ANALISIS KOMPONEN UTAMA DENGAN DATA HILANG MENGGUNAKAN METODE *NONLINEAR ITERATIVE PARTIAL LEAST SQUARES* (NIPALS)

(Skripsi)

Oleh

ANINDIA PUTRI

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2018

ABSTRACT

PRINCIPAL COMPONENT ANALYSIS WITH MISSING DATA USING *NONLINEAR ITERATIVE PARTIAL LEAST SQUARES* (NIPALS) METHOD

By

Anindia Putri

Principal component analysis (PCA) is a multivariate statistical technique which is used to reduce large dimension and correlated data into smaller dimension and uncorrelated data. The conventional PCA (i.e. by using the eigen vectors of the covariance or correlation matrix as the coefficient of the principal components) cannot be used to analyze data with missing values. One method that can be used to solve missing values problem in PCA is by using the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm. NIPALS algorithm decomposes data matrix **X** into scores matrix **T** and loadings matrix **P**. In this research, investigation of PCA using NIPALS algorithm was conducted by using data with missing values 1%, 3%, 5%, 10%, 15%, 20%, 25% and 28%. Based on the bias, Mean Square Errors (MSE) and spectral norm, the result shows that NIPALS algorithm still can overcome the missing values as much as 28%. However, for the missing values as much as 25% or more, the resulted principal components were not accurate.

Keyword : Principal component analysis, missing data, *nonlinear iterative partial least squares* (NIPALS) algorithm

ABSTRAK

ANALISIS KOMPONEN UTAMA DENGAN DATA HILANG MENGGUNAKAN METODE *NONLINEAR ITERATIVE PARTIAL LEAST SQUARES* (NIPALS)

Oleh

Anindia Putri

Analisis komponen utama (AKU) adalah salah satu teknik statistika multivariat yang digunakan untuk mereduksi dimensi data yang berukuran besar dan saling berkorelasi menjadi dimensi data yang berukuran lebih kecil dan tidak saling berkorelasi. AKU konvensional (yaitu dengan menggunakan vektor eigen dari matriks varian kovarian atau korelasi sebagai koefisien dari komponen utama) tidak dapat digunakan untuk menganalisis suatu data yang mengandung nilai yang hilang. Salah satu cara untuk mengatasi data hilang dalam AKU yaitu dengan menggunakan algoritma *Nonlinear Iterative Partial Least Squares* (NIPALS). Algoritma NIPALS mendekomposisikan matriks $\boldsymbol{X}$ menjadi matriks *scores* $\boldsymbol{T}$ dan matriks *loadings* $\boldsymbol{P}$. Dalam penelitian ini, AKU menggunakan data hilang dengan algoritma NIPALS yang memuat nilai hilang sebanyak 1%, 3%, 5%, 10%, 15%, 20%, 25%, dan 28%. Berdasarkan biasnya, *Mean Square Errors* (MSE) dan *spectral norm*, menyatakan bahwa algoritma NIPALS masih dapat mengatasi nilai hilang sebanyak 28%, namun untuk nilai hilang sebanyak lebih dari atau sama dengan 25% hasil komponen utama yang didapat sudah tidak akurat.

Kata Kunci : **analisis komponen utama, data hilang, algoritma *nonlinear iterative partial least squares* (NIPALS)**

ANALISIS KOMPONEN UTAMA DENGAN DATA HILANG MENGGUNAKAN METODE *NONLINEAR ITERATIVE PARTIAL LEAST SQUARES* (NIPALS)

Oleh

ANINDIA PUTRI

Skripsi

Sebagai Salah Satu Syarat Untuk Mencapai Gelar
SARJANA SAINS

pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2018

Judul Skripsi : **ANALISIS KOMPONEN UTAMA DENGAN DATA HILANG MENGGUNAKAN METODE *NONLINEAR ITERATIVE PARTIAL LEAST SQUARES* (NIPALS)**

Nama Mahasiswa : *Anindia Putri*

Nomor Pokok Mahasiswa : 1417031014

Program Studi : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Dr. Khoirin Nisa, S.Si., M.Si.	**Agus Sutrisno, S.Si., M.Si.**
NIP. 19740726 200003 2 001	NIP. 19700831 199903 1 002

2. Ketua Jurusan Matematika

Prof. Dra. Wamiliana, M.A, Ph.D.
NIP. 19631108 198902 2 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Dr. Khoirin Nisa, S.Si., M.Si.** ....

Sekretaris : **Agus Sutrisno, S.Si., M.Si.** ....

Penguji
Bukan Pembimbing : **Prof. Drs. Mustofa Usman, M.A., Ph.D.** ....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

KEMENTERIAN RISET, TEKNOLOGI DAN PENDIDIKAN TINGGI
UNIVERSITAS LAMPUNG
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

Prof. Warsito, S.Si., D.E.A., Ph.D.
NIP. 19710212 199512 1 001

Tanggal Lulus Ujian Skripsi : **14 Mei 2018**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini :

Nama : **Anindia Putri**

Nomor Pokok Mahasiswa : **1417031014**

Jurusan : **Matematika**

Judul Skripsi : **Analisis Komponen Utama Dengan Data Hilang Menggunakan Metode *Nonlinear Iterative Partial Least Squares* (NIPALS)**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 14 Mei 2018

Yang Menyatakan

METERAI TEMPEL

TGL 20

96AA0AFF070339541

6000

ENAM RIBU RUPIAH

Anindia Putri

RIWAYAT HIDUP

Penulis bernama lengkap Anindia Putri, anak pertama dari dua bersaudara dari pasangan Bapak Muhamad dan Ibu Titik Puspa Wati yang dilahirkan di Lampung Selatan pada tanggal 27 Agustus 1996.

Penulis menempuh pendidikan di RA Anisa pada tahun 2000-2002 dan dilanjutkan sekolah dasar di SDN Banjarsari pada tahun 2002-2005, SDN Pegadungan 02 pada tahun 2005-2006, SDN 4 Tanjung Priok pada tahun 2006-2008, kemudian bersekolah di SMPN 2 Wonosobo pada tahun 2008-2011, dan bersekolah di SMAN 1 Pringsewu pada tahun 2011-2014.

Pada tahun 2014 penulis melanjutkan pendidikan di perguruan tinggi dan terdaftar sebagai mahasiswa S1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

Selama menjadi mahasiswa, pada tahun 2015-2017 penulis dipercaya menjadi anggota Bidang Minat dan Bakat Himatika Unila.

Pada tahun 2017 penulis melaksanakan Kerja Praktik (KP) di BPS Provinsi Lampung dan pada tahun yang sama penulis melakukan Kuliah Kerja Nyata (KKN) di Desa Sumur Kumbang, Kecamatan Kalianda, Kabupaten Lampung Selatan, Provinsi Lampung.

Kata Inspirasi

"Kemudian apabila kamu telah membulatkan tekad, maka bertawakallah kepada Allah, sesungguhnya Allah menyukai orang – orang yang bertawakal (kepada-Nya)"
(QS. Ali Imraan : 159)

"Karena sesungguhnya sesudah kesulitan itu ada kemudahan"
(QS. Alam Nasyroh : 5-6)

"Maka nikmat Tuhanmu manakah yang kau dustakan?"
(QS. Ar Rahman : 13)

"Perbaiki shalatmu, maka akan baik hidupmu"

"Dream it. Wish it. Do it"

"Orang yang paling bahagia di dunia ini adalah orang yang senantiasa bersyukur dalam segala keadaan, dan menyadari betapa berharganya hidup yang diberikan padanya."

Dengan mengucapkan Alhamdulillah,
Puji dan syukur kepada Allah Subhanahu Wata'ala atas segala nikmat dan karunia-Nya, dan suri tauladan Nabi Muhammad Shallallahu 'Alaihi Wasallam yang menjadi contoh dan panutan untuk kita semua.

Kupersembahkan sebuah karya sederhana ini untuk :

Ayahanda Muhamad, Ibunda Titik Puspa Wati, dan Nenek Rasmi
Terimakasih atas limpahan kasih sayang, pengorbanan, doa, dan seluruh motivasi di setiap langkahku. Karena atas doa dan ridho kalian, Allah memudahkan tiap perjalanan hidup ini.
Terimalah bukti kecil ini sebagai kado keseriusanku untuk membalas semua pengorbanan, keikhlasan, dan jerih payah yang selama ini kalian lakukan.

Adikku Tercinta
Maulana Putra
Semoga apa yang telah kakak lakukan selalu bisa menjadi contoh dan motivasi untuk kamu.

Almamaterku Tercinta Universitas Lampung

SANWACANA

Puji syukur penulis haturkan kepada Allah SWT yang selalu memberi rahmat dan karunia tak terhingga kepada umat-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul **"Analisis Komponen Utama Dengan Data Hilang Menggunakan Metode *Nonlinear Iterative Partial Least Squares* (NIPALS)"**. Shalawat serta salam tak lupa penulis haturkan kepada junjungan dan suri tauladan Nabi Muhammad SAW, semoga seluruh umat islam termasuk dalam golongan yang akan mendapatkan syafaatnya di Yaumil Akhir Kelak.

Dalam penulisan skripsi ini, penulis dibantu oleh banyak pihak baik dalam hal bimbingan, dukungan, doa, maupun saran. Pada kesempatan ini penulis berterimakasih kepada :

1. Ibu Dr. Khoirin Nisa, S.Si., M.Si., selaku pembimbing I yang telah bersedia memberi arahan, bimbingan, kritik dan saran bagi penulis selama proses pembuatan skripsi ini.
2. Bapak Agus Sutrisno, S.Si., M.Si., selaku pembimbing II yang telah memberikan arahan serta dukungan bagi penulis.
3. Bapak Prof. Drs. Mustofa Usman, MA., Ph.D., selaku pembahas yang telah bersedia memberikan ide, kritik dan saran sehingga terselesaikannya skripsi ini.

4. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Pembimbing Akademik yang telah membimbing penulis dalam menyelesaikan permasalahan seputar akademik.
5. Ibu Prof. Dra. Wamiliana, MA., Ph.D., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam.
6. Bapak Prof. Warsito, S.Si., D.E.A., Ph.D., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam.
7. Seluruh dosen, staf dan karyawan Jurusan Matematika FMIPA Universitas Lampung yang telah memberikan ilmu dan bantuan kepada penulis.
8. Mama, Papa dan keluarga yang selalu menjadi penyemangat, memberikan doa dan dukungan yang tak terhingga bagi penulis.
9. Sahabat – sahabat tercinta penulis Nanda Arsy Syafitri Islami, Annisa’ul Mufidah, Ratna Puspita Sari yang selalu menemani dalam suka maupun duka dan selalu menjadi penyemangat penulis.
10. Raka, Kiki, Ira, Camel, Julian, Ega, Tika, Eca, Dhea, Rara, Dea, Alvin, Kodir, Arif, Yunda Karina, Rifka dan Bella yang selalu menemani, membantu dan memberi dukungan kepada penulis.
11. Pule, Maget, Yola, Poppy, Ecy, Amoy, Olin, Wika, Syafa, Fajar, Aldo, Redi, Yona, Rium, Agus, Acong, Uti, Tiara, Andan, Vivin, dan Yunika yang telah memberi kebahagiaan dan keceriaan bagi penulis.
12. Teman-teman Matematika 2014 yang selalu menjadi semangat bagi penulis.
13. Bidang Minat dan Bakat HIMATIKA FMIPA UNILA.
14. Keluarga Besar HIMATIKA FMIPA UNILA.

15. Seluruh pihak yang terlibat dalam penyelesaian skripsi ini yang tidak dapat penulis sebutkan satu persatu.

Tentunya, Penulis menyadari bahwa masih ada kekurangan dari skripsi ini, akan tetapi besar harapan semoga skripsi ini dapat bermanfaat bagi kita semua. Sekian dan terimakasih.

Bandar Lampung, 18 Mei 2018
Penulis,

Anindia Putri

DAFTAR ISI

Halaman

DAFTAR ISI.. i

DAFTAR TABEL iii

DAFTAR GAMBAR... v

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah 1
1.2 Tujuan Penelitian.. 4
1.3 Manfaat Penelitian 5

II. TINJAUAN PUSTAKA

2.1 Data Hilang 6
2.2 Matriks.. 7
2.2.1 Jenis-Jenis Matriks 8
2.2.2 Operasi Matriks 9
2.2.3 Nilai Eigen dan Vektor Eigen... 12
2.3 Analisis Komponen Utama... 13
2.4 *Partial Least Squares* (PLS).. 16
2.5 Algoritma NIPALS... 19
2.6 Ukuran Akurasi 20

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian 22
3.2 Data Penelitian.. 22
3.3 Metode Penelitian 23

IV. HASIL DAN PEMBAHASAN

4.1 Hasil dan Pembahasan 25
4.2 Studi Kasus 32
4.3.1 Analisis Komponen Utama Data Lengkap 34
4.3.2 Analisis Komponen Utama Menggunakan Algoritma NIPALS 36

V. KESIMPULAN

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR TABEL

Halaman

Tabel 1. Rata-rata bias *loadings* dan *scores* komponen utama data simulasi 26

Tabel 2. MSE *loadings* dan *scores* komponen utama data simulasi 28

Tabel 3. *Spectral norm* data simulasi 31

Tabel 4. Data Volume Ekspor yang Tercatat pada Dinas Perindustrian (ton) 33

Tabel 5. Hasil analisis komponen utama data lengkap 34

Tabel 6. Koefisien tiap komponen utama 35

Tabel 7. *Loadings* komponen utama data lengkap sampai data hilang 10. 37

Tabel 8. Bias *loadings* komponen utama data volume ekspor.................. 38

Tabel 9. Bias *scores* komponen utama data volume ekspor 39

Tabel 10. MSE *loadings* komponen utama data volume ekspor................. 40

Tabel 11. MSE *scores* komponen utama data volume ekspor 42

Tabel 12. *Spectral* norm data volume ekspor... 43

DAFTAR GAMBAR

Halaman

Gambar 1. Rata-rata bias *loadings* komponen utama data simulasi 26

Gambar 2. Rata-rata bias *scores* komponen utama data simulasi 27

Gambar 3. MSE *loadings* komponen utama data simulasi 29

Gambar 4. MSE *scores* komponen utama data simulasi 30

Gambar 5. *Spectral norm* data simulasi 31

Gambar 6. Bias *loadings* komponen utama data volume ekspor 38

Gambar 7. Bias *scores* komponen utama data volume ekspor..................... 40

Gambar 8. MSE *loadings* komponen utama data volume ekspor 41

Gambar 9. MSE *scores* komponen utama data volume ekspor 42

Gambar 10. *Spectral norm* data volume ekspor.. 43

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis multivariat merupakan salah satu teknik statistika yang digunakan untuk memahami struktur data dalam dimensi tinggi. Variabel-variabel itu saling terkait satu sama lain. Disinilah letak perbedaan antara multivariabel dan multivariat. Multivariat pasti melibatkan multivariabel tetapi tidak sebaliknya. Multivariabel yang saling berkorelasilah yang dikatakan multivariat. Menurut Santoso (2004), analisis multivariat dapat didefinisikan secara sederhana sebagai metode pengolahan variabel dalam jumlah banyak untuk mencari pengaruhnya terhadap suatu objek secara simultan. Dalam analisis multivariat kita juga belajar analisis komponen utama.

Analisis komponen utama (AKU) merupakan salah satu analisis multivariat yang digunakan untuk mereduksi dimensi data dari yang berukuran besar dan saling berkorelasi menjadi dimensi yang lebih kecil dan tidak saling berkorelasi. Namun, walaupun dimensi data menjadi lebih kecil, kita tidak akan kehilangan banyak informasi karena variasi data tetap dipertahankan minimal 80% (Johnson & Wichern, 1992).

Perkembangan analisis komponen utama dimulai sejak diperkenalkan pertama kali oleh Pearson pada tahun 1901. Sejalan dengan perkembangan teknologi komputer dan kemajuan di bidang matematika, analisis komponen utama hingga kini masih terus mengalami perkembangan. Perkembangan selanjutnya, diperkenalkan generalisasi dari analisis komponen utama oleh Loeve pada tahun 1963. Analisis komponen utama dihadapkan dengan masalah yang muncul saat terdapat gugusan data yang tidak lengkap, misalnya saat ada beberapa nilai yang hilang. Ketidaklengkapan data dapat mengakibatkan kurang akurat dalam menganalisis dan mengevaluasi hasil penelitian.

Dalam kasus nyata, misalnya saja survei atau sensus. Salah satu permasalahan yang sering dihadapi dalam survei atau sensus dengan cara memberikan kuesioner adalah ada beberapa variabel yang *non* respon. Penyebab data tidak lengkap (data hilang) ini yaitu responden yang tidak menjawab beberapa pertanyaan atau bahkan tidak mengisi kuesioner pada saat proses pengambilan data, atau bisa juga data hilang terjadi karena kesalahan pada saat pelaksanaan *entry* data. Sehingga dalam kasus survei atau sensus ini sering terdapat data tidak lengkap atau data hilang.

Little dan Rubin (1987) membagi tiga tipe mekanisme dari data hilang. Pertama, *missing completely at random* (MCAR) yang berarti bahwa terjadinya data hilang tidak berkaitan dengan nilai semua variabel, apakah itu variabel nilai hilang atau variabel yang terobservasi. Hal ini berarti data hilang terjadi secara acak. Kedua, *missing at random* (MAR) yaitu terjadinya data hilang hanya berkaitan dengan variabel respon atau variabel pengamatan. Contohnya seseorang yang memiliki rasa waswas yang tinggi cenderung tidak akan melaporkan pendapatan mereka, rasa waswas akan berhubungan pada pelaporan pendapatan. Sedangkan tipe ketiga adalah *missingness not at random* (MNAR) bahwa terjadinya *missing data* pada suatu variabel berkaitan dengan variabel itu sendiri, sehingga ini tidak bisa diprediksi dari variabel lain pada suatu dataset.

Terdapat berbagai metode untuk menangani masalah data hilang dalam analisis statistik. Metode yang pertama yaitu prosedur berbasis unit yang lengkap. Metode ini cukup memuaskan jika jumlah data hilang tidak terlalu besar, tapi prosedur ini menjadi tidak efisian jika persentase data hilang meningkat atau mengelompok, hal ini akan menyebabkan hasil yang sangat bias. Metode yang kedua yaitu prosedur berbasis imputasi. Terdapat beberapa macam pendekatan untuk imputasi ini antara lain yaitu *hot deck imputation, cold deck imputation, mean imputation, regression imputation*, prosedur pembobotan, dan prosedur berbasis model.

Salah satu prosedur berbasis model yaitu suatu prosedur yang dibentuk dengan menentukan suatu model sebagian data yang hilang dan selanjutnya melakukan estimasi *partial least square* dengan iteratif maximum. *Partial least squares* (PLS) adalah suatu teknik statistik multivariat yang bisa menangani banyak variabel respon dan variabel eksplanatori sekaligus.

Prosedur selanjutnya untuk mengatasi masalah data hilang yaitu dengan algoritma *Nonlinear Iterative Partial Least Squares* (NIPALS). NIPALS pertama kali diperkenalkan oleh ekonom dan statistikawan Herman Wold pada tahun 1966. Konsep ini tidak sepenuhnya baru, sebab temuan serupa telah diterbitkan pada tahun 1923 oleh Fischer dkk. Akan tetapi, orisinalitas dalam karya Wold merupakan interpretasi dari *partial least squares* dan memiliki kemampuan untuk mencari data hilang dengan analisis komponen utama. Kajian tentang metode NIPALS telah dilakukan selama beberapa dasawarsa, diantaranya oleh : Geladi dan Bruce (1986), Andrecut (2009), dan Pawestri (2012).

Dengan demikian, metode ini menjadi menarik jika diterapkan pada studi kasus mencari data hilang dalam analisis komponen utama. Sehingga dalam penelitian ini, penulis mengkaji tentang penggunaan algoritma NIPALS untuk mengatasi data hilang pada analisis komponen utama.

1.2 Tujuan Penelitian

Berdasarkan latar belakang diatas, adapun tujuan penelitian ini adalah :

1. Mengkaji bagaimana mengatasi data hilang pada analisis komponen utama dengan menggunakan algoritma NIPALS
2. Mengevaluasi tingkat keterandalan algoritma NIPALS dalam mengatasi data hilang pada analisis komponen utama dengan simulasi data.

1.3 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

1. Menambah wawasan pengetahuan kepada penulis khususnya tentang metode NIPALS
2. Sebagai salah satu bahan dalam menambah referensi tentang mengatasi data hilang dengan algoritma NIPALS pada analisis komponen utama.
3. Mempopulerkan salah satu teknik estimasi dalam statistika yaitu algoritma NIPALS.
4. Sebagai bahan tinjauan pustaka yang berguna bagi setiap pihak yang memerlukan.

II. TINJAUAN PUSTAKA

2.1 Data Hilang

Data hilang merupakan informasi yang tidak tersedia pada sebuah objek atau kasus, yang terjadi disebabkan informasi untuk sesuatu tentang objek tidak diberikan, sulit dicari, atau memang informasi tersebut tidak ada. Data hilang dapat mengakibatkan kurang akurat dalam menganalisis dan mengevaluasi hasil penelitian. Untuk mengatasi data hilang, maka selanjutnya perlu dilakukan pengujian apakah data statistik yang mengandung banyak data hilang tersebut masih layak diproses lebih lanjut atau tidak. Pada penelitian sosial, data hilang kerap terjadi dikarenakan responden tidak memberikan jawaban pada alternatif jawaban yang disediakan.

Mekanisme data hilang diklasifikasikan sebagai berikut :

1) *Missing completely at random* (MCAR): mekanisme kasus saat pola nilai hilang pada variabel tidak berkaitan dengan variabel lain atau terhadap dirinya. Pola terjadi secara tidak sistematik dan dapat dianggap sebagai *random subsample* dari hipotesis data saat lengkap.
2) *Missing at random* (MAR): mekanisme kasus saat terjadi kaitan antara variabel data yang memuat nilai hilang dengan variabel data yang tidak

terdapat nilai hilang. Pola terjadi secara sistematik menjelaskan kecenderungan terdapat korelasi.

3) *Nonignorable*: mekanisme nilai hilang dengan nilai yang hilang jelas tergantung dengan variabel yang tidak lengkap tersebut. Dikenal juga sebagai NMAR "*Not Missing at Random*" atau MNAR "*Missing Not at Random*" (Little & Rubin, 1987).

2.2 Matriks

Matriks adalah susunan bilangan atau fungsi yang diletakkan atas baris dan kolom serta diapit oleh dua kurung siku. Bilangan atau fungsi tersebut disebut entri atau elemen matriks. Elemen matriks terdiri dalam baris dan kolom. Lambang matriks dilambangan dengan huruf besar, sedangkan entri (elemen) matriks dilambangkan dengan huruf kecil.

Matriks **A** yang memiliki m baris dan n kolom dapat ditulis sebagai berikut :

$$\boldsymbol{A} = \quad a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix} \qquad (2.1)$$

Dimana a_{ij} menyatakan elemen matriks **A** yang berada pada baris ke-*i* dan kolom ke-*j* (Anton, 1987).

2.2.1 Jenis-jenis Matriks

Terdapat beberapa jenis matriks, yaitu :

a. Matriks simetris

Matriks simetris adalah matriks persegi yang elemennya simetris secara diagonal. Matriks **A** dikatakan simetris jika $a_{ij} = a_{ji}$ untuk semua *i* dan *j*, dengan a_{ij} menyatakan unsur pada baris ke *i* dan kolom ke *j*. Matriks yang simetris dapat dikatakan pula sebagai matriks yang transposenya sama dengan dirinya sendiri. Contoh matriks simetris yaitu:

$$\boldsymbol{A} = \begin{bmatrix} 1 & -5 & 4 \\ -5 & 2 & 0 \\ 4 & 0 & 3 \end{bmatrix}$$

b. Matriks ortogonal

Matriks ortogonal ialah matriks yang apabila dikalikan dengan matriks tranposenya menghasilkan matriks satuan (identitas).

Matriks ortogonal dapat dituliskan sebagai:

$$\boldsymbol{A^T A} = \boldsymbol{A A^T} = \boldsymbol{I} \tag{2.2}$$

Sifat matriks ortogonal :

1) Invers matriks ortogonal juga matriks ortogonal
2) Hasil kali matriks-matriks ortogonal juga matriks ortogonal
3) Jika **A** matriks ortogonal, maka $det(\boldsymbol{A}) = 1$ atau $det(\boldsymbol{A}) = -1$

c. Matriks definit positif

Suatu matriks simetris $\boldsymbol{A}$ berorder $n \times n$ disebut matriks definit positif apabila $\boldsymbol{X^T A X} > 0$ untuk semua vektor tak nol $\boldsymbol{X}$. Suatu matriks $\boldsymbol{A}$ yang berukuran

$n \times n$ dikatakan definit positif jika dan hanya jika $\boldsymbol{A^T} = \boldsymbol{A}$ dan $\boldsymbol{X^T A X} > 0$ untuk setiap vektor $\boldsymbol{X}$ tidak nol.

d. Matriks semi definit positif

Matriks $\boldsymbol{A}$ yang berukuran $n \times n$ dikatakan semi definit positif jika dan hanya jika kondisi-kondisi berikut terpenuhi :

1) $\boldsymbol{A^T} = \boldsymbol{A}$ ($\boldsymbol{A}$ matriks simetris).
2) $\boldsymbol{X^T A X} \geq 0$ untuk setiap vektor **X** tidak nol.
3) $\boldsymbol{X^T A X} = 0$ untuk paling sedikit satu vektor tidak nol.

(Anton, 1987).

2.2.2 Operasi Matriks

Terdapat beberapa operasi matriks diantaranya yaitu :

a. Jumlah unsur diagonal suatu matriks (*trace*)

Bila $\boldsymbol{A}$ adalah suatu matriks persegi dengan ukuran $n \times n$, maka jumlah unsur diagonal matriks $\boldsymbol{A}$ dilambangkan $tr(\boldsymbol{A})$, adalah

$$tr(\boldsymbol{A}) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}$$

Lambang *tr* adalah singkatan dari *trace* dalam bahasa inggris.

b. Perkalian skalar

Jika **A** adalah suatu matriks dan *c* adalah suatu *scalar*, maka hasil kali *c***A** adalah matriks yang diperoleh dengan mengalikan setiap anggota dari **A** dan *c*. Perkalian matriks dengan skalar menghasilkan sebuah matriks baru yang elemennya adalah hasil perkalian setiap elemen matriks aslinya dengan skalar.

Dalam notasi matriks :

Jika $\boldsymbol{A} = \; a_{ij}]$, maka $c\boldsymbol{A} = c[a_{ij}]$

c. Perkalian matriks

Jika $\boldsymbol{A}$ adalah matriks berukuran $m \times r$ dan $\boldsymbol{B}$ adalah matriks berukuran $r \times n$, maka hasil kali $\boldsymbol{AB}$ adalah matriks $\boldsymbol{C}$ berukuran $m \times n$ yang unsur-unsurnya adalah:

$$\begin{aligned} \boldsymbol{C}_{ij} &= Baris_i(\boldsymbol{A})\ Kolom_j(\boldsymbol{B}) \\ &= a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} \\ &= \textstyle\sum_{k=1}^{n} a_{ik}b_{kj} \end{aligned}$$

Perkalian matriks hanya bisa dilakukan jika banyaknya kolom matriks yang pertama sama dengan banyaknya baris matriks yang kedua.

d. Transpose matriks

Transpose dari suatu matriks $\boldsymbol{A}$ berorde $m \times n$ yang dinotasikan $[a_{ij}]$ adalah matriks $\boldsymbol{B}$ berorde $n \times m$ yang dinotasikan $[b_{ji}]$ yang didefinisikan oleh:

$$b_{ji} = a_{ij}$$

Untuk $\boldsymbol{i} = 1,2, \dots, m$ dan $j = 1,2, \dots, n$. Transpose dari matriks $\boldsymbol{A}$ dinyatakan oleh $\boldsymbol{A^T}$.

Beberapa sifat transpose matriks:

a) $(\boldsymbol{A^T})^T = \boldsymbol{A}$

b) $(\boldsymbol{A} + \boldsymbol{B})^T = \boldsymbol{A^T} + \boldsymbol{B^T}$

c) $(k\boldsymbol{A})^T = k\boldsymbol{A}^{\boldsymbol{T}}$, dengan k sembarang skalar

d) $(\boldsymbol{AB})^T = \boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{A}^{\boldsymbol{T}}$

e. Determinan matriks

Misalkan $\boldsymbol{A} = \ a_{ij}]$ adalah matriks berukuran $n \times n$, determinan dari $\boldsymbol{A}$ didefinisikan sebagai berikut :

$$\det(A) = \sum a_{ij}(-1)^{i+j} M_{ij} \tag{2.3}$$

Dengan M_{ij} adalah minor dari a_{ij} yaitu determinan sub matriks $\boldsymbol{A}$ yang diperoleh dengan cara membuang semua entri pada baris ke-i dan semua entri pada kolom ke-j.

Jika $\boldsymbol{A} = \ a_{ij}]$ adalah matriks berukuran $n \times n$ yang mengandung sebaris bilangan nol, maka $|\boldsymbol{A}| = 0$. Jika $\boldsymbol{A} = \ a_{ij}]$ matriks segitiga, maka $|\boldsymbol{A}|$ adalah hasil kali elemen-elemen diagonal utama, yaitu $\boldsymbol{A} = a_{11}a_{22}a_{33} \ldots a_{mm}$. Jika $\boldsymbol{A} = \ a_{ij}]$ adalah sebarang matriks persegi, maka $|\boldsymbol{A}| = \ |\boldsymbol{A}^{\boldsymbol{T}}|$, dimana $\boldsymbol{A}^{\boldsymbol{T}}$ adalah transpose dari **A**. Jika $\boldsymbol{A} = \ a_{ij}]$ dan $\boldsymbol{B} = \ b_{ij}]$ adalah matriks persegi yang ordonya sama, maka $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$.

f. Invers matriks

Jika $\boldsymbol{A}$ adalah suatu matriks persegi berukuran $n \times n$ dan jika suatu matriks $\boldsymbol{B}$ yang berukuran sama $n \times n$ disebut invers (balikan) dari $\boldsymbol{A}$ jika dipenuhi $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$, maka $\boldsymbol{A}$ bisa dibalik dan $\boldsymbol{B}$ disebut invers dari $\boldsymbol{A}$. Invers dari $\boldsymbol{A}$ dilambangkan dengan $\boldsymbol{A}^{-\boldsymbol{1}}$.

Jika $\boldsymbol{A}$ dan $\boldsymbol{B}$ adalah matriks-matriks yang dapat dibalik yang ukurannya sama, maka :

1) $\mathbf{AB}$ dapat dibalik
2) $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$

Jika $\mathbf{A}$ adalah sebuah matriks yang dapat dibalik, maka

1) $\boldsymbol{A}^{-\mathbf{1}}$ dapat dibalik dan $(\boldsymbol{A}^{-\mathbf{1}})^{-1} = \boldsymbol{A}$
2) $(\boldsymbol{A}^{n})^{-1} = (\boldsymbol{A}^{-1})^{n}$ untuk $n = 0,1,2, \ldots$
3) Untuk skalar k, dimana $k \neq 0$, maka $k\mathbf{A}$ dapat dibalik dan $(k\boldsymbol{A})^{-1} = \frac{1}{k}\boldsymbol{A}^{-1}$.

(Anton, 1987).

2.2.3 Nilai Eigen dan Vektor Eigen

Misalkan $\boldsymbol{A}$ adalah matriks berordo $n \times n$, vektor $x \in R^{n}$ dan $x \neq 0$, disebut vektor eigen, jika terdapat bilangan riil λ yang disebut nilai eigen, sehingga memenuhi persamaan :

$$\boldsymbol{A}x = \lambda x \tag{2.4}$$

Dari pernyataan diatas, dapat diketahui persyaratan-persyaratan untuk nilai eigen maupun vektor eigen. Nilai eigen λ merupakan bilangan riil, yang berarti dapat bernilai nol, negatif dan juga positif. Sedangkan vektor eigen x merupakan anggota dari R^{n} untuk $\boldsymbol{A}$ dan x bukan vektor nol (Anton, 1987).

2.3 Analisis Komponen Utama

Analisis Komponen Utama (AKU) adalah analisis multivariat yang mentransformasi variabel-variabel asal yang saling berkorelasi menjadi variabel-variabel baru yang tidak saling berkorelasi dengan mereduksi sejumlah variabel tersebut sehingga mempunyai dimensi yang lebih kecil namun dapat menerangkan sebagian besar keragaman variabel aslinya. Tujuan utamanya ialah menjelaskan sebanyak mungkin jumlah keragaman data asli dengan sedikit mungkin komponen utama.

Analisis multivariat sendiri merupakan metode dalam melakukan penelitian terhadap lebih dari dua variabel secara bersamaan. Dengan menggunakan teknik analisis ini maka kita dapat menganalisis pengaruh beberapa variabel terhadap variabel lainnya dalam waktu yang bersamaan.

Komponen utama merupakan kombinasi linear dari peubah yang diamati, informasi yang terkandung pada komponen utama merupakan gabungan dari semua peubah dengan bobot tertentu. Kombinasi linear yang dipilih merupakan kombinasi linear dengan ragam paling besar yang memuat informasi paling banyak . Antar komponen utama bersifat ortogonal, tidak berkorelasi dan informasinya tidak tumpang tindih. Hasil dari prosedur ini nantinya digunakan pada analisis lebih lanjut, seperti analisis pengelompokan dan regresi komponen utama.

AKU tidak selalu bermanfaat digunakan untuk mereduksi banyaknya peubah asal menjadi beberapa peubah baru yang dapat menjelaskan dengan baik keragaman data asal. Bila tidak ada korelasi antara peubah asal, AKU tidak akan memberikan hasil yang diinginkan, karena peubah baru yang diperoleh hanyalalh peubah asal yang ditata berdasarkan besar keragamannya. Makin erat korelasi (baik positif maupun negative) antar peubah, makin baik pula hasil yang diperoleh dari AKU (Mattjik & Sumertajaya, 2011).

Dalam analisis komponen utama misalkan $x_1, x_2, \dots, x_p$ memiliki sebaran peubah ganda dengan vektor rataan μ dan matriks peragam $\sum$. Komponen utama seperti telah dijelaskan diatas merupakan kombinasi linier dari *p* peubah asal, atau dapat ditulis :

$$Y = AX$$

dimana:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix}, A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}, X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

Komponen utama dapat dibentuk dari matriks ragam-peragam (Σ) maupun matriks korelasi. Kedua matriks tersebut berguna dalam penghitungan nilai eigen λ_i dan vektor eigen γ_i. Dimana $\lambda_1, \lambda_2, \dots, \lambda_p$ merupakan nilai eigen yang diperoleh dari persamaan $|\Sigma - \lambda \mathrm{I}| = 0$, sedangkan $\gamma_1, \gamma_2, \dots, \gamma_p$ merupakan vektor eigen yang diperoleh dari persamaan $(\Sigma - \lambda_i I)\gamma_i = 0$; $i = 1,2, \dots, p$. Misalkan

$X1, X2, \ldots\ldots.., Xp$ adalah peubah acak yang menyebar menurut sebaran tertentu dengan vektor rataan μ serta memiliki pasangan nilai eigen dan vektor eigen yang saling ortonormal $(\lambda_1, \gamma_1), (\lambda_2, \gamma_2), \ldots, (\lambda_p, \gamma_p)$, maka komponen utama ke-i dapat didefinisikan sebagai berikut :

$$Y_i = \gamma_{i1}{}^T X_1 + \cdots + \gamma_{ip}{}^T X_p \qquad (2.5)$$

Berdasarkan definisi diatas ragam dari komponen utama ke-i adalah

$$\sigma_{Y1}^2 = Var(Y_i) = \lambda_i = \gamma_1{}^T \sum \gamma_1 = \sum_{i=1}^{p} \sum_{j=1}^{p} \gamma_{1i}\gamma_{1j}\sigma_{ij} \ \ ; j = 1,2, \ldots, p \qquad (2.6)$$

Hasil penurunan persamaan langrange menunjukkan bahwa λ_i merupakan nilai eigen terbesar yang memaksimumkan ragam $Y1$ dan γ_1 merupakan vektor eigen yang berpadanan dengan λ_i. Selanjutnya $Y2$ adalah komponen utama ke-2 yang memaksimumkan nilai $\gamma_2{}^T \sum \gamma_2 = \lambda_2$. Begitu seterusnya sampai Yp adalah komponen utama ke-p yang memaksimumkan ragam Yp dengan memaksimumkan $\gamma_p{}^T \sum \gamma_p = \lambda_p$. Urutan $Y1, Y2, \ldots, Yp$ harus memenuhi persyaratan $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Sementara itu, kontribusi keragaman dari setiap komponen utama ke-*k* terhadap total keragaman disebut dengan proporsi dilambangkan dengan R^2 dengan persamaan sebagai berikut :

$$R^2 = \frac{\lambda_k{}^2}{\sum \lambda_k{}^2} \qquad (2.7)$$

Matriks peragam digunakan bila semua peubah yang diamati diukur dalam satuan pengukuran yang sama, tetapi bila peubah yang diamati mempunyai satuan pengukuran yang berbeda, maka digunakan matriks korelasi, dalam hal ini veriabel bebas perlu dibakukan terlebih dahulu dalam variabel baku sebagai berikut :

$$z_p = \frac{x_p - \mu_p}{\sigma_p}$$

Korelasi antara peubah ke-i dengan komponen utama ke-j jika diturunkan berdasarkan matriks peragam dinyatakan sebagai $r_{x_i,y_j} = \frac{\gamma_i\sqrt{\lambda_j}}{s_i}$ dengan λ_j adalah akar ciri matriks peragam dan s_i adalah simpangan baku peubah ke-i. Sedangkan jika diturunkan berdasarkan matriks korelasi maka $r_{x_i,y_j} = \gamma_i\sqrt{\lambda_j}$ (Mattjik & Sumertajaya, 2011).

2.4 *Partial Least Squares* (PLS)

PLS merupakan pemodelan ‘lunak’ tanpa asumsi sebaran yang dapat menjelaskan struktur keragaman data dan metode umum untuk prediksi model variabel laten (*latent variable*) yang diukur tidak langsung oleh variabel penjelas. Model yang diperoleh dengan PLS dapat mengoptimalkan hubungan prediksi antara kelompok variabel Y dengan kelompok variabel X.

$$Y = \alpha + \beta X + \varepsilon \tag{2.8}$$

Proses penentuan model PLS dilakukan secara iterasi, masing-masing vektor laten yang diperoleh dihubungkan dengan respon dan setiap pengurangan variabel terikat juga dihubungkan dengan respon. Pendugaan pada vektor berikutnya dilakukan dengan menghubungkan perhitungan secara simultan (Wold, 1982).

PLS mempunyai struktur ragam dalam Y yang mempengaruhi perhitungan komponen kombinasi linier dalam X dan sebaliknya, struktur ragam dalam X berpengaruh terhadap kombinasi linier dalam Y. Asumsi dasar pemodelan PLS adalah semua informasi dari variabel penjelas ditujukan pada variabel-variabel laten.

Model PLS menggambarkan hubungan antara variabel laten ξ dari variabel X dengan variabel laten η dari variabel Y. model tersebut adalah :

1) Model hubungan internal

$$\eta = \beta_0 + \beta_1 \, \xi + \vartheta$$

dengan β_0 dan β_1 adalah keofisien model hubungan internal, dan ϑ adalah faktor acak.

2) Model hubungan eksternal

$$X = \pi_0 + \pi_1 \eta + \varepsilon$$

$$Y = (\pi_0 + \pi_1 \beta_0) + \pi_1 \beta_1 \xi + (\varepsilon + \mu_1 \vartheta)$$

dengan π_0 dan π_1 adalah koefisien model hubungan eksternal (*loading*) bagi X, dan ε adalah faktor acak.

Pendugaan parameter pada setiap model PLS adalah sebagai berikut :

1) Variabel laten ξ dan η diduga berdasarkan jumlah X dan Y yang masing-masing diboboti dengan pembobot Y (w_y) dan pembobot X (w_x), yaitu :

$$\xi = X = \sum w_y x \;\; \text{dan} \;\; \eta = Y = \sum w_x y \qquad (2.9)$$

2) Model hubungan internal

$$Y = b_0 + b_1 X + u \qquad (2.10)$$

3) Model hubungan eksternal

$$X = p_0 + p_1 Y + e \quad (2.11)$$

$$Y = (p_0 + p_1 b_0) + p_0 b_1 X + (e + p_1 u) \quad (2.12)$$

Dalam variabel ganda, PLS digunakan untuk memodelkan hubungan antara variabel terikat Y dengan satu kelompok variabel bebas X. Modelnya adalah :

$$X = t_1 p_1 + t_2 p_2 + \cdots + t_a p_a + E_a \quad (2.13)$$

dengan t_a adalah vektor berdimensi n, sebagai vektor skor (*score vector*), sedangkan p_a adalah vektor berdimensi k, sebagai vektor muatan (*loading vector*). Matriks E_a berukuran $n \times k$ adalah matriks residu. Dasar model regresi PLS adalah hubungan antara X dan Y melalui vektor-vektor tersebut sehingga :

$$Y = u_1 q_1 + u_1 q_1 + \cdots + u_a q_a + F_a \quad (2.14)$$

dengan q_a berupa skalar muatan dan F_a adalah vektor residu.

Pendugaan parameter model diatas berdasarkan pada dua tahap berikut:

1) Pendugaan variabel laten ξ dan η dilakukan secara iteratif
2) Pendugaan parameter-parameter b_0, b_1, p_0, dan p_1 dilakukan dengan regresi sederhana.

Proses iterasi dilakukan untuk mengoptimalkan model hubungan antar variabel y dan x berdasarkan masing-masing variabel latennya (Geladi & Bruce , 1986).

2.5 Algoritma NIPALS

Algoritma NIPALS adalah metode iteratif seperti Jacobi yang digunakan untuk memperkirakan unsur-unsur analisis komponen utama (vektor acak) berdimensi terbatas. Algoritma ini bisa diadaptasi untuk data dengan data hilang. Dalam konteks ini, NIPALS memperkirakan faktor utama dan komponen utama dalam analisis komponen utama sebagai metode imputasi untuk data hilang.

Algoritma NIPALS dapat digunakan untuk menemukan komponen utama dengan dekomposisi

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^{T} \tag{2.15}$$

Dimana $\boldsymbol{T}$ dan $\boldsymbol{P}$ harus orthogonal dan kolom dari $\boldsymbol{T}$ disebut *scores* dan kolom dari $\boldsymbol{P}$ (baris dari $\boldsymbol{P}^{\boldsymbol{T}}$) disebut *loadings.*

Untuk menjalankan algoritma NIPALS dimulai dengan menginisialisasi $h = 1$ dan $\boldsymbol{X}_h = \boldsymbol{X}$, kemudian dilanjutkan melalui langkah-langkah dasar berikut :

1) Pilih t_h sebagai kolom apapun dari $\boldsymbol{X_h}$
2) Hitung *loading* $p_h = \frac{{X_h}^T t_h}{{t_h}^T t_h}$
3) Misal $p_h = \frac{p_h}{\sqrt{{p_h}^T p_h}}$
4) Hitung *scores* $t_h = \frac{X_h t_h}{{p_h}^T p_h}$

Ulangi langkah ke (3) dan (4) sampai konvergen untuk komponen utama *h*.

Misalkan $\boldsymbol{X_{h+1}} = \boldsymbol{X}_h - t_h {p_h}^T$ dan nilai eigen $\lambda_h = {t_h}^T t$. Kenaikan $h = h + 1$

dan ulangi untuk komponen utama selanjutnya. Kumpulan kolom $\boldsymbol{T}$ dari t_h dan kolom $\boldsymbol{P}$ dari vektor p_h.

Algoritma NIPALS dapat juga dimodifikasi untuk mengakomodasi nilai yang hilang. Jika untuk variabel k (kolom dari $\boldsymbol{X}$) tertentu, nilai yang hilang ditemui di $\boldsymbol{X}$ untuk objek tertentu i (baris $\boldsymbol{X}$), maka elemen yang sesuai di t_{ih} juga harus dilewati dalam menghitung *loadings*, untuk ***X*** dengan variabel *k* adalah sebagai berikut :

$$p_{hk} = \boldsymbol{X}_{k,h-1} \frac{t_h^T}{t_h^T t_h} \tag{2.16}$$

Demikian juga, jika untuk sampel tertentu *i* (baris dari ***X***), dengan nilai yang hilang ditemukan pada ***X*** untuk variabel tertentu *k* (kolom dari ***X***), maka elemen yang sesuai pada p_{kh} harus juga dilewati untuk menghitung *scores*, dimana untuk sampel *i* adalah

$$t_{ih} = X_{i,h-1} \frac{p_h}{p_h^T p_h} \tag{2.17}$$

Metode ini mungkin memiliki masalah konvergensi jika terlalu banyak nilai yang hilang (Wright, 2017).

2.6 Ukuran Akurasi

Indikator-indikator yang umum digunakan untuk mengukur keakuratan suatu ramalan yaitu rata-rata penyimpangan absolut (*Mean Absolute Deviation*) dan selisih nilai eigen maksimum (*Spectral Norm).*

a) *Mean Squared Error* (MSE)

MSE adalah metode untuk mengevaluasi hasil dugaan yaitu dengan selisih masing-masing kesalahan atau sisa dikuadratkan. Nilai MSE dapat dihitung dengan menggunakan rumus sebagai berikut :

$$MSE = \frac{1}{n}\sum_{I=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

Dimana :

Y_i = nilai awal pada periode *i*

$\hat{Y}_i$ = nilai peramalan periode *i*

n = jumlah data

(Gaspersz, 1998).

b) *Spectral Norm*

Spectral norm adalah salah satu alat ukur untuk menunjukan seberapa dekat perbedaan suatu nilai dalam konteks matriks yaitu dengan melihat selisih nilai eigen maksimum suatu matriks. *Spectral norm* dapat dihitung dengan menggunakan rumus sebagai berikut :

$$|\Delta\psi| = \left|\psi - \hat{\psi}\right|$$

Dimana :

$\Delta\psi$ = *spectral norm* (selisih nilai eigen maksimum)

ψ = nilai eigen maksimum data lengkap

$\hat{\psi}$ = nilai eigen maksimum data hilang

(Padmanabhan dkk, 2016).

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil Tahun Ajaran 2017/2018 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data yang dibangkitkan dengan simulasi pada *software* R, yaitu data berukuran $n = 20$ dan $p = 5$ variabel dari sebaran normal multivariat serta tentukan terlebih dahulu vektor *mean* dan matriks *variance-covariance* yang diulang sebanyak 1000 kali. Berdasarkan perhitungan, untuk membangkitkan data pada R digunakan vektor *mean* dan matriks *variance-covariance* sebagai berikut :

$$\mu = \begin{bmatrix} 5 \\ 4 \\ 3 \\ 2 \\ 4 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 16 & -4 & 8{,}25 & 2{,}75 & -0{,}5 \\ -4 & 19{,}3 & -0{,}75 & -4{,}25 & 12{,}5 \\ 8{,}25 & -0{,}75 & 13 & 0{,}75 & 2{,}25 \\ 2{,}75 & -4{,}25 & 0{,}75 & 6 & -3{,}25 \\ -0{,}5 & 12{,}5 & 2{,}25 & -3{,}25 & 19{,}3 \end{bmatrix} \qquad (3.1)$$

3.3 Metode Penelitian

Penelitian ini dilakukan dengan menggunakan studi literatur secara sistematis yang diperoleh dari buku-buku penunjang, jurnal dan juga media lain seperti internet untuk mendapatkan informasi sebanyak mungkin kemudian melakukan simulasi sebagai aplikasi untuk menerapkan teori yang telah didapat. Untuk mempermudah perhitungan digunakan perangkat lunak (*software*) R 3.4.1.

Adapun tahapan yang dilakukan pada penelitian ini adalah sebagai berikut :

1. Membangkitkan data lengkap yang berukuran $n = 20$ dan $p = 5$ variabel dari sebaran normal multivariat dengan vektor mean dan matriks *variance-covariance* pada persamaan (3.1) dengan pengulangan sebanyak 1000 kali.
2. Menghilangkan elemen dari matriks lengkap dengan presentasi 1%, 3%, 5%, 10%, 15%, 20%, 23%, 25%, dan 28%
3. Menghitung *loadings* dan *scores* komponen utama dari data lengkap dan masing-masing data hilang
4. Menghitung bias *loadings* dan bias *scores* komponen utama pada masing-masing data hilang
5. Menghitung MSE *loadings* dan *scores* komponen utama pada masing-masing data hilang
6. Menghitung *spectral norm* yaitu selisih nilai eigen maksimum untuk masing-masing data hilang

7. Mengevaluasi bias dan MSE dari *loadings* dan *scores* komponen utama dari algoritma NIPALS dari masing-masing data hilang
8. Mengevaluasi *spectral norm* dari nilai eigen pada masing-masing data hilang
9. Melakukan studi kasus pada data *real* yaitu data volume ekspor yang tercatat pada dinas perindustrian.

V. KESIMPULAN

Berdasarkan proses penelitian mengatasi data hilang dengan algoritma NIPALS dalam analisis komponen utama, dapat diambil kesimpulan sebagai berikut :

1. Algoritma NIPALS mengatasi data hilang dalam komponen utama dengan cara mendekomposisikan matriks ***X*** menjadi matriks *scores* ***T*** dan matriks *loadings* ***P***.
2. Penggunaan algoritma NIPALS untuk mengatasi data hilang pada data simulasi berdasarkan bias dan MSE masih cukup baik karena bias dan MSE sama-sama rendah pada semua data hilang. Namun, berdasarkan *spectral norm* hasil algoritma NIPALS sudah tidak akurat pada data hilang 25%.
3. Pada studi kasus penggunaan algoritma NIPALS untuk mengatasi data hilang pada data volume ekspor yang tercatat pada Dinas Perindustrian berdasarkan bias dan *spectral norm* bahwa hasil algoritma NIPALS sudah tidak akurat pada data hilang 10 (30,3%). Namun, berdasarkan MSE hasil algoritma NIPALS sudah tidak akurat pada data hilang 9 (27,27%).
4. Berdasarkan hasil penelitian dengan data simulasi dan studi kasus, penulis menyimpulkan bahwa algoritma NIPALS dapat mengatasi data hilang sebesar 30% atau lebih dari jumlah data. Namun, untuk data hilang sebanyak 25% hasil analisis yang didapat sudah tidak akurat.

DAFTAR PUSTAKA

Andrecut, M. 2009. Parallel GPU Implementation of Iterative PCA Algorithms. *Journal Of Computational Biology*. **16** (11): 1593-1599.

Anton, H. 1987. *Aljabar Linier Elementer*. Jakarta: Erlangga.

Johnson, R. A. dan Wichern, D. W. 1992. *Applied Multivariate Analysis*, Third Edition, Prentice Hall Inc, New Jersey.

Geladi, P. dan Bruce R. K. 1986. *Partial Least Square Regression: A Tutorial*. Analytica Chimica Acta. 185:1-17.

Gaspersz, V. 1998. *Production Planning and Inventory Control.* PT. Sun, Jakarta.

Little, R. J. A., dan Rubin, D. A. 1987. *Statistical Analysis with Missing Data*. New York : John Willey and Sons.

Mattjik, A. A., dan Sumertajaya, I. M. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS*. IPB PRESS, Bogor.

Padmanabhan, N., White, M., Zhou, H.H., dan O'Connell, R. 2016. *Estimating Sparse Precicion Matrices*. Oxford University Press on behalf of the Royal Astronomical Society.

Pawestri, A.P. 2012. Penaksiran Parameter Univariate *Partial Least Squares Regression* Menggunakan Algoritma NIPALS. *Skripsi*. Universitas Indonesia, Jakarta.

Santoso, S. 2004. *SPSS Statistik Multivariat*. Jakarta : PT. Elex Media Komputindo.

Wold, H. 1982. *Soft modeling: the basic desing and some extensions*, I*n: Systems under Indirect Observation*, Part 2, Jöreskog K.G., Wold H. (eds). North Hollad, 1-5.

Wright, K. 2017. *The NIPALS Algorithm*. Diakses pada 7 Maret 2018. https://cran.r-project.org/web/packages/nipals/vignettes/nipals_algorithm.pdf