

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Dewasa ini, polusi cahaya sudah menjadi isu global yang memberikan dampak negatif dari sebagian besar di dunia ini, menyelimuti dunia dengan kabut bercahaya yang menghalangi manusia untuk mengamati galaksi (Falchi et al., 2016). Polusi cahaya dapat diartikan sebagai peningkatan cahaya di malam hari yang disebabkan oleh sumber cahaya antropogenik, yaitu sumber cahaya yang disebabkan oleh tindakan dari manusia (Cinzano et al., 2000). Adapun komponen yang membentuk adanya polusi cahaya tersebut meliputi: *light trespass*, *over-illumination*, *glare*, *clutter*, dan *sky glow* (Rajkhowa, 2014).

Banyaknya cahaya artifisial yang terpancar dari bumi menghambat pengamatan astronomi dari bumi. Polusi cahaya berdampak buruk terhadap lingkungan, meliputi dunia umat manusia, dunia hewan, dan dunia tumbuhan. Bagi manusia, polusi cahaya dapat menimbulkan dampak buruk bagi kesehatan seperti sering sakit kepala, kelelahan, peningkatan stres, penurunan libido, dan peningkatan kecemasan. Kecerahan cahaya buatan di malam hari mengganggu hewan yang bermigrasi, aktif, maupun terbang di malam hari. Bagi tumbuhan, meningkatnya pencahayaan artifisial di malam hari mempengaruhi siklus tumbuhan dalam pola pembungaan dan perkembangannya. Beberapa ilmuwan bahkan berpendapat bahwa polusi cahaya yang terus meningkat dapat menyebabkan terganggunya jaring makanan dan mempengaruhi keseluruhan ekosistem (Rajkhowa, 2014).

*Sky Quality Meter* (SQM) merupakan salah satu alat yang dapat digunakan untuk mengukur tingkat kecerahan langit malam. *Sky Quality Meter* (SQM) telah menjadi perangkat yang paling umum digunakan untuk melacak evolusi kecerahan langit dari daerah yang tercemar sehingga dapat menentukan suatu daerah menjadi observatorium astronomi kelas satu (De Miguel et al., 2017). Fotometer SQM memiliki ketidakpastian sistematis sebesar 10 persen ( $0,1 \text{ mag arcsec}^{-2}$ ). Hasil dari pemantauan menggunakan sensor SQM didefinisikan dalam istilah kecerahan langit malam yaitu MPSAS, *magnitude per unit square arc second* (Kurniawati et

al., 2024). Penggunaan SQM untuk penelitian sangat populer di kalangan peneliti, astronom, serta asosiasi yang tertarik dengan polusi cahaya (De Miguel et al., 2017).

Semakin berkembang zaman, semakin banyak pula database hasil pengukuran SQM dari berbagai tempat di dunia. Salah satu lembaga nasional Badan Riset dan Inovasi Nasional (BRIN) telah melakukan pemantauan kecerahan langit malam di lokasi Timau, Nusa Tenggara Timur dan dapat diakses datanya dalam Repositori Ilmiah Nasional (RIN) Dataverse (Priyatikanto et al., 2022). Selain itu juga terdapat lembaga internasional yang rutin melakukan kegiatan pengukuran kecerahan langit malam yaitu *NSF's National Optical-Infrared Astronomy Research Laboratory* (NSF's NOIRLab) dengan proyeknya *Globe at Night Monitoring Network* (GaN-MN). Proyek *Globe at Night* sudah dimulai sejak tahun 2006, diikuti oleh relawan dari berbagai belahan bumi di dunia yang mengirimkan hasil pemantauan oleh sensor SQM, memberikan ukuran kecerahan langit malam (Barringer et al., 2011). Semua data yang dimiliki oleh GaN-MN tersedia secara gratis untuk publik.

SQM mengukur tingkat kegelapan atau kualitas langit malam. Satuan dari SQM yaitu MPSAS bersifat logaritmik, semakin besar angkanya maka malam tersebut merupakan malam yang gelap (Espey & McCauley, 2014; Hänel et al., 2018). Sebaliknya, semakin kecil angkanya maka malam tersebut merupakan malam yang cerah atau dapat dianggap memiliki polusi cahaya (Riza et al., 2024). Data SQM dengan kualitas langit malam yang bagus kemungkinan besar dihasilkan dari lokasi pedesaan, jauh dari lokasi dengan populasi yang besar (Espey & McCauley, 2014). Normalnya, setiap pusat observatorium mengalami langit malam yang gelap. Jika terdapat malam yang cerah, maka itu dapat dianggap menjadi sebuah anomali atau pencilan. Anomali ini dapat mengindikasikan beberapa peristiwa atau kondisi khusus yang berdampak pada kualitas langit malam pada pusat observatorium tersebut. Anomali dalam data SQM sangat memungkinkan untuk terjadi karena SQM mengambil data kegelapan langit malam secara sekuens sehingga menghasilkan data deret waktu (*time series*).

Deteksi pencilan atau yang seringkali disebut dengan istilah *outlier detection* merupakan metode untuk mencari rangkaian abnormal dari sebuah *series* (Esling &

Agon, 2012). Dalam mendeteksi pencilan, pendekatan yang biasa dilakukan adalah membuat model dari sebuah *series* dan menggolongkan rangkaian berikutnya yang menyimpang terlalu jauh sebagai sebuah pencilan atau anomali. Deteksi pencilan ini juga sering disebut sebagai *anomaly detection* atau deteksi anomali. Sebelumnya, deteksi pencilan ini dimotivasi dari *data cleansing*: menghilangkan *outlier* atau pencilan dari kumpulan data sehingga dapat meningkatkan kualitas *data training*. Namun, kini pencilan semakin diperhatikan karena seringkali mewakili informasi yang menarik dan penting, seperti adanya serangan siber, kesalahan mekanis dari peralatan industri yang rusak, dan sebagainya (Boukerche et al., 2020). Deteksi pencilan ini mulai diminati di kalangan para peneliti di berbagai bidang penelitian dalam *time series data mining* (Blázquez-García et al., 2021). Oleh karena itu, telah banyak penelitian yang dikembangkan untuk pendeteksian anomali pada berbagai skenario kehidupan nyata, di antaranya adalah sistem deteksi penyusupan (Yeung & Chow, 2002; Gwadera et al., 2005; García-Teodoro et al., 2009), *fraud detection* (Bolton & Hand, 2001; Thiprungsri & Vasarhelyi, 2011; Phua et al., 2004), diagnosis anomali medis (Wong et al., 2003), deteksi kesalahan atau cacat peralatan dan produk industri (Fujimaki et al., 2005; Vercruyssen et al., 2018), dan deteksi pencilan dalam arus lalu lintas (Djenouri et al., 2019). Adanya pencilan dapat menandakan perubahan negatif pada sistem, seperti fluktuasi frekuensi rotasi turbin pada mesin jet yang mungkin mengindikasikan kegagalan yang akan segera terjadi (Ahmad et al., 2017).

Deteksi pencilan merupakan hal yang menantang, di antaranya adalah karena kurangnya data berlabel sehingga jarang terjadi anomali dan kemajuan teknologi yang pesat di berbagai bidang membuat data yang tersedia dalam jumlah besar dengan kompleksitas yang tinggi. Tantangan dalam deteksi pencilan adalah untuk dapat mengidentifikasi adanya pencilan dalam data dengan dimensi yang sangat tinggi, dalam aliran data dengan jumlah besar yang tak terbatas, dan data terdistribusi dalam skala besar (Boukerche et al., 2020). Seperti halnya yang telah disampaikan oleh Boukerche et al. (2020), sensor *Sky Quality Meter* (SQM) merekam data secara *real-time* dan terus-menerus sehingga akan ada saat di mana terdapat anomali atau outlier dari data yang telah diambil. Adanya anomali pada

data SQM dapat mengindikasikan terjadinya peristiwa fisis tertentu saat langit sedang direkam oleh sensor.

Penelitian terdahulu terkait deteksi pencilan pada data SQM ini telah dilakukan oleh Putra (2023). Pada penelitiannya, Putra (2023) menggunakan data polusi cahaya langit malam untuk dideteksi anomalnya menggunakan algoritma *Probabilistic Exponential Weighted Moving Average* (PEWMA) dengan *big data platform* Apache Kafka sebagai layanan untuk *data streaming*. Namun, terdapat beberapa kelemahan dalam penelitiannya yaitu proses *streaming* hanya berjalan paling lama selama lima hari, kurangnya analisis dari sisi astrofisika, dan hasil akurasi dari algoritmanya belum optimal. Oleh karena itu, penelitian ini bertujuan untuk meningkatkan kualitas dari penelitian yang telah dilakukan oleh Putra (2023) dengan mengimplementasikan proses *streaming* untuk data SQM yang lebih panjang dan menggunakan algoritma lain yang dikhususkan untuk deteksi pencilan dengan cepat secara *real-time*.

## 1.2 Rumusan Masalah

Rumusan masalah dari latar belakang yang telah dipaparkan pada subbab sebelumnya adalah sebagai berikut:

1. Bagaimana model komputasi algoritma exact-STORM dalam mendeteksi pencilan pada *data stream* polusi cahaya langit malam hasil pemantauan *Sky Quality Meter* (SQM) yang disimulasikan menggunakan *platform* Apache Kafka?
2. Bagaimana implementasi dari algoritma exact-STORM dalam mendeteksi pencilan pada *data stream* polusi cahaya langit malam hasil pemantauan *Sky Quality Meter* (SQM) yang disimulasikan menggunakan *platform* Apache Kafka?
3. Bagaimana hasil deteksi dan performa dari algoritma exact-STORM dalam mendeteksi pencilan pada *data stream* polusi cahaya langit malam hasil pemantauan *Sky Quality Meter* (SQM) yang disimulasikan menggunakan *platform* Apache Kafka?

### 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan pada sub-bab sebelumnya, maka didapatkan tujuan penelitian sebagai berikut:

1. Merancang model komputasi algoritma exact-STORM dalam mendeteksi pencilan pada data *stream* polusi cahaya langit malam hasil sensor *Sky Quality Meter* (SQM) yang disimulasikan menggunakan *platform* Apache Kafka.
2. Mengimplementasikan model komputasi algoritma exact-STORM dalam mendeteksi pencilan pada *data stream* polusi cahaya langit malam hasil pemantauan *Sky Quality Meter* (SQM) yang disimulasikan menggunakan *platform* Apache Kafka.
3. Menganalisis hasil eksperimen dan evaluasi model algoritma exact-STORM dalam mendeteksi pencilan pada *data stream* polusi cahaya langit malam hasil pemantauan *Sky Quality Meter* (SQM) yang disimulasikan menggunakan *platform* Apache Kafka.

### 1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Memperoleh model komputasi dari program deteksi pencilan secara *real-time* menggunakan algoritma exact-STORM dan *platform streaming* Apache Kafka yang tepat untuk proses pemantauan langit malam oleh sensor *Sky Quality Meter* (SQM) sehingga dapat memberikan peringatan secara *real-time*.
2. Membantu pengguna untuk memiliki sistem peringatan pencilan dari pemantauan langit malam secara *real-time* sehingga dapat segera menyadari adanya anomali pada malam tersebut.
3. Mendapatkan wawasan baru mengenai model algoritma exact-STORM yang tepat untuk digunakan dalam deteksi pencilan secara *real-time streaming* pada data polusi cahaya langit malam berdasarkan hasil analisis pada performa model.
4. Memproleh data hasil deteksi pencilan dari data hasil pemantauan polusi cahaya langit malam yang dapat digunakan kembali untuk dianalisis lebih lanjut.

## 1.5 Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut:

1. Data yang digunakan merupakan data polusi cahaya langit malam hanya berasal dari satu lokasi pemantauan yaitu Observatorium Nasional Timau, Nusa Tenggara Timur dengan rentang waktu data tersedia adalah selama tahun 2020-2021
2. Proses *streaming* masih berupa simulasi *streaming* menggunakan Apache Kafka.
3. Proses *streaming* belum terintegrasi secara langsung dengan sensor SQM, namun proses *streaming* dilakukan melalui Apache Kafka dengan *input file batch* hasil sensor SQM yang disimulasikan menjadi proses *streaming*.
4. Antarmuka yang dibuat hanya sebatas tampilan pada Windows *Command Prompt*.

## 1.6 Sistematika Penulisan

Dalam penulisan dokumen skripsi ini terdapat sistematika yang digunakan sebagai pedoman penulisan. Sistematika tersebut dibagi ke dalam lima bab, yaitu:

### BAB I PENDAHULUAN

Bab ini membahas mengenai latar belakang masalah yang dihadapi dalam bidang astrofisika khususnya tentang polusi cahaya pada langit malam. Kemudian, bab ini juga menyampaikan pendekatan solutif untuk mendeteksi pencilan dari data polusi cahaya kecerahan langit malam yang dihasilkan oleh sensor *Sky Quality Meter* (SQM). Pada bab ini juga dijelaskan mengenai tujuan dari dilakukannya penelitian tersebut, serta berbagai manfaat yang akan didapat setelah penelitian dilakukan.

### BAB II KAJIAN PUSTAKA

Bab ini membahas berbagai teori pendukung dari penelitian yang dilakukan. Berbagai teori yang dikaji meliputi permasalahan dalam polusi cahaya beserta alat yang digunakan dalam mengukur kecerahan langit malam, kemudian teori yang membahas tentang *big data* meliputi *platform*, analisis-*analisis* yang ada dalam *big data* serta teori mengenai *data stream* dalam *big data*, lalu teori mengenai *platform*

Apache Kafka yang digunakan sebagai *platform* untuk *real-time streaming*, teori mengenai deteksi pencilan dan aplikasinya di berbagai disiplin ilmu, dan terakhir adalah kajian teori mengenai algoritma yang akan digunakan dalam penelitian untuk mendeteksi pencilan. Teori-teori tersebut didapatkan dari penelitian yang telah dilakukan sebelumnya, studi literatur, dan situs resmi yang berkaitan.

### BAB III METODOLOGI PENELITIAN

Bab ini membahas mengenai desain penelitian yang akan dilakukan yang akan digunakan sebagai acuan dalam melakukan penelitian ini. Desain penelitian yang dirancang memiliki berbagai tahapan yang perlu dilakukan dimulai dari tahapan studi literatur hingga tahapan terakhir adalah penarikan kesimpulan dari hasil penelitian. Selain itu, pada bab ini juga dibahas mengenai kebutuhan-kebutuhan yang perlu dipersiapkan sebelum memulai penelitian, seperti kebutuhan perangkat lunak dan kebutuhan perangkat keras.

### BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas mengenai implementasi dari proses dan hasil dari setiap tahapan penelitian yang dilakukan. Bab ini diawali dari proses pengumpulan data yang diperlukan dalam penelitian yaitu dengan mengunduh dataset dari RIN Dataverse. Kemudian, data tersebut akan dihitung menggunakan metode Chebyshev sehingga menghasilkan nilai yang akan digunakan sebagai salah satu parameter dalam algoritma. Selanjutnya, *dataset* hasil pengumpulan data akan dialirkan secara *streaming* menggunakan Apache Kafka, dan diproses kembali dalam program deteksi pencilan sehingga menghasilkan data hasil deteksi pencilan. Pada bab ini juga dirumuskan berbagai skenario eksperimen dan pembahasan dari setiap skenario eksperimen, serta dilakukan pula evaluasi terhadap performa algoritma yang dibandingkan performanya dengan algoritma lain.

### BAB V KESIMPULAN DAN SARAN

Pada bab ini berisikan kesimpulan secara keseluruhan yang dibuat dari hasil penelitian yang telah dilakukan. Selain itu, pada bab ini juga akan merumuskan beberapa saran yang dapat digunakan dan diterapkan pada penelitian berikutnya.