

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Dalam menyelesaikan masa perkuliahan dan untuk mendapatkan gelar sarjana (S1), mahasiswa di seluruh Perguruan Tinggi Negeri maupun Perguruan Tinggi Swasta memiliki kewajiban untuk menyelesaikan sebuah penelitian yang kemudian ditulis ke dalam sebuah laporan skripsi (Nurdin, 2017). Pada dasarnya dalam menulis skripsi, mahasiswa perlu menerapkan teori dan pengetahuan yang telah mereka dapatkan selama perkuliahan untuk menciptakan suatu solusi dari permasalahan sosial yang terjadi (Cahyadi et al., 2021).

Program studi Sistem Informasi UPN Veteran Jawa Timur memiliki dua bidang minat skripsi yang diantaranya bidang minat Solusi Sistem Informasi dan Manajemen Sistem Informasi. Dimana kedua bidang minat tersebut memiliki fokus pada topik penelitian yang berbeda-beda. Bidang minat Solusi Sistem Informasi lebih berfokus pada penelitian dengan topik seperti rancang bangun sistem informasi, *data mining*, sistem pendukung keputusan, dan lain sebagainya. Sedangkan bidang minat Manajemen Sistem Informasi lebih berfokus pada penelitian dengan topik seperti tata kelola teknologi informasi, analisis kepuasan pengguna, audit sistem informasi, dan lain sebagainya (*Buku Panduan Akademik Fakultas Ilmu Komputer Tahun Akademik 2019/2020*, 2019).

Setiap tahunnya jumlah data skripsi yang ada di program studi Sistem Informasi UPN Veteran Jawa Timur semakin bertambah, selain itu tren topik atau judul penelitian skripsi setiap tahunnya juga semakin bervariasi. Dengan bertambahnya penelitian skripsi pada masing-masing bidang minat skripsi yang

ada pada program studi Sistem Informasi, tak jarang membuat beberapa dosen bahkan mahasiswa itu sendiri kurang memperhatikan ketika ingin mengambil penelitian dengan tema, topik, metode, bahkan objek penelitian yang mirip atau sama dengan penelitian yang telah dilakukan oleh mahasiswa terdahulu.

Dengan bertambahnya jumlah skripsi dan topik penelitian skripsi yang ada di program studi Sistem Informasi, terdapat beberapa topik penelitian skripsi pada masing-masing bidang minat lebih banyak diminati dibanding dengan topik lainnya yang belum pernah digunakan dan/atau kurang diminati oleh mahasiswa. Hal tersebut menyebabkan adanya ketimpangan dalam masing-masing bidang minat penelitian skripsi.

Selain itu, selama ini di program studi Sistem Informasi UPN Veteran Jawa Timur belum memiliki sebuah sistem yang dapat mendokumentasikan data skripsi mahasiswa dan tren judul penelitian skripsi setiap tahunnya. Mengingat skripsi merupakan kegiatan ilmiah mahasiswa yang bersifat akuntabel, otentik serta dapat dipertanggungjawabkan (Saptono et al., 2019). Maka, dalam menjaga data skripsi agar tetap akuntabel dan otentik perlu dilakukan antisipasi terhadap penelitian yang akan diusulkan oleh mahasiswa agar tidak mengambil penelitian dengan tema, topik, metode, maupun objek penelitian yang mirip atau sama dengan penelitian terdahulu.

Oleh karena itu, penelitian ini bermaksud untuk melakukan pengelompokan (*clustering*) data skripsi di program studi Sistem Informasi UPN Veteran Jawa Timur dengan menggunakan *text mining*. *Dataset* yang digunakan adalah data skripsi mahasiswa yang didapat dari koordinator skripsi program studi Sistem Informasi UPN Veteran Jawa Timur. *Dataset* tersebut terdiri dari data

skripsi mahasiswa program studi Sistem Informasi UPN Veteran Jawa Timur dengan rentang tahun antara tahun 2019 sampai dengan awal tahun 2023.

Atribut yang akan dikelompokkan adalah atribut judul skripsi dan kata kunci yang nantinya akan diolah dengan melakukan *text preprocessing* kemudian memberikan bobot nilai pada masing-masing *term* yang telah terbentuk. Dari hasil pembobotan tersebut kemudian dilakukan pengelompokan (*clustering*) menggunakan algoritma *Agglomerative Hierarchical Clustering* yang dinilai mampu mengelompokkan data berupa teks atau dokumen dan algoritma ini tidak perlu menentukan jumlah *cluster* yang diinginkan

*Text mining* merupakan salah satu cabang ilmu *data mining* yang menganalisis data berupa dokumen atau teks dengan tujuan untuk menemukan suatu informasi atau tren baru yang sebelumnya belum pernah ditemukan dengan melakukan proses dan analisis data tekstual dengan jumlah besar. (Rachman et al., 2020). Adapun tujuan dari *text mining* adalah untuk mencari dan menemukan kata-kata yang mewakili isi dari suatu dokumen sehingga dapat dilakukan analisa hubungan atau keterkaitan antar dokumen (Langgeni et al., 2010).

*Clustering* merupakan algoritma *unsupervised classification* yang bertujuan mengelompokkan data ke dalam suatu *cluster*. *Clustering* sendiri telah digunakan pada banyak disiplin ilmu. Sehingga menyebabkan algoritma *clustering* sering digunakan sebagai salah satu langkah dalam melakukan analisis dan eksplorasi data (Calvillo et al., 2013). Pada proses pengelompokan (*clustering*), data akan dikelompokkan menjadi beberapa *cluster* dengan kemiripan antara satu data dengan data yang lain (Februariyanti & Santoso, 2017).

Dengan dilakukannya penelitian ini, harapannya para pemangku kepentingan seperti koordinator program studi, kepala laboratorium, dan/atau koordinator skripsi program studi Sistem Informasi dapat memberikan kebijakan dan arahan kepada mahasiswa agar dapat mengambil topik penelitian skripsi yang belum pernah atau jarang digunakan sebelumnya. Sehingga penelitian skripsi yang dilakukan oleh mahasiswa di program studi Sistem Informasi lebih merata, selain itu dapat memperluas topik penelitian skripsi yang ada di setiap bidang minat penelitian, serta meminimalisir terjadinya ketimpangan penelitian skripsi pada masing-masing bidang minat penelitian.

Adapun luaran dari penelitian ini adalah sebuah sistem informasi berbasis *website* yang berisikan kumpulan judul penelitian skripsi mahasiswa sekaligus visualisasi *clustering* judul dan kata kunci penelitian skripsi di Program Studi Sistem Informasi UPN Veteran Jawa Timur. Dengan adanya *website* tersebut, harapannya dapat mempermudah dosen maupun mahasiswa dalam mencari referensi judul penelitian skripsi yang akan dilakukan sekaligus memastikan bahwa judul yang akan diambil tidak memiliki kesamaan dengan penelitian skripsi sebelumnya.

## **1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang tersebut maka dapat dihasilkan beberapa rumusan permasalahan pada penelitian ini antara lain.

1. Bagaimana pengelompokan (*clustering*) judul skripsi dan kata kunci yang ada di program studi Sistem Informasi UPN Veteran Jawa Timur?

2. Bagaimana tingkat kemunculan serta visualisasi dari *term* atau kata yang sering digunakan pada penelitian skripsi di program studi Sistem Informasi UPN Veteran Jawa Timur?
3. Bagaimana cara memvisualisasikan hasil *clustering* yang telah dilakukan?

### **1.3 Batasan Masalah**

Batasan masalah merupakan hal yang penting dalam sebuah penelitian, agar penelitian yang dilakukan berfokus pada permasalahan yang telah dirumuskan dan tidak menimbulkan kerancuan dalam menginterpretasikan hasil penelitian yang didapatkan. Batasan masalah pada penelitian ini adalah sebagai berikut.

1. Data didapatkan dari koordinator skripsi program studi Sistem Informasi UPN Veteran Jawa Timur.
2. Data yang digunakan merupakan data skripsi mahasiswa program studi Sistem Informasi mulai dari tahun 2019 sampai dengan awal tahun 2023.
3. Metode yang digunakan untuk *clustering* yaitu *text mining* menggunakan algoritma *agglomerative hierarchical clustering*.
4. Proses pengolahan data menggunakan bahasa pemrograman Python.
5. Hasil visualisasi akan ditampilkan ke dalam sebuah sistem berbasis *website* dengan menggunakan bahasa pemrograman HTML, Python dan kerangka kerja Flask.

### **1.4 Tujuan**

Berdasarkan rumusan masalah yang telah dipaparkan maka didapatkan tujuan penelitian yang akan dicapai antara lain.

1. Mendapatkan informasi serta persebaran dari hasil pengelompokan (*clustering*) judul skripsi dan kata kunci di program studi Sistem Informasi UPN Veteran Jawa Timur.
2. Mendapatkan informasi dan visualisasi dari *term* atau kata yang sering digunakan pada penelitian skripsi di program studi Sistem Informasi UPN Veteran Jawa Timur.
3. Mendapatkan visualisasi dari hasil *clustering* judul skripsi dan kata kunci di program studi Sistem Informasi UPN Veteran Jawa Timur yang ditampilkan ke dalam sebuah sistem informasi berbasis *website*.

## **1.5 Sistematika Penulisan**

Sistematika penulisan disajikan untuk memberikan gambaran menyeluruh mengenai uraian penelitian yang akan dilakukan. Adapun sistematika penulisan skripsi disajikan sebagai berikut.

## **BAB I PENDAHULUAN**

Bab Pendahuluan memuat hal-hal yang berkaitan mengenai gambaran umum dari penelitian yang akan dilakukan mengenai pengelompokan topik skripsi menggunakan algoritma *agglomerative hierarchical clustering*. Bab ini terdiri dari uraian latar belakang, rumusan masalah, batasan masalah, tujuan, dan sistematika penulisan dari penelitian yang akan dilakukan.

## **BAB II TINJAUAN PUSTAKA**

Bab Tinjauan Pustaka memuat mengenai dasar teori serta penelitian terdahulu yang relevan sebagai pendukung penelitian yang akan dilakukan. Adapun dasar-dasar teori yang menunjang dalam penelitian ini antara lain tinjauan mengenai skripsi, *data mining*, *text mining*, Python, *text preprocessing*, *term*

*weighting, dimensionality reduction, clustering, agglomerative hierarchical clustering, silhouette coefficient, Flask, dan black-box testing.*

### **BAB III METODOLOGI PENELITIAN**

Bab Metodologi Penelitian membahas mengenai alur tahapan yang digunakan dalam penelitian untuk mendapatkan solusi dari permasalahan yang telah dirumuskan sehingga mencapai tujuan penelitian yang diharapkan. Adapun alur yang digunakan adalah studi literatur, analisis kebutuhan, pengumpulan data, pembuatan model *data mining*, serta pembuatan dan pengujian dari sistem yang telah dibuat.

### **BAB IV HASIL DAN PEMBAHASAN**

Bab Hasil dan Pembahasan memuat mengenai hasil penelitian yang telah dilakukan. Hasil yang disertakan adalah hasil dari proses pengumpulan data, pembuatan model *data mining*, pembuatan sistem berbasis *website*, serta pengujian dari sistem yang telah dibuat. Dari hasil pengelompokan tersebut, kemudian dapat dilakukan analisis untuk mengetahui topik skripsi apa saja yang populer di program studi Sistem Informasi.

### **BAB V PENUTUP**

Bab Penutup memuat mengenai kesimpulan dari hasil penelitian yang telah dilakukan serta saran yang disampaikan sebagai pertimbangan untuk penelitian berikutnya.

### **DAFTAR PUSTAKA**

### **LAMPIRAN**