

BAB 1

PENDAHULUAN

1.1. Latar Belakang Masalah

Pada era modern yang serba digital ini, metode yang dipakai perusahaan dalam melakukan pemilihan karyawan juga tengah mengalami digitalisasi. Misalnya dari lamaran kerja yang awalnya selalu berupa kertas-kertas dalam amplop kini dapat diwakilkan dengan teks ataupun surat digital yang dikirim melalui internet. Selain itu, metode yang digunakan untuk menilai dan menganalisa kandidat pekerja pun dapat dilakukan dengan bantuan teknologi. Masalah yang ditimbulkan apabila perusahaan tidak berhasil mempekerjakan kandidat yang tepat dapat berupa kerugian waktu, kerugian materil, dan penurunan kinerja perusahaan.

Peran teknologi pun diharapkan dapat meningkatkan efisiensi dan kinerja proses pemilihan karyawan. Kerugian waktu dalam proses pemilihan karyawan dapat berupa waktu yang diperlukan untuk mencari kandidat hingga melatih dan mempekerjakan kandidat. Sedangkan kerugian materil dapat berupa biaya yang diperlukan untuk menari kandidat baru, biaya pelatihan karyawan. Selain itu kerugian dalam bentuk penurunan kinerja perusahaan juga dapat terjadi karena perusahaan tidak mampu memenuhi kebutuhan karyawan yang berkualitas. Oleh karena itu, pemilihan kandidat yang tepat adalah hal yang cukup penting.

Penelitian ini menggunakan Count Vectorizer untuk memproses data teks surat lamaran pekerjaan. Beberapa penelitian terkini juga telah memanfaatkan teknologi ini untuk mempelajari data pelamar kerja. Misalnya adalah implementasi Machine Learning untuk menganalisa video lamaran pekerjaan [1], prediksi surat riwayat hidup pelamar kerja [2], dan prediksi lamaran pekerjaan [3]. Count Vectorizer adalah teknik untuk memproses kata-kata agar dapat diproses ataupun dimengerti oleh computer dengan mengubah data kata-kata menjadi data biner.

Penelitian ini menerapkan Count Vectorizer dari library Natural Language Toolkit (NLTK) untuk memproses data berupa teks. Metode serupa juga telah digunakan untuk meneliti lamaran pekerjaan dengan algoritma Machine Learning

Latent Dirichlet Allocation (LDA) yang menghasilkan akurasi sebesar 97% [4]. Selain NLTK, terdapat library lainnya yang digunakan untuk meneliti lamaran pekerjaan antara lain seperti Text Blob [5], dan Core NLP [6]. Keunggulan NLTK adalah dipakai oleh peneliti dari jurusan komputer dan linguistik sehingga terdapat keseimbangan antara kedua ruang lingkup tersebut. Selain itu NLTK juga mengutamakan kemudahan dalam penggunaan dan mendukung *object-oriented programming* yang dikembangkan dengan performa yang cepat dan dapat diandalkan.

Selanjutnya, penelitian ini juga menggunakan Logistic Regression sebagai algoritma Machine Learning untuk menganalisa dan memprediksi data yang sudah diproses dengan Count Vectorizer. Logistic Regression adalah algoritma Machine Learning yang menggunakan metode regresi. Teknik Machine Learning atau Pembelajaran Mesin adalah teknik yang meniru cara pembelajaran manusia untuk menganalisa data. Beberapa algoritma Machine Learning yang sebelumnya telah digunakan untuk meneliti bersamaan dengan teknik Count Vectorizer antara lain Random Forest [7], Logistic Regression [8] dan Multinomial Naïve Bayes [9].

Jumlah data yang diteliti pada penelitian ini adalah sebanyak 2.155 data yang didapatkan pada tahun 2020 hingga 2021. Sedangkan jenis klasifikasi yang dilakukan pada penelitian ini adalah *Binary Classification* atau klasifikasi dengan dua output. Kedua output yang diteliti pada penelitian ini adalah berhasil atau tidaknya seorang pelamar kerja yang disimbolkan dengan nilai 1 (berhasil) dan 0 (gagal). Data berupa lamaran kerja juga telah diteliti untuk meneliti keberhasilan pelamar kerja [10], dan juga untuk memprediksi jenis karir [11]. Pada penelitian ini data yang diteliti didapatkan dari sebuah situs layanan online untuk melamar pekerjaan. Secara garis besar, data tersebut berupa teks atau kata-kata yang kemudian diteliti dengan metode Count Vectorizer dan Logistic Regression untuk memprediksi keberhasilan pelamar kerja.

1.2. Rumusan Masalah

Rumusan masalah dalam penelitian ini, antara lain:

1. Masalah sulitnya menilai pelamar kerja secara manual berdasarkan surat lamaran kerjanya

2. Masalah sulitnya menggunakan tipe data berupa teks atau paragraf untuk diprediksi menggunakan algoritma Machine Learning

1.3. Tujuan

Penelitian ini bertujuan untuk menganalisa keakuratan algoritma Logistic Regression yang digabungkan dengan metode Count Vectorizer untuk memprediksi surat lamaran pekerjaan.

1.4. Manfaat

Manfaat yang hendak dicapai pada penelitian ini antara lain:

1. Menghasilkan algoritma Machine Learning yang dapat memprediksi karyawan yang berkualitas melalui surat lamaran pekerjaannya
2. Membantu perusahaan dalam mempercepat proses perekrutan karyawan

1.5. Batasan Masalah

Ruang lingkup penelitian ini, adalah sebagai berikut:

1. 2.155 data surat lamaran pekerjaan dengan format teks
2. Melakukan analisa dan eksperimen untuk memprediksi surat lamaran pekerjaan
3. Algoritma machine learning yang digunakan adalah Logistic Regression
4. Metode Count Vectorizer yang digunakan berasal dari library NLTK
5. Bahasa pemrograman yang digunakan adalah Python

1.6. Keterbaruan

Count Vectorizer telah diteliti bersamaan dengan berbagai algoritma machine learning [7]–[9]. Meskipun demikian, belum banyak penelitian yang menggunakan Machine Learning dengan Logistic Regression dan Count Vectorizer untuk meneliti surat lamaran pekerjaan. Penelitian ini memberikan kontribusi berupa eksperimen menggunakan algoritma Logistic Regression dan Count Vectorizer dengan 2.155 data teks lamaran pekerjaan pada tahun 2020-2021.