

BAB I PENDAHULUAN

1.1. Latar Belakang

Di Provinsi Jawa Timur jumlah pencari kerja setiap tahun mengalami peningkatan dengan angka mencapai 200.388 orang pada tahun 2023 [1]. Pencari kerja tersebut berasal dari berbagai latar belakang pendidikan, usia, pengalaman dan kondisi yang berbeda, Hal ini dapat memengaruhi ketidakseimbangan dalam penyaluran tenaga kerja khususnya di Kota Surabaya. Di Kota Surabaya, Dinas Perindustrian dan Tenaga Kerja melalui *website* ASSIK (Arek Suroboyo Siap Kerjo) berperan sebagai perantara untuk memberikan informasi lowongan pekerjaan kepada masyarakat serta membantu perusahaan dalam menemukan kandidat yang sesuai dengan kualifikasi yang dibutuhkan.

Syarat kualifikasi dalam *website* tersebut adalah pencari kerja harus memasukkan data berupa biodata diri dan lampiran kualifikasi yang sudah disediakan. Data tersebut akan ditampilkan pada *website* sehingga industri yang sedang membuka lowongan pekerjaan dapat memilih pencari kerja yang memenuhi kualifikasi industri. Namun, pada realitanya pencari kerja di kota Surabaya pada *website* ASSIK (Arek Suroboyo Siap Kerjo) mempunyai karakteristik yang berbeda – beda dari segi kondisi dan latar belakang seperti sebagian ada yang sudah menikah dan mempunyai anak, ada yang disabilitas dan ada yang belum mempunyai pengalaman kerja seperti lulusan baru.

Lulusan baru di Kota Surabaya juga mendapatkan informasi mengenai pelatihan dan sertifikasi yang disediakan oleh Dinas Perindustrian dan Tenaga Kerja Kota Surabaya melalui *website* serta akun instagram resmi. Namun, tidak semua profesi mempunyai pelatihan dan sertifikasi secara berkala, sehingga calon peserta pelatihan seringkali harus menunggu jadwal pelatihan dan uji sertifikasi yang sesuai dengan profesi yang diinginkan.

Berdasarkan hal tersebut, perlu dilakukan analisis data untuk mengelompokkan pencari kerja di Kota Surabaya berdasarkan karakteristik yang sama menggunakan proses pengelompokkan implementasi dari *data mining*.

Dalam *data mining*, terdapat metode yang dapat menemukan pola yang

menarik serta tidak terlihat dalam kumpulan data [2]. Metode tersebut adalah metode klastering yang dapat memetakan atau mengelompokkan data berdasarkan karakteristik kesamaan data. Klastering mempunyai keunggulan yang tidak dimiliki oleh metode lain dari data mining yaitu dapat mengelompokkan data tanpa pengetahuan sebelumnya dan didalam proses klastering terdapat algoritma yang sering digunakan yaitu *K-Means*. *K-Means* adalah algoritma pendekatan yang digunakan untuk membentuk cluster di sebuah kelompok yang sama berdasarkan tingkat kesamaan anggotanya [3]. *K-Means* juga relatif efisien dan cepat, sehingga cocok untuk mengatasi jumlah data yang cukup besar, dan mempunyai fleksibilitas dalam menangani beragam jenis data, salah satunya adalah pada data pencari kerja yang mempunyai karakteristik yang berbeda – beda [4]. Selain menggunakan algoritma pada proses klastering metode PCA juga membantu dalam proses klastering untuk mengambil informasi penting dari dataset yang kompleks dengan meminimalisir dimensi pada data, kombinasi PCA dengan *K-Means* digunakan bersama untuk mengelompokkan data yang mempunyai dimensi tinggi [5].

Berdasarkan penelitian terdahulu yang dilakukan oleh peneliti sebelumnya [6] yaitu dengan topik “Penerapan Algoritma *K-Means Clustering* dalam Memetakan Produktivitas Lokasi Perkebunan Nanas PT. *Great Giant Pineapple*”. Tujuan penelitian ini adalah untuk mengelompokkan lokasi – lokasi yang terdapat pada PG 1 (*Plantation Group*), PG 2 (*Plantation Group*) dan PG 3 (*Plantation Group*) menggunakan algoritma *K-Means* klastering dan selanjutnya menganalisa strategi untuk masing – masing klaster. Metode penelitian ini menggunakan metode *elbow* dan metode *silhouette coefficient*. Hasil dari penelitian ini menunjukkan bahwa jumlah klaster yang paling optimal menggunakan metode *elbow* dan *silhouette coefficient* adalah 2 dan didapatkan hasil anggota untuk masing – masing ke tiga PG (*Plantation Group*) yaitu Pada PG 1 klaster 1 berisikan 134 lokasi dan klaster 2 berisikan 41 lokasi, Pada PG 2 klaster 1 berisikan 166 lokasi dan klaster 2 berisikan 49 lokasi, Pada PG 3 klaster 1 berisikan 115 lokasi dan klaster 2 berisikan 79 lokasi. Klaster 1 mempunyai arti lokasi dengan produktivitas tinggi dan klaster 2 mempunyai arti lokasi dengan produktivitas rendah. Dari hasil evaluasi berdasarkan nilai *Silhouette Coefficient* pada algoritma *K-Means* mendapatkan 0,51 berdasarkan nilai dari *silhouette coefficient* maka termasuk dalam

kategori struktur baik. Strategi yang didapatkan dari hasil klastering adalah perlu dilakukannya pemerataan produktivitas setiap lokasi pada klaster yang mempunyai produktivitas rendah yaitu klaster 2 pada PG 1, PG 2 dan PG 3 yaitu dengan melakukan manajemen air yang lebih efektif, pengendalian hama dan penyakit untuk meminimalisir populasi yang terkena penyakit dan strategi untuk klaster 1 yaitu lokasi dengan produktivitas tinggi pada PG 1, PG 2, PG 3 dengan melakukan strategi pengawasan lebih lanjut agar terhindar dari pencurian sehingga produktivitas tidak akan mengalami penurunan. Selanjutnya penelitian terkait metode klasterisasi dengan algoritma *K-Means* yang dilakukan peneliti sebelumnya [7] dengan judul “*Clustering Of Popular Spotify Songs In 2023 Using K-Means Method and Silhouette Coefficient*” digunakan untuk mengelompokkan data lagu – lagu Spotify pada tahun 2023 yang dapat membantu dalam pemahaman yang lebih mendalam terkait preferensi pendengar ataupun tren musik. Metode dalam penelitian ini menggunakan teknik klastering dengan algoritma *K-Means*. Berdasarkan pengujian menunjukkan terdapat 2 klaster dengan mendapatkan nilai *Silhouette Score* sejumlah 0.81 dengan arti kategori struktur klaster kuat. Dari hasil setelah dilakukannya pengujian klastering maka dapat disarankan untuk memberikan rekomendasi musik yang lebih personal dan sesuai dengan pengguna aplikasi *Spotify*. Penelitian terdahulu dengan topik [5] “Implementasi *Principal Component Analysis* Pada *K-Means* Untuk Klasterisasi Tingkat Pendidikan Penduduk Kabupaten Semarang” penelitian ini mempunyai tujuan untuk mengelompokkan tingkat pendidikan berdasarkan variabel jenis kelamin, usia dan status individu dalam keluarga dengan menerapkan algoritma *K-Means* yang didahului oleh proses reduksi dimensi menggunakan PCA (*Principal Component Analysis*). Hasil dari PCA menunjukkan dua komponen utama yang mampu menjelaskan hingga 70% proporsi kumulatif dari keseluruhan data. Setelah proses reduksi dimensi tersebut, dilakukan analisis klaster menggunakan algoritma *K-Means* yang menghasilkan empat klaster dengan karakteristik yang berbeda – beda untuk masing – masing kelompok atau klaster.

Oleh karena itu, dalam skripsi ini menggunakan implementasi data mining metode klasterisasi menggunakan algoritma *K-Means* untuk membuat sebuah kelompok berdasarkan data pencari kerja dengan atribut usia, tingkat pendidikan,

pengalaman kerja, status disabilitas, status pernikahan dan jumlah anak yang berjudul “Klastering Pencari Kerja Menggunakan Metode *K-Means* dan PCA Di Dinas Perindustrian dan Tenaga Kerja Kota Surabaya” dengan skripsi ini diharapkan memberikan rekomendasi strategi untuk memberikan bantuan pelatihan yang disesuaikan dengan karakteristik pencari kerja di Kota Surabaya. Hasil dari skripsi ini adalah sebuah sistem representasi hasil klastering dalam bentuk *dashboard* analitik berbasis web, dapat memetakan pencari kerja yang mendaftar melalui *website* ASSIK berdasarkan karakteristik yang serupa. Diharapkan sistem ini dapat membantu meningkatkan keseimbangan dalam penyaluran tenaga kerja di Kota Surabaya.

1.2. Rumusan Masalah

Dari permasalahan diatas dapat dirumuskan masalah dalam skripsi ini yaitu sebagai berikut :

1. Bagaimana menganalisis klasterisasi pencari kerja menggunakan metode *K-Means* dan PCA ?
2. Bagaimana mengimplementasikan hasil klasterisasi dari metode terbaik antara algoritma *K-Means* serta kombinasi *K-Means* dan PCA dalam bentuk *dashboard* analitik berbasis web?

1.3. Batasan Masalah

Batasan masalah dalam penelitian ini meliputi :

1. Data yang digunakan merupakan data pencari kerja pada *website* ASSIK (Arek Suroboyo Siap Kerjo) Dinas Perindustrian dan Tenaga Kerja Kota Surabaya dari tahun 2022 – 2024
2. Dalam skripsi ini algoritma yang digunakan adalah algoritma Kmeans dengan teknik reduksi dimensi menggunakan PCA (Principal Component Analysis)
3. Merepresentasikan hasil model berdasarkan algoritma terbaik K-Means dengan PCA atau tanpa menggunakan PCA yang digunakan untuk pembuatan *dashsboard* berbasis web

1.4. Tujuan

Berdasarkan rumusan masalah yang telah dijabarkan sebelumnya maka dapat dirumuskan tujuan skripsi sebagai berikut :

1. Menganalisis klasterisasi pencari kerja menggunakan metode *K-Means* dan PCA
2. Mengimplementasikan hasil klasterisasi dari metode terbaik antara algoritma *K-Means* serta kombinasi *K-Means* dan PCA dalam bentuk *dashboard* analitik berbasis web

1.5. Sistematika Penulisan

Adapun sistematika pada penulisan skripsi ini, ditampilkan dalam limabab sebagai berikut :

BAB I PENDAHULUAN

BAB I berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan sistematika penulisan skripsi ini

BAB II TINJAUAN PUSTAKA

BAB II berisi dasar teori yang menjelaskan maksud arti dari istilah – istilah tertentu yang digunakan dalam skripsi, serta mencakup pembahasan penelitian sebelumnya sebagai referensi dan sumber literatur untuk skripsi ini.

BAB III METODOLOGI PENELITIAN

BAB III menjelaskan langkah – langkah yang diambil dalam skripsi, dimulai dari studi literatur, proses data mining, hingga penyusunan laporan. Proses data mining dilakukan dengan berpedoman pada kerangka kerja CRISP-DM.

BAB IV HASIL DAN PEMBAHASAN

BAB IV menjelaskan hasil dan analisa dari tahap – tahap yang telah dilakukan dalam skripsi. Bab ini mencakup pembahasan mengenai proses klasterisasi pencari kerja pada Dinas Perindustrian dan Tenaga Kerja Kota

Surabaya yang mendaftar melalui *website* ASSIK dengan menggunakan algoritma terbaik antara K-Means dengan PCA atau K-Means tanpa PCA.

BAB V PENUTUP

BAB V menjelaskan kesimpulan dari skripsi yang sudah dilakukan, serta rekomendasi dari penulis untuk pengembangan penelitian di masa yang akan datang.

1.6. Manfaat Penelitian

Skripsi yang dilakukan mempunyai manfaat bagi DISPERINAKER (Dinas Perindustrian dan Tenaga Kerja) Kota Surabaya. Adapun beberapa manfaat dari skripsi ini, hasil skripsi representasi klastering dari metode terbaik yang dipilih antara algoritma *K-Means* serta kombinasi *K-Means* dan PCA dalam bentuk *dashboard* analitik ini dapat digunakan untuk memahami pola dan karakteristik dari pencari kerja Kota Surabaya sehingga hal tersebut dapat bermanfaat untuk menentukan strategi rekomendasi dalam menyusun program pelatihan yang disesuaikan dengan karakteristik pencari kerja yang ada di Kota Surabaya.

Skripsi ini juga memberikan manfaat yang besar bagi penulis dalam hal pengembangan wawasan dan keterampilan teknis, khususnya dalam penerapan algoritma data mining seperti K-Means dan PCA dalam konteks nyata. Melalui proses penelitian ini, penulis memperoleh pengalaman langsung dalam pengolahan data, analisis klastering, visualisasi data serta pengembangan dashboard analitik berbasis web menggunakan *python* dan *flask*. Selain itu skripsi ini juga membantu dalam memperkuat kemampuan penulis dalam berpikir kritis, menyusun metodologi ilmiah dan mengambil kesimpulan berdasarkan hasil analisis data yang objektif. Proses ini memperluas pemahaman penulis terhadap permasalahan ketenagakerjaan dan pentingnya pemanfaatan data dalam mendukung kebijakan publik.

Selanjutnya skripsi ini juga mempunyai manfaat bagi penelitian selanjutnya karena dapat menjadi referensi awal dan dasar pengembangan bagi penelitian – penelitian selanjutnya yang ingin menganalisis permasalahan ketenagakerjaan atau

karakteristik pencari kerja dengan pendekatan data mining. Metodologi yang diterapkan yaitu, perbandingan antara algoritma K-Means dengan kombinasi K-Means dengan PCA, dapat dijadikan acuan untuk pengujian algoritma klusterisasi lainnya seperti *DBSCAN*, *Hierarchical Clustering* atau algoritma berbasis pembelajaran mesin lainnya guna memperoleh hasil yang lebih optimal.

Halaman ini sengaja dikosongkan