# AKASH KUMAR SINGH

+91-7970976675 | akash27590@gmail.com | Jamshedpur, India | linkedin.com/in/akash-kumar-277750337 | github.com/razesoni

## PROFESSIONAL SUMMARY

Motivated Computer Science & Engineering undergraduate specializing in data science, predictive modeling, and end-to-end data processing pipelines. Proficient in Python, SQL, exploratory data analysis (EDA), and machine learning classification algorithms using Scikit-Learn, Pandas, and NumPy. Hands-on experience analyzing large-scale datasets (32,500+ records), feature engineering, statistical modeling, and deploying AI solutions such as Retrieval-Augmented Generation (RAG). Strong foundation in core CS principles (DSA, DBMS, OS) with demonstrated analytical thinking and data-driven problem-solving abilities. Eager to leverage technical expertise to deliver actionable insights as a Graduate Engineer Trainee (GET) or Associate Data Scientist.

## EDUCATION

**Bhilai Institute Of Technology** — **Raipur, Chhattisgarh**
*Bachelor of Technology (B.Tech) in Computer Science and Engineering* — *2023 – 2027 (Expected) — CGPA: 6.8*

**Vig English School** — **Jamshedpur, Jharkhand**
*Senior Secondary Examination (Class XII)* — *2023 — Percentage: 88.0%*

**Vig English School** — **Jamshedpur, Jharkhand**
*Secondary School Examination (Class X)* — *2021 — Percentage: 70.0%*

## TECHNICAL SKILLS

**Programming Languages:** Python
**Core Computer Science Concepts:** Object-Oriented Programming (OOP), Operating Systems (OS), Database Management Systems (DBMS), Computer Networks (CN)
**Data Structures & Algorithms Knowledge:** Data Structures & Algorithms (DSA), Space & Time Complexity Analysis
**Databases:** SQL, Relational Database Management, Data Modeling
**Machine Learning & Statistical Modeling:** Predictive Modeling, Classification Algorithms, Regression Algorithms, Data Science Libraries, Hyperparameter Tuning, Feature Engineering
**Generative AI & Natural Language Processing (NLP):** Retrieval-Augmented Generation (RAG), LLM Integration & Local Inference, Audio & Speech Processing
**Web / Application Development:** FastAPI, Flask, Role-Based Access Control (RBAC), RESTful APIs, HTML, CSS
**Tools & Platforms:** Advanced Excel (Pivot Tables, Power Query, Power Pivot), Postman, Jupyter Notebook, Google Colab, FFmpeg, VS Code
**Cloud / DevOps Exposure:** Local LLM Inference Engines (Ollama), Streaming Pipeline Frameworks (Pathway)
**Testing / Version Control:** Git, GitHub, API Testing
**Soft Skills:** Analytical Problem Solving, Continuous Learning, Technical Documentation, Logical Thinking, Visual Communication

## INTERNSHIPS / INDUSTRIAL TRAINING / TECHNICAL EXPERIENCE

**RAYS IT & Design World Pvt. Ltd.** — **Raipur, Chhattisgarh**
*Vocational Trainee – AI Powered Data Analytics with Python & Advanced Excel* — *June 2026 – July 2026*

- Acquired hands-on experience in data cleaning, preprocessing, and Exploratory Data Analysis (EDA) using Python libraries including NumPy and Pandas.
- Utilized Advanced Excel, Power Query, Power Pivot, and Pivot Tables to model unstructured datasets and automate business analytics reporting workflows.
- Developed and executed a capstone project titled **"Freemium Gaming Monetization & Player Retention"**, building interactive analytics dashboards to track key player behavior metrics.

**RAYS IT & Design World Pvt. Ltd.** — **Raipur, Chhattisgarh**
*Vocational Trainee – Web Development with Python & Django* — *June 2025 – July 2025*

- Gained practical experience in full-stack web application development leveraging Python, Django framework, and relational database management systems.
- Implemented Object-Relational Mapping (ORM) and RESTful API patterns to handle database operations and secure data routing.
- Constructed scalable backend models, views, and template structures following standard SDLC practices.

## PROJECTS

**Real-Time Air Quality Monitoring System** **July 2026 – Present**
*Academic / Major Project — Python, Pathway, OpenAQ API, RAG* GitHub

- Engineered a high-throughput, pure-Python streaming data pipeline to continuously ingest and process environmental metrics from external REST endpoints using the OpenAQ API.
- Architected a real-time data processing layer using Pathway, optimizing low-latency streaming performance and continuous computation without external web UI dependencies.
- Integrated a Retrieval-Augmented Generation (RAG) layer, enabling natural language querying and real-time contextual analysis over continuous streaming data.
- Applied software design practices to build an extensible, scalable backend pipeline using Python libraries.

**Freemium Gaming Monetization & Player Retention** **June 2026 – July 2026**
*Data Analytics Project — Python, Pandas, NumPy, Advanced Excel, Power Query* GitHub

- Conducted comprehensive Exploratory Data Analysis (EDA) on player engagement datasets using Python (NumPy, Pandas) to identify churn triggers and in-game purchase patterns.
- Leveraged Advanced Excel, Power Query, Power Pivot, and interactive Pivot Tables to aggregate multi-source game telemetry data and automate reporting pipelines.
- Engineered structured business dashboards highlighting retention curves, cohort metrics, and player lifetime value (LTV) drivers to inform monetization strategies.

**Census Income Predictive Modeling** **Feb 2026 – Mar 2026**
*Data Science / ML Project — Python, Pandas, NumPy, Scikit-Learn, Seaborn* GitHub

- Conducted exploratory data analysis (EDA) and data preprocessing on 32,500+ records from the Adult Census Income dataset using Pandas, NumPy, Seaborn, and Matplotlib.
- Implemented data cleaning pipelines involving duplicate removal, mode-based categorical imputation, label encoding, and feature scaling via Scikit-Learn.
- Trained, evaluated, and benchmarked 5 supervised machine learning models (including Random Forest, Support Vector Machines, and Logistic Regression) to classify income tiers.
- Optimized predictive capabilities, selecting Random Forest as the optimal algorithm based on an evaluation metrics achieving 85.3% accuracy and a 0.898 ROC-AUC score.

**Generative AI Course Context Engine** **Aug 2025 – Oct 2025**
*Academic / Personal Project — Python, FastAPI, Ollama, Whisper, Scikit-Learn* GitHub

- Developed a local Retrieval-Augmented Generation (RAG) backend to process educational video assets using FFmpeg for audio extraction and OpenAI Whisper for automated transcription.
- Designed an offline embedding and vector retrieval workflow using Ollama (bge-m3 model), Pandas, Joblib, and Scikit-Learn cosine similarity metrics for semantic query matching.
- Deployed modular FastAPI REST endpoints integrated with llama3.2 to deliver context-aware Q&A, content summarization, and automated JSON-formatted quiz generation.
- Handled end-to-end backend integration, structured JSON data responses, and local LLM execution environments.

## ADDITIONAL INFORMATION

- **Certifications:** CodeWithHarry – The Ultimate Job Ready Data Science Course — CodeWithHarry – Complete 2025 Python Bootcamp
- **Coding Platforms Link:** GitHub profile (github.com/razesoni) showcasing open-source backend, machine learning, and data analytics projects.
- **Achievements / Co-Curricular Activities:** Expressive sketching and illustration focused on structural layout and visual communication; continuous learner of deep narrative literature.