

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Bekerja merupakan hak bagi setiap warga negara sebagaimana diatur dalam UUD 1945 pasal 27 ayat 2, "Tiap-tiap warga negara berhak atas pekerjaan dan penghidupan yang layak bagi kemanusiaan". Pekerjaan yang dijalani seseorang dilatarbelakangi oleh berbagai hal seperti motivasi, faktor kebutuhan, mencari pengalaman, dan lain-lain. Namun, banyaknya pilihan pekerjaan tidak menjamin seseorang dapat memperoleh pekerjaan yang diharapkan.

Terdapat kualifikasi tertentu yang menjadi persyaratan bagi sebuah pekerjaan yaitu, keterampilan, tingkat pendidikan, dan pengalaman kerja yang merupakan modal dasar bagi seorang pekerja untuk dapat bersaing dengan layak di sektor kerja yang sama maupun berbeda (Frankiano, 2011). Di sisi lain, pelamar pekerjaan jumlahnya meningkat tiap tahunnya, sehingga tidak semua pelamar dapat terserap di satu jenis pekerjaan. Diharapkan dari kualitas yang terus meningkat, waktu tunggu lulusan dalam bekerja dapat terserap dengan cepat (Setyaningsih & Abrori, 2013).

Salah satu upaya untuk mendapatkan pekerjaan yang lebih layak adalah dengan menempuh jenjang pendidikan tertentu. Hal ini selaras dengan tujuan institusi pendidikan dalam mencetak lulusannya. Pada program pembangunan nasional yang diatur dalam Undang-Undang No. 25 tahun 2000 pembangunan pendidikan tidak hanya dikaitkan secara erat dengan pembangunan ekonomi, melainkan juga dengan tantangan globalisasi. Oleh karena itu setiap institusi pendidikan dituntut untuk selalu *uptodate* terhadap tuntutan dunia pekerjaan.

Bagi sebuah institusi pendidikan, mengevaluasi keterserapan lulusan di bidang pekerjaan yang sesuai merupakan pekerjaan rutin dalam rangka meningkatkan mutu lulusan. Alumni dan survei merupakan salah satu bentuk studi empiris yang bisa memberikan informasi berharga untuk mengevaluasi hasil pendidikan dan pelatihan yang telah di tempuh pada masa pendidikan (Eschborn &

Kassel, 2003). Informasi yang telah di berikan perlu suatu tindak lanjut berupa *Tracer Study* dan akan menjadikan suatu jaminan kualitas sebuah institusi dalam memberikan kepercayaan kegiatan untuk perencanaan masa depan. Maka dari itu setiap universitas berusaha untuk memperbaiki kualitas lulusannya.

Namun dalam praktiknya penyebaran kuesioner tidaklah sesuai dengan yang diharapkan. Pengisian kuesioner yang dilakukan oleh alumni ilmu komputer dan pendidikan ilmu komputer cenderung diisi oleh alumni yang sama, atau hanya sebagian alumni yang benar-benar dapat meluangkan waktu untuk mengisi kuesioner. Jika pengisian hanya dilakukan oleh sebagian alumni maka hasil yang didapatkan dari *tracer study* pekerjaan tidak akan mewakili keberagaman jenis pekerjaan yang dilakukan alumni, sehingga tidak dapat tercapainya perbaikan pada kurikulum.

Kebutuhan tanggapan dari hasil *tracer study* nantinya akan digunakan sebagai keperluan perbaikan kurikulum. Jika terjadi ketidakseimbangan pengisian, maka beberapa jenis pekerjaan tidak akan terwakili ketika dilakukannya pengumpulan kuesioner. Dari paparan diatas maka penelitian ini akan menggunakan algoritma *k-means* dan perhitungan menggunakan *hamming distance*. Dalam data mining ada dua jenis metode yaitu metode *clustering* dan *non-hierarchical clustering* (Santosa, 2009). Data yang didapat biasanya dalam bentuk representasi simbol atau dikenal sebagai data kategori. Data yang diperoleh melalui penelitian harus akurat yaitu data yang memenuhi kriteria valid, reliabel dan objektif.

Permasalahan yang terkait mengenai segmentasi data teknik *clustering* saat ini telah banyak digunakan. Tujuan utama dari metode *clustering* adalah tentang pengelompokan sejumlah data atau objek ke dalam *cluster* sehingga dalam setiap *cluster* akan berisi data yang memiliki kesamaan karakteristik dari data tersebut. Pada pengelompokan dokumen, sekumpulan dokumen yang belum diberi label kelasnya akan dikelompokan sesuai dengan karakteristik-karakteristik kata yang dimiliki setiap dokumen tersebut.

Algoritma *K-Means* adalah algoritma yang sering digunakan dalam melakukan proses *clustering*, karena sifatnya yang relatif cepat dan dapat beradaptasi. Adapun algoritma *K-Means* ini merupakan algoritma iteratif yang saat melakukan partisi set data ke dalam sejumlah *k cluster* yang sudah ditetapkan di awal. Pemilihan *k* sebagai pusat *cluster* awal juga mempengaruhi hasil *clustering*. Sifat tersebut adalah menjadikan karakteristik alami dari *K-Means* yang dapat mengakibatkan hasil *cluster* yang didapat pada percobaan berbeda dengan hasil telah proses *clustering*. Kondisi tersebut dikenal sebagai solusi *local optimum*, yang artinya algoritma *K-Means* sangat sensitif terhadap lokasi awal pusat *cluster* (Prasetyo, 2012).

Seperti pada penelitian (Nabilah, 2017) dalam penelitiannya dilakukan pengelompokan terhadap *supplier*. Dimana dalam prosesnya perusahaan juga ikut andil dalam melakukan segmentasi *supplier* berdasarkan daerah para *supplier* tersebut tinggal. Solusi yang ditawarkan adalah mengelompokkan perilaku bisnis yang dinamis. Maka dalam penelitiannya dilakukan pengelompokan perilaku bisnis yang sama, segmentasi *supplier* berdasarkan *frekuensi pengiriman, rentan waktu pembelian dan nilai barang dalam bentuk uang* yang disebut dengan analisis *RFM*. Pada penelitiannya menggunakan metode *K-Means clustering* untuk memproses hasil analisis *RFM*.

Seperti pada penelitian (Jinchao, 2013) yang menggabungkan *clustering* kategorial data dengan algoritma *clustering k-prototypes* yaitu sebuah algoritma *k-prototipe* yang meningkatkan untuk mengelompokkan data campuran. Beberapa hasil kontribusi yang dihasilkan antara lain, pertama algoritma yang dihasilkan dapat menangkap karakteristik kluster atribut campuran dengan lebih efektif. Kedua algoritma menganggap bahwa kesalahan sebagai hal penting dalam proses pengelompokan atribut, ketiga dalam partisi objek algoritma pada algoritma ini akan secara otomatis menghitung atribut yang berbeda menuju proses pengelompokan dengan menggunakan strategi Huang.

Implementasi *automatic clustering* menggunakan *differential evolution* dan *cs measure* untuk analisis data kemahasiswaan dalam *papernya* menjelaskan penerapan algoritma *automatic clustering differential evolution* digunakan untuk

mengetahui pemetaan hasil belajar dengan sampel 166 mahasiswa (Yasid, 2014). Penerapan metode *clustering* dengan *K-Means* untuk memetakan potensi tanaman padi di kota Semarang dalam jurnalnya menjelaskan tentang pengelompokan daerah potensial penghasil padi dengan jumlah terbanyak dan penghasil padi sedikit (Felicia, 2014).

Berdasarkan masalah data, maka penulis melakukan penelitian mengenai pengelompokan profil pekerjaan alumni. Formula untuk menentukan perhitungan dengan menggunakan *hamming distance*. Perbedaan profil pekerjaan yang belum dikelompokkan dan sesudah dikelompokkan akan terlihat pada setiap cluster, untuk memastikan nilai k terbaik adalah dengan *elbow method* yang sebelumnya telah dihitung *sum square of error* nya.

Dengan melakukan penelitian ini, diharapkan mampu menghasilkan suatu cluster pekerjaan yang nantinya bisa dimanfaatkan untuk kebutuhan kurikulum maupun kebutuhan lainnya. Data yang di dapat dari sebaran kuesioner kepada alumni haruslah tepat sasaran. Pada dasarnya program studi di setiap universitas itu sebenarnya berbeda, sehingga penggunaan teknik *clustering* lebih cocok untuk *tracer study*.

## **1.2 Rumusan Masalah**

Berdasarkan permasalahan-permasalahan yang muncul, maka dirumuskan beberapa masalah yang ingin diselesaikan, yaitu:

1. Bagaimana cara menyiapkan data agar siap diproses dalam pengelompokan?
2. Bagaimana menentukan formula yang tepat untuk rumus perhitungan jarak agar dapat diproses menggunakan algoritma k-means?
3. Bagaimana hasil dari visualisasi *clustering* profil pekerjaan alumni berdasarkan data *tracer study* yang digunakan?

## **1.3 Tujuan Penelitian**

Setelah rumusan masalah dibuat, maka akan muncul tujuan penelitian ini dilakukan, tujuan tersebut adalah sebagai berikut:

1. Menganalisis dan menyiapkan data agar siap diproses untuk pengelompokan.
2. Membuktikan formula hasil perhitungan yang tepat untuk di proses oleh algoritma k-means.
3. Mengimplementasi dan mengevaluasi hasil visualisasi *clustering* profil pekerjaan alumni berdasarkan data *tracer study* yang digunakan.

#### **1.4 Manfaat Penelitian**

Adapun manfaat yang diharapkan dari penelitian ini yaitu dapat bermanfaat bagi pihak yang memiliki kepentingan yang sama sebagai berikut:

1. Bagi Peneliti
  - a. Memberikan wawasan dan pengetahuan baru mengenai pentingnya melakukan pengelompokan profil pekerjaan di lingkungan organisasi.
  - b. Mengetahui cara pembagian *cluster* sesuai dengan tingkat kemiripan untuk membangun sebuah pemodelan.
2. Bagi Universitas
  - a. Penelitian ini dapat menjadi acuan bagi peneliti lain untuk dapat mengembangkan atau melanjutkan penelitian pada topik pengelompokan profil pekerjaan alumni pada ruang lingkup kecerdasan buatan dan *data mining*.
  - b. Memberikan sumbangan karya ilmiah mengenai disiplin ilmu kecerdasan buatan dan *data mining*.

#### **1.5 Batasan Masalah**

Agar pembahasan masalah tidak meluas, maka untuk memfokuskan sasaran penelitian diperlukan adanya batasan masalah untuk penelitian ini. Batasan masalah tersebut adalah sebagai berikut:

1. Data yang digunakan pada penelitian ini adalah data Ilmu Komputer UPI dengan format kuesioner.
2. Algoritma yang dipakai adalah Algoritma.
3. *File* yang digunakan merupakan *file* yang memiliki ekstensi *.csv*.

## 1.6 Sistematika Penulisan

Pada bagian sistematika penulisan ini penulis akan menguraikan gambaran umum dari skripsi yang disusun. Sistematika penulisan skripsi ini adalah sebagai berikut:

### BAB I PENDAHULUAN

Bab ini menjelaskan gagasan penulis dan beberapa sumber terkait untuk menguatkan gagasan yang diajukan dalam melakukan penelitian ini. Di dalam bab ini terdapat bagian latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, definisi operasional dan struktur organisasi skripsi.

### BAB II TINJAUAN PUSTAKA

Bab ini berisi sumber atau literatur-literatur berupa *paper*, *ebook* dan jurnal untuk memperoleh teori tentang *Data Mining*, *Clustering*, *K-Means*, *Hamming Distance*, *Sum Squared Error*, *Tracer Study* serta sumber lainnya yang dapat mendukung penelitian.

### BAB III METODOLOGI PENELITIAN

Bab ini berisikan langkah-langkah penelitian, mulai dari desain penelitian, fokus penelitian, kemudian alat dan bahan yang akan digunakan selama penelitian dan metode penelitian yang digunakan untuk penelitian ini.

### BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil penelitian yang telah dilakukan beserta analisisnya. Pertanyaan penelitian yang dibahas dalam rumusan masalah, dibahas pada bab ini. Pembahasan meliputi penjabaran hasil studi literatur melalui hasil *cluster iterasi*, hasil eror, dan tabel perbandingan.

### BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang menjawab pertanyaan penelitian dan saran dapat dijadikan rujukan bagi peneliti selanjutnya dari hasil penelitian yang telah dilakukan.