

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam perkembangan teknologi di era revolusi 5.0 yang cukup pesat, informasi dapat ditemukan dengan sangat efisien, terutama dalam mengumpulkan informasi untuk menganalisis fenomena yang terjadi di dunia kerja. Pembahasan yang cukup signifikan di dunia kerja adalah dinamika perpindahan pekerjaan [1]. Aktifitas berpindah kerja sebagian besar dikaitkan dengan karyawan generasi muda [2]. Fenomena yang terjadi saat ini pada karyawan generasi muda yang lahir di antara tahun 1997 hingga 2012 di Indonesia, menunjukkan kecenderungan untuk memiliki tingkat loyalitas yang rendah terhadap pekerjaan dan lebih mengutamakan fleksibilitas dalam pekerjaan serta gaya hidup. Menurut data sebanyak 49% dari mereka berencana untuk meninggalkan pekerjaan dalam waktu dua tahun, dan 61% dalam tiga tahun kedepan [3]. Karyawan yang berpindah pekerjaan dapat dipicu oleh beberapa faktor, diantaranya perihal kompensasi, kepuasan kerja, aktualisasi karyawan dan hubungan kerja [4]. Karyawan cenderung kehilangan fokus pada pekerjaan, yang dapat menyebabkan kerugian besar bagi organisasi perusahaan, seperti produktivitas menurun, kurangnya disiplin, merosotnya moral kerja, berkurangnya motivasi, dan bahkan peningkatan kecelakaan kerja [5]. Penting sebuah perusahaan memerlukan data statistik berupa analisis sentimen yang signifikan pada faktor-faktor perpindahan karir untuk dapat mengerti karakteristik karyawan milenial dalam menyamakan nilai-nilai dan ekspektasi yang dimiliki antara karyawan dan perusahaan [6], [7].

Aspect-Based Sentiment Analysis (ABSA) dapat digunakan sebagai sistem yang potensial untuk memahami persepsi individu dengan menganalisis sentimen pada faktor – faktor pergantian karir dari kalimat ulasan di media sosial [8]. *Aspect-Based Sentiment Analysis* (ABSA) sebagai pendekatan analisis sentimen yang lebih terperinci, melibatkan ekstraksi dari ketiga komponen pokok, yaitu kategori aspek, istilah aspek, dan polaritas [9]. *Aspect-Based Sentiment Analysis* (ABSA) menganalisis sentimen lebih spesifik pada setiap aspek yang telah diidentifikasi dalam setiap kalimat sentimen dengan menghitung polaritas sentimen pada masing-masing aspek yang telah diidentifikasi [10]. ABSA mengklusterisasi berdasarkan aspek menggunakan model *K-Means* sebagai metode klusterisasi dengan mengidentifikasi kategori aspek dan istilah aspek pada setiap kalimat sentimen dalam sekumpulan dataset yang sudah dilakukan ekstraksi fitur dengan merepresentasikan teks dalam bentuk vektor menggunakan *Term Frequency-Inverse Document Frequency* [11],

[12]. Model perhitungan yang digunakan dalam menganalisis sentimen berupa sentimen positif, negatif atau netral pada masing-masing klaster menggunakan model *Logistic Regression* [13].

Terdapat penelitian yang menggunakan perhitungan *Logistic Regression* pada klasifikasi ABSA dengan menganalisis setiap sentimen. *Split data* yang digunakan pada model dengan proporsi 80% data latih dan 20% data uji pada 1564 ulasan memberikan evaluasi kinerja model pada parameter akurasi sebesar 0.888 [13]. Penelitian pada 349 data ulasan menggunakan *Logistic Regression* dengan *split data* 80% data latih dan 20% data uji memberikan nilai akurasi sebesar 78,57%, presisi 76,92%, *recall* 83,3% dan nilai *F1-Score* sebesar 80% [14]. Penelitian yang menggunakan *K-Means* sebagai klusterisasi mengelompokkan penyebaran virus *corona* di Indonesia. *K-Means* mendapatkan nilai *Davies Bouldin Index* terkecil sebesar 0.064 pada salah satu klaster, dapat dikatakan metode *K-Means* baik dalam melakukan pengelompokan [15]. Penelitian yang serupa dilakukan menggunakan metode *K-Means* sebagai klusterisasi nilai ujian nasional di Indonesia sebanyak 3 klaster. Hasil evaluasi menghasilkan nilai metrik *connectivity* 11.916, *dunn* 0.246 dan *silhouette* 0.464 dikatakan bahwa algoritma *K-Means* melakukan pengelompokan dengan cukup baik, dan masih ada ruang untuk perbaikan dalam pemisahan antar klaster [16]. Penelitian selanjutnya menggunakan metode *K-Means* sebagai klusterisasi topik pada survei dilakukan pada *platform Kudata*. Pengukuran nilai terbaik untuk k menggunakan *silhouette score* dengan rentang nilai k dari 2 hingga 7 menghasilkan jumlah klaster yang optimal dan nilai metrik *silhouette score* tertinggi untuk setiap sentimen dan kombinasi dataset [17].

Penelitian ini mengidentifikasi sentimen karyawan yang menjadi alasan utama dalam melakukan perpindahan karir dengan menganalisis sentimen karyawan pada faktor-faktor perpindahan karir yang dapat dijadikan sebagai data konkret dalam strategi retensi karyawan pada perusahaan untuk menghadapi tren pergeseran karir dikalangan generasi muda. Faktor-faktor perpindahan karir meliputi, kompensasi, kepuasan kerja, aktualisasi karyawan, hubungan kerja. Analisis sentimen dilakukan menggunakan metode *Aspect-Based Sentiment Analysis* (ABSA) dan pengelompokan sentimen karyawan pada ke-empat aspek yang dapat dilakukan dengan mendefinisikan sebuah istilah aspek berupa *value dictionary* sebagai nilai *centroid* awal pada model *K-Means* di dalam kategori aspek (*key dictionary*).

Tahapan penelitian meliputi pengumpulan data teks ulasan media sosial X terkait perpindahan karir dengan teknik *crawling data*, *preprocessing*, pelabelan sentimen pada

dataset menggunakan kamus leksikon, ekstraksi fitur menggunakan TD-IDF (*Term Frequency-Inverse Document Frequency*), implementasi model *K-Means* dalam menentukan kluster aspek dan mengelompokkan dokumen berdasarkan aspek. Melakukan pengujian model *Logistic Regression* yang dilatih dengan menentukan proporsi *split data* dan penggunaan algoritma *Synthetic Minority Over-Sampling Technique* (SMOTE) sehingga mendapatkan nilai akurasi model yang terbaik berdasarkan *data testing*. Mengimplementasikan model *Logistic Regression* pada masing-masing kluster. Mengevaluasi model *Logistic Regression* menggunakan parameter akurasi, presisi, *recall* dan *F1-score* dengan *Confusion Matrix*, serta evaluasi model *K-Means* menggunakan *Confusion Matrix* dan *Davies Bouldin Index* (DBI).

Berdasarkan uraian sebelumnya, maka pada tugas akhir ini diangkat topik penelitian yang berjudul **"ANALISIS SENTIMEN PADA FAKTOR-FAKTOR YANG MEMPENGARUHI PERPINDAHAN KARIR DENGAN MENGGUNAKAN METODE ASPECT-BASED SENTIMENT ANALYSIS DAN K-MEANS"**. Diharapkan dapat memberikan kontribusi dalam pengembangan kebijakan perusahaan serta manajemen sumber daya manusia yang lebih efektif.

1.2 Rumusan Masalah

- 1) Melakukan analisis sentimen pada setiap faktor-faktor perpindahan karir menggunakan metode klasifikasi *Aspect-Based Sentiment Analysis* (ABSA) dan metode klasterisasi *K-Means*.
- 2) Mengukur kinerja model klasifikasi *Aspect-Based Sentiment Analysis* (ABSA) dengan *Logistic Regression* dan klasterisasi *K-Means* menggunakan *Confusion Matrix*, akurasi, presisi, *recall* dan *F1-score*, serta *Davies Bouldin Index* (DBI).

1.3 Tujuan

Tujuan penelitian ini adalah menganalisis faktor-faktor yang mempengaruhi perpindahan karir menggunakan metode *Aspect-Based Sentiment Analysis* (ABSA) dengan menggunakan model *K-Means* sebagai klasterisasi dan model *Logistic Regression* sebagai model klasifikasi sentimen pada ulasan karyawan serta mengevaluasi model *Logistic Regression* dan *K-Means*.

1.4 Manfaat

- 1) Penyempurnaan pemahaman tentang sentimen karyawan dan faktor yang mempengaruhi perpindahan karir berdasarkan analisis sentimen yang berfokus pada aspek yang diidentifikasi.
- 2) Penelitian dapat dijadikan sebagai bahan pertimbangan bagi perusahaan untuk melakukan retensi karyawan dalam menghadapi tren pergeseran karir karyawan serta memberikan kontribusi bagi literatur akademik dibidang sains dan teknologi.

1.5 Ruang Lingkup

Beberapa ruang lingkup yang akan dibatasi oleh keterbatasan dan asumsi yang harus dipertimbangkan:

1. Data teks yang digunakan merupakan hasil dari proses *crawling data* ulasan dari *platform* media sosial X dengan konten terkait perpindahan karir dengan jumlah sentimen sebanyak 2985 komentar antara 1 Januari 2021 hingga 31 Desember 2024. Frasa yang digunakan yakni, “Alasan pindah kerja”, “Alasan pindah perusahaan”, “Alasan resign kerja”, “Pengaruh pindah kerja”, dan “*Toxic* tempat kerja”.
2. Data teks yang akan dianalisis merupakan teks bahasa Indonesia.
3. *Preprocessing* untuk memastikan tahapan pembersihan teks meliputi *case folding*, penghapusan karakter khusus, normalisasi, tokenisasi, *stemming*, *stopwords* dan ekstraksi fitur menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*).
4. Klasterisasi sentimen berdasarkan aspek yang telah ditentukan yakni, kompensasi, kepuasan kerja, aktualisasi karyawan dan hubungan kerja menggunakan metode *K-Means*.
5. Analisis sentimen berupa sentimen positif, negatif, atau netral pada masing-masing dokumen menggunakan model *Logistic Regression*.
6. Mengukur kinerja model *Logistic Regression* menggunakan evaluasi *Confusion Matrix*, akurasi, presisi, *recall*, dan *F1-score*..
7. Mengukur kinerja model klasterisasi sentimen *K-Means* menggunakan evaluasi *Confusion Matrix* dan *Davies Bouldin Index* (DBI).