

Proposal Stage 0

Agi Rahmawandi
Shan Ramadhan
Muhammad Muqorrobin
I Gusti Ngurah Agung Hari Vijaya Kusuma

August 26, 2025

Submission Links

- Repository: github.com/4Kings-Rakamin

1 Pendahuluan

1.1 Latar Belakang

Dalam dunia perekrutan karyawan, pengambilan keputusan yang tepat sangat penting untuk memastikan bahwa perusahaan mendapatkan kandidat yang paling sesuai dengan kebutuhan dan budaya organisasi. Dengan kemajuan teknologi dan analisis data, perusahaan kini dapat memanfaatkan data historis untuk meningkatkan proses perekrutan mereka. Dataset yang digunakan dalam proyek ini berisi informasi tentang berbagai kandidat yang telah melamar pekerjaan di sebuah perusahaan, termasuk fitur-fitur seperti umur, jenis kelamin, tingkat pendidikan, pengalaman kerja, dan skor wawancara. Dengan menganalisis data ini, kita dapat mengidentifikasi pola dan faktor-faktor yang mempengaruhi keputusan perekrutan, sehingga membantu perusahaan dalam membuat keputusan yang lebih baik di masa depan.

1.2 Riset Bisnis

Dalam konteks bisnis, proses perekrutan yang efisien dan efektif sangat penting untuk keberhasilan jangka panjang perusahaan. Dengan menggunakan data historis dari proses perekrutan sebelumnya, perusahaan dapat mengidentifikasi karakteristik kandidat yang paling berhasil dan sesuai dengan kebutuhan organisasi. Hal ini tidak hanya membantu dalam mengurangi biaya dan waktu yang dihabiskan untuk proses perekrutan, tetapi juga meningkatkan kualitas karyawan yang direkrut. Dengan demikian, analisis data ini dapat memberikan wawasan berharga bagi tim HR dan manajemen dalam mengoptimalkan strategi perekrutan mereka.

Adapun penggunaan model machine learning dalam proses perekrutan ini dapat membantu perusahaan dalam mengelola biaya secara lebih efisien. Dengan memprediksi kandidat

yang memiliki kemungkinan besar untuk diterima berdasarkan data historis, perusahaan dapat mengurangi jumlah pengiriman email penolakan. Hal ini tidak hanya menghemat waktu dan sumber daya, tetapi juga memungkinkan tim HR untuk lebih fokus pada aspek-aspek lain dari proses perekrutan, seperti pengembangan karyawan dan retensi. Dengan demikian, implementasi model machine learning dalam proses perekrutan dapat memberikan manfaat ekonomi yang signifikan bagi perusahaan.

1.3 Problem Statement

Berdasarkan analisis awal terhadap dataset perekrutan, ditemukan bahwa terdapat ketidakseimbangan yang signifikan antara jumlah kandidat yang diterima dan ditolak. Dari total 1500 kandidat, sekitar 69% diterima sementara 31% ditolak. Hal ini menyebabkan banyaknya email penolakan yang harus dikirimkan oleh perusahaan, yang pada akhirnya dapat meningkatkan biaya operasional dan mengurangi efisiensi proses perekrutan. Oleh karena itu, penting untuk mengembangkan model machine learning yang dapat memprediksi keputusan perekrutan dengan akurasi tinggi, sehingga perusahaan dapat mengurangi jumlah email penolakan yang perlu dikirimkan. Dengan demikian, tujuan utama dari proyek ini adalah untuk menciptakan model prediktif yang tidak hanya akurat tetapi juga mampu menangani ketidakseimbangan kelas dalam dataset, sehingga membantu perusahaan dalam mengoptimalkan proses perekrutan mereka.

1.4 Goals and Objectives

Adapun Tujuan dan Objektif dari proyek ini adalah sebagai berikut:

1. Mengembangkan model machine learning yang dapat memprediksi keputusan perekrutan (HiringDecision) dengan akurasi tinggi.
2. Mengurangi jumlah email penolakan yang perlu dikirimkan oleh perusahaan, sehingga mengoptimalkan biaya operasional.
3. Menekan Biaya perusahaan dalam proses perekrutan dengan memanfaatkan analisis data historis.

Business metric yang relevan untuk ditingkatkan adalah sebagai berikut:

1. Akurasi Model: Meningkatkan akurasi prediksi model untuk memastikan keputusan perekrutan yang lebih tepat.
2. Precision dan Recall: Meningkatkan precision untuk mengurangi jumlah false positives (kandidat yang diprediksi diterima tetapi sebenarnya ditolak) dan recall untuk memastikan bahwa kandidat yang layak tidak terlewatkan.
3. Reduksi Biaya Operasional: Mengukur pengurangan biaya yang terkait dengan pengiriman email penolakan dan proses perekrutan secara keseluruhan.
4. Waktu Proses Perekrutan: Mengurangi waktu yang dibutuhkan untuk menyelesaikan proses perekrutan dengan menggunakan model prediktif.

2 Tinjauan Pustaka

2.1 Proses Rekrutmen

Proses rekrutmen adalah serangkaian langkah yang diambil oleh perusahaan untuk menarik, menilai, dan memilih kandidat yang paling sesuai untuk mengisi posisi yang tersedia. Proses ini biasanya dimulai dengan identifikasi kebutuhan tenaga kerja, diikuti oleh pencarian kandidat melalui berbagai saluran seperti iklan lowongan kerja, agen perekrutan, dan media sosial. Setelah kandidat ditemukan, mereka akan melalui tahap seleksi yang melibatkan penilaian kualifikasi, wawancara, dan tes keterampilan. Akhirnya, keputusan perekrutan dibuat berdasarkan evaluasi menyeluruh dari semua kandidat yang telah melalui proses seleksi. Proses rekrutmen yang efektif tidak hanya memastikan bahwa perusahaan mendapatkan karyawan yang berkualitas, tetapi juga membantu dalam membangun budaya organisasi yang positif dan mendukung tujuan bisnis jangka panjang. (Mathis et al., 2017)

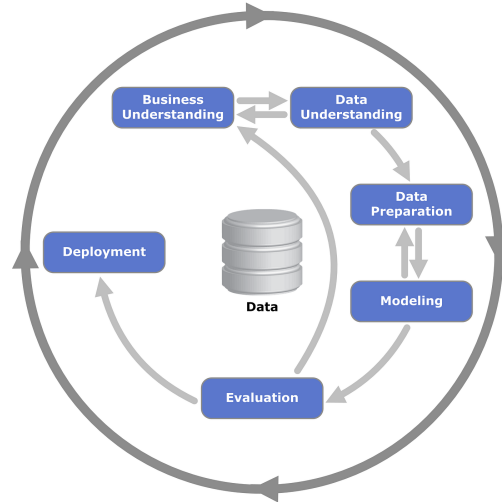
Adapun beberapa faktor yang mempengaruhi keputusan perekrutan meliputi:

1. Demografis: usia, gender, latar belakang pendidikan. (Ng & Burke, 2005)
2. Pengalaman kerja: jumlah tahun pengalaman dan variasi perusahaan sebelumnya. (Ployhart, 2006)
3. Kompetensi teknis dan soft skills: hasil tes keterampilan (skill score), wawancara, dan penilaian kepribadian. (Schmidt & Hunter, 1998)
4. Faktor eksternal: jarak tempat tinggal ke kantor sering menjadi pertimbangan dalam retensi. (Hausknecht et al., 2009)
5. Strategi rekrutmen: pendekatan organisasi (job fairs, rekrutmen online, campus hiring) dapat memengaruhi kualitas kandidat. (Breaugh, 2013).

Berdasarkan studi literatur, faktor-faktor di atas secara signifikan mempengaruhi keputusan perekrutan. Dimana hal tersebut dapat menjadi pertimbangan penting dalam pengembangan model prediktif untuk proses perekrutan.

2.2 CRISP-DM

CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah metodologi standar yang digunakan dalam proyek data mining dan analisis data. Metodologi ini terdiri dari enam fase utama yang membantu dalam mengorganisir dan mengelola proyek data secara sistematis. (Chumbar, 2020)



Gambar 1: Alur Kerja CRISP-DM

Gambar 1 menggambarkan alur kerja CRISP-DM yang terdiri dari enam fase utama, yaitu:

1. Business Understanding: Memahami tujuan bisnis dan kebutuhan proyek.
2. Data Understanding: Mengumpulkan dan memahami data yang tersedia.
3. Data Preparation: Membersihkan dan mempersiapkan data untuk analisis.
4. Modeling: Membangun model prediktif menggunakan teknik machine learning.
5. Evaluation: Mengevaluasi model untuk memastikan bahwa tujuan bisnis tercapai.
6. Deployment: Menerapkan model dalam lingkungan produksi untuk digunakan dalam pengambilan keputusan bisnis.

2.3 EDA (Exploratory Data Analysis)

Exploratory Data Analysis (EDA) adalah proses awal dalam analisis data yang bertujuan untuk memahami karakteristik dan pola dalam dataset. EDA melibatkan berbagai teknik visualisasi dan statistik untuk mengeksplorasi data, mengidentifikasi outlier, dan menemukan hubungan antara variabel. Proses ini sangat penting karena membantu dalam mengarahkan langkah-langkah selanjutnya dalam analisis data, seperti pemilihan fitur dan pemodelan. (Tukey, 1977)

2.4 T-Test

T-Test adalah metode statistik yang digunakan untuk membandingkan rata-rata dari dua kelompok data. Uji ini membantu menentukan apakah perbedaan antara kedua kelompok tersebut signifikan secara statistik atau hanya terjadi secara kebetulan. T-Test dapat digunakan dalam berbagai konteks, seperti membandingkan hasil tes antara dua kelompok siswa

atau mengevaluasi efektivitas dua metode pengajaran yang berbeda. Hasil dari uji T-Test memberikan nilai p-value yang digunakan untuk menilai signifikansi perbedaan antara kedua kelompok. Rumus untuk menghitung nilai T-Test (t) adalah sebagai berikut:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (1)$$

Dimana:

- t = nilai T-Test
- \bar{X}_1 = rata-rata kelompok pertama
- \bar{X}_2 = rata-rata kelompok kedua
- s_1^2 = varians kelompok pertama
- s_2^2 = varians kelompok kedua
- n_1 = ukuran sampel kelompok pertama
- n_2 = ukuran sampel kelompok kedua

Singkatnya uji t-test ini memudahkan kita dalam membandingkan dua kelompok data untuk menentukan apakah ada perbedaan yang signifikan antara keduanya. Apabila p-value lebih kecil dari tingkat signifikansi (misalnya 0,05), maka kita menolak hipotesis nol dan menyimpulkan bahwa ada perbedaan yang signifikan antara kedua kelompok tersebut. Ini berguna apabila fitur numerik ingin dibandingkan terhadap target kategorikal.

2.5 Chi Square Test

Chi Square Test adalah metode statistik yang digunakan untuk menguji hubungan antara dua variabel kategorikal. Uji ini membandingkan frekuensi yang diamati dalam setiap kategori dengan frekuensi yang diharapkan jika tidak ada hubungan antara variabel. Hasil dari uji Chi Square memberikan nilai p-value yang digunakan untuk menentukan apakah hubungan antara variabel tersebut signifikan secara statistik.

Rumus untuk menghitung nilai Chi Square (χ^2) adalah sebagai berikut:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (2)$$

Dimana:

- χ^2 = nilai Chi Square
- O_i = frekuensi yang diamati dalam kategori ke-i
- E_i = frekuensi yang diharapkan dalam kategori ke-i

- \sum = penjumlahan untuk semua kategori

Apabila p-value lebih kecil dari tingkat signifikansi (misalnya 0,05), maka kita menolak hipotesis nol dan menyimpulkan bahwa ada hubungan yang signifikan antara kedua variabel tersebut. Ini berguna apabila fitur kategorikal ingin dibandingkan terhadap target kategorikal.

2.6 Standard Scaler

Standard Scaler adalah teknik normalisasi data yang digunakan untuk mengubah fitur numerik sehingga memiliki rata-rata (mean) nol dan standar deviasi (standard deviation) satu. Proses ini membantu dalam mengurangi skala variabilitas antar fitur, sehingga model machine learning dapat belajar lebih efektif. Standard Scaler sangat berguna ketika fitur-fitur dalam dataset memiliki rentang nilai yang berbeda-beda, karena dapat meningkatkan konvergensi dan kinerja model. Rumus untuk menghitung nilai yang telah dinormalisasi (z) menggunakan Standard Scaler adalah sebagai berikut:

$$z = \frac{(X - \mu)}{\sigma} \quad (3)$$

Dimana:

- z = nilai yang telah dinormalisasi
- X = nilai asli dari fitur
- μ = rata-rata (mean) dari fitur
- σ = standar deviasi (standard deviation) dari fitur
- $X - \mu$ = selisih antara nilai asli dan rata-rata
- $\frac{(X-\mu)}{\sigma}$ = hasil pembagian selisih dengan standar deviasi

3 Metodologi Penelitian

Daftar Pustaka

- Breaugh, J. A. (2013). Employee recruitment. *Annual Review of Psychology*, 64, 389–416. <https://doi.org/10.1146/annurev-psych-113011-143757>
- Chumbar, S. (2020). *The crisp-dm process: A comprehensive guide* [Accessed: 2025-08-26]. <https://medium.com/@shawn.chumbar/the-crisp-dm-process-a-comprehensive-guide-4d893aecb151>
- Hausknecht, J. P., Rodda, J., & Howard, M. J. (2009). Targeted employee retention: Performancebased and jobrelated differences in reported reasons for staying. *Human Resource Management*, 48(2), 269–288. <https://doi.org/10.1002/hrm.20279>
- Mathis, R., Jackson, J., Valentine, S., & Meglich, P. (2017). *Human resource management*. Cengage Learning.
- Ng, E. S. W., & Burke, R. J. (2005). Person–organization fit and the war for talent: Does diversity management make a difference? *The International Journal of Human Resource Management*, 16(7), 1195–1210. <https://doi.org/10.1080/09585190500144038>
- Ployhart, R. E. (2006). Staffing in the 21st century: New challenges and strategic opportunities. *Journal of Management*, 32(6), 868–897. <https://doi.org/10.1177/0149206306293625>
- Schmidt, F., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology. *Psychological Bulletin*, 124(2), 262–274.
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison–Wesley.