

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pada saat ini perkembangan data sudah tidak dapat dihentikan, hampir setiap orang akan menciptakan atau menggunakan data baru setiap harinya. Sebanyak 5 *exabytes* (10^{18} bytes) data diciptakan oleh manusia hingga 2003. Hari ini jumlah informasi ini dibuat dalam dua hari. Pada tahun 2012, dunia data digital diperluas menjadi 2,72 *zettabyte* (10^{21} byte). Diperkirakan akan berlipat ganda setiap dua tahun, mencapai sekitar 8 *zettabyte* data pada tahun 2015 (Intel IT Center, 2012). Sebuah survei mengungkapkan bahwa Facebook menangani lebih dari 250 juta *upload* foto dan interaksi 800 juta pengguna aktif dengan lebih dari 900 juta objek (*pages, groups, dll.*) setiap harinya (Troester, 2015). Sudah ada 6 miliar langganan seluler di dunia dan pada setiap harinya sebanyak 10 miliar pesan teks dikirim. Pada tahun 2020, 50 miliar perangkat akan terhubung ke jaringan dan internet (Gerhardt, Griffin, & Klemann, 2012). Fenomena tentang data pada saat ini dapat dikembangkan dan dimanfaatkan secara maksimal oleh manusia dengan ilmu *Big Data*.

Big Data adalah istilah yang merepresentasikan fenomena data yaitu aset informasi bervolume tinggi, berkecepatan tinggi, dan beraneka ragam yang menuntut pemrosesan informasi yang hemat biaya dan bentuk-bentuk inovatif dari pemrosesan informasi untuk peningkatan wawasan dan pengambilan keputusan (Beyer & Laney, 2012). Dalam *Big Data* terdapat 3 isu terbesar yaitu 3V, *volume* (kapasitas), *velocity* (kecepatan), dan *variety* (keragaman). Untuk menyelesaikan masalah masalah tersebut para ilmuwan data dan *programmer* pun mengembangkan teknologi *Big Data* agar data yang sangat banyak itu dapat diolah secara cepat dan mendapatkan pengetahuan dari hasil olahan data tersebut.

Fenomena *Big Data* ini pun tidaklah hanya terjadi pada bidang teknologi informasi saja, melainkan terjadi pada bidang lainnya. Contohnya pada bidang biologi dan astronomi dimana bidang tersebut adalah penghasil data yang sangat banyak setelah teknologi informasi. Informasi tersebut dapat dilihat pada tabel 1.1. Permasalahan pada bidang astronomi selain datanya yang sangat banyak adalah

kurangnya efisiensi tenaga yang masih membutuhkan proses manual dalam memproses data dan waktu yang dibutuhkan untuk mengolah data masih terhitung lama. Salah satu data yang memiliki jumlah yang sangat banyak pada bidang astronomi adalah data *time series* resonansi partikel asteroid, yang melibatkan 6 parameter penting dalam komponen Kepler, yaitu a (*Semi-major axis*), e (Eksentrisitas), i (Inklinasi), ω (Argumen dari pericenter), Ω (*Longitude* dari *ascending node*), dan M (rata-rata anomali) (Riza, Utama, & Putra, 2018).

Tabel 1.1 Empat domain Big Data pada 2025 (Stephens et al., 2015).

<i>Data Phase</i>	<i>Astronomy</i>	<i>Twitter</i>	<i>YouTube</i>	<i>Genomics</i>
<i>Acquisition</i>	25 zetta-bytes/year	0,5-15 billion tweets/year	500-900 million hours/year	1 zetta-bases/year
<i>Storage</i>	1 EB/year	1-17 PB/year	1-2 EB/year	2-40 EB/year
<i>Analysis</i>	<i>In situ data reduction</i>	<i>Topic and sentiment mining</i>	<i>Limited requirements</i>	<i>Heterogeneous data and analysis</i>
	<i>Real-time processing</i>	<i>Meta data analysis</i>		<i>Variant calling, ~2 trillion central processing unit (CPU) hours</i>
	<i>Massive volumes</i>			<i>All-pairs genome alignments, ~10.000 trillion CPU hours</i>
<i>Distribution</i>	<i>Dedicated lines from antennae to server (600 TB/s)</i>	<i>Small units of distribution</i>	<i>Major component of modern user's bandwidth (10 MB/s)</i>	<i>Many small (10 MB/s) and fewer massive (10 TB/s) data movement</i>

Dalam ilmu astronomi, diantara orbit Mars dan Jupiter, terdapat populasi asteroid. Mekanisme yang berkontribusi untuk mengantarkan asteroid di *Main Belt* menuju ruang dekat Bumi bisa menjadi tabrakan antara asteroid yang dengan kecepatan tinggi. Sementara di ruang dekat-Bumi, asteroid bisa mengalami pertemuan dekat dengan planet tertentu (Merkurius, Venus, Bumi, dan Mars). Sebagai hasil dari pertemuan dekat ini, asteroid dapat terfragmentasi karena kekuatan pasang yang kuat, atau jika mereka dapat bertahan hidup, orbit mereka dapat berubah secara drastis dalam waktu singkat. Perubahan drastis pada orbit terjadi dapat mengubah nasib asteroid di masa depan, yaitu apakah mereka masih mengorbit Matahari atau bahkan bertabrakan dengan benda-benda besar di tata surya (Riza, Utama, & Putra, 2018).

Untuk memprediksi kapan terjadinya pertemuan dekat asteroid dengan benda-benda besar di Tata Surya di masa depan dapat diketahui dengan mencari 1:1 *mean motion resonance* dari partikel x dengan planet y. *Mean motion resonance* adalah keadaan suatu partikel langit beresonansi dengan suatu planet di sekitarnya. Resonansi yang terjadi oleh partikel langit dengan planet dapat ditentukan perbandingan jaraknya dan kapan partikel tersebut sedang beresonansi. Hal ini dapat ditentukan oleh parameter a (*semi-major axis*) dalam komponen Kepler yang menunjukkan jarak antara pusat partikel asteroid ke pusat matahari sebagai inti dari orbit di tata surya. Sisa parameter lainnya dapat diteliti lebih lanjut dalam fokus lain, seperti e dapat diteliti untuk menemukan resonansi kozai dan dapat diteliti untuk *astro mining*. Ada beberapa cara untuk mendapatkan *mean motion resonance* pada partikel langit, salah satunya dengan metode *motif discovery* pada data *time series*. Setelah waktu resonansi didapatkan maka dapat diolah lebih lanjut apakah partikel tersebut akan menabrak, berputar, atau terlempar dengan planet yang sedang beresonansi.

Permasalahan pada kasus ini adalah belum adanya program dan model komputasi yang tepat, cepat, dan efisien untuk mendapatkan waktu dan partikel yang mengalami 1:1 *mean motion resonance* dengan planet lain. Selama ini peneliti astrofisika menggunakan aplikasi SwiftVis dalam mencari 1:1 *mean motion resonance*. Namun proses pencarian *mean motion resonance* pada data partikel asteroid masih lambat dan cara pencarian resonansi masih menggunakan *plot* visual dari keluaran SwiftVis.

Oleh karena itu, penelitian ini difokuskan pada membangun model komputasi, modifikasi, dan implementasi algoritma *motif discovery* pada *big data platform* yang mampu memberikan hasil pencarian motif dan hasil resonansi dalam keadaan 1:1 yang ditemukan pada data *time series* resonansi orbital asteroid dengan tepat dan cepat pada data yang besar. Penelitian ini menggunakan dua algoritma karena dalam kasus ini membutuhkan algoritma yang dapat mendeteksi motif yang terjadi pada data *time series* tanpa masukan motif dari pengguna. Algoritma yang akan digunakan pada penelitian ini adalah algoritma SAX (Lin et al., 2002) untuk mengubah representasi *time series* dan algoritma *random projection* (Jones & Pevzner, 2004) untuk *motif discovery*. Algoritma SAX memungkinkan untuk

mengubah *time series* menjadi sebuah representasi string dengan cepat. Algoritma ini berkesinambungan dengan algoritma *random projection* dikarenakan *random projection* memerlukan data berbentuk string untuk mendeteksi data motif yang ditemukan pada suatu data. *Random projection* dapat mendeteksi motif tanpa membutuhkan data masukan motif yang ingin dicari oleh pengguna secara cepat.

Untuk mengolah data, penelitian ini akan menggunakan Apache Hadoop dan Apache Spark sebagai *big data platform* dan menggunakan hasil modifikasi algoritma SAX dan algoritma *random projection* untuk memproses *big time series data*. Apache Hadoop akan digunakan sebagai media penyimpanan data yang menggunakan HDFS (*Hadoop Distributed File System*) dikarenakan HDFS dapat menyimpan data yang besar dengan disebar ke dalam beberapa *node* pada *cluster* (White, 2012). Kemudian Apache Spark akan digunakan sebagai media komputasi untuk penelitian ini agar dapat memproses data yang besar secara cepat dengan membagi proses ke dalam beberapa *node* pada *cluster* (Zaharia et. al, 2015).

1.2. Rumusan Masalah

Rumusan masalah yang menjadi fokus pada pembahasan dalam skripsi ini adalah:

1. Bagaimana cara menemukan motif dan 1:1 *mean motion resonance* dalam *time series* dengan efektif?
2. Bagaimana cara menemukan motif dan 1:1 *mean motion resonance* dalam *time series* pada *big data platform*?
3. Bagaimana implementasi algoritma *motif discovery* pada *big data platform*?
4. Bagaimana hasil keluaran program dan hasil analisa perbandingan kecepatan dan akurasi dari setiap eksperimen yang telah dirancang sesuai model yang dibuat dan diimplementasikan?

1.3. Tujuan Penelitian

Dari rumusan masalah yang telah dituliskan sebelumnya, terdapat beberapa tujuan dari penelitian ini adalah sebagai berikut:

1. Merancang model komputasi pada *standalone* Python untuk *motif discovery* dengan menggunakan algoritma SAX dan *random projection*.

2. Memodifikasi model komputasi pada *standalone* untuk *cluster* pada *big data platform* menggunakan Apache Hadoop dan Apache Spark.
3. Mengimplementasikan model komputasi *standalone* dan *cluster* pada Apache Hadoop dan Apache Spark.
4. Melakukan eksperimen dan analisa untuk mendeteksi motif dan 1:1 *mean motion resonance* pada resonansi elemen orbital asteroid yang terkait kecepatan dan akurasi.

1.4. Manfaat Penelitian

Adapun manfaat dalam penelitian ini, diantaranya adalah sebagai berikut:

1. Membuat sebuah program dan model komputasi *cluster* pada *big data platform* untuk dikembangkan oleh peneliti selanjutnya.
2. Mendapatkan algoritma baru dari hasil modifikasi algoritma SAX dan algoritma *random projection* pada Apache Spark.
3. Mengefektifkan dan mengefisienkan waktu serta sumber daya yang dibutuhkan untuk mendeteksi motif dan mencari 1:1 *mean motion resonance* pada resonansi elemen orbital asteroid.

1.5. Batasan Masalah

Batasan masalah sangat diperlukan agar pembahasan tidak terlalu luas, karena studi kasus yang diteliti begitu luas maka dalam hal penelitian menggunakan batas-batas sebagai berikut:

1. Data yang akan diteliti adalah data evolusi orbit benda langit berupa asteroid.
2. Penelitian ini hanya mencari 1:1 *mean motion resonance* berdasarkan motif yang ditemukan.
3. Program ini bekerja untuk data dengan format standar JPLNASA *Small Body Database*.

1.6. Sistematika Penulisan

Pada bagian sistematika penulisan ini akan diuraikan mengenai penjelasan tiap bab.

BAB I PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang diadakannya penelitian, disusun berdasarkan latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan mengenai teori yang dipelajari selama melaksanakan penelitian. Teori yang terdapat pada bab ini ialah mengenai *Big Data*, *time series motif discovery*, algoritma SAX dan *Random projection*, *cloud computing*, elemen orbital asteroid dan *mean motion resonance*, dan penelitian terkait.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan langkah-langkah penelitian yang akan dilakukan, dimulai dari desain penelitian, fokus penelitian, kemudian alat dan bahan yang digunakan untuk penelitian dan yang terakhir adalah metode penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjabarkan hasil penelitian yang telah dilakukan dan analisisnya. Semua pertanyaan mengenai masalah yang diangkat dalam tema skripsi dibahas di sini. Yaitu tentang data penelitian, pengembangan model, implementasi sistem, desain eksperimen, dan hasil dan analisa.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran bagi peneliti selanjutnya dari hasil penelitian yang telah dilakukan.