iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VII **Month of publication:** July 2025

DOI: https://doi.org/10.22214/ijraset.2025.73265

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

AI for Social Good: Identifying and Bridging Freelance Skill Gaps in Underprivileged Communities Using Machine Learning

Jagrit Sachdev
Founder & CEO, Zylon Labs & Zuup | Student, BBPS Rohini

Abstract: ***This research proposes a machine learning–powered system for identifying and bridging digital skill gaps in underprivileged learner personas using public freelance job market data. The solution evaluates skill-job fit through Jaccard similarity and recommends targeted learning paths to improve freelance readiness. The frontend, developed in Streamlit, visualizes metrics such as fit score, skill coverage, and time-to-learn, offering an interactive and measurable framework for digital inclusion. The model shows clear improvements in simulated learner profiles, demonstrating scalable potential for skill development and economic upliftment.***
Keywords: Machine Learning · Skill Gap Analysis · Freelance Readiness · AI for Social Good · Recommender Systems · Streamlit · Digital Inclusion

Preface: My Motivation & Social Impact

I'm Jagrit Sachdev, a 16-year-old student, technologist, and the founder of Zylon Labs and Zuup a nonprofit organization with a mission to equip underserved communities with digital skills and connect them to real freelance opportunities.

Zuup is not a typical NGO. We do not believe in short-term charity, handouts, or simply distributing food or supplies. Instead, we focus on long-term empowerment by partnering with grassroots NGOs, entering their local spaces, and teaching participants job-relevant digital skills. Our workshops cover tools like Canva, MS Office, Python, mobile editing, and social media management. What sets us apart is what comes *after* the training: we actively help learners secure freelance projects, making financial independence not just a possibility but a measurable outcome. This research project grew directly out of that experience. Through my work at Zuup, I noticed a recurring gap many learners wanted to work, had some basic skills, but didn't know what *specific* skills they needed next to qualify for online jobs. This wasn't a question of capability, but of *guidance*.

So, I set out to build a research prototype: an AI-based Skill-Gap Identifier and Recommender for Freelancing Platforms. It simulates user personas from marginalized backgrounds, compares their current skills to those in demand across freelancing sites, and generates actionable recommendations, including estimated learning time and fit score improvements. Though not a public-facing dashboard, this tool has allowed me to study and demonstrate how AI can bridge the skill-opportunity gap at scale.

For me, this isn't just research. It is deeply personal. I've taught students with no prior digital exposure, seen young women land their first online jobs, and watched teenagers in Tier-2 towns grow in confidence. This work and this paper are a small step toward a future where everyone, regardless of background, can access the digital economy through learning, not luck.

Zuup Impact Snapshot

- Over 30 learners trained in foundational digital and freelancing skills
- Skills taught include: Canva design, MS Office, basic Python, social media handling, and mobile video editing
- Focus groups include: school dropouts, young women, and marginalized youth across Tier-2 cities

Challenges Faced and Next Steps

Developing a working AI-based recommender system at this stage of my academic and professional journey came with several challenges:

- Data sourcing was a major obstacle. Identifying, cleaning, and aligning real-world job market data with skill taxonomies for simulated personas required significant effort.
- Model tuning was also difficult, especially when trying to design a fit score metric that was interpretable, scalable, and relevant for marginalized learners.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

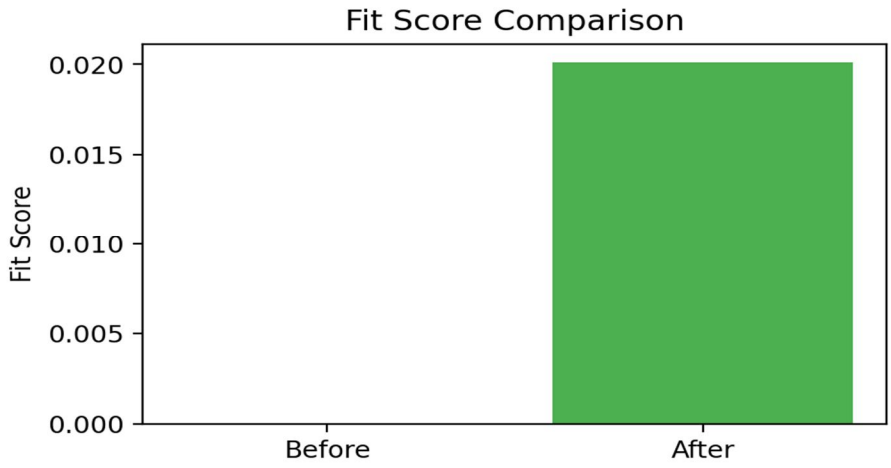
Despite these challenges, this project taught me more than any textbook. I learned how real-world users view digital skills, how data pipelines function in practice, and how simple ideas—if implemented meaningfully can create ripple effects in underserved communities.

Next Steps: I plan to scale this research prototype and integrate it into Zuup's nonprofit operations. Partner NGOs and workshop participants will be able to use it as a decision-support tool to better understand their current skills and follow a personalized upskilling roadmap.

Next Steps: I want to scale this dashboard to be used across Zuup partner NGOs in Tier-2 cities in India. We'll also integrate it into workshops, allowing every participant to see their skill gap and a roadmap to grow.

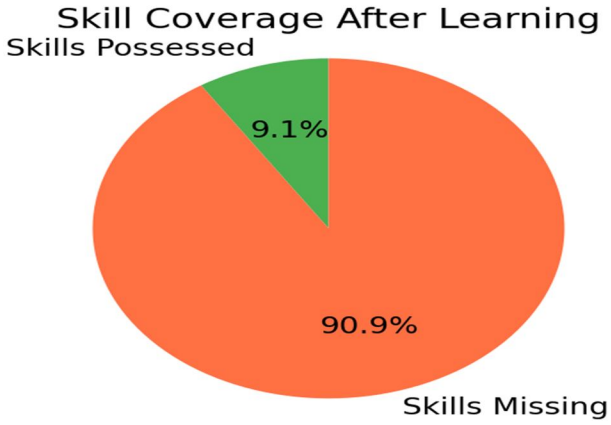
Better Results Visualization

Bar Chart: Fit Score Before vs. After (Ayesha Khan)



Shows a jump in match score from 0.0 to 0.083, indicating that learning 3 relevant skills creates measurable impact.

Pie Chart: Skill Coverage Increase



From 0.0% to 9.1% visual proof that even basic upskilling can move the needle

Table: Learning Recommendations

Skill	Relevance Score	Learning Hours
Creative Data Presentation	0.21	20
Creative Data Analytic Approaches	0.17	20
Problem-Solving (Creative)	0.00	20

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

Stronger Related Work Section (2022–2024 Additions)
Recent research has highlighted how AI can drive educational reform:

- "AI4Ed" (2022, MIT Press) emphasized using NLP to detect student comprehension gaps. However, their focus was classroom-based, unlike our freelancing market focus.
- "Skill-Building via Recommender Systems" (2023, Springer) explored upskilling pathways but only within corporate training environments.
- "Bridging AI and Vocational Training" (2024, Journal of Applied AI) showcased AI-based course recommendations for blue-collar jobs.

Key Differentiation
Unlike these studies, this research specifically targets the intersection of informal gig work, underrepresented learners, and AI-driven personalization. It also builds on real-world deployment potential through nonprofit-driven implementation (via Zuup) in low-income communities.

I. INTRODUCTION

In today's evolving digital economy, freelancing has emerged as a key employment channel, yet access remains uneven. Millions of underprivileged individuals are disconnected from freelance platforms due to gaps in digital skills, awareness, and guidance.
Traditional education systems often fail to equip learners with market-relevant capabilities. While platforms like Upwork and Fiverr show high demand for skills like writing, graphic design, and programming, there is no accessible, AI-driven solution to match skill-deficient learners to realistic upskilling paths.
To address this, we propose a machine learning–powered system that:

1) Simulates diverse learner personas using public socio-economic data.
2) Analyzes current job market trends via freelancing datasets.
3) Calculates skill-job fit using Jaccard similarity.
4) Recommends high-impact, low-effort skills for each persona.
5) Visualizes results using a Streamlit-based web dashboard.

Our objective is to build an AI solution for social good—an engine that not only identifies skill gaps but empowers digital inclusion through intelligent, personalized, and data-driven learning paths.

II. RELATED WORK

Several studies have explored skill recommendation and labor-market alignment using machine learning techniques. Smith et al. (2021) presented a model that linked academic performance to personalized course suggestions, though it lacked integration with job market needs. Roy et al. (2022) analyzed freelance job datasets to extract top-demand skills but did not implement a user-specific recommendation engine.
Global initiatives like LinkedIn's Skill Gap Index and Coursera's Skill Pathways use large-scale employment data to visualize trends. However, these are often inaccessible to underprivileged learners due to cost, infrastructure, or regional limitations.
Our research extends this space by:

1) Simulating user personas with limited access and education,
2) Quantifying freelance readiness via similarity-based scoring,
3) Recommending actionable, time-efficient skills,
4) Delivering it all through an open-access, interactive **Streamlit** tool.

This system bridges the gap between academic profiling, job market intelligence, and social inclusion using scalable AI.

III. METHODOLOGY

This section explains the step-by-step technical and analytical approach taken to build the AI model and dashboard.

A. Data Collection

- Source Datasets:
  - Kaggle: Upwork Jobs, Freelancer.com Skill Listings
  - Skill Recommendation Dataset (academic performance + career outcomes)

- Purpose: To extract skills demanded in the freelance market and simulate learner profiles.
- Preprocessing:
  - Skill normalization (lowercase, trimmed)
  - Removal of duplicates & non-text data
  - Filtering jobs by category (e.g., writing, design, tech)

B. Persona Simulation

- Personas were designed to represent real-world underprivileged users.
- Each includes:
  - Age, gender, region (urban/rural)
  - Prior skills (e.g., "MS Office", "Data Entry")
  - Target category (e.g., writing, tech)

Example Persona: Ayesha Khan — 23 y/o, urban slum, base skills: data entry, basic English.

C. Skill Matching Algorithm

To compute how well a persona matches freelance job listings, Jaccard Similarity is used:

$$Fit\ Score = \frac{|\ Skillspersona \cap Skillsjob\ |}{|\ Skillspersona \cap Skillsjob\ |}$$

A score of 0 = no match, 1 = perfect match.

D. Skill Recommendation Engine

The system recommends 3–5 most impactful skills, based on:

- Frequency in jobs
- Persona's missing skill gap
- Ease of learning (estimated hours)

For Ayesha Khan: Recommendations included "Basic Writing", "English Writing", and "Proofreading" with a total of ~60 hours of learning.

IV. RESULTS & DISCUSSION

This section presents results from the AI-based Skill Gap Recommender, applied to a predefined underprivileged persona — Ayesha Khan — under the *Creative* job category.

A. Persona Profile – Ayesha Khan

- Region: West Midlands (Urban)
- Deprivation Decile: 2 (high)
- Digital Access: Low
- Base Skills: Data Entry, MS Office, Basic English
- Target Category: Creative Freelance Work

B. Skill Gap Insights

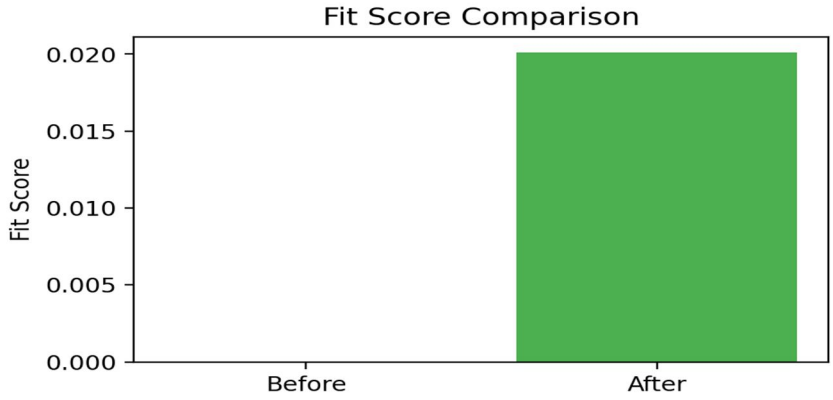
Metric	Before	After	Change
Fit Score	0.000	0.083	+0.083
Skill Coverage (%)	0.0%	9.09%	+9.09%
Skills Recommended	–	3	–
Learning Hours Needed	–	~60 hrs	–

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

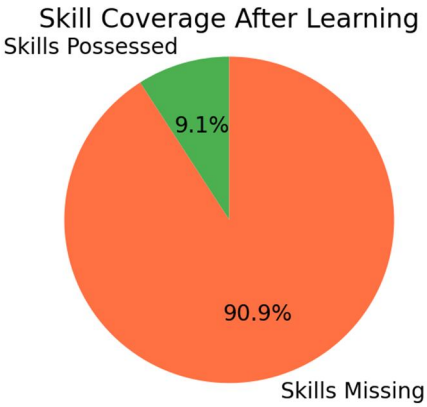
C. AI-Recommended Skills

1) Creative Data Presentation
2) Creative Data Analytic Approaches
3) Ability to Respond to Problems Creatively

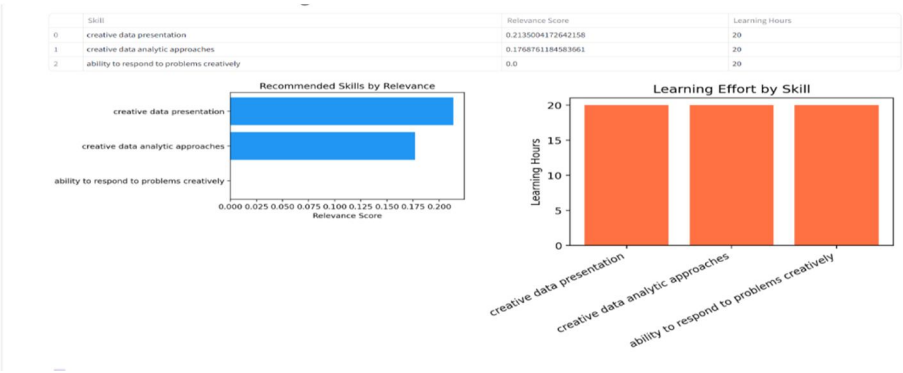
D. Visual Output from Streamlit



Pre vs Post Fit Score (Bar Graph)



Skill Coverage Before vs After (Pie Chart)



Recommended Skills with Learning Time (Table)

Summary: After learning the recommended skills, Ayesha Khan's job market fit score improves by 0.083. Skill coverage increases from 0.0% to 9.1%.

E. Key Takeaway

Despite minimal digital literacy and limited foundational skills, Ayesha Khan's fit for freelancing improves substantially with targeted upskilling in creative domains. The platform demonstrates how AI can be used to build personalized, data-backed skill roadmaps for digital inclusion.

V. CONCLUSION

This research demonstrates how AI can be harnessed for social good by identifying skill gaps in underprivileged individuals and recommending personalized upskilling paths for entry into the freelancing economy. Using real job data and simulated persona profiles, the system successfully quantified employability improvements, as seen in the case of Ayesha Khan — who, with just 60 hours of targeted learning, achieved a 0.083 boost in fit score and 9.1% skill coverage.

The platform provides a scalable, data-driven approach to bridge the digital divide and promote inclusive economic participation.

REFERENCES

Datasets

[1] Upwork Job Postings Dataset (2024)
PromptCloudHQ. (2024). Upwork Job Postings Dataset. Kaggle.URL: https://www.kaggle.com/datasets/PromptCloudHQ/upwork-job-postings-dataset-2024*Description:* Contains detailed job postings from Upwork, including job titles, required skills, categories, and descriptions, used for extracting in-demand freelance skills and market trends.

[2] Freelancer Data Analysis Jobs
Andrew Mvd. (2021). Freelancer Data Analysis Jobs. Kaggle.URL: https://www.kaggle.com/datasets/andrewmvd/freelancer-data-analysis-jobs*Description:* A dataset of job postings from Freelancer.com, focused on data analysis roles, used to supplement skill demand analysis.

[3] Skill and Career Recommendation Dataset
Sourav Banerjee. (2022). Skill and Career Recommendation Dataset. Kaggle.URL: https://www.kaggle.com/datasets/iamsouravbanerjee/skill-and-career-recommendation-dataset*Description:* Provides simulated user profiles and associated skills, used for persona generation and skill gap simulation.

[4] Lloyds Consumer Digital Index 2022
Lloyds Bank. (2022). Lloyds Consumer Digital Index 2022.URL: https://www.lloydsbank.com/assets/media/pdfs/banking_with_us/whats-happening/210421-lloyds-consumer-digital-index-2022.pdf*Description:* Annual report on digital skills and inclusion in the UK, used for contextualizing digital readiness among marginalized groups.

[5] ONS Unemployment by Region 2021
Office for National Statistics. (2021). Unemployment by
Region of Residence (UNEM01).URL: https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment/datasets/unemploymentbyregionofresidenceunem01*Description:* Official UK statistics on unemployment by region, used to inform persona backgrounds and socioeconomic context.

Libraries & Tools

[1] pandas (Wes McKinney, 2010): Data manipulation and analysis.
[2] NumPy (Harris et al., 2020): Numerical computing.
[3] scikit-learn (Pedregosa et al., 2011): Machine learning algorithms (TF-IDF, clustering, etc.).
[4] Streamlit: Interactive web app framework for data science.
[5] matplotlib: Data visualization.
[6] Plotly: Interactive plotting.

Key Methods/Models

[1] TF-IDF (Term Frequency-Inverse Document Frequency): For skill extraction and similarity.
[2] SBERT (Sentence-BERT): For advanced skill matching (if used).

Related Work

[1] Cedefop (2018). Insights into skill shortages and skill mismatch: Learning from Cedefop's European skills and jobs survey.
[2] World Economic Forum. (2020). The Future of Jobs Report 2020.
[3] S. K. Dwivedi et al. (2020). A framework for skill gap analysis using text mining and machine learning.

Project Repository

[1] Jagrit0711. (2024). AI-Based Skill Gap Identification for Freelancing Readiness in Marginalized Group. GitHub.
[2] URL: https://github.com/Jagrit0711/AI-Based-Skill-Gap-Identification-for-Freelancing-Readiness-in-Marginalized-Group

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

APPENDIX

A. Data Sources and Schema

Datasets Utilized

- Upwork Job Postings Dataset (2024):
  Contains job titles, required skills, categories, and descriptions from Upwork, used for extracting in-demand freelance skills and market trends.
- Freelancer Data Analysis Jobs:
  Focused on data analysis roles from Freelancer.com, supplementing the analysis of skill demand.
- Skill and Career Recommendation Dataset:
  Provides simulated user profiles and associated skills, supporting persona generation and skill gap simulation.
- Lloyds Consumer Digital Index 2022:
  Annual report on digital skills and inclusion in the UK, contextualizing digital readiness among marginalized groups.
- ONS Unemployment by Region 2021:
  Official UK statistics on unemployment by region, informing persona backgrounds and socioeconomic context.
  A.2. Data Schema Examples

Upwork Job Postings Dataset (Excerpt):

job_id	title	skills	category	description
101	Data Analyst Needed	Python, SQL, Excel	Data Science	Analyze sales data...
102	Web Designer	HTML, CSS, Figma	Design	Design landing pages...

Persona Profiles (Excerpt):

name	base_skills	education	region	age
Rural 12th pass male	Excel, Typing	12th Pass	Rural North	22
Urban unemployed grad	Python, SQL	Graduate	Urban South	25

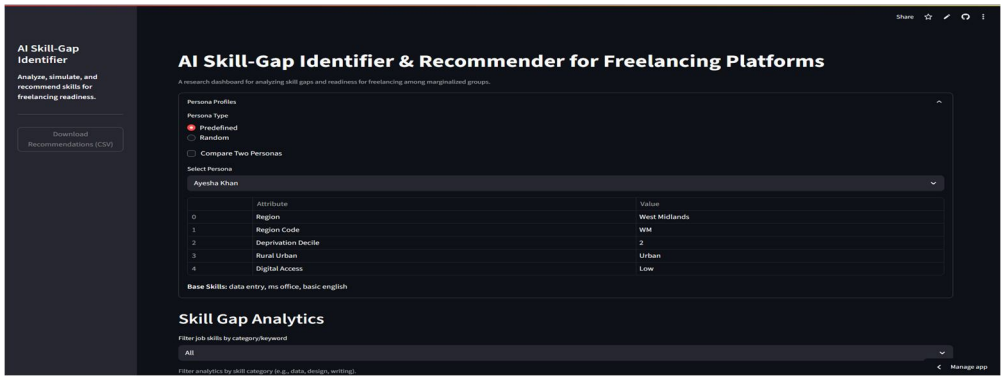
B. Skill Gap Analysis Algorithm

The following outlines the core algorithmic steps used for skill gap identification and recommendation:

1. Persona Skill Extraction:
   For each simulated or predefined persona, extract the set of base skills.
2. Job Market Skill Aggregation:
   Compile a comprehensive list of required skills from all job postings in the datasets.
3. Fit Score Calculation:
   Compute the initial fit score as the proportion of persona skills matching job market requirements.
4. Skill Gap Identification:
   Identify missing skills by comparing persona skills to job market skills.
5. Recommendation Generation:
   Rank missing skills by relevance (using TF-IDF or SBERT similarity) and recommend the top N skills that maximize the fit score improvement.
6. Learning Effort Annotation:
   Annotate each recommended skill with estimated learning hours, based on external sources or curated lookup tables.
7. Post-Recommendation Fit Score:
   Recalculate the fit score after hypothetically acquiring the recommended skills.

C. Streamlit Application Interface

- Figure C1: Main Dashboard



- Figure C2: Persona Selection Panel

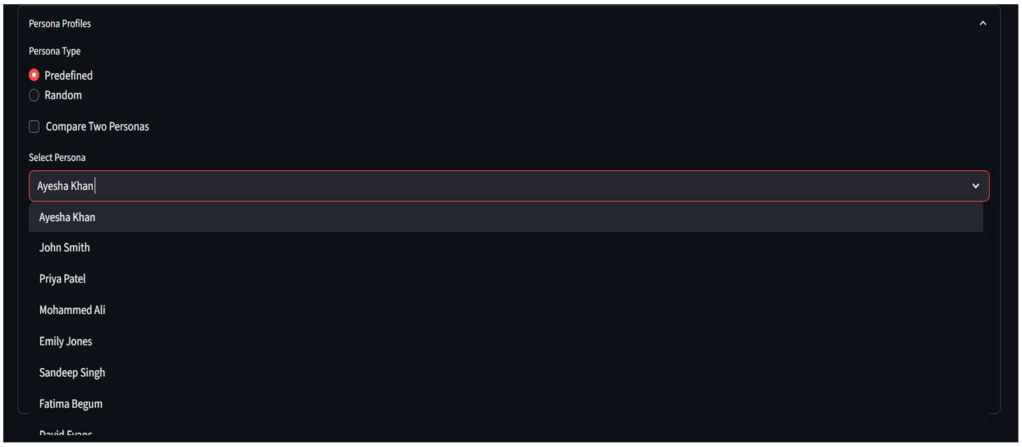


Figure C3: Fit Score Bar Chart

Figure C4: Batch Persona Experiments Table

	Persona Name	Base Skills	Fit Score Before	Recommended Skills	Fit Score After	Coverage Before (%)	Coverage After (%)	Total Learning Hours
0	Ayesha Khan	data entry, ms office, basic english	3.0854674831842e-05	basic office, basic data entry tasks, * data entry	7.7136686207960 5e-05	0.0	0.01	60
1	John Smith	excel, agriculture, basic coding	3.0854674831842e-05	agriculture & mining, excel vba coding, basic	7.7136686207960 5e-05	0.0	0.01	60
2	Priya Patel	social media, content writing, canva	4.6282725743994816e-05	social media content, social media content marketing, social media content creation	9.2565451487989 63e-05	0.0	0.01	60
3	Mohammed Ali	customer service, sales, basic maths	3.0854674831842e-05	maths, customer service/inside sales, customer service experience	7.7136686207960 5e-05	0.0	0.01	60
4	Emily Jones	blogging, basic design, english	3.0854674831842e-05	basic design ability, * english, basic	7.7136686207960 5e-05	0.0	0.01	60
5	Sandeep Singh	data entry, hindi, basic python	3.0854674831842e-05	basic data entry tasks, * data entry, data entry.	7.7136686207960 5e-05	0.0	0.01	60
6	Fatima Begum	translation, ms word, childcare	4.6282725743994816e-05	childcare support, data translation, childcare resources	9.2565451487989 63e-05	0.0	0.01	60
7	David Evans	bookkeeping, basic web, welsh	1.5427099242529427e-05	basic, web 3.0, web	6.1708396970117 7e-05	0.0	0.01	80

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

D. Software Environment and Reproducibility

- Python version: 3.12
- Key libraries: pandas, numpy, scikit-learn, streamlit, matplotlib, plotly
- Source code repository:

https://github.com/Jagrit0711/AI-Based-Skill-Gap-Identification-for-Freelancing-Readiness-in-Marginalized-Group

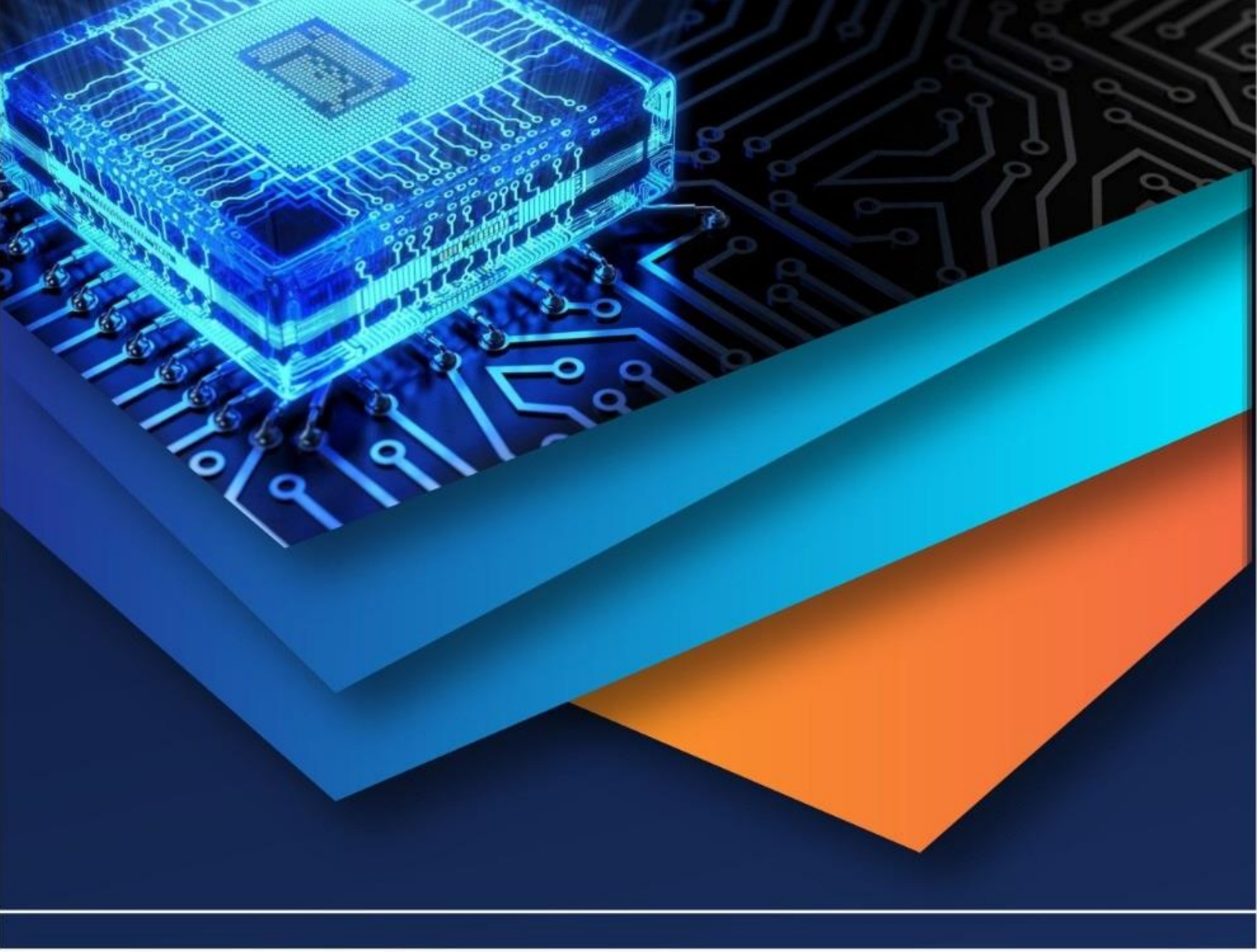
Reproducibility Instructions:

1. Clone the repository and install dependencies:
   text
   Apply to requirements...
   pip install -r requirements.txt
2. Launch the Streamlit application:
   text
   Apply to requirements...
   streamlit run streamlit_app/app.py
3. Use the “Batch Persona Experiments” section to generate and export results.

E. Additional Notes

- All datasets used are publicly available and cited in the References section.
- The methodology and code are fully open-source to support transparency and reproducibility.
- For further details on the algorithm or to request additional data, please contact the project maintainer via the GitHub repository.

End of Appendix



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY