

**IMPLEMENTASI ALGORITMA SAX DAN *RANDOM PROJECTION* UNTUK  
*TIME SERIES MOTIF DISCOVERY* PADA *BIG DATA PLATFORM*  
STUDI KASUS: RESONANSI ELEMEN ORBITAL ASTEROID**

SKRIPSI

Diajukan untuk Memenuhi Sebagian dari  
Syarat Memperoleh Gelar Sarjana Komputer  
Program Studi Ilmu Komputer



oleh  
Muhammad Naufal Fazanadi  
1505439

**PROGRAM STUDI ILMU KOMPUTER  
DEPARTEMEN PENDIDIKAN ILMU KOMPUTER  
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS PENDIDIKAN INDONESIA  
2019**

**IMPLEMENTASI ALGORITMA SAX DAN *RANDOM PROJECTION* UNTUK  
*TIME SERIES MOTIF DISCOVERY* PADA *BIG DATA PLATFORM*  
STUDI KASUS: RESONANSI ELEMEN ORBITAL ASTEROID**

Oleh  
Muhammad Naufal Fazanadi  
NIM 1505439

Sebuah Skripsi yang Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh  
Gelar Sarjana Komputer di Fakultas Pendidikan Matematika dan Ilmu  
Pengetahuan Alam

© Muhammad Naufal Fazanadi 2019  
Universitas Pendidikan Indonesia  
Juni 2019

Hak Cipta Dilindungi Undang-Undang  
Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak  
ulang, difoto kopi, atau cara lainnya tanpa izin dari penulis

**MUHAMMAD NAUFAL FAZANADI**

1505439

**IMPLEMENTASI ALGORITMA SAX DAN *RANDOM PROJECTION* UNTUK  
*TIME SERIES MOTIF DISCOVERY* PADA *BIG DATA PLATFORM*  
STUDI KASUS: RESONANSI ELEMEN ORBITAL ASTEROID**

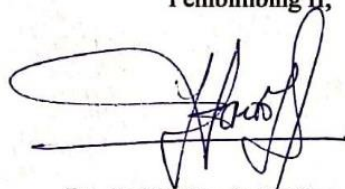
DISETUJUI DAN DISAHKAN OLEH PEMBIMBING:

Pembimbing I,



**Lala Septem Riza, M.T., Ph.D.**  
NIP. 197809262008121001

Pembimbing II,



**Dr. Judhistira Aria Utama, M.Si.**  
NIP. 197703312008121001

Mengetahui,

Ketua Departemen Pendidikan Ilmu Komputer



**Prof. Dr. H. Munir, M.IT.**  
NIP. 19660325200112100

## PERNYATAAN

Dengan ini penulis menyatakan bahwa skripsi dengan judul “Implementasi Algoritma SAX dan *Random projection* untuk *Time series Motif discovery* pada *Big Data Platform* Studi Kasus: Resonansi Elemen Orbital Asteroid” ini beserta seluruh isinya adalah benar-benar karya penulis sendiri. Penulis tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, penulis siap menanggung risiko/sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya penulis ini.

Bandung, Juni 2019

Yang Membuat Pernyataan,

Muhammad Naufal Fazanadi

NIM 150549

**IMPLEMENTASI ALGORITMA SAX DAN *RANDOM PROJECTION* UNTUK  
*TIME SERIES MOTIF DISCOVERY* PADA *BIG DATA PLATFORM*  
STUDI KASUS: RESONANSI ELEMEN ORBITAL ASTEROID**

Oleh

Muhammad Naufal Fazanadi — mnfnaufal@gmail.com

1505439

**ABSTRAK**

Fenomena *Big Data* telah terjadi pada banyak bidang pengetahuan, salah satunya adalah pada bidang astronomi. Data yang memiliki jumlah yang sangat banyak pada bidang astronomi salah satunya adalah data resonansi elemen orbital asteroid. Data tersebut dapat diolah sehingga para ilmuwan dapat mencari *mean motion resonance* pada suatu partikel asteroid untuk mengetahui pada tahun keberapa asteroid akan beresonansi dengan planet planet tertentu. Untuk itu, penelitian ini membuat sebuah model komputasi untuk mendapatkan *mean motion resonance* secara cepat dan efektif dengan memodifikasi dan mengimplementasikan algoritma SAX dan algoritma *motif discovery random projection* pada *Big Data Platform* yaitu Apache Hadoop dan Apache Spark. Hasil penelitian ini menunjukkan adanya percepatan yang sangat signifikan antara penggunaan *stand alone* dan penggunaan *Big Data platform* dengan perancangan 2 skenario. Skenario pertama yaitu penggunaan cluster dengan 4 *cores* dan beberapa *worker node* dan skenario kedua yaitu penggunaan cluster dengan 2 *worker node* dan beberapa jumlah *core*. Penelitian ini juga membuktikan bahwa model komputasi yang dibangun dibandingkan dengan kelauran software SwiftVis mendapatkan rata-rata akurasi sebesar 83%.

Kata Kunci: *Big Data*, algoritma SAX, algoritma *random projection*, *time series*, *motif discovery*, *mean motion resonance*

**IMPLEMENTATION OF SAX AND RANDOM PROJECTION ALGORITHM FOR  
TIME SERIES MOTIF DISCOVERY ON BIG DATA PLATFORM  
CASE STUDY: RESONANCE ELEMENT ORBITAL ASTEROID**

*Arranged by*

*Muhammad Naufal Fazanadi — mnfnaufal@gmail.com*

*1505439*

**ABSTRACT**

*The Big Data phenomenon has occurred in many fields of knowledge, one of them is in the field of astronomy. One of data that has a very large amount in the astronomy field is the resonance data of asteroid orbital elements. The data can be processed so that scientists can find the mean motion resonance in an asteroid particle to find out in what year the asteroid will resonate with a particular planet. But now, it cannot be denied that to process the data with a large amount of data will take a lot of time. For this reason, this research makes a computational model to get mean motion resonance quickly and effectively by modifying and implementing the SAX algorithm and motif discovery random projection algorithm on the Big Data platform, using Apache Hadoop and Apache Spark. The results of this study indicate a very significant acceleration between standalone use and the use of Big Data platforms by designing 2 scenarios. The first scenario is the use of clusters with 4 cores and several worker nodes and the second scenario is the use of clusters with 2 worker nodes and a number of cores. This study also proves that the computational model compared to the result from SwiftVis software gets an average accuracy of 83%.*

*Keywords: Big Data, SAX Algorithm, random projection algorithm, time series, motif discovery, mean motion resonance.*

## KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Allah swt. karena hanya dengan kehendak, berkat, serta karunia-Nya lah penulis dapat menyelesaikan skripsi yang berjudul “Implementasi Algoritma SAX dan *Random projection* untuk *Time series Motif discovery* pada *Big Data Platform* Studi Kasus: Resonansi Elemen Orbital Asteroid” ini dapat terselesaikan.

Penyusunan skripsi ini ditunjukan untuk memenuhi dan melengkapi salah satu syarat untuk penyusunan skripsi yang merupakan syarat untuk mendapatkan gelar sarjana komputer atas jenjang studi S1 pada Program Studi Ilmu Komputer Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam Universitas Pendidikan Indonesia.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan dan keterbatasan yang perlu disempurnakan. Oleh karena itu, penulis sangat mengharapkan saran maupun kritik yang membangun agar tidak terjadi kesalahan yang sama dikemudian hari dan dapat meningkatkan kualitas ke tahap lebih baik.

Bandung, Juni 2019

Penulis

## UCAPAN TERIMAKASIH

Alhamdulillahirabilalamin, puji dan syukur kehadirat Allah SWT Yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis diberikan kelancaran dalam menyelesaikan penulisan skripsi ini. Dalam proses menyelesaikan penelitian dan penyusunan skripsi ini, peneliti banyak mendapat bimbingan, dorongan, serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini peneliti mengucapkan terimakasih serta penghargaan yang setinggi-tingginya, kepada:

1. Kedua orang tua yaitu Ruswandi dan Rina Fatimatusari yang selalu memberikan doa dan dukungan moral dan materil, serta selalu menjadi penyemangat utama dalam menempuh pendidikan tinggi sehingga penulis dapat menyelesaikan skripsi ini.
2. Bapak Lala Septem Riza, M.T., Ph.D. selaku pembimbing I atas segala waktu yang dicurahkan untuk membimbing penulis demi terselesaikannya skripsi ini.
3. Bapak Dr. Judhistira Aria Utama, M.Si., selaku pembimbing II yang telah memberikan saran kepada penulis selama proses penyelesaian penelitian dan penulisan skripsi.
4. Bapak Prof. Dr. H. Munir, M.IT., selaku Kepala Departemen Pendidikan Ilmu Komputer FPMIPA Universitas Pendidikan Indonesia.
5. Bapak Eddy Prasetyo Nugroho, M.T., selaku Ketua Program Studi Ilmu Komputer.
6. Ibu Rosa Ariani Sukamto, M.T., selaku dosen pemrograman yang telah memberikan motivasi untuk penulis pada saat awal kuliah.
7. Bapak dan Ibu Dosen Prodi Pendidikan Ilmu Komputer dan Ilmu Komputer yang telah berbagi ilmu yang sangat bermanfaat kepada penulis.
8. Sahabat kelas C, Ammar, Farhan, Burhan, Rendy, dan Rizki yang senantiasa memberikan dukungan, semangat, canda dan tawa kepada penulis baik selama proses perkuliahan maupun selama proses pengerjaan skripsi ini.
9. Farhan dan teh Fidel selaku grup bimbingan dan topik *Big Data* pertama dengan pak Lala yang telah melalui pahit manis penelitian ini bersama, juga selalu memberi dukungan, semangat, dan melengkapi satu sama lain.



10. Pengurus Inti BEM KEMAKOM 2017/2018 yang telah berjuang bersama sama dalam menyebarkan kebaikan untuk KEMAKOM.
11. DEPKOMINFO BEM KEMAKOM 2016/2017, yang telah membentuk keluarga baru di KEMAKOM pada tahun pertama penulis berorganisasi, serta memberikan kedekatan luar biasa hingga saat ini.
12. Kelas C 2015, yang sama-sama berjuang dari awal perkuliahan dari awal hingga ke titik akhir perkuliahan.
13. Tim Variable, Teddy, Desphi, Ammar, Ziady, Fawaz, dan Dzikri yang selalu memberikan semangat dan pekerjaan tambahan selama masa skripsi.
14. Semua pihak yang telah membantu peneliti dalam menyelesaikan skripsi ini yang tidak dapat peneliti sebutkan satu persatu.

Semoga semua amal baik yang telah diberikan kepada penulis mendapatkan balasan yang berlipat dari Allah SWT. Aamiin.

Bandung, Juni 2019

Muhammad Naufal Fazanadi

## DAFTAR ISI

ABSTRAK .....	i
<i>ABSTRACT</i> .....	ii
KATA PENGANTAR .....	iii
UCAPAN TERIMAKASIH.....	iv
DAFTAR ISI.....	vi
DAFTAR TABEL.....	x
DAFTAR GAMBAR .....	xii
BAB I PENDAHULUAN .....	<b>Error! Bookmark not defined.</b>
1.1 Latar Belakang .....	<b>Error! Bookmark not defined.</b>
1.2 Rumusan Masalah .....	<b>Error! Bookmark not defined.</b>
1.3 Tujuan Penelitian.....	<b>Error! Bookmark not defined.</b>
1.4 Manfaat Penelitian.....	<b>Error! Bookmark not defined.</b>
1.5 Batasan Masalah.....	<b>Error! Bookmark not defined.</b>
1.6 Sistematika Penulisan.....	<b>Error! Bookmark not defined.</b>
BAB II KAJIAN PUSTAKA .....	<b>Error! Bookmark not defined.</b>
2.1 <i>Big Data</i> .....	<b>Error! Bookmark not defined.</b>
2.1.1 Definisi <i>Big Data</i> .....	<b>Error! Bookmark not defined.</b>
2.1.2 Karakteristik <i>Big Data</i> .....	<b>Error! Bookmark not defined.</b>
2.1.3 Apache Hadoop.....	<b>Error! Bookmark not defined.</b>
2.1.4 Apache Spark .....	<b>Error! Bookmark not defined.</b>
2.1.5 Spark SQL.....	<b>Error! Bookmark not defined.</b>
2.2 <i>Time series</i> .....	<b>Error! Bookmark not defined.</b>
2.3 <i>Motif discovery</i> .....	<b>Error! Bookmark not defined.</b>
2.3.1 <i>Time series Motif discovery</i> .....	<b>Error! Bookmark not defined.</b>

2.3.2	Algoritma SAX .....	<b>Error! Bookmark not defined.</b>
2.3.3	Algoritma <i>Random Projection</i> .....	<b>Error! Bookmark not defined.</b>
2.4	<i>Cloud Computing</i> .....	<b>Error! Bookmark not defined.</b>
2.4.1	Definisi <i>Cloud Computing</i> .....	<b>Error! Bookmark not defined.</b>
2.4.2	Karakteristik dan Jenis Layanan <i>Cloud Computing</i> .....	<b>Error! Bookmark not defined.</b>
2.4.3	Google Cloud Platform .....	<b>Error! Bookmark not defined.</b>
2.5	Astronomi.....	<b>Error! Bookmark not defined.</b>
2.5.1	<i>Big Data</i> pada Astronomi .....	<b>Error! Bookmark not defined.</b>
2.5.2	Elemen Orbital Asteroid .....	<b>Error! Bookmark not defined.</b>
2.5.3	<i>Mean Motion Resonance</i> .....	<b>Error! Bookmark not defined.</b>
2.6	Penelitian Terkait .....	<b>Error! Bookmark not defined.</b>
BAB III METODOLOGI PENELITIAN.....		<b>Error! Bookmark not defined.</b>
3.1	Desain Penelitian .....	<b>Error! Bookmark not defined.</b>
3.2	Alat dan Bahan Penelitian .....	<b>Error! Bookmark not defined.</b>
3.2.1	Alat Penelitian.....	<b>Error! Bookmark not defined.</b>
3.2.2	Bahan Penelitian.....	<b>Error! Bookmark not defined.</b>
3.3	Metode Penelitian.....	<b>Error! Bookmark not defined.</b>
3.3.1	Metode Pengumpulan Data.....	<b>Error! Bookmark not defined.</b>
3.3.2	Metode Pengembangan Perangkat Lunak.....	<b>Error! Bookmark not defined.</b>
BAB IV HASIL DAN PEMBAHASAN .....		<b>Error! Bookmark not defined.</b>
4.1	Pengumpulan Data .....	<b>Error! Bookmark not defined.</b>
4.2	Perancangan Model .....	<b>Error! Bookmark not defined.</b>
4.2.1	<i>Standalone Python</i> .....	<b>Error! Bookmark not defined.</b>
4.2.2	<i>Cluster</i> dengan <i>Big Data Platform</i> ....	<b>Error! Bookmark not defined.</b>

4.3	Pengembangan Perangkat Lunak <i>Standalone</i>	<b>Error! Bookmark not defined.</b>
4.3.1	Analisa.....	<b>Error! Bookmark not defined.</b>
4.3.2	Desain.....	<b>Error! Bookmark not defined.</b>
4.3.3	Implementasi .....	<b>Error! Bookmark not defined.</b>
4.3.4	Pengujian.....	<b>Error! Bookmark not defined.</b>
4.4	Pengembangan Perangkat Lunak <i>Cluster</i>	<b>Error! Bookmark not defined.</b>
4.4.1	Analisa.....	<b>Error! Bookmark not defined.</b>
4.4.2	Desain.....	<b>Error! Bookmark not defined.</b>
4.4.3	Implementasi .....	<b>Error! Bookmark not defined.</b>
4.4.4	Pengujian.....	<b>Error! Bookmark not defined.</b>
4.5	Rancangan Skenario Eksperimen .....	<b>Error! Bookmark not defined.</b>
4.5.1	Skenario 4 <i>Core</i> dan Beberapa <i>Worker Nodes</i>	<b>Error! Bookmark not defined.</b>
4.5.2	Skenario Beberapa <i>Core</i> dan 2 <i>Worker Nodes</i>	<b>Error! Bookmark not defined.</b>
4.6	Langkah Menjalankan Eksperimen .....	<b>Error! Bookmark not defined.</b>
4.7	Hasil Eksperimen .....	<b>Error! Bookmark not defined.</b>
4.7.1	Hasil Eksperimen 4 <i>Core</i> dan Beberapa <i>Worker Nodes</i> .....	<b>Error! Bookmark not defined.</b>
4.7.2	Hasil Eksperimen Beberapa <i>Core</i> dan 2 <i>Worker Nodes</i> .....	<b>Error! Bookmark not defined.</b>
4.7.3	Hasil Keluaran Program .....	<b>Error! Bookmark not defined.</b>
4.8	Analisa.....	<b>Error! Bookmark not defined.</b>
4.8.1.	Perbandingan Kecepatan dengan 4 <i>core</i> dan Beberapa <i>Worker Nodes</i> .....	<b>Error! Bookmark not defined.</b>

4.8.2.	Perbandingan Kecepatan dengan Beberapa <i>Core</i> dan 2 <i>Worker Nodes</i> .....	<b>Error! Bookmark not defined.</b>
4.8.3.	Perbandingan Kecepatan dengan Software SwiftVis.....	<b>Error! Bookmark not defined.</b>
4.8.4.	Perbandingan Akurasi dengan Software Swiftvis	<b>Error! Bookmark not defined.</b>
4.8.5.	Perbandingan Fitur dengan Software SwiftVis	<b>Error! Bookmark not defined.</b>
BAB V KESIMPULAN DAN SARAN.....		<b>Error! Bookmark not defined.</b>
5.1	Kesimpulans .....	<b>Error! Bookmark not defined.</b>
5.2	Saran .....	<b>Error! Bookmark not defined.</b>
DAFTAR PUSTAKA .....		144
LAMPIRAN .....		<b>Error! Bookmark not defined.</b>

## DAFTAR TABEL

Tabel 1.1 Empat domain <i>Big Data</i> pada 2025 .....	2
Tabel 2.1 Fungsi <i>transformations</i> dalam Spark .....	25
Tabel 2.2 Fungsi <i>actions</i> dalam Spark .....	27
Tabel 2.3 Contoh tabel hasil konversi k-mers .....	46
Tabel 2.4. Elemen Orbit .....	57
Tabel 4.1 Daftar data resonansi elemen orbital asteroid .....	74
Tabel 4.2 Daftar data yang digunakan pada <i>standalone</i> .....	75
Tabel 4.3 Contoh beberapa baris data yang akan digunakan .....	76
Tabel 4.4 Nama pada 10 partikel asteroid pertama .....	77
Tabel 4.5 Contoh hasil konversi k-mers .....	83
Tabel 4.6 Perancangan fungsi beserta kegunaan .....	93
Tabel 4.7 Pengujian <i>error handling</i> pada program di <i>standalone</i> .....	100
Tabel 4.8 Perancangan fungsi beserta kegunaan .....	102
Tabel 4.9 Pengujian <i>error handling</i> pada program di <i>cluster</i> .....	108
Tabel 4.10 Skenario eksperimen 4 <i>core</i> dan beberapa <i>worker nodes</i> .....	110
Tabel 4.11 Skenario eksperimen beberapa <i>core</i> dan 2 <i>worker nodes</i> .....	110
Tabel 4.12 Hasil eksperimen 4 <i>core</i> dan beberapa <i>worker nodes</i> seluruh <i>file</i> ....	113
Tabel 4.13 Detail hasil eksperimen 4 <i>core</i> dan beberapa <i>worker nodes</i> setiap <i>file</i> .....	114
Tabel 4.14 Hasil eksperimen beberapa <i>core</i> dan 2 <i>worker nodes</i> seluruh <i>file</i> ....	115
Tabel 4.15 Detail hasil eksperimen beberapa <i>core</i> dan 2 <i>worker nodes</i> setiap <i>file</i> .....	115
Tabel 4.16 Hasil keluaran program .....	117
Tabel 4.17 Beban <i>job</i> pada setiap <i>worker nodes</i> dan <i>core</i> pada penggunaan 15 <i>worker nodes</i> .....	127
Tabel 4.18 Beban <i>job</i> pada setiap <i>worker nodes</i> setelah selesai pada penggunaan 15 <i>worker nodes</i> .....	128
Tabel 4.19 Beban <i>job</i> pada setiap <i>worker nodes</i> dan <i>core</i> pada penggunaan 48 <i>core</i> .....	132

Tabel 4.20 Beban <i>job</i> pada setiap <i>worker nodes</i> setelah selesai pada penggunaan 48 <i>core</i> .....	133
Tabel 4.21 Perbandingan kecepatan skenario pertama dengan SwiftVis .....	134
Tabel 4.22 Perbandingan kecepatan skenario kedua dengan SwiftVis .....	134
Tabel 4.23 Perbandingan hasil eksperimen dengan SwiftVis .....	137
Tabel 4.24 Perbandingan fitur program dengan SwiftVis .....	140

## DAFTAR GAMBAR

Gambar 2.1 3V <i>Big Data</i> .....	9
Gambar 2.2 Komponen dan arsitektur Hadoop .....	11
Gambar 2.3 Contoh proses MapReduce .....	12
Gambar 2.4 Ekosistem Hadoop .....	15
Gambar 2.5 HDFS <i>web UI</i> .....	21
Gambar 2.6 <i>Resources manager web UI</i> .....	21
Gambar 2.7 Komponen Spark .....	22
Gambar 2.8 Struktur Spark .....	23
Gambar 2.9 <i>Lazy evaluation</i> pada Apache Spark .....	24
Gambar 2.10 <i>Web UI</i> pada Apache Spark .....	30
Gambar 2.11 Antarmuka ke Spark SQL, dan interaksi dengan Spark .....	33
Gambar 2.12 Seri indeks harga gandum Beveridge dari tahun 1810 hingga 1864 .....	36
Gambar 2.13 <i>Time series Subsequence Motif</i> .....	38
Gambar 2.14 <i>Time series</i> astronomikal yang mengandung 3 <i>subsequences</i> yang hampir identik .....	39
Gambar 2.15 <i>Flowchart</i> algoritma SAX .....	40
Gambar 2.16 Plot probabilitas normal dari distribusi nilai dari urutan panjang 128 dari 8 dataset berbeda.....	41
Gambar 2.17 <i>Lookup table</i> yang berisi <i>breakpoints</i> yang membagi distribusi Gaussian dari 3 hingga 10 .....	41
Gambar 2.18 <i>Pseudocode</i> algoritma SAX .....	42
Gambar 2.19 <i>Time series</i> yang telah didiskritisasi menjadi simbol SAX .....	43
Gambar 2.20 <i>Flowchart</i> algoritma <i>random projection</i> .....	44
Gambar 2.21 <i>Pseudocode</i> algoritma <i>random projection</i> .....	45
Gambar 2.22 Contoh 10 <i>bucket</i> 1-mers pada <i>random projection</i> .....	48
Gambar 2.23 Karakteristik penting dari <i>cloud computing</i> .....	50
Gambar 2.24 Macam macam layanan pada <i>cloud computing</i> .....	52
Gambar 2.25 Ilustrasi elemen-elemen sudut dalam orbit Keplerian .....	58
Gambar 2.26 Contoh data <i>time series</i> dari resonansi elemen orbital asteroid .....	60



Gambar 2.27 Celah Kirkwood di sabuk asteroid yang menandai lokasi orbit yang tidak stabil .....	62
Gambar 2.28 Populasi asteroid koorbital (MMR 1:1) yang menjadi Trojan planet Jupiter .....	63
Gambar 2.29 Arsitektur sistem pada penelitian .....	64
Gambar 3.1 Desain penelitian .....	67
Gambar 3.2 Model <i>Waterfall</i> .....	72
Gambar 4.1 Model komputasi pada <i>standalone</i> Python .....	78
Gambar 4.2 <i>Pseudocode</i> keseluruhan pada model komputasi <i>standalone</i> Python .....	79
Gambar 4.3 <i>Pseudocode</i> proses <i>preprocessing</i> .....	80
Gambar 4.4 <i>Pseudocode</i> proses normalisasi data .....	80
Gambar 4.5 Hasil plot dari 2 data <i>time series</i> .....	81
Gambar 4.6 Plot dari <i>time series</i> indeks ke 69 hingga 82 .....	82
Gambar 4.7 <i>Bucket</i> l-mers pada <i>random projection</i> .....	84
Gambar 4.8 Contoh penggabungan motif yang berderet dari plot <i>time series</i> .....	85
Gambar 4.9 <i>Pseudocode</i> proses <i>preprocessing</i> .....	85
Gambar 4.10 Model komputasi pada <i>cluster</i> dengan <i>big data platform</i> .....	86
Gambar 4.11 <i>Pseudocode</i> keseluruhan pada model komputasi <i>cluster</i> dengan <i>big data platform</i> .....	87
Gambar 4.12 Arsitektur sistem dengan <i>big data platform</i> dan Google Cloud Platform .....	90
Gambar 4.13 Kode program untuk <i>import library</i> dan fungsi .....	93
Gambar 4.14 Kode program untuk proses utama hingga <i>output</i> program .....	94
Gambar 4.15 Kode program untuk proses <i>input</i> dan praproses data .....	95
Gambar 4.16 Kode program untuk fungsi z normalisasi .....	95
Gambar 4.17 Kode program untuk fungsi konversi indeks menjadi char .....	96
Gambar 4.18 Kode program untuk fungsi daftar <i>breakpoints</i> pada SAX .....	96
Gambar 4.19 Kode program untuk fungsi SAX .....	97
Gambar 4.20 Kode program untuk fungsi <i>random projection</i> .....	97
Gambar 4.21 Kode program untuk fungsi pengecekan resonansi 1:1 .....	98
Gambar 4.22 Kode program untuk fungsi <i>postprocessing</i> data .....	99

Gambar 4.23 Kode program untuk inisiasi <i>library</i> dan <i>SparkSession</i> .....	103
Gambar 4.24 Kode program untuk inisiasi informasi <i>cluster</i> dan variabel <i>input</i> .....	104
Gambar 4.25 Kode program untuk mendapatkan semua nama <i>file</i> dalam HDFS .....	104
Gambar 4.26 Kode program untuk <i>import</i> semua <i>file</i> dalam HDFS .....	104
Gambar 4.27 Kode program untuk <i>preprocessing data</i> di Apache Spark .....	105
Gambar 4.28 Kode program untuk proses pembuatan UDF untuk fungsi <i>motif discovery</i> .....	106
Gambar 4.29 Kode program proses utama dan menyimpan <i>output</i> .....	106
Gambar 4.30 Kode program untuk menyimpan data ke dalam HDFS .....	107
Gambar 4.31 Plot <i>time series</i> dengan ParticleID 3 beserta dengan <i>breakpoints</i> karakter SAX.....	122
Gambar 4.32 Plot <i>time series</i> dengan ParticleID 3 beserta lokasi resonansi yang yang ditemukan .....	122
Gambar 4.33 Plot <i>time series</i> dengan ParticleID 6 beserta dengan <i>breakpoints</i> karakter SAX.....	123
Gambar 4.34 Plot <i>time series</i> dengan ParticleID 6 beserta lokasi resonansi yang yang ditemukan .....	123
Gambar 4.35 Profil <i>cost time</i> antara <i>workers</i> terhadap waktu .....	124
Gambar 4.36 Profil <i>cost time</i> untuk <i>file</i> Batch1_1000Asteroid_20JutaTahun ...	125
Gambar 4.37 Profil <i>cost time</i> untuk <i>file</i> Batch2_1000Asteroid_5JutaTahun .....	126
Gambar 4.38 Profil <i>cost time</i> untuk <i>file</i> Batch3_1000Asteroid_8JutaTahun .....	126
Gambar 4.39 Profil <i>cost time</i> untuk <i>file</i> Batch4_372Asteroid_10,5JutaTahun ..	127
Gambar 4.40 Profil <i>cost time</i> antara total <i>cores</i> terhadap waktu .....	129
Gambar 4.41 Profil <i>cost time</i> untuk <i>file</i> Batch1_1000Asteroid_20JutaTahun ...	130
Gambar 4.42 Profil <i>cost time</i> untuk <i>file</i> Batch2_1000Asteroid_5JutaTahun .....	131
Gambar 4.43 Profil <i>cost time</i> untuk <i>file</i> Batch3_1000Asteroid_8JutaTahun .....	131
Gambar 4.44 Profil <i>cost time</i> untuk <i>file</i> Batch4_372Asteroid_10,5JutaTahun ..	132
Gambar 4.45 Profil untuk perbandingan kecepatan skenario pertama dengan SwiftVis .....	135

Gambar 4.46 Profil untuk perbandingan kecepatan skenario kedua dengan SwiftVis  
.....136

## DAFTAR PUSTAKA

- Armbrust, M., Xin, R. S., Lian, C., Huai, Y., Liu, D., Bradley, J. K., ... Berkeley, U. C. (2015). Spark SQL : Relational Data Processing in Spark, 1383–1394.
- Ashraf, F. B., Abir, A. I., Salekin, M. S., & Mottalib, M. A. (2017, February). RPPMD (Randomly projected possible *motif discovery*): An efficient *bucketing* method for finding DNA planted Motif. In E
- Baldini, G., Giuliani, R., Steri, G., Sanchez, I., & Gentile, C. (2017, November). The application of the symbolic aggregate approximation algorithm (SAX) to radio frequency fingerprinting of IoT devices. In 2017 IEEE Symposium on Communications and Vehicular Technology (SCVT) (pp. 1-6). IEEE.
- Beyer, M. A., & Laney, D. (2012). The importance of 'big data': a definition. Stamford,CT:Gartner,2014-2018.
- Bottke, W. F., Morbidelli, A., Jedicke, R., Petit, J. M., Levison, H. F., Michel, P., & Metcalfe, T. S. (2002). Debaised orbital and absolute magnitude distribution of the near-Earth objects. *Icarus*, 156(2), 399-433.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- Buhler, J., & Tompa, M. (2002). Finding motifs using *random projections*. *Journal of computational biology*, 9(2), 225-242.
- Chandrasekaran, K. (2015). *Essentials of Cloud computing*. CRC Press.
- Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified Data Processing on Large Clusters. *Proceedings of 6th Symposium on Operating Systems Design and Implementation*, 137–149. <https://doi.org/10.1145/1327452.1327492>
- Dhiba, Tyas Farrah. (2018). Planted Motif Search dalam Sekuens DNA

Menggunakan Algoritma *Random projection* pada R High Performance Computing Package. Universitas Pendidikan Indonesia.

Fechner, G. T., Howes, D. H., & Boring, E. G. (1966). Elements of psychophysics (Vol. 1). New York: Holt, Rinehart and Winston.

Feigelson, E. D., & Babu, G. J. (2012). Big data in astronomy. *Significance*, 9(4), 22–25. <https://doi.org/10.1111/j.1740-9713.2012.00587.x>

Forgács-Dajka, E., Sándor, Z., & Érdi, B. (2018). A fast method to identify mean motion resonances. *Monthly Notices of the Royal Astronomical Society*, 477(3), 3383-3389.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>

Gerhardt, B., Griffin, K., & Klemann, R. (2012). Unlocking Value in the Fragmented World of Big Data Analytics. *How Information Infomediaries Will Create a New Data Ecosystem*, Cisco Internet Business Solutions Group (IBSG), (June), 2013–2017.

Ghemawat, S., Gobioff, H., & Leung, S.-T. (2003). *The Google file system. Proceedings of the nineteenth ACM symposium on Operating systems principles - SOSP '03*. <https://doi.org/10.1145/945449.945450>

Intel IT Center. (2012). Steps IT Managers Can Take to Move Forward with Apache Hadoop\* Software. *Planning Guide : Getting Started with Big Data Steps*, (June). Retrieved from <http://www.intel.com/content/www/us/en/big-data/getting-started-with-hadoop-planning-guide.html>

Jones, N. C., & Pevzner, P. (2004). An introduction to bioinformatics algorithms. MIT press.

Levison, H.F. dan Duncan, M.J. (1994): The long-term dynamical behavior of short-period comets, *Icarus*, **108**, 18 – 36.

- Lin, J., Keogh, E., Lonardi, S., & Chiu, B. (2003). A Symbolic Representation of *Time series*, with Implications for Streaming Algorithms, 2–11.
- Lin, J., Keogh, E., Lonardi, S., & Patel, P. (2002). Finding motifs in *time series*. *Proc. of the 2nd Workshop on Temporal Data Mining*, 53–68. <https://doi.org/10.1.1.19.6629>
- Lin, J., & Lonardi, S. (2017). Experiencing SAX : a Novel Symbolic Representation of *Time series*.
- Liu, B., Li, J., Chen, C., Tan, W., Member, S., Chen, Q., & Zhou, M. (2015). Efficient *Motif discovery* for Large-Scale *Time series* in Healthcare, *11*(3), 583–590.
- Martinez, H. M. (1983). An efficient method for finding repeats in molecular sequences. *Nucleic acids research*, *11*(13), 4629-4634.
- Mehta, S., & Mehta, V. (2016). Hadoop Ecosystem : An Introduction, *5*(6), 557–562. <https://doi.org/10.21275/v5i6.NOV164121>
- Mishra, A. K., Hellerstein, J. L., Cirne, W., & Das, C. R. (2010). Towards characterizing *cloud* backend workloads: insights from Google compute clusters. *ACM SIGMETRICS Performance Evaluation Review*, *37*(4), 34-41.
- Mueen, A., Keogh, E., Zhu, Q., Cash, S., & Westover, B. (2009). Exact Discovery of *Time series* Motifs. *Proceedings of the 2009 SIAM International Conference on Data Mining*, 473–484.
- Murray, C.D. dan Dermott, S.F. (1999): *Solar System Dynamics*. New York: Cambridge University Press
- Ohri, A. (2014). *R for Cloud computing*. springer.
- Pariwatthanasak, K., & Ratanamahatana, C. A. (n.d.). *Approximated Matrix Profile*. Springer Singapore. <https://doi.org/10.1007/978-981-13-1165-9>
- Pevzner, P. A., & Sze, S. H. (2000, August). Combinatorial approaches to finding

subtle signals in DNA sequences. In ISMB (Vol. 8, pp. 269-278).

- Rani, C. S. (2017). MapReduce with Hadoop for Simplified Analysis of Big Data, 8(5), 2015–2018.
- Riza, L. S., Utama, J. A., & Putra, S. M. (2018). SCIENCE & TECHNOLOGY Parallel Exponential Smoothing Using the Bootstrap Method in R for Forecasting Asteroid 's Orbital Elements, 26(1), 441–462.
- Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 42–47.
- Scott, J. A. (2015). Getting Started with Apache Spark, 84.
- Sommerville, I. (2011). Software Engineering. Addison-Wesley.
- Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., ... Robinson, G. E. (2015). Big data: Astronomical or genetical? *PLoS Biology*, 13(7), 1–11. <https://doi.org/10.1371/journal.pbio.1002195>
- Tanaka, Y., Iwamoto, K., & Uehara, K. (2005). Discovery of time-series motif from multi-dimensional data based on MDL principle. *Machine Learning*, 58(2–3), 269–300. <https://doi.org/10.1007/s10994-005-5829-2>
- Troester, M. (2015). Big Data Meets Big Data Analytics. Retrieved from [http://www.informatik-aktuell.de/entwicklung/methoden/big-data-meets-big-data.html?utm\\_source=CleverReach&utm\\_medium=email&utm\\_campaign=02-11-2015+Newsletter+19-2015&utm\\_content=Mailing\\_6367514](http://www.informatik-aktuell.de/entwicklung/methoden/big-data-meets-big-data.html?utm_source=CleverReach&utm_medium=email&utm_campaign=02-11-2015+Newsletter+19-2015&utm_content=Mailing_6367514)
- White, T. (2012). Hadoop: The definitive guide. " O'Reilly Media, Inc."
- Zaharia, M., Karau, H., Konwinski, A., & Wendell, P. (2015). *Learning spark: lightning-fast big data analysis*. O'Reilly Media, Inc.
- Zhu, Y., Mueen, A., & Keogh, E. (2018). Admissible *Time series Motif discovery* with Missing Data, (February).