# Bryan Ozora / C14220070

Daud Christian / C14220206

Janssen Nathanael Siswanto / C14220279

Ruben Emmanuel / C14220320

# **Laporan JobGenie**

## Gambaran Umum

JobGenie adalah aplikasi pencocokan pekerjaan yang cerdas menggunakan machine learning dan natural language processing untuk mencocokkan pencari kerja dengan peluang pekerjaan yang sesuai. Aplikasi ini menganalisis resume, mengekstrak keterampilan dan kategori, serta memberikan rekomendasi pekerjaan yang dipersonalisasi berdasarkan kesamaan semantik, kesamaan kategori, dan pencocokan keterampilan.

## **Fitur Utama**

### **Untuk Pencari Kerja**

* **Analisis Resume**: Upload resume PDF untuk ekstraksi dan analisis teks otomatis
* **Dukungan Multibahasa**: Mendukung resume dalam bahasa Inggris dan Indonesia
* **Prediksi Kategori**: Secara otomatis mengkategorikan resume ke dalam kategori pekerjaan
* **Ekstraksi Keterampilan**: Mengidentifikasi keterampilan teknis dan profesional dari resume
* **Pencocokan Pekerjaan Cerdas**: Mencocokkan resume dengan lowongan pekerjaan menggunakan beberapa algoritma penilaian
* **Pelacakan Aplikasi**: Melamar pekerjaan langsung melalui platform

### **Untuk Perusahaan**

* **Posting Pekerjaan**: Mengirimkan deskripsi pekerjaan detail dengan ekstraksi fitur otomatis
* **Dashboard Aplikasi**: Melihat dan mengelola aplikasi dari pencari kerja
* **Analitik Pelamar**: Mengakses skor pencocokan detail dan wawasan pelamar

## C. **Arsitektur Sistem**

1. **Sistem Autentikasi**
   * Registrasi dan login pengguna untuk pencari kerja, dan akun bisnis
   * Penyimpanan database pengguna dan akun bisnis berbasis CSV
2. **Model Machine Learning**
   * **Klasifikasi Resume Bahasa Inggris**: Model berbasis BERT untuk kategorisasi resume bahasa Inggris
   * **Klasifikasi Resume Bahasa Indonesia**: Model berbasis IndoBERT untuk kategorisasi resume bahasa Indonesia
   * **Ekstraksi Keterampilan**: Model spaCy NER dengan pola keterampilan khusus
   * **Pencocokan Semantik**: paraphrase-multilingual-MiniLM-L12-v2 untuk membandingkan embedding antar dokumen
3. **Sistem Pencocokan Pekerjaan**
   * Algoritma penilaian multi-faktor yang menggabungkan:
     + Pencocokan kategori (bobot 50%)
     + Tumpang tindih keterampilan (bobot 30%)
     + Kesamaan semantik (bobot 20%)
4. **Pipeline Data**
   * Scraping pekerjaan dari API eksternal
   * Prekomputasi fitur untuk pencocokan yang efisien (simpan hasil embedding di csv)
   * Pelacakan dan penyimpanan aplikasi

## 

## D. **Implementasi Teknis**

#### **Model Bahasa Inggris (BERT)**

* Model Dasar: bert-base-uncased
* Data Pelatihan: Dataset resume dari kaggle
* Konfigurasi Pelatihan:
  + Epoch: 20
  + Batch Size: 8
  + Learning Rate: 5e-5
  + Panjang Sekuens Maksimal: 512

#### **Model Bahasa Indonesia (INDOBERT)**

* Model Dasar: indobenchmark / indobert-base-p1
* Data Pelatihan: Dataset resume dari kaggle yang di translate dengan machine translation
* Konfigurasi Pelatihan:
  + Epoch: 20
  + Batch Size: 8
  + Learning Rate: 3e-5
  + Panjang Sekuens Maksimal: 512

#### **Model Pembanding Kesamaan Semantik (paraphrase-multilingual-MiniLM-L12-v2)**

* Model Pretrain yang digunakan untuk membandingkan kesamaan semantik dari job dan resume
* Digunakan dalam mengkonversi job description menjadi embedding dan mengkonversi resume menjadi embedding

### **Algoritma Pencocokan Pekerjaan**

Sistem pencocokan menggunakan pendekatan penilaian berbobot:

skor\_akhir = (0.5 \* skor\_kategori) + (0.3 \* skor\_keterampilan) + (0.2 \* skor\_semantik)

Dimana:

* **Skor Kategori**: Pencocokan biner antara kategori resume dan pekerjaan
* **Skor Keterampilan**: Jumlah Skill yang ada di resume dan job / jumlah skill yang ada di job
* **Skor Semantik**: Kesamaan kosinus antara embedding resume dan pekerjaan

### **Penyimpanan Data**

* users.csv: Akun pencari kerja
* business\_users.csv: Akun bisnis
* applications.csv: list aplikasi kerja
* jobs.csv: list pekerjaan yang mencakup deskripsi, embedding dan lainnya

## E. Alur Penggunaan

### **Alur Pencari Kerja**

1. **Autentikasi**: Login atau buat akun
2. **Upload Resume**: Upload resume PDF
3. **Analisis**: Pemrosesan otomatis menampilkan:
   * Bahasa yang terdeteksi
   * Kategori yang diprediksi
   * Keterampilan yang diekstrak
   * Metrik waktu pemrosesan
4. **Rekomendasi Pekerjaan**: Lihat 10 pekerjaan teratas yang cocok dengan:
   * Skor kompatibilitas
   * Metrik pencocokan detail
   * Deskripsi pekerjaan

**2. Alur Bisnis**

1. **Autentikasi**: Login/registrasi bisnis
2. **Posting Pekerjaan**: Kirim pekerjaan dengan:
   * Detail pekerjaan (judul, perusahaan, lokasi, gaji)
   * Pemilihan kategori dari daftar yang telah ditentukan
   * Deskripsi pekerjaan
3. **Manajemen Aplikasi**: Lihat detail pelamar:
   * Rincian skor pencocokan
   * Keterampilan yang diekstrak
   * Konten resume lengkap

## F. **Kategori Pekerjaan**

Sistem mendukung 24 kategori pekerjaan yang telah ditentukan:

* "accountant", "advocate", "agriculture", "apparel", "arts", "automobile", "aviation", "banking", "bpo", "business development", "chef", "construction", "consultant", "designer", "digital marketing", "engineering", "finance", "fitness", "healthcare", "hr", "information technology", "public relations", "sales", "teacher"

## G. **Sumber Data**

### **Scraping Pekerjaan**

* **API**: JSearch RapidAPI untuk data pekerjaan
* **Cakupan**: 4 halaman per kategori
* **Pemrosesan**: Pembersihan HTML dan ekstraksi data terstruktur

### **Dataset Kaggle**

<https://www.kaggle.com/datasets/snehaanbhawal/resume-dataset>

### **Dataset Bahasa Indonesia**

<https://www.kaggle.com/datasets/snehaanbhawal/resume-dataset>

(dataset kaggle yang sama di translate ke bahasa indonesia menggunakan

MarianMT Model dari hugginface)

## 

## H. **Kesimpulan**

JobGenie merupakan solusi komprehensif untuk pencocokan pekerjaan yang cerdas, menggabungkan teknik NLP modern dengan desain pengalaman pengguna yang praktis. Aplikasi ini berhasil menjembatani kesenjangan antara pencari kerja dan pemberi kerja melalui pencocokan otomatis yang cerdas sambil mempertahankan kesederhanaan dan kemudahan penggunaan.

Arsitektur modular memungkinkan ekspansi dan perbaikan yang mudah, membuatnya cocok untuk deployment skala kecil maupun implementasi tingkat enterprise dengan strategi scaling yang tepat.