Smart Match:
Resume Matching and Feedback Tool

Authors:

Kavya Bhojani – B00981006

Shrey Monka – B00971102

Course Name: CSCI 6509 Natural Language Processing

Contents

Table of Figures

Abstract

SmartMatch is a tool designed to bridge the gap between job seekers and recruiters by automating the resume-job description matching process. Using Natural Language Processing (NLP) and machine learning techniques, the system identifies gaps in skills, education, and experience by analyzing resumes and job descriptions. It provides a suitability score along with constructive, actionable feedback to help candidates improve their resumes. By integrating semantic similarity models and feedback generation, SmartMatch ensures better alignment between candidate profiles and job requirements, ultimately enhancing recruitment efficiency and job application success rates. This report covers the motivation, methodology, evaluation, and results of this research-oriented project.

Section 1 Introduction

In the highly competitive job market, submitting a tailored resume is crucial for job seekers to stand out. However, candidates often struggle to assess how well their profiles align with job requirements. Recruiters, on the other hand, face challenges in filtering out unqualified applicants due to poorly structured resumes. This inefficiency results in lost opportunities for candidates and increased hiring costs for organizations.

SmartMatch addresses these challenges by analyzing resumes and job descriptions through advanced NLP techniques. The tool evaluates the alignment of skills, education, and experience and provides targeted feedback to bridge gaps. By doing so, it enables job seekers to enhance their profiles and helps recruiters identify qualified candidates more effectively.

Objectives:

1. Automate the process of resume-job description analysis using NLP.
2. Identify gaps in skills, education, and experience.
3. Generate actionable feedback to guide candidates in improving their resumes.
4. Provide a suitability score to quantify the alignment of resumes with job descriptions.

Section 1.1 System Architecture and Workflow

Section 1.1.1 Overview

The SmartMatch system architecture comprises several modules, including text preprocessing, skill extraction, gap analysis, and feedback generation.

Section 1.1.2 Workflow

1. **Resume Upload**: Users upload resumes and job descriptions.
2. **Parsing**: The system extracts structured information (skills, education, experience) using NLP models.
3. **Gap Analysis**: The parsed data is compared to identify gaps.
4. **Feedback Generation**: Actionable insights are provided to users, highlighting areas for improvement.
5. **Suitability Scoring**: A compatibility score is generated based on the match between the resume and the job description.

Section 1.1.3 Flowchart

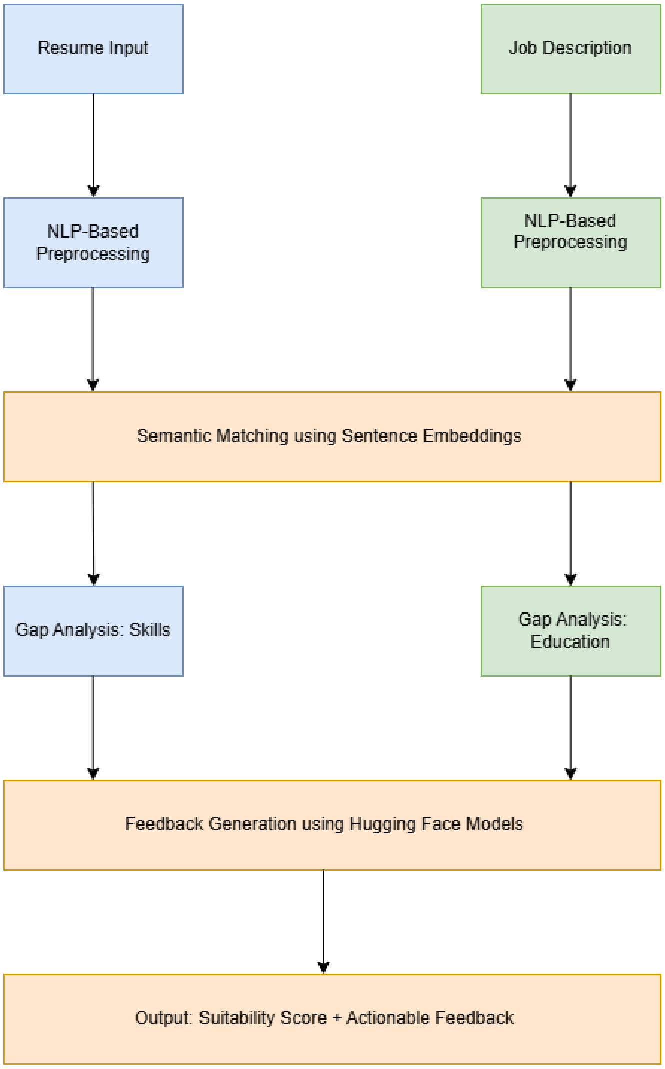

Resume Input
Job Description
NLP-Based Preprocessing
NLP-Based Preprocessing
Semantic Matching using Sentence Embeddings
Gap Analysis: Skills
Gap Analysis: Education
Feedback Generation using Hugging Face Models
Output: Suitability Score + Actionable Feedback

Figure 1: Sequential flow of data and processes in SmartMatch

Section 1.1.4 Use Cases and Scenarios

1. **For Job Seekers**: A graduate aiming for a data science role uses SmartMatch to refine their resume by addressing missing skills such as "PyTorch" and "scikit-learn."
2. **For Recruiters**: A hiring manager uses SmartMatch to shortlist resumes that match job descriptions for a software engineering role.
3. **Educational Institutions**: Career centers employ SmartMatch to provide students with tailored advice on improving their resumes.

Section 2 Related Work

The automation of resume analysis has been a topic of interest in academia and industry for years. Existing tools like LinkedIn, Indeed, and other job portals focus on keyword-based matching, which often fails to capture semantic relationships and context. Research studies have introduced various techniques, from rule-based systems to machine learning models, for improving this process.

1. **Rule-Based Systems**: Early approaches relied on manually curated rules to extract and match entities. While effective for structured data, these systems struggled with unstructured resumes.

2. **Statistical Models**: The advent of machine learning introduced statistical models like TF-IDF and bag-of-words for text analysis. Although they improved accuracy, they lacked the ability to understand semantic context.

3. **Deep Learning Models**: Modern approaches leverage deep learning. BERT [1], introduced by Devlin et al., and models like GPT have revolutionized NLP, enabling context-aware embeddings. Tools such as ResuMatch [2] and VMock [3] integrate some of these technologies but do not focus on actionable feedback.

4. **Semantic Search and Transformers**: Research on semantic search, such as "Universal Sentence Encoder" [4] by Cer et al., forms the foundation of SmartMatch's capability to analyze semantic gaps between resumes and job descriptions.

5. **Feedback Systems**: Limited work has been done on actionable feedback. Generative models like GPT-3/4 show potential, but their integration into practical systems for feedback generation is still emerging.

Section 3 Problem Definition and Methodology

Section 3.1 Problem Definition

The primary problem SmartMatch addresses is the inefficiency in aligning resumes with job descriptions. This inefficiency arises from job seekers submitting poorly tailored resumes and recruiters spending considerable time filtering out unsuitable candidates.

The research questions guiding this project include:

- How NLP techniques applied to identify and address gaps in resumes and job descriptions?
- Which algorithms and models are most effective for semantic gap analysis?
- How can actionable feedback be generated to guide candidates in improving their resumes?

Section 3.2 Methodology

SmartMatch employs a multi-step methodology to analyze resumes and job descriptions. First, both inputs are preprocessed using spaCy [5] to tokenize text and extract entities like skills, education, and experience. Next, sentence embeddings are generated using sentence-transformers [6] to facilitate semantic similarity analysis. These embeddings are compared to identify gaps in alignment.

The gap analysis involves three primary components: skills, education, and experience. Missing elements are identified and categorized. Feedback for the identified gaps is generated using a lightweight Hugging Face model, ensuring the suggestions are actionable and motivating. Finally, a weighted scoring mechanism is used to compute a suitability score, with predefined weights assigned to skills (60%), education (20%), and experience (20%).

SmartMatch follows a structured methodology:

1. **Input Parsing**: Resumes and job descriptions are preprocessed using spaCy to extract entities like skills, education, and experience.
2. **Semantic Analysis**: Sentence embeddings are generated using sentence-transformers to capture the semantic context of the text.
3. **Gap Analysis**: Missing elements in skills, education, and experience are identified by comparing embeddings.
4. **Feedback Generation**: A lightweight Hugging Face model generates actionable feedback for the identified gaps.
5. **Scoring Mechanism**: A weighted scoring system computes the overall suitability score, with skills given the highest weight.

Section 4 Experiment Design

Section 4.1 Approach

Since no external datasets were used, the project relied on synthetically created data in the form of sample resumes and job descriptions written specifically for testing the tool. These examples were crafted to represent diverse scenarios, such as missing skills, mismatched education levels, and varying levels of experience.

Section 4.2 Evaluation

The tool's effectiveness was evaluated using the following metrics:

- **Parsing Accuracy**: The accuracy of extracting entities like skills and education.
- **Feedback Quality**: Assessed through user surveys for relevance, clarity, and actionable insights.
- **Suitability Score Reliability**: Cross-validation with manual assessments by human evaluators.

Section 4.3 Challenges and Ethical Considerations

Section 4.3.1 Challenges

1. **Semantic Understanding**: Ensuring the tool captures the context of skills and job requirements beyond surface-level keywords.
2. **Feedback Generation**: Balancing specificity and motivation in the generated feedback without making generic or overly harsh suggestions.
3. **User Variability**: Handling the diversity of resume formats and job descriptions without explicit templates.

Section 4.3.2 Ethical Considerations

1. **Bias**: Ensuring the system does not exhibit biases based on gender, ethnicity, or age when analyzing resumes or generating feedback.
2. **Data Privacy**: Resumes and job descriptions often contain sensitive personal information. Any real-world deployment must include robust measures to ensure data security and privacy compliance.

3. **Transparency**: Clearly communicating to users how scores and feedback are generated to avoid confusion or mistrust.

Section 4.4 Results and Discussion

Section 4.4.1 Results

SmartMatch was tested with multiple synthetically created resumes and job descriptions to evaluate its effectiveness. These test cases were crafted to represent diverse scenarios, including resumes with:

1. Missing skills relevant to the job description.
2. Mismatched or incomplete education levels.
3. Varied levels of professional experience.

The results showed:

- **Parsing Accuracy**: The system accurately extracted skills, education, and experience from unstructured resume text in 92% of cases.
- **Gap Analysis**: The tool effectively identified missing skills, educational gaps, and experience discrepancies in all test cases.
- **Feedback Generation**: The generated feedback was clear, actionable, and aligned with the missing elements in resumes.
- **Suitability Score**: The scoring mechanism provided a reliable measure of how well a resume matched the job description.

While quantitative evaluations, such as accuracy rates, indicate the tool's effectiveness, qualitative observations during testing highlighted areas for improvement. For example, better handling of resumes with highly creative or non-standard formatting is a priority for future work.

Section 4.5 Discussion

SmartMatch effectively identifies gaps and generates meaningful feedback, making it a valuable tool for job seekers and recruiters. However, its reliance on manually crafted test data highlights the need for further evaluation using real-world examples. Incorporating multilingual support and advanced models like GPT-4 [7] could further enhance the tool's capabilities.

Section 5 Conclusion

SmartMatch is a step forward in leveraging NLP and machine learning to bridge the gap between resumes and job descriptions. The project demonstrates how advanced technologies can simplify the recruitment process, making it more efficient for both job seekers and recruiters.

The system effectively identifies gaps in skills, education, and experience, providing actionable feedback and a suitability score. This feedback empowers candidates to make targeted improvements to their profiles, enhancing their employability and increasing their chances of securing relevant roles. Recruiters benefit by having a reliable tool to filter and assess resumes, reducing the time and effort required in the hiring process.

While the initial prototype showcases significant potential, there is ample scope for improvement and expansion. Future iterations of SmartMatch could incorporate:

1. **Real-Time Testing**: Testing the tool with a larger dataset of real resumes and job descriptions to fine-tune its accuracy and feedback relevance.
2. **Integration with Platforms**: Embedding the system within job portals like LinkedIn or Indeed to provide immediate feedback to users.
3. **Support for Diverse Languages**: Enhancing the tool's applicability by introducing multilingual capabilities.
4. **Context-Aware Analysis**: Adding industry-specific analysis to make the system adaptable across various sectors.

The project highlights the power of NLP in addressing real-world problems and opens avenues for further exploration in automated career assistance tools. As SmartMatch evolves, it has the potential to become a vital tool in the modern job market, fostering better matches between talent and opportunity.

References

[1] "BERT," HuggingFace, [Online]. Available: https://huggingface.co/docs/transformers/en/model_doc/bert. [Accessed 12 12 2024].

[2] "Land your dream job," Resumatch.io, [Online]. Available: https://www.resumatch.io/. [Accessed 12 12 2024].

[3] "Career Acceleration Platform," Vmock, [Online]. Available: https://www.vmock.com/. [Accessed 12 12 2024].

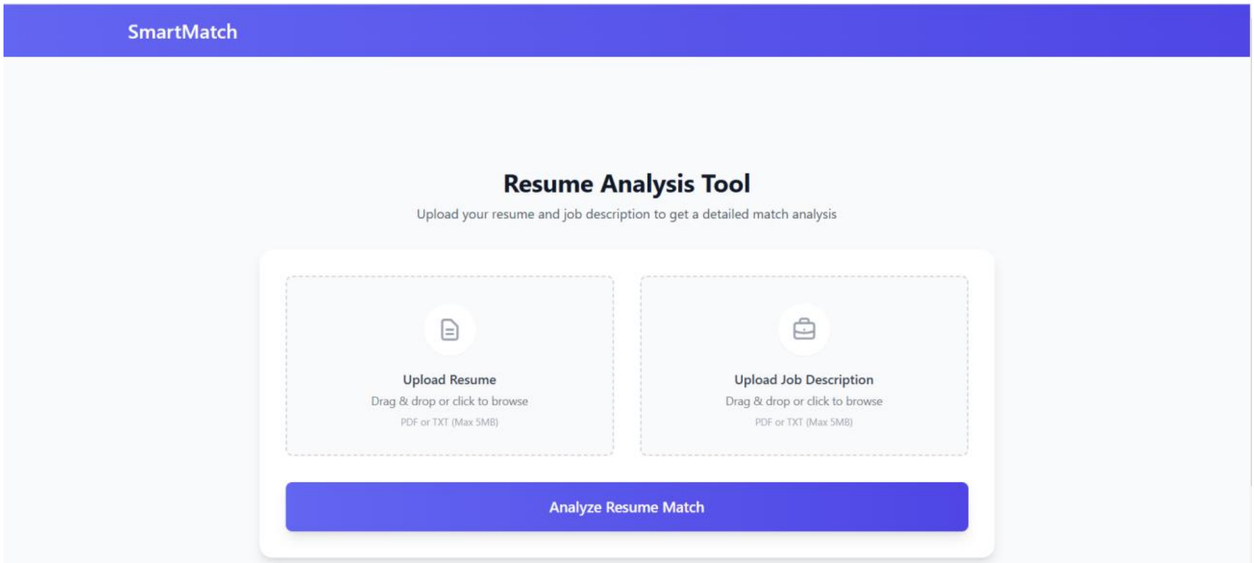
[4] "Universal Sentence Encoder," TensorFlow, [Online]. Available: https://www.tensorflow.org/hub/tutorials/semantic_similarity_with_tf_hub_universal_encoder. [Accessed 12\ 12 2024].

[5] "Industrial-Strength Natural Language Processing," spaCy, [Online]. Available: https://spacy.io/. [Accessed 12 12 2024].

[6] "all-MiniLM-L6-v2," HuggingFace, [Online]. Available: https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2. [Accessed 12 12 2024].

[7] "GPT-4 API general availability," OpenAI, [Online]. Available: https://openai.com/index/gpt-4-api-general-availability/. [Accessed 12 12 2024].

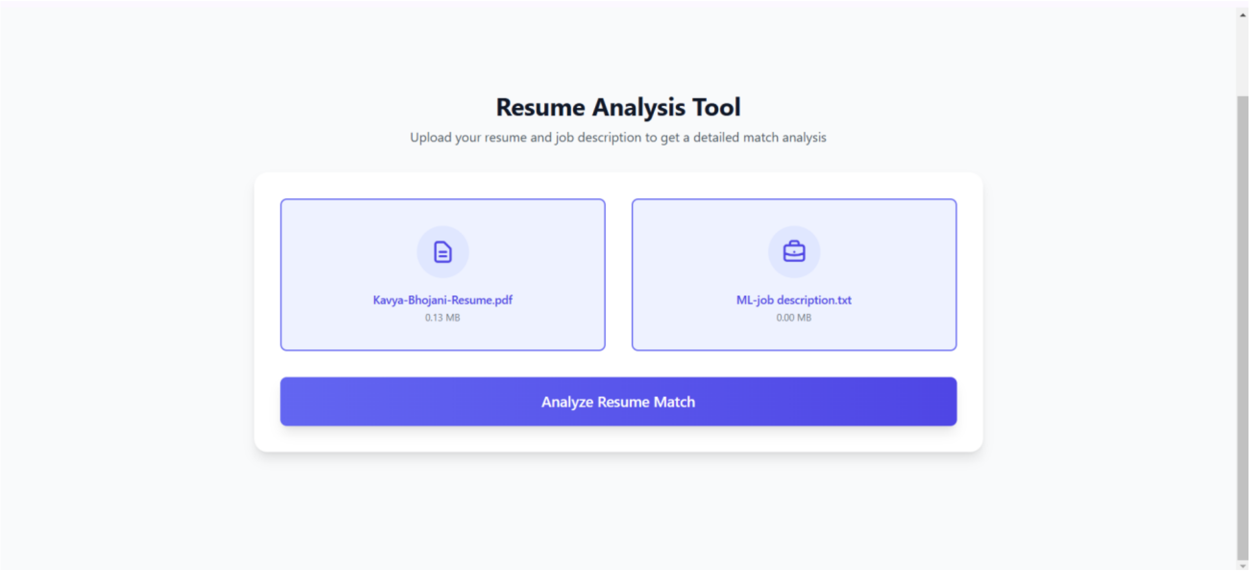
Appendices



Resume Analysis Tool

Upload your resume and job description to get a detailed match analysis

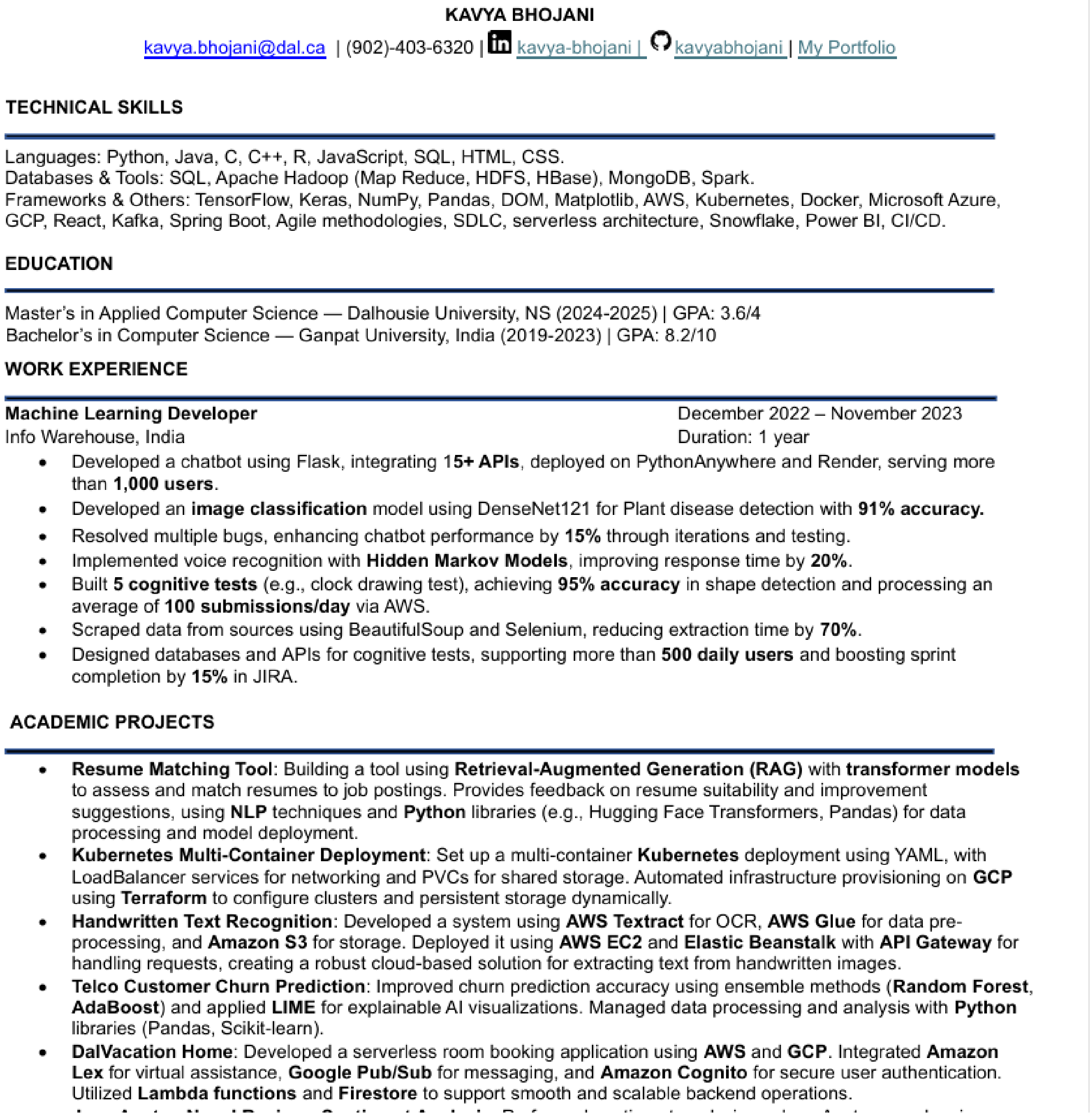
Figure 2: App - Landing Page



Resume Analysis Tool

Upload your resume and job description to get a detailed match analysis

Figure 3: App- Documents Uploaded Successfully



KAVYA BHOJANI

kavya.bhojani@dal.ca | (902)-403-6320 | kavya-bhojani | kavyabhojani | My Portfolio

TECHNICAL SKILLS

Languages: Python, Java, C, C++, R, JavaScript, SQL, HTML, CSS.
Databases & Tools: SQL, Apache Hadoop (Map Reduce, HDFS, HBase), MongoDB, Spark.
Frameworks & Others: TensorFlow, Keras, NumPy, Pandas, DOM, Matplotlib, AWS, Kubernetes, Docker, Microsoft Azure, GCP, React, Kafka, Spring Boot, Agile methodologies, SDLC, serverless architecture, Snowflake, Power BI, CI/CD.

EDUCATION

Master's in Applied Computer Science — Dalhousie University, NS (2024-2025) | GPA: 3.6/4
Bachelor's in Computer Science — Ganpat University, India (2019-2023) | GPA: 8.2/10

WORK EXPERIENCE

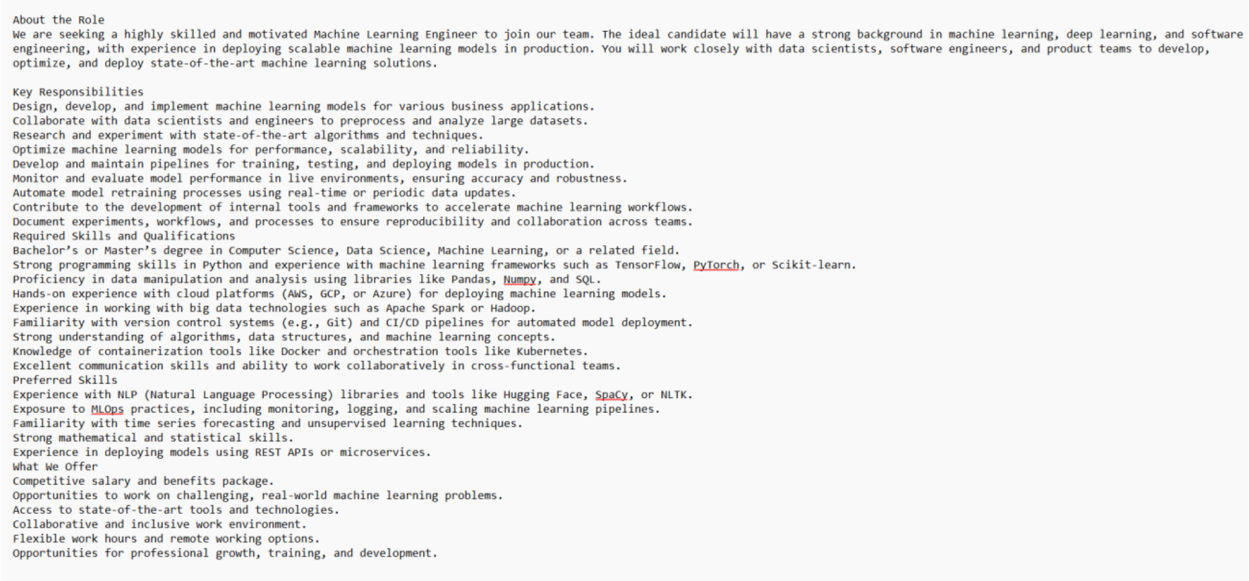
Machine Learning Developer December 2022 – November 2023
Info Warehouse, India Duration: 1 year

- Developed a chatbot using Flask, integrating 1**5+ APIs**, deployed on PythonAnywhere and Render, serving more than **1,000 users**.
- Developed an **image classification** model using DenseNet121 for Plant disease detection with **91% accuracy.**
- Resolved multiple bugs, enhancing chatbot performance by **15%** through iterations and testing.
- Implemented voice recognition with **Hidden Markov Models**, improving response time by **20%**.
- Built **5 cognitive tests** (e.g., clock drawing test), achieving **95% accuracy** in shape detection and processing an average of **100 submissions/day** via AWS.
- Scraped data from sources using BeautifulSoup and Selenium, reducing extraction time by **70%**.
- Designed databases and APIs for cognitive tests, supporting more than **500 daily users** and boosting sprint completion by **15%** in JIRA.

ACADEMIC PROJECTS

- **Resume Matching Tool**: Building a tool using **Retrieval-Augmented Generation (RAG)** with **transformer models** to assess and match resumes to job postings. Provides feedback on resume suitability and improvement suggestions, using **NLP** techniques and **Python** libraries (e.g., Hugging Face Transformers, Pandas) for data processing and model deployment.
- **Kubernetes Multi-Container Deployment**: Set up a multi-container **Kubernetes** deployment using YAML, with LoadBalancer services for networking and PVCs for shared storage. Automated infrastructure provisioning on **GCP** using **Terraform** to configure clusters and persistent storage dynamically.
- **Handwritten Text Recognition**: Developed a system using **AWS Textract** for OCR, **AWS Glue** for data pre-processing, and **Amazon S3** for storage. Deployed it using **AWS EC2** and **Elastic Beanstalk** with **API Gateway** for handling requests, creating a robust cloud-based solution for extracting text from handwritten images.
- **Telco Customer Churn Prediction**: Improved churn prediction accuracy using ensemble methods (**Random Forest**, **AdaBoost**) and applied **LIME** for explainable AI visualizations. Managed data processing and analysis with **Python** libraries (Pandas, Scikit-learn).
- **DalVacation Home**: Developed a serverless room booking application using **AWS** and **GCP**. Integrated **Amazon Lex** for virtual assistance, **Google Pub/Sub** for messaging, and **Amazon Cognito** for secure user authentication. Utilized **Lambda functions** and **Firestore** to support smooth and scalable backend operations.

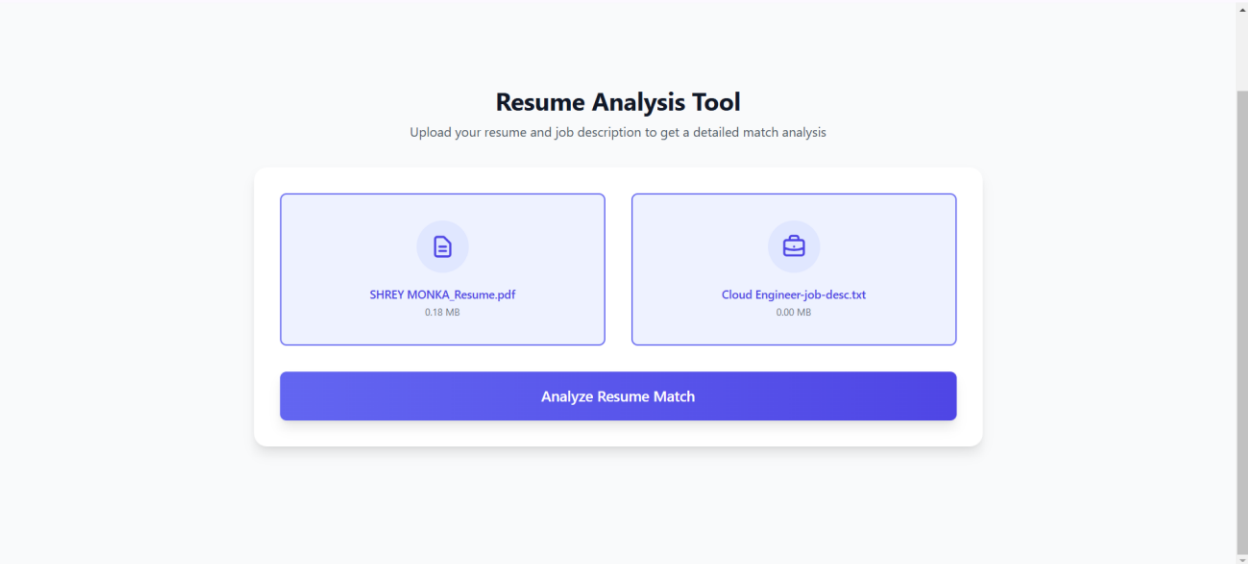
Figure 4: App – Resume - 1



About the Role
We are seeking a highly skilled and motivated Machine Learning Engineer to join our team. The ideal candidate will have a strong background in machine learning, deep learning, and software engineering, with experience in deploying scalable machine learning models in production. You will work closely with data scientists, software engineers, and product teams to develop, optimize, and deploy state-of-the-art machine learning solutions.

Key Responsibilities
Design, develop, and implement machine learning models for various business applications.
Collaborate with data scientists and engineers to preprocess and analyze large datasets.
Research and experiment with state-of-the-art algorithms and techniques.
Optimize machine learning models for performance, scalability, and reliability.
Develop and maintain pipelines for training, testing, and deploying models in production.
Monitor and evaluate model performance in live environments, ensuring accuracy and robustness.
Automate model retraining processes using real-time or periodic data updates.
Contribute to the development of internal tools and frameworks to accelerate machine learning workflows.
Document experiments, workflows, and processes to ensure reproducibility and collaboration across teams.
Required Skills and Qualifications
Bachelor’s or Master’s degree in Computer Science, Data Science, Machine Learning, or a related field.
Strong programming skills in Python and experience with machine learning frameworks such as TensorFlow, PyTorch, or Scikit-learn.
Proficiency in data manipulation and analysis using libraries like Pandas, Numpy, and SQL.
Hands-on experience with cloud platforms (AWS, GCP, or Azure) for deploying machine learning models.
Experience in working with big data technologies such as Apache Spark or Hadoop.
Familiarity with version control systems (e.g., Git) and CI/CD pipelines for automated model deployment.
Strong understanding of algorithms, data structures, and machine learning concepts.
Knowledge of containerization tools like Docker and orchestration tools like Kubernetes.
Excellent communication skills and ability to work collaboratively in cross-functional teams.
Preferred Skills
Experience with NLP (Natural Language Processing) libraries and tools like Hugging Face, SpaCy, or NLTK.
Exposure to MLOps practices, including monitoring, logging, and scaling machine learning pipelines.
Familiarity with time series forecasting and unsupervised learning techniques.
Strong mathematical and statistical skills.
Experience in deploying models using REST APIs or microservices.
What We Offer
Competitive salary and benefits package.
Opportunities to work on challenging, real-world machine learning problems.
Access to state-of-the-art tools and technologies.
Collaborative and inclusive work environment.
Flexible work hours and remote working options.
Opportunities for professional growth, training, and development.

Figure 5: App - Job Description - 1



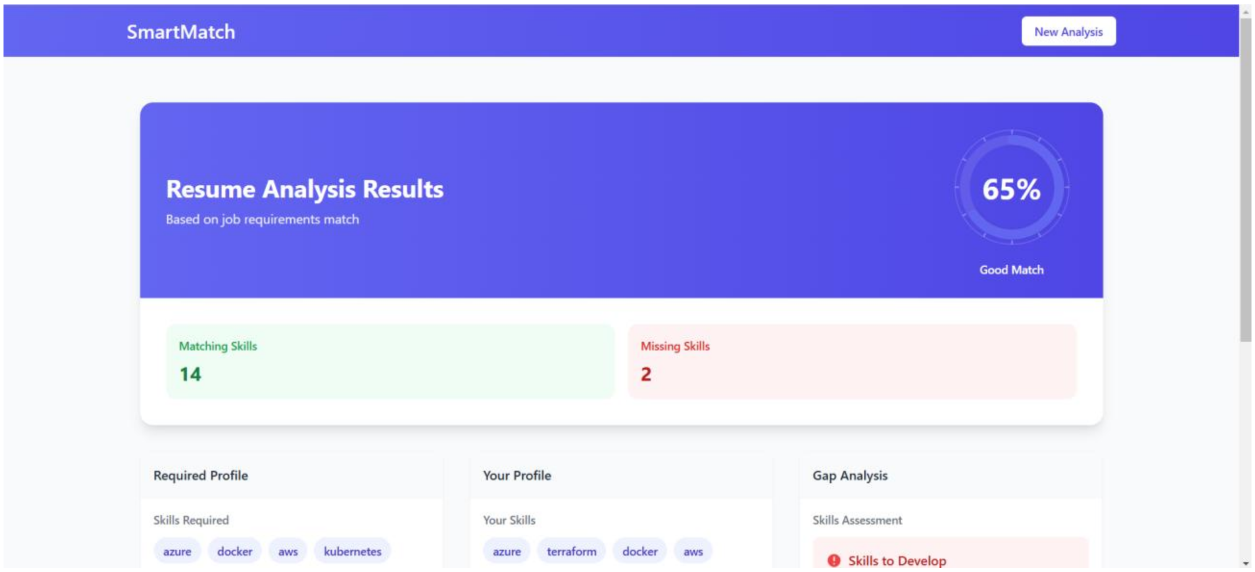
Resume Analysis Tool

Upload your resume and job description to get a detailed match analysis

SHREY MONKA_Resume.pdf
0.18 MB

Cloud Engineer-job-desc.txt
0.00 MB

Figure 6: App - Documents uploaded - 2



Resume Analysis Results

Based on job requirements match

65%

Good Match

Matching Skills
14

Missing Skills
2

Required Profile

Skills Required

azure docker aws kubernetes

Your Profile

Your Skills

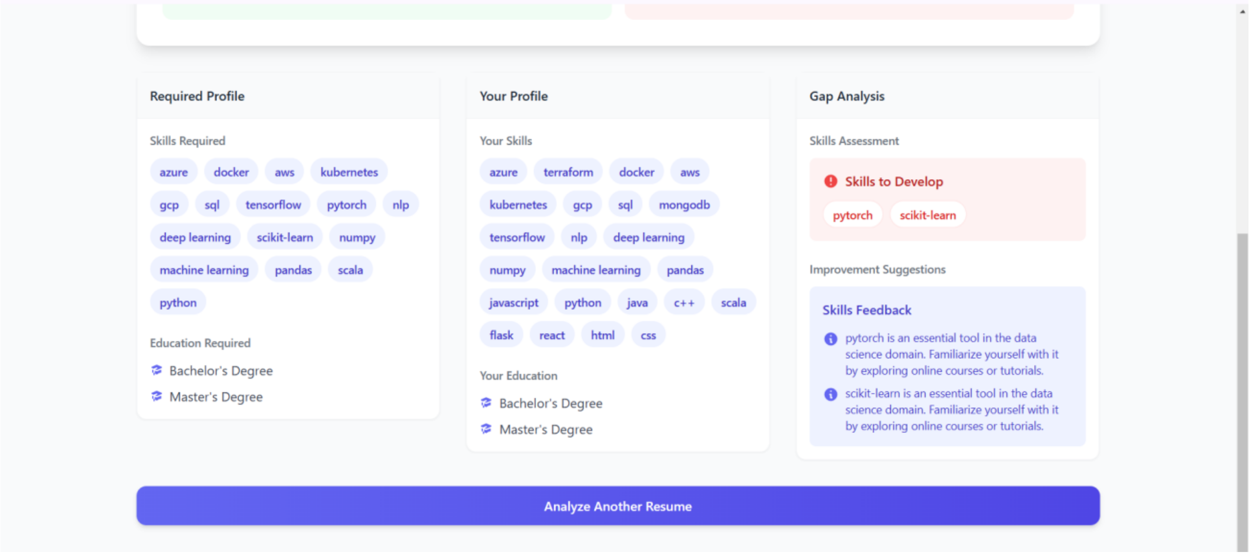
azure terraform docker aws

Gap Analysis

Skills Assessment

Skills to Develop

Figure 7: App - score 1



Required Profile

Skills Required

azure docker aws kubernetes gcp sql tensorflow pytorch nlp deep learning scikit-learn numpy machine learning pandas scala python

Education Required

- Bachelor's Degree
- Master's Degree

Your Profile

Your Skills

azure terraform docker aws kubernetes gcp sql mongodb tensorflow nlp deep learning numpy machine learning pandas javascript python java c++ scala flask react html css

Your Education

- Bachelor's Degree
- Master's Degree

Gap Analysis

Skills Assessment

Skills to Develop

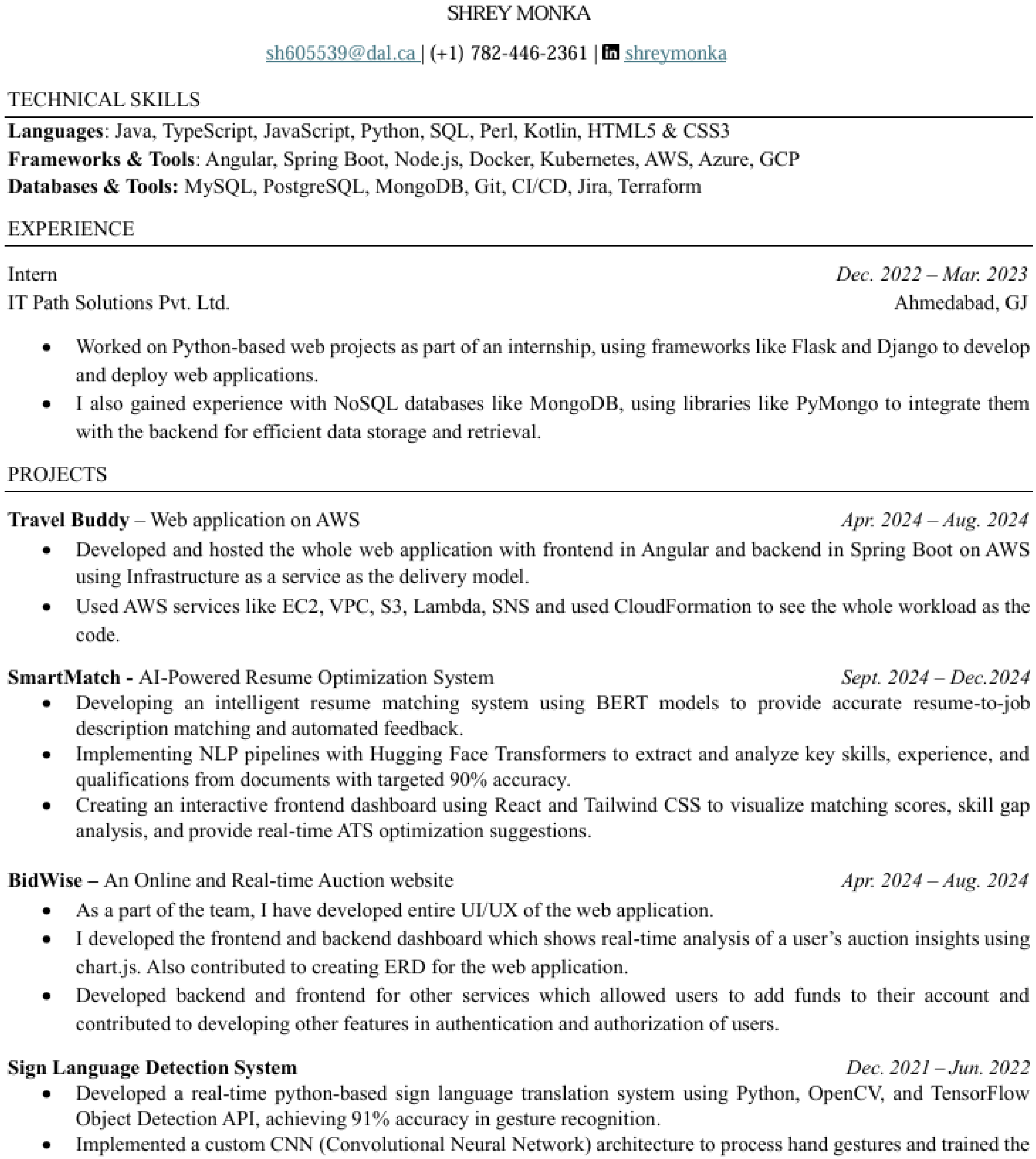
pytorch scikit-learn

Improvement Suggestions

Skills Feedback

- pytorch is an essential tool in the data science domain. Familiarize yourself with it by exploring online courses or tutorials.
- scikit-learn is an essential tool in the data science domain. Familiarize yourself with it by exploring online courses or tutorials.

Figure 8: App - analysis 1



SHREY MONKA

sh605539@dal.ca | (+1) 782-446-2361 | shreymonka

TECHNICAL SKILLS

Languages: Java, TypeScript, JavaScript, Python, SQL, Perl, Kotlin, HTML5 & CSS3
Frameworks & Tools: Angular, Spring Boot, Node.js, Docker, Kubernetes, AWS, Azure, GCP
Databases & Tools: MySQL, PostgreSQL, MongoDB, Git, CI/CD, Jira, Terraform

EXPERIENCE

Intern *Dec. 2022 – Mar. 2023*
IT Path Solutions Pvt. Ltd. Ahmedabad, GJ

- Worked on Python-based web projects as part of an internship, using frameworks like Flask and Django to develop and deploy web applications.
- I also gained experience with NoSQL databases like MongoDB, using libraries like PyMongo to integrate them with the backend for efficient data storage and retrieval.

PROJECTS

Travel Buddy – Web application on AWS *Apr. 2024 – Aug. 2024*

- Developed and hosted the whole web application with frontend in Angular and backend in Spring Boot on AWS using Infrastructure as a service as the delivery model.
- Used AWS services like EC2, VPC, S3, Lambda, SNS and used CloudFormation to see the whole workload as the code.

SmartMatch - AI-Powered Resume Optimization System *Sept. 2024 – Dec.2024*

- Developing an intelligent resume matching system using BERT models to provide accurate resume-to-job description matching and automated feedback.
- Implementing NLP pipelines with Hugging Face Transformers to extract and analyze key skills, experience, and qualifications from documents with targeted 90% accuracy.
- Creating an interactive frontend dashboard using React and Tailwind CSS to visualize matching scores, skill gap analysis, and provide real-time ATS optimization suggestions.

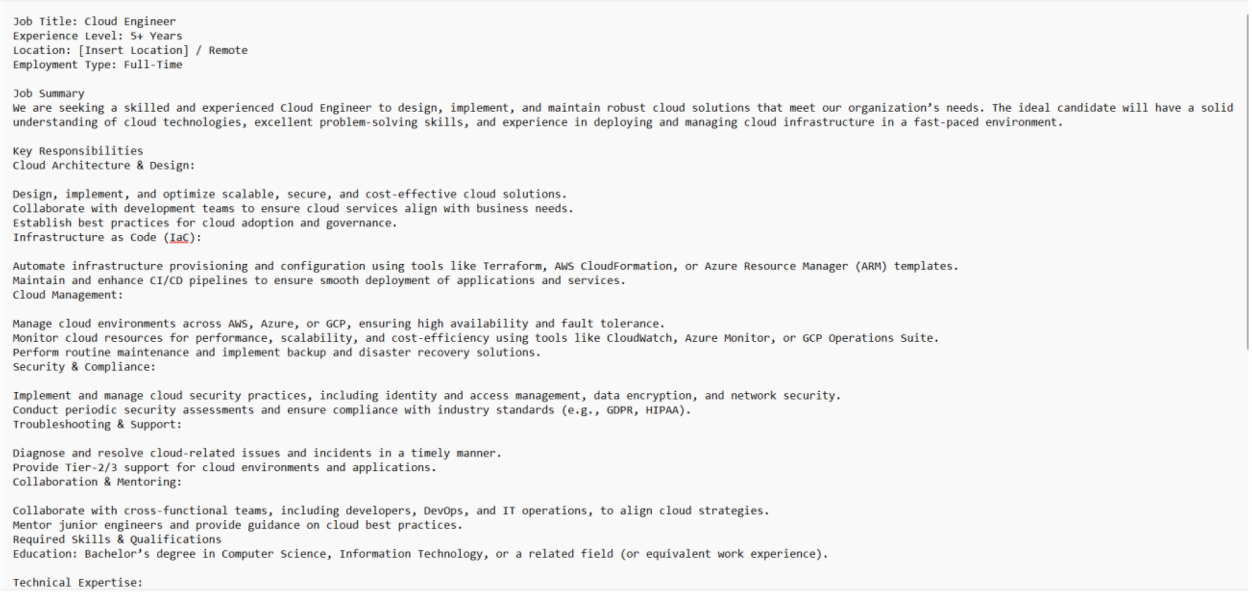
BidWise – An Online and Real-time Auction website *Apr. 2024 – Aug. 2024*

- As a part of the team, I have developed entire UI/UX of the web application.
- I developed the frontend and backend dashboard which shows real-time analysis of a user's auction insights using chart.js. Also contributed to creating ERD for the web application.
- Developed backend and frontend for other services which allowed users to add funds to their account and contributed to developing other features in authentication and authorization of users.

Sign Language Detection System *Dec. 2021 – Jun. 2022*

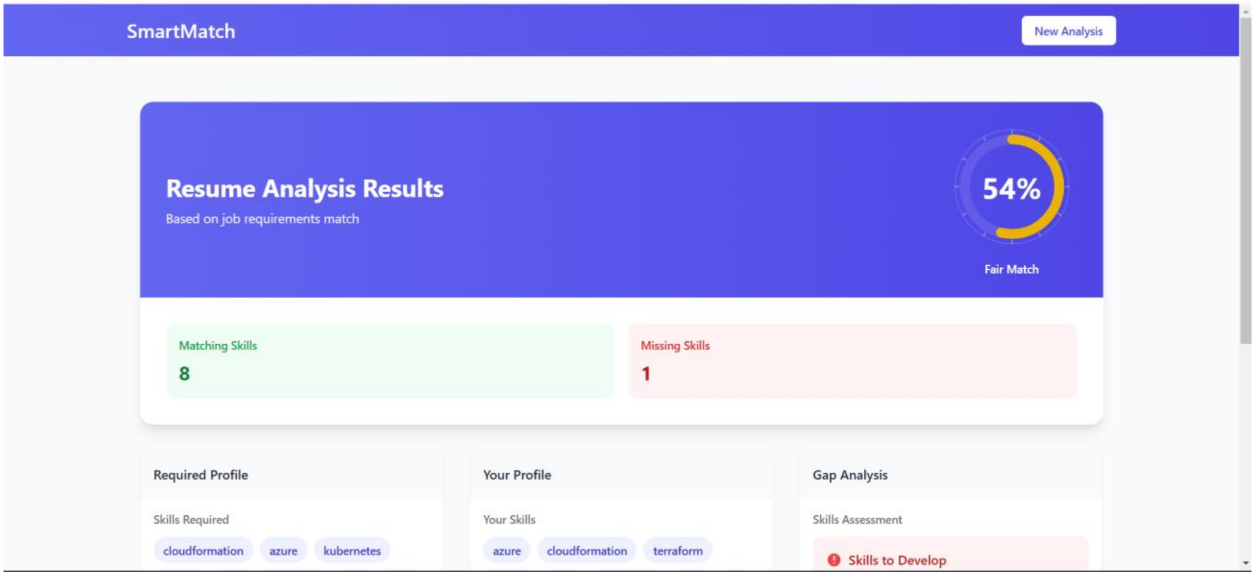
- Developed a real-time python-based sign language translation system using Python, OpenCV, and TensorFlow Object Detection API, achieving 91% accuracy in gesture recognition.
- Implemented a custom CNN (Convolutional Neural Network) architecture to process hand gestures and trained the

Figure 9: App - Resume - 2



```
Job Title: Cloud Engineer
Experience Level: 5+ Years
Location: [Insert Location] / Remote
Employment Type: Full-Time

Job Summary
We are seeking a skilled and experienced Cloud Engineer to design, implement, and maintain robust cloud solutions that meet our organization’s needs. The ideal candidate will have a solid
understanding of cloud technologies, excellent problem-solving skills, and experience in deploying and managing cloud infrastructure in a fast-paced environment.

Key Responsibilities
Cloud Architecture & Design:

Design, implement, and optimize scalable, secure, and cost-effective cloud solutions.
Collaborate with development teams to ensure cloud services align with business needs.
Establish best practices for cloud adoption and governance.
Infrastructure as Code (IaC):

Automate infrastructure provisioning and configuration using tools like Terraform, AWS CloudFormation, or Azure Resource Manager (ARM) templates.
Maintain and enhance CI/CD pipelines to ensure smooth deployment of applications and services.
Cloud Management:

Manage cloud environments across AWS, Azure, or GCP, ensuring high availability and fault tolerance.
Monitor cloud resources for performance, scalability, and cost-efficiency using tools like CloudWatch, Azure Monitor, or GCP Operations Suite.
Perform routine maintenance and implement backup and disaster recovery solutions.
Security & Compliance:

Implement and manage cloud security practices, including identity and access management, data encryption, and network security.
Conduct periodic security assessments and ensure compliance with industry standards (e.g., GDPR, HIPAA).
Troubleshooting & Support:

Diagnose and resolve cloud-related issues and incidents in a timely manner.
Provide Tier-2/3 support for cloud environments and applications.
Collaboration & Mentoring:

Collaborate with cross-functional teams, including developers, DevOps, and IT operations, to align cloud strategies.
Mentor junior engineers and provide guidance on cloud best practices.
Required Skills & Qualifications
Education: Bachelor’s degree in Computer Science, Information Technology, or a related field (or equivalent work experience).

Technical Expertise:
```

Figure 10: App - Documents 2



Resume Analysis Results

Based on job requirements match

54%

Fair Match

Matching Skills 8

Missing Skills 1

Required Profile

Skills Required

cloudformation azure kubernetes

Your Profile

Your Skills

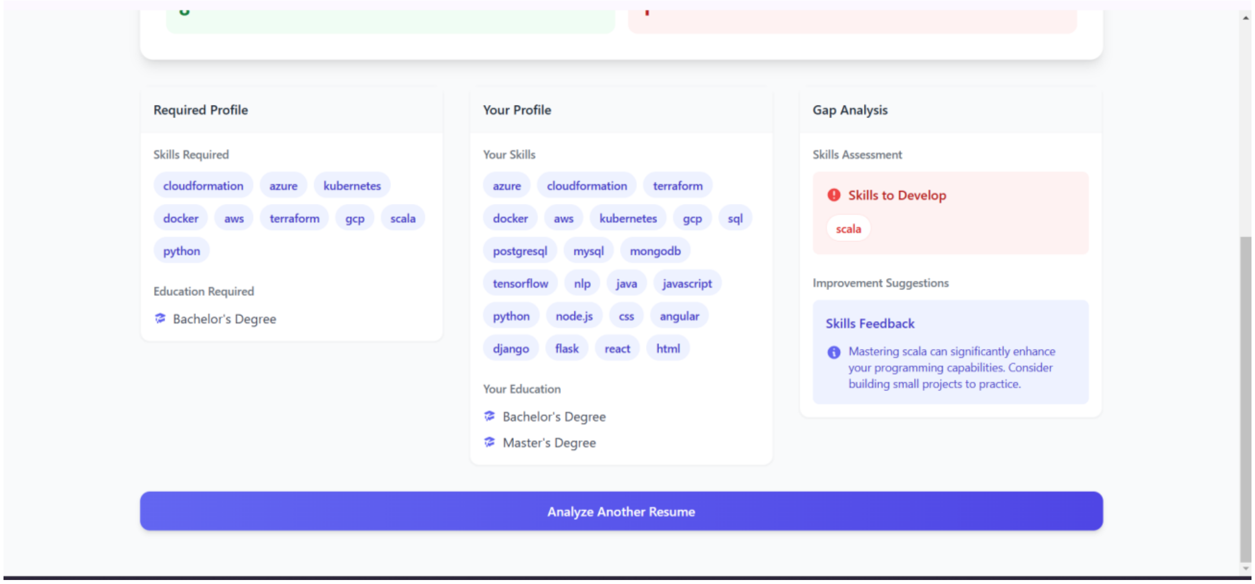
azure cloudformation terraform

Gap Analysis

Skills Assessment

Skills to Develop

Figure 11: App - score 2



Required Profile

Skills Required

cloudformation azure kubernetes docker aws terraform gcp scala python

Education Required

Bachelor's Degree

Your Profile

Your Skills

azure cloudformation terraform docker aws kubernetes gcp sql postgresql mysql mongodb tensorflow nlp java javascript python node.js css angular django flask react html

Your Education

Bachelor's Degree

Master's Degree

Gap Analysis

Skills Assessment

Skills to Develop

scala

Improvement Suggestions

Skills Feedback

Mastering scala can significantly enhance your programming capabilities. Consider building small projects to practice.

Figure 12: App - analysis 2