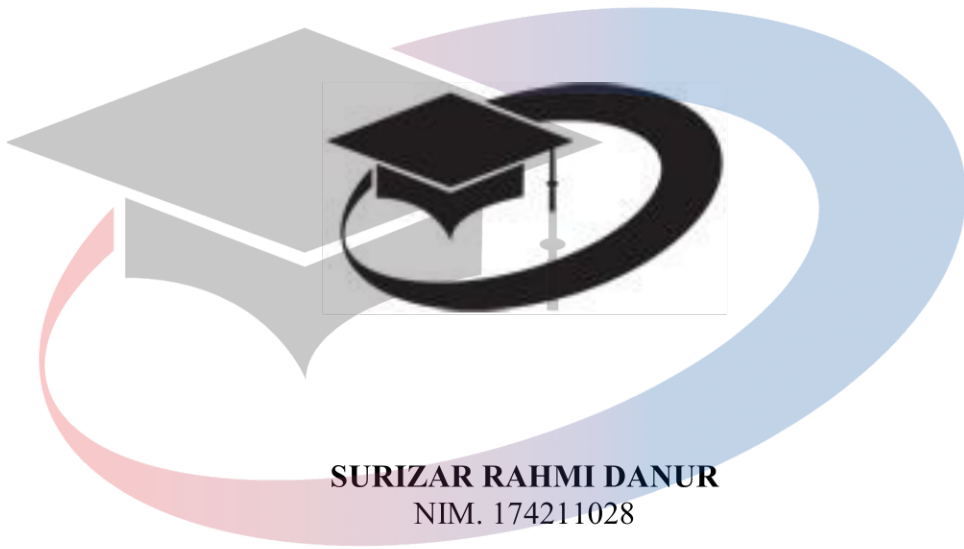


**OPTIMASI KLASIFIKASI BAYESIAN NETWORK MELALUI
REDUKSI ATTRIBUTE MENGGUNAKAN METODE
PRINCIPAL COMPONENT ANALYSIS**

TESIS



SURIZAR RAHMI DANUR
NIM. 174211028

**UNIVERSITAS
MIKROSKIL**

**PROGRAM STUDI MAGISTER TEKNOLOGI INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
MIKROSKIL
MEDAN
2020**

**OPTIMIZATION OF BAYESIAN NETWORK THROUGH
CLASSIFICATION ATTRIBUTE REDUCTION USING
PRINCIPAL COMPONENT ANALYSIS**

THESIS

By :

Surizar Rahmi Danur
ID. 174211028



**UNIVERSITAS
MIKROSKIL**

**INFORMATION TECHNOLOGY MASTER'S PROGRAM
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER**

MIKROSKIL

MEDAN

2020

LEMBARAN PENGESAHAN

**OPTIMASI KLASIFIKASI BAYESIAN NETWORK MELALUI REDUKSI
ATTRIBUTE MENGGUNAKAN METODE PRINCIPAL COMPONENT
ANALYSIS**

TESIS

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Magister
Program Studi Magister Teknologi Informasi

Oleh :

SURIZAR RAHMI DANUR

NIM: 174211028


Disetujui Oleh:

Dosen Pembimbing I,


Dr. Pahala Sirait, S.T., M.Kom.

Medan, 30 Januari 2020
Diketahui dan Disahkan Oleh:

Ketua Program Studi
Magister Teknologi Informasi,


Dr. Ronsen Purba, M.Sc.

UNIVERSITAS
MIKROSKIL

KATA PENGANTAR

Segala puji syukur penulis ucapkan kepada Allah SWT atas berkat rahmat dan karunia-Nya sehingga penelitian yang berjudul “OPTIMASI KLASIFIKASI BAYESIAN NETWORK MELALUI REDUKSI ATRIBUT MENGGUNAKAN PRINCIPAL COMPONENT ANALYSIS” dapat diselesaikan dengan baik. Tesis ini merupakan salah satu syarat untuk mendapatkan gelar Magister Teknologi Informasi (M.T.I) pada program studi magister Sekolah Tinggi Manajemen Informatika dan Komputer Mikroskil Medan.

Dalam penulisan tesis ini penulis banyak menerima dukungan, nasihat, bimbingan dan dorongan semangat dari berbagai pihak. Maka dengan kerendahan hati penulis mengucapkan terima kasih kepada:

1. Bapak Dr. Pahala Sirait, S.T., M.Kom. selaku Ketua Sekolah Tinggi Manajemen Informatika dan Komputer Mikroskil Medan dan selaku Dosen Pembimbing I penulis, atas arahan, bimbingan dan motivasi yang diberikan penulis dapat menyelesaikan laporan tesis ini dengan baik.
2. Bapak Ir. Erwin Setiawan Panjaitan, MMSI., Ph.D. selaku Dosen Pendamping yang telah memberikan arahan dan masukan kepada penulis.
3. Bapak Dr. Ronsen Purba, M.Sc. selaku Ketua Program Studi Magister Teknologi Informasi Sekolah Tinggi Manajemen Informatika dan Komputer Mikroskil Medan atas arahan, dan motivasi dalam penyelesaian laporan tesis ini.
4. Seluruh Dosen dan Staf Pegawai Program Studi Magister Teknologi Informasi Sekolah Tinggi Manajemen Informatika dan Komputer Mikroskil Medan yang telah memberikan ilmu serta mendidik penulis selama di bangku kuliah.
5. Teristimewa kepada orang tua tercinta ayahanda Nursaman dan Ibu Elizar Jusda atas doa dan dukungan serta nasihat yang membangun sehingga penulis dapat menyelesaikan pendidikan S2 ini.

6. Teruntuk kekasih tersayang Ahmad Pardi Hasibuan yang telah sabar memberikan dukungan luar biasa tanpa henti menyemangati penulis dalam menyelesaikan pendidikan S2 ini.
7. Sahabat dan teman seperjuangan M.T.I atas kebersamaan dan kekompakan dalam perkuliahan dan tetap semangat sampai akhir, serta semua pihak yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari penelitian tesis ini masih jauh dari kata sempurna, oleh karena itu mohon kiranya para pembaca memberikan masukan, kritik dan saran yang bersifat membangun untuk perbaikan penulisan selanjutnya. Akhir kata penulis berharap semoga karya ilmiah ini bermanfaat bagi kemajuan ilmu pengetahuan dan pendidikan.

Medan, 31 Januari 2020
Penulis

Surizar Rahmi Danur
174211028

UNIVERSITAS MIKROSKIL

LEMBAR PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Jurusan/Program Studi S-2 Magister Teknologi Informasi STMIK Mikroskil dengan identitas mahasiswa sebagai berikut:

Nama : Surizar Rahmi Danur

Nim : 174211028

Peminatan : Teknologi Informasi

Saya telah melaksanakan penelitian dan penulisan Tesis dengan judul "OPTIMASI KLASIFIKASI BAYESIAN NETWORK MELALUI REDUKSI ATTRIBUT MENGGUNAKAN PRINCIPAL COMPONENT ANALYSIS", dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Tesis tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh STMIK Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada STMIK Mikroskil Medan Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty Free Right*) atas Tesis saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, STMIK Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tesis saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Tesis saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 31 Januari 2020

Saya yang membuat pernyataan,



Surizar Rahmi Danur

ABSTRAK

Reduksi dimensionalitas merupakan topik yang sedang hangat diperbincangkan dalam perkembangannya telah dilakukan diberbagai bidang penelitian salah satunya *machine learning* dengan melakukan reduksi dapat menurunkan kapasitas dimensi tanpa mengurangi (menghilangkan) informasi yang terkandung pada data tersebut. *Principal Component Analysis* merupakan salah satu teknik mereduksi yang telah teruji mampu mengurangi kapasitas data tanpa menghilangkan informasi yang terkandung pada dataset secara signifikan. Pada penelitian ini dilakukan reduksi atribut menggunakan *Principal Component Analysis* dengan menggunakan dataset faktor-faktor yang mempengaruhi ketidakhadiran karyawan diambil dari *Repository University of California di Irvine (UCI)*. Kombinasi dengan *Bayesian Network* untuk mengklasifikasi data sebagai perbandingan antara sebelum dan sesudah dilakukan reduksi atribut. Hal tersebut dapat terlihat pada hasil akurasi awal sebelum dilakukan reduksi dengan akurasi sebesar 100% dan setelah dilakukan reduksi atribut ke lima terjadi penurunan akurasi sebesar 89,7%.

Kata Kunci: Reduksi Atribut, *Principal Component Analysis*, Klasifikasi *Bayesian Network*

UNIVERSITAS
MIKROSKIL

ABSTRACT

Dimensionality reduction is a hot topic being discussed in its development has been carried out in various fields of research one of which is machine learning by reducing can reduce the capacity of dimensions without reducing (eliminating) information contained in the data. Principal Component Analysis is one of the proven reduction techniques capable of reducing data capacity without significantly eliminating the information contained in the dataset. In this research attribute reduction using principal component analysis using a dataset of factors affecting employee absence was taken from the University of California repository at Irvine (UCI). Combination with Bayesian Network to classify data as a comparison between before and after attribute reduction. This can be seen in the initial results before the reduction with an accuracy of 100% and after the fifth attribute reduction there is a decrease in accuracy by 89,7%.

Keywords: Attribute Reduction, Principal Component Analysis, Bayesian Network Classification



UNIVERSITAS
MIKROSKIL

DAFTAR ISI

ABSTRAK	i
ABTRACT	ii
KATA PENGANTAR	iii
DAFTAR ISI	v
DAFTAR GAMBAR	vii
DAFTAR TABEL	viii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Masalah Penelitian.....	4
1.3 Tujuan Dan Manfaat Penelitian.....	4
1.4 Batasan Masalah	5
1.5 Metodologi Penelitian	5
1.6 Sistematika Penulisan	6
BAB II KAJIAN LITERATUR	7
2.1 Reduksi Data	7
2.1.1 Pendekatan Reduksi.....	8
2.2 Principal Component Analysis	10
2.2.1 Prosedur Pengerjaan Reduksi Menggunakan PCA	11
2.2.2 Pendekatan Principal Component Analysis.....	13
2.3 Klasifikasi	13
2.3.1 Bayesian Network	15
2.4 RMSE (<i>Root Mean Square Error</i>).....	21
2.5 Kerangka Pikir / Pola Pemecahan Masalah	21
BAB III METODOLOGI PENELITIAN	23
3.1. Analisis Masalah	23
3.2 Metodologi Penelitian	23
3.2.1 Dataset yang Digunakan.....	25
3.2.2 Preprocessing Data	26

3.2.3 Reduksi Data	30
3.2.4 Klasifikasi.....	32
3.2.4.1 Pembentukan Struktur Bayesian Network.....	34
3.2.4.2 Penentuan Parameter	34
3.2.4.3 Conditional Probability Tabel (CTP)	35
3.2.4.4 Joint Probability Tabel (JPT)	37
3.2.4.5 Hitung Posterior Probability	39
3.2.4.6 Inferensi Probabilistik	41
3.3 Alat-Alat Penelitian	42
3.4 Metode Analisi	42
BAB IV HASIL DAN PEMBAHASAN	45
4.1 Hasil Pengujian	45
4.1.1 Data Training.....	45
4.1.2 Reduksi Atribut	47
4.1.2.1 Reduksi Atribut Dengan 5% Dataset	48
4.1.2.2 Reduksi Atribut Dengan 10% Dataset	55
4.1.2.3 Reduksi Atribut Dengan 15% Dataset	62
4.1.2.4 Reduksi Atribut Dengan 20% Dataset	68
4.1.3 Hasil Pengukuran Akurasi	75
4.2 Pembahasan	75
4.2.1 Hasil Reduksi Atribut	76
4.2.2 Hasil Klasifikasi Atribut	77
BAB V KESIMPULAN DAN SARAN	78
5.1 Kesimpulan.....	78
5.2 Saran	78

REFERENSI

DAFTAR GAMBAR

Gambar 2.1. Langkah-Langkah Analisis Reduksi <i>Big Data</i> Berbasis <i>IoT</i>	10
Gambar 2.2. Skema Klasifikasi Validasi Prosedur	14
Gambar 2.3. Contoh <i>Bayesian Network</i> dengan Lima Variabel (<i>Node</i>).....	16
Gambar 2.4. Faktor-Faktor yang Mempengaruhi Penurunan Kinerja Karyawan	17
Gambar 2.5. Contoh <i>Bayesian Network</i>	19
Gambar 3.1 Metode Penelitian	24
Gambar 3.2 Reduksi Data dengan Metode PCA.....	30
Gambar 3.3 Langkah-Langkah Algoritma <i>Bayesian Network</i>	33
Gambar 3.4 Pembentukan Struktur <i>Bayesian Network</i>	34
Gambar 4.1 Data Training Faktor yang Mempengaruhi Ketidakhadiran Karyawan	46
Gambar 4.2 Normalisasi Dataset	47
Gambar 4.3 Hasil Akurasi Klasifikasi (5%)	49
Gambar 4.4 Reduksi 1 Atribut	44
Gambar 4.5 Klasifikasi Reduksi 1 Atribut	50
Gambar 4.6 Reduksi 2 Atribut	51
Gambar 4.7 Klasifikasi Reduksi 2 Atribut	51
Gambar 4.8 Reduksi 3 Atribut	52
Gambar 4.9 Klasifikasi Reduksi 3 Atribut	53
Gambar 4.10 Reduksi 4 Atribut	54
Gambar 4.11 Klasifikasi Reduksi 4 Atribut	54
Gambar 4.12 Hasil Akurasi Klasifikasi (10%)	55
Gambar 4.13 Reduksi 1 Atribut	56
Gambar 4.14 Klasifikasi Reduksi 1 Atribut	57
Gambar 4.15 Reduksi 2 Atribut	58
Gambar 4.16 Klasifikasi Reduksi 2 Atribut	58
Gambar 4.17 Reduksi 3 Atribut	59
Gambar 4.18 Klasifikasi Reduksi 3 Atribut	60
Gambar 4.19 Reduksi 4 Atribut	61
Gambar 4.20 Klasifikasi Reduksi 4 Atribut	61

Gambar 4.21 Hasil Akurasi Klasifikasi (15%)	62
Gambar 4.22 Reduksi 1 Atribut	63
Gambar 4.23 Klasifikasi Reduksi 1 Atribut	64
Gambar 4.24 Reduksi 2 Atribut	65
Gambar 4.25 Klasifikasi Reduksi 2 Atribut	65
Gambar 4.26 Reduksi 3 Atribut	66
Gambar 4.27 Klasifikasi Reduksi 3 Atribut	67
Gambar 4.28 Reduksi 4 Atribut	67
Gambar 4.29 Klasifikasi Reduksi 4 Atribut	68
Gambar 4.30 Hasil Akurasi Klasifikasi (20%)	69
Gambar 4.31 Reduksi 1 Atribut	70
Gambar 4.32 Klasifikasi Reduksi 1 Atribut	70
Gambar 4.33 Reduksi 2 Atribut	71
Gambar 4.34 Klasifikasi Reduksi 2 Atribut	72
Gambar 4.35 Reduksi 3 Atribut	73
Gambar 4.36 Klasifikasi Reduksi 3 Atribut	73
Gambar 4.37 Reduksi 4 Atribut	74
Gambar 4.38 Klasifikasi Reduksi 4 Atribut	75

UNIVERSITAS MIKROSKIL

DAFTAR TABEL

Tabel 3.1 Kriteria Atribut Dataset	25
Tabel 3.2 Nilai Attribute Maksimum dan Minimum	27
Tabel 3.3 Hasil Normalisasi	29
Tabel 3.4 Penentuan Parameter.....	35
Tabel 3.5 Conditional Probability Tabel Faktor Ketidakhadiran Karyawan	36
Tabel 3.6 Joint Probability Tabel Faktor Ketidakhadiran Karyawan	38
Tabel 3.7 Posterior Probability Faktor Ketidakhadiran Karyawan	40
Tabel 3.8 Confusion Matrix	43
Tabel 4.1. Data Training Faktor yang Mempengaruhi Ketidakhadiran Karyawan....	46
Tabel 4.2 Hasil Klasifikasi Dataset Sebelum Pengujian 5%	48
Tabel 4.3 Hasil Klasifikasi Dataset Sebelum Pengujian 10%.....	55
Tabel 4.4 Hasil Klasifikasi Dataset Sebelum Pengujian 15%.....	63
Tabel 4.5 Hasil Klasifikasi Dataset Sebelum Pengujian 20%.....	69
Tabel 4.6 Hasil Klasifikasi Dataset 5%.....	77
Tabel 4.7 Hasil Klasifikasi Dataset 10%.....	77
Tabel 4.8 Hasil Klasifikasi Dataset 15%.....	78
Tabel 4.9 Hasil Klasifikasi Dataset 20%.....	78

UNIVERSITAS
MIKROSKIL