

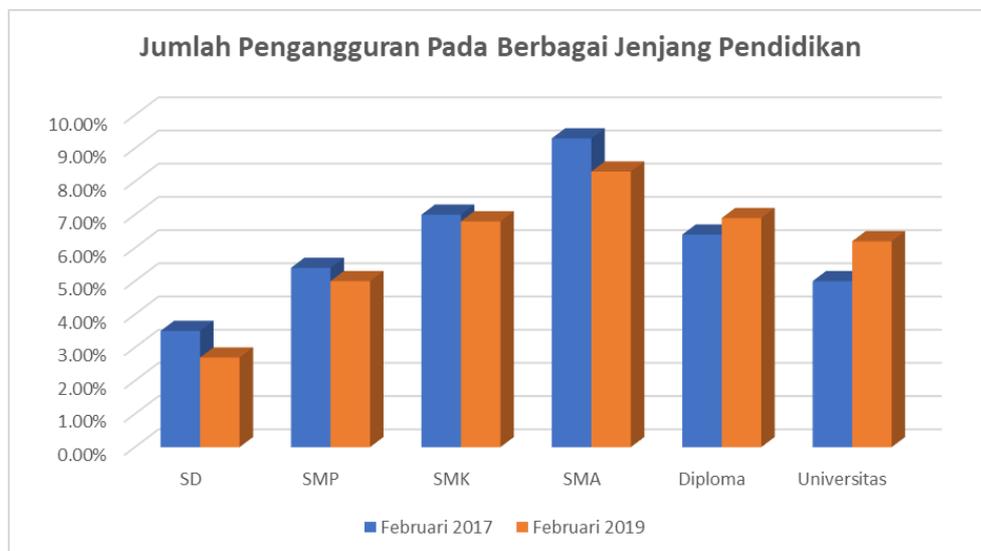
# BAB I

## PENDAHULUAN

Bab ini akan membahas hal yang melatarbelakangi dilakukannya penelitian ini. Lalu akan dijabarkan pula mengenai rumusan masalah, tujuan, manfaat dan batasan masalah beserta sistematika penulisan pada penelitian ini.

### 1.1 Latar Belakang Masalah

Pengangguran menjadi salah satu masalah terbesar di Indonesia. Menurut (Statistik, 1999) pengangguran didefinisikan sebagai angkatan kerja yang tidak bekerja tetapi sedang mencari pekerjaan atau orang yang *full timer* dalam mencari pekerjaan. Data dari (Statistik, 2019) pada bulan Agustus 2019 menunjukkan tingkat pengangguran terbuka (TPT) menurut pendidikan pada jenjang universitas sebesar 5.67% dari keseluruhan pengangguran. Data statistik menurut (Pusparisa, 2019) menunjukkan perbandingan jumlah pengangguran per Februari 2019 dengan per Februari 2017 yang ditunjukkan pada Gambar 1.1 di bawah ini.



**Gambar 1.1** Jumlah pengangguran pada berbagai jenjang pendidikan

Pada tabel 1.1 di atas terlihat pada jenjang SD, SMP, SMK, dan SMA mengalami penurunan jumlah pengangguran, sedangkan jenjang Diploma dan Sarjana mengalami peningkatan dan peningkatan terbesar terdapat pada lulusan Sarjana.

Berdasarkan data hasil survei jumlah lowongan pekerjaan yang tersedia pada kuartal II 2018 oleh Bank Indonesia (BI) yang dipaparkan oleh Direktur Eksekutif Departemen Statistik BI Yani Kurniati (Sugianto, 2018) terjadi pertumbuhan angka lowongan pekerjaan pada kuartal II 2018 sebesar 19,2% dibandingkan periode pada kuartal sebelumnya. Peningkatan jumlah lowongan ini terjadi di beberapa sektor usaha seperti perdagangan besar dan eceran (25,1%), pertambangan dan penggalian (53,5%), jasa keuangan dan asuransi (47,7%), penyedia akomodasi dan makanan dan minuman (16,4%) serta sektor konstruksi (48,5%). Sementara dari sisi wilayah, pertumbuhan lowongan kerja paling banyak terjadi di luar Pulau Jawa, terutama di Sumatera dan Kalimantan.

Peningkatan jumlah lowongan pekerjaan ternyata tidak membuat jumlah pengangguran menurun, dan yang terjadi adalah sebaliknya pada jenjang Sarjana. Menurut (Iswahyuni, 2018) salah satu faktor yang mempengaruhi lulusan universitas sulit mendapatkan pekerjaan yaitu dipengaruhi oleh kesenjangan pengetahuan antara teknologi yang diterapkan di dunia kerja dengan apa yang dipelajari di dunia pendidikan tinggi. Perubahan yang cepat di dunia kerja sebagai akibat dari globalisasi dan revolusi di bidang teknologi informasi, dan sains telah menuntut antisipasi dan evaluasi terhadap kompetensi yang dibutuhkan oleh dunia kerja. Sehingga evaluasi menjadi penting untuk dilakukan agar dunia pendidikan tinggi tidak terpisah dan berjarak dengan dunia kerja yang riil yang ada di masyarakat (Handayani, 2015).

Terjadinya ketimpangan pengetahuan yang diterapkan di dunia kerja dengan yang dipelajari selama pendidikan tinggi menyebabkan ketidaksesuaian antara perusahaan dan tenaga kerja dalam mendapatkan pekerja yang sesuai dengan keahlian (Soleh, 2017). Untuk mengatasi ini, berbagai macam keahlian yang diperlukan oleh perusahaan harus dapat tersampaikan dengan baik kepada calon pelamar kerja kelak. Salah satu media untuk menyampaikan berbagai kebutuhan perusahaan adalah melalui lowongan pekerjaan.

Lowongan pekerjaan yang dibagikan kepada calon pelamar memiliki berbagai macam bentuk dan media, salah satunya adalah dengan membuat sebuah *posting* dalam sebuah media sosial. Lowongan yang dibagikan berupa nama pekerjaan, *job desc*, dan kebutuhan *skill* yang diinginkan perusahaan. Dalam sebuah

*posting* lowongan pekerjaan, informasi seperti ini dapat diolah dan dibuat menjadi lebih berguna, yaitu *information retrieval* atau pengambilan informasi menggunakan *clustering*.

Dari beberapa lowongan pekerjaan yang telah di-*posting*, akan diambil dan dikumpulkan untuk dilakukan pengambilan informasi (*information retrieval*). *Clustering* atau pengelompokan pada konteks lowongan pekerjaan, merupakan salah satu metode untuk pengambilan informasi (*information retrieval*) dengan cara mengumpulkan lowongan pekerjaan yang mirip ke dalam satu kelompok dan memisahkan kelompok dengan lowongan pekerjaan yang berbeda. Tujuan dari pengelompokan ini untuk mendapatkan lowongan pekerjaan sejenis dan memudahkan calon pelamar pekerjaan mencari pekerjaan sesuai dengan kelompok yang dicari.

Metode *clustering* yang akan digunakan adalah metode *k-Means*. *K-Means* pertama kali dipopulerkan oleh (MacQueen, 1967). *K-Means* menghitung dan mencari *cluster* dengan mengutamakan nilai rata-rata sebagai proses utamanya. Namun tidak semua kasus dapat dilakukan dengan menggunakan rata-rata. Seperti data yang menggunakan tipe kategorikal tidak akan bisa menggunakan algoritma ini. Pun halnya dengan data teks seperti lowongan pekerjaan, perlu dikonversikan ke dalam bentuk numerik agar dapat diproses. Setelah data teks dapat direpresentasikan ke dalam bentuk yang lain, maka pengelompokan akan dapat diproses.

Setelah melakukan pengelompokan, maka perlu dilakukan evaluasi kualitas dari *cluster* yang telah dibuat. Tujuan dari evaluasi ini adalah untuk mendapatkan jumlah *cluster* terbaik berdasarkan dari lowongan pekerjaan yang didapatkan. Salah satu metode penentuan kualitas dari *clustering* yang populer digunakan adalah *Silhouette Coefficient*. Penelitian oleh (Thinsungnoen dkk., 2015) berhasil melakukan pengelompokan menggunakan algoritma *k-Means* dan mengevaluasi menggunakan SSE dan *Silhouette*. Metode *Silhouette* dan SSE tersebut akan penulis gunakan untuk mengevaluasi kinerja pengelompokan pada penelitian kali ini.

Metode evaluasi pengelompokan menggunakan metrik SSE (*Sum of Squared Error*) menggunakan jumlah kuadrat galat dari setiap *cluster* yang terbentuk. Untuk mendapatkan jumlah *cluster* optimal pada metrik ini dibutuhkan

metode tambahan, yaitu metode *elbow* yang diimplementasikan pada penelitian (Bholowalia & Kumar, 2014). Metode *elbow* didefinisikan sebagai metode yang mengambil nilai  $k$  optimal pada titik yang menyiku pada salah satu titik jumlah  $k$  terhadap nilai SSE.

Penelitian mengenai *clustering* sudah banyak dilakukan sebelumnya. Pada studi kasus yang berbeda namun menggunakan metode serupa yaitu pengelompokan surah-surah dari Al-Qur'an menggunakan algoritma  $k$ -Means yang dilakukan oleh (Slamet, Rahman, Ramdhani, & Dharmalaksana, 2016) yang membahas mengenai tahap-tahap pemrosesan teks menggunakan kerangka kerja *text mining* seperti *case folding*, *tokenization*, *stopwords removal*, dan *stemming*, lalu dilanjutkan dengan *clustering* menggunakan  $k$ -Means. Teknik *text mining* ini penulis gunakan sebagai referensi pemrosesan teks yang akan diimplementasikan pada penelitian ini.

Penelitian berikutnya adalah pengelompokan barang dalam sebuah minimarket dengan menggunakan metode algoritma  $k$ -Means yang dilakukan oleh (Kusrini, 2015). Perbedaan utama penelitian ini adalah data yang digunakan tidak sepenuhnya teks, sehingga tidak diperlukan praproses teks terlebih dahulu. Penelitian lain yang serupa namun memiliki perbedaan dalam penentuan jarak antar objek clustering, yang pada umumnya menggunakan *euclidean* namun penelitian ini menggunakan *cosine similarity* oleh (Hudin, Fauzi, & Adinugroho, 2018).

Oleh karena itu penulis bermaksud untuk melakukan penelitian mengenai lowongan pekerjaan yang ditawarkan oleh alumni Departemen Pendidikan Ilmu Komputer UPI untuk mendapatkan kelompok-kelompok pekerjaan yang sejenis dengan menggunakan algoritma  $k$ -Means serta pemrosesan teks menggunakan *text mining*. Penelitian ini dilakukan dengan menggunakan bahasa pemrograman Python 3 sebagai basisnya, dan *Jupyter Notebook* sebagai *web application* untuk berinteraksi dengan program yang akan dibuat.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah, maka dapat dirumuskan masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut.

1. Bagaimana implementasi representasi teks pada lowongan pekerjaan ke dalam pengelompokan  $k$ -Means?

2. Berapa jumlah *cluster* optimal dan kualitas yang dihasilkannya?
3. Bagaimana manfaat *clustering* dalam pengambilan informasi (*information retrieval*) yang dilakukan terhadap lowongan pekerjaan?

### 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan dilakukannya penelitian ini adalah sebagai berikut.

1. Mengimplementasikan metode *text mining* terhadap lowongan pekerjaan agar dapat dilakukan pengelompokan menggunakan algoritma *k-Means*.
2. Melakukan evaluasi terhadap kualitas *cluster* yang telah dibuat untuk mendapatkan jumlah *cluster* optimal.
3. Memberikan *label* kepada *cluster* optimal agar dapat dilakukan pengambilan informasi terhadap data lowongan pekerjaan.

### 1.4 Manfaat Penelitian

Dengan dilakukannya penelitian ini, diharapkan dapat memberikan beberapa manfaat, diantaranya:

1. Dapat memberikan kemudahan di kemudian hari ketika pengguna aplikasi ingin mencari pekerjaan sesuai dengan kelompok yang dicari.
2. Dapat memberikan data *input* untuk penelitian berikutnya, seperti ekstraksi fitur. Pengelompokan membuat data yang sesuai kelompok terkumpul dan memudahkan di penelitian berikutnya.
3. Dapat memberikan gambaran mengenai pekerjaan serta keahlian yang dibutuhkan oleh industri saat ini.

### 1.5 Batasan Masalah Penelitian

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 digunakan merupakan data yang didapatkan dari media sosial yang berisikan lowongan pekerjaan dari alumni di bidang IT, yaitu *Facebook*.
2. Pengumpulan data dilakukan secara manual mengambil dari beberapa sumber tersebut.

3. Bahasa yang digunakan adalah Bahasa Indonesia. Dengan digunakannya Bahasa Indonesia, maka beberapa lowongan pekerjaan yang menggunakan Bahasa Inggris akan diterjemahkan terlebih dahulu menjadi Bahasa Indonesia.

## 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 kerangka kerja *text mining*, algoritma Nazief dan Andriani, algoritma *k-Means*, dan algoritma *silhouette coefficient*.

### 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, implementasi sistem, hasil dan pembahasan model.

### BAB V KESIMPULAN DAN SARAN

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