BAB I
PENDAHULUAN

1.1. Latar Belakang

Perkembangan dunia digital menjadi salah satu hal yang membawa banyak perubahan dalam kehidupan manusia. Semua aspek dalam kehidupan manusia telah didigitalisasi sehingga kehidupan manusia sangat dipengaruhi dengan adanya teknologi. Perkembangan teknologi inilah yang mempermudah manusia untuk memperoleh informasi dari berbagai sumber. Informasi dinilai sangat penting, karena dengan adanya informasi manusia mendapat pengetahuan mengenai data, fakta, maupun berita yang dapat menunjang kehidupan manusia. Informasi-informasi yang didapat manusia dari berbagai sumber dapat dikumpulkan dan diolah untuk mendukung kehidupan manusia dalam berbagai aspek seperti bisnis dan pendidikan. Informasi-informasi yang dikumpulkan dan diolah tersebut disebut sebagai data (Sutabri, 2012).

Menurut Longkutoy (dalam Sutabri, 2012) data tidak hanya mengandung gambar, namun bisa juga mengandung angka, huruf, atau simbol yang menunjukkan suatu ide, objek, kondisi, atau situasi lainnya. Data dapat berukuran sangat besar apabila berisi sejumlah besar informasi, sehingga diperlukan tempat penyimpanan modern untuk menyimpan data berukuran besar. Seiring berjalannya waktu, data yang dikumpulkan akan mengalami pertumbuhan yang sangat signifikan. Pertumbuhan data yang sangat cepat ini menjadi hal utama dalam konsep *big data* (Narendra, 2015).

Dalam Gartner IT Glossary dijelaskan bahwa *big data* adalah asset informasi bervolume tinggi dan/atau beraneka ragam yang menuntut bentuk pemrosesan informasi yang efektif dan inovatif yang memungkinkan peningkatan wawasan, pengambilan keputusan, dan otomatisasi proses. *Big data* juga memiliki jenis yang sangat beragam dan frekuensi perubahan data sangat besar. Narendra (2015) menjelaskan bahwa kecepatan dalam pertambahan dan semakin bervariasinya data

akan menciptakan tantangan baru untuk mengelola dan memahami data tersebut. Oleh karena itu *big data* akan selalu berkaitan dengan volume, keberagaman data, serta kecepatan data tersebut mengalir. *Big data* biasanya disimpan untuk menganalisis permasalahan yang sedang dihadapi, contohnya adalah permasalahan bisnis yang dihadapi oleh suatu perusahaan. Salah satu contoh dari *big data* adalah database profil pengguna *Facebook* dan daftar produk pada aplikasi *Shopee*.

Data dikumpulkan oleh perusahaan biasanya bersifat rahasia karena berisi berbagai data pribadi milik konsumennya. Sehingga perusahaan memerlukan pengamanan ekstra untuk data yang dimilikinya. Keamanan data sangat diperlukan supaya data tersebut tidak jatuh ke pihak yang salah atau pun bocor ke publik. Kriptografi menurut Menez (1996) adalah ilmu yang mempelajari teknik-teknik matematika yang berhubungan dengan aspek keamanan informasi seperti kerahasiaan, integritas data, serta otentikasi (Munir, 2019). Penggunaan kriptografi dalam kehidupan sehari-hari bukanlah hal yang asing bagi manusia. Salah satu contoh implementasi kriptografi dalam kehidupan sehari-hari adalah *double authentication* yang biasanya dapat ditemui ketika login ke dalam akun e-mail.

Algoritma kriptografi yang saat ini dinilai cukup aman (Yuniati, dkk., 2011) karena memiliki skor pengujian tertinggi pada uji coba yang dilakukan oleh *National Institute of Standard and Technology* (NIST) adalah *Advanced Encryption Standard* (AES). NIST mempublikasikan AES sebagai standar algoritma kriptografi terbaru menggantikan *Data Encryption Standard* (DES) yang masa penggunaannya telah berakhir (Yuniati, dkk. 2011) karena telah ditemukannya teknik kriptanalisis oleh Biham dan Shamir pada awal 1990-an (Stinson dan Paterson, 2019). AES memiliki panjang kunci yang bervariasi, mulai dari 128 bit, 192 bit, hingga 256 bit. AES 256 dipilih oleh penulis sebagai salah satu algoritma untuk mengenkripsi data berukuran besar karena AES merupakan algoritma yang cukup rumit sehingga dapat dinilai aman untuk melindungi data. Data biasanya bersifat sangat rahasia, sehingga diperlukan algoritma yang tepat untuk mengenkripsinya supaya data tersebut akan terjamin keamanan dan kerahasiaannya.

Sebagai algoritma yang memiliki skor pengujian tertiggi berdasarkan uji coba yang dilakukan oleh NIST (Jamil, 2004), AES pernah menjadi sasaran dari berbagai serangan *cryptanalysis* seperti *Differential Cryptanalisis and Linear Cryptanalysis*, *Truncated Differentials*, *The Square Attacks*, *Interpolation Attacks*, dan *Biclique Cryptanalysis* (Asriyanik, 2017). Kekuatan AES yang didasari oleh operasi matematis yang kompleks menjadi salah satu kelemahannya juga karena dengan berhasilnya dipecahkan persamaan matematis yang mendasari sistem AES, maka pertahanan AES dapat ditembus begitu saja (Asriyanik, 2017), sehingga perlu ditambahkan sebuah algoritma lain seperti SHA-256 untuk memperkuat keamanan AES-256.

Algoritma SHA-256 merupakan algoritma hash yang menghasilkan *message digest* berukuran 256 bit. Algoritma ini biasa digunakan untuk melakukan pengecekan integritas data, pembuatan *digital signature* dan lain-lain. SHA-256 bisa menerima input pesan hingga 2^64 bit yang akan diproses melalui blok dengan ukuran 512 bit.

Menurut Nainggolan (2022), SHA-256 tergolong aman karena didesain sedemikian rupa sehingga tidak memungkinkan mendapatkan pesan yang berhubungan dengan *message digest* yang sama. Pada SHA-256 belum pernah ditemukan adanya kolisi sehingga algoritma ini cukup aman untuk mengamankan data bersifat rahasia.

Pada Yuniati (2011), hasil enkripsi dari AES akan terjadi penambahan ukuran *file*, sedangkan ukuran *file* dekripsi akan kembali ke ukuran semula. Hal ini ditegaskan pula pada Lusiana (2011), menyatakan bahwa ukuran *file* hasil enkripsi dipengaruhi oleh ukuran awal *file* lampiran dan panjang kunci yang digunakan. Waktu yang dibutuhkan untuk enkripsi dan dekripsi berbeda dikarenakan adanya pemakaian *resource* komputer (Yuniati, dkk. 2011) yang juga ditegaskan pada Lusiana (2011), bahwa proses dekripsi membutuhkan komputasi yang lebih banyak sehingga memerlukan waktu proses yang lebih lama dibanding proses enkripsi.

Data yang berukuran besar akan memerlukan kunci enkripsi yang berukuran besar pula. Algoritma SHA-256 pada penelitian ini akan berfungsi untuk membuat kunci yang berukuran sangat besar menjadi sebuah kunci yang memiliki ukuran tetap yaitu 256 bit. *Message digest* yang dihasilkan pada proses SHA-256 berupa huruf-huruf yang diacak, hal ini mirip dengan konsep pada algoritma AES-256 ketika mengubah kunci asli menjadi bilangan heksadesimal. Setiap huruf pada kunci yang digunakan untuk proses enkripsi AES-256 akan diubah ke dalam bentuk heksadesimal yang apabila digabungkan akan menjadi sebuah kalimat acak tanpa arti. Apabila *message digest* yang dihasilkan diubah kedalam bentuk *character* sesuai dengan aturan heksadesimal pada tabel ASCII, akan ada kemungkinan terdapat beberapa karakter asing yang tidak dapat ditemukan pada aturan *password* yang biasanya hanya gabungan dari karakter huruf, angka, dan simbol. Beberapa karakter asing seperti *escape* dan *space* merupakan karakter yang tidak digunakan ketika membuat suatu *password* namun dapat ditemukan pada kasus ini apabila terdapat kombinasi heksadesimal yang sesuai dalam *message digest*. Sehingga dapat diasumsikan bahwa dengan mengenkripsi kunci AES-256 dengan menggunakan SHA-256 akan menghasilkan algoritma yang lebih aman di banding AES-256 yang asli.

Sebelumnya telah dilakukan penelitian untuk mengimplementasikan AES-256 dan SHA-256 sebagai pengamanan data *file* oleh Fathurrozi dan Selviyani (2021). Pada penelitian tersebut digunakan *file* dokumen, *file* suara, *file* video, serta *file* gambar sebagai *plaintext* yang akan dienkripsi. Program yang dihasilkan penelitian tersebut berhasil mengenkripsi berbagai jenis *file* dan kemudian mendekripsikannya kembali. *File* yang digunakan memiliki ukuran yang kecil, namun hasil enkripsi menunjukan adanya perbedaan ukuran *file* asli dengan *file* hasil enkripsi karena adanya proses *padding* dalam proses enkripsi tersebut. Berdasarkan pemaparan tersebut, penulis tertarik untuk mengkaji keamanan *big data* yang memiliki ukuran yang besar dengan menggunakan kriptografi algoritma SHA-256 dan AES-256 dengan menggunakan program *python*. Pada penelitian kali ini, penulis akan menggunakan *file* dengan format CSV (*Comma Separated*

Value) yang dirancang untuk dapat menampung dan transfer data dalam jumlah besar yang tidak dapat ditampung oleh *file* XLS/XLSX.

1.2. Rumusan Masalah

Berdasarkan pemaparan latar belakang sebelumnya, permasalahan dapat dirumuskan sebagai berikut:

1. Bagaimana implementasi penyandian algoritma AES-256 dengan kunci *hashing* SHA-256 pada pengamanan *big data*?
2. Bagaimana konstruksi program AES-256 dengan kunci *hashing* SHA-256 dalam pengamanan *big data* dengan menggunakan program *python*?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan dari penulisan penelitian ini adalah:

1. Mendeskripsikan implementasi algoritma AES-256 dengan kunci *hashing* SHA-256 pada *big data*.
2. Membuat suatu *prototype* program untuk implementasi AES-256 dengan kunci *hashing* SHA-256 dalam pengamanan *big data* dengan menggunakan program *python*.

1.4. Batasan Masalah

Batasan masalah yang digunakan pada penelitian ini adalah

1. Format *file* yang digunakan adalah format *.csv
2. Ukuran *file* yang digunakan berukuran minimal 1 *gigabyte* dan maksimal 10 *gigabyte*.
3. Laptop yang digunakan ditenagai dengan prosesor *Intel Core i3* dan memiliki RAM sebesar 8GB (DDR4).

1.5. Manfaat Penelitian

Adapun manfaat dari penulisan penelitian ini adalah:

1. Menambah alternatif pengamanan *big data*.
2. Memberikan kontribusi pada bidang matematika terapan serta mempermudah pengguna dalam proses enkripsi dan dekripsi pengamanan

big data menggunakan algoritma SHA-256 dan AES-256 dengan program *python*.