

Shivam Kumar

+91-7061726881 | shivam799kumar@gmail.com

Summary

AI Engineer and Software Developer with over 2 years of experience in building and deploying scalable AI-driven backend systems and machine learning applications. Skilled in designing and optimizing solutions across cloud platforms (AWS), with a proven record of developing high-impact applications like recommendation engines, retrieval-augmented generation (RAG) and LLM fine tuning. Proficient in Python, SQL, and AWS, with hands-on expertise in creating automated workflows and enhancing data retrieval efficiency. Driven to apply advanced machine learning and data analysis skills to solve complex challenges and contribute to innovative, data-focused projects in a dynamic tech environment.

Professional Experience

Estrel AI

Associate Consultant - Software Engineer

Client Location

Dec 2024 - Present

Bihar One Project - An Initiative by Government of Bihar (GoB)

- Contributed to the development of a unified citizen database, integrating multiple government datasets for better beneficiary identification and service delivery.
- Assisted in building a scalable and secure data architecture, ensuring compliance with government data protection policies.
- Led stakeholder meetings and conducted requirement gathering sessions with government departments to identify and document key project needs, ensuring alignment with policy and operational goals.

SemantiQ - Advanced RAG Application

- Designed and deployed a scalable backend architecture on AWS, ensuring efficient retrieval-augmented generation workflows.
- Implemented Dockerization and integrated Kafka for a queue-based file training system, enabling seamless asynchronous processing of large-scale documents.
- Developed RESTful APIs for chat interactions and file training, optimizing response times and enhancing system reliability.

YogLeads

Software Engineer

Noida, India

June 2023 - Oct 2024

- Engineered a Graph-RAG application using Neo4j, reducing query response time by 40% and streamlining data retrieval.
- Implemented NER in Neo4j to categorize 1,000+ nodes, enhancing data search and categorization.
- Created a RAG application with Llama Index and Quadrant DB, improving document retrieval efficiency by 30%.
- Deployed and fine-tuned Automatic Speech Recognition using Whisper and Nemo, achieving 95% accuracy in speech-to-text conversion for enhanced data processing.
- Spearheaded backend development, creating and deploying 100+ RESTful APIs, reducing data latency by 25% for web and mobile platforms using FAST API framework.
- Built and managed WhatsApp automation using Selenium and AWS EC2 with a distributed, queue-based architecture, automating message dispatch and reducing manual intervention by 90%.
- Led team development efforts by conducting in-depth code reviews, improving code quality and reducing bugs.

Blackcoffer

Remote

Data Scientist Associate

April 2022 - June 2022

- Engaged with international clients to identify and resolve data challenges, facilitating 5+ project reviews and achieving a 90% client satisfaction score.
- Executed wine sales analysis, designing and implementing a predictive recommendation system that increased client revenue by 15% through targeted marketing.
- Designed and integrated a cross-platform Slack chatbot with Telegram, streamlining team communications and reducing response time by 20%.

Strata Unison, New York

Remote

Data Annotator Intern

Sep 2021 - Feb 2022

- Annotated over 5,000 images with 98% accuracy, delivering high-quality datasets that enhanced model training for image recognition projects.
- Employed Supervisely for efficient image annotation, reducing annotation time per image by 15% and ensuring consistent data quality for analysis.

Ananta Resource Management

Remote

App Developer Intern

May 2021 - Aug 2021

- Constructed and launched the Naukri portal and Yatri Bhavan applications on Microsoft PowerApps, enhancing user engagement by 25% through efficient, mobile-responsive design.
- Streamlined intuitive and visually engaging user interfaces with Figma, resulting in a 30% increase in positive user feedback.
- Conducted technical presentations on app development strategies, improving cross-functional team understanding and contributing to a 15% boost in project workflow efficiency.

Projects

Mood-based Music Recommendation System Using Brainwaves (Research Project)

- Crafted an EEG-based music recommendation system to enhance emotional well-being, leveraging real-time brainwave data from the Emotiv EPOC headset.
- Utilized Python for robust system development and integration. Applied statistical quantile methods for accurate mood classification from EEG datasets.
- Integrated Spotify dataset for personalized music recommendations aligned with user emotions. Used Selenium to automate recommended music playing on YouTube.
- Research paper can be accessed here: <https://iirpr.com/uploads/V4/ISSUE6/IJRPR14262.pdf>

Targeted Advertisement Using PySpark

- Devised a targeted advertisement system with PySpark, achieving a 20% improvement in processing efficiency by leveraging parallel data processing for faster, more accurate ad targeting.
- Formulated a recommendation engine that analyzes user watch time data, enabling personalized ad delivery and increasing click-through rates by 15%.
- Leveraged Python and PySpark for real-time data manipulation and analysis, supporting advanced personalization algorithms that matched user interests to ads effectively.
- Implemented collaborative filtering algorithms to align ads with user interests, resulting in a 12% engagement increase through tailored ad recommendations.

Education

D Y Patil International University
2019 - June 2023

Aug

B.Tech in Computer Science and Engineering

- Percentage: 8 CGPA
- Specialization: Data Science and Machine Learning

Skills

Programming Languages: Python, Java, C++

Database Technologies: MySQL, Redis, DynamoDB, MongoDB

Frameworks & Libraries: Flask, FastAPI, Django, TensorFlow, PyTorch, Langchain, Llama Index

Cloud Platforms & Services: AWS (EC2, S3, CloudWatch, SES, OpenSearch, S3, Lambda)

DevOps & Monitoring Tools: Git, Docker, New Relic, PM2

Data Science & Analytics: Power BI, Tableau, Data Analysis, Data Engineering, Machine Learning, Neural Networks

Soft Skills: Analytical Thinking, Leadership, Communication, Collaboration

Languages: English, Hindi, Japanese

Achievements

Data Science Code-a-thon Contest (TECHGIG)

Ranked 1st among competitors in TechGig's Data Science Code-a-thon, recognized for outstanding contributions and innovative solutions in data science.

Gold Badge – Python, SQL, and Problem Solving (HackerRank)

Awarded 5-Star Gold Badge in Python and SQL, and 6-Star Gold Badge in Problem Solving on HackerRank, demonstrating advanced skills in complex query-solving, analytical reasoning, and programming proficiency.

LeetCode Performance

Successfully solved 500+ problems, showcasing consistency in daily learning and a strong commitment to honing algorithmic and problem-solving skills.

Certifications

Microsoft AI Classroom Series - Microsoft 

Design Thinking and Global Startup - Korea Advanced Institute of Science and Technology 

Design Thinking for Innovation - University of Virginia 

Socials

GitHub: <https://github.com/data-scientist-shivam799>

Linkedin: <https://www.linkedin.com/in/entrepreneur-shivam/>

HackerRank: <https://www.hackerrank.com/profile/shivam799kumar>