Kumar Rohit - Professional Profile Summary

Professional Summary

Kumar Rohit is a Data Engineer and AI/ML specialist with extensive experience in building scalable data pipelines, ETL processes, and implementing advanced machine learning solutions. Currently working as an Analyst at StoneX Group Inc., he has progressed from a Data Engineering Intern to a full-time analyst role, demonstrating strong technical capabilities and problem-solving skills in production environments.

With a B.Tech in Artificial Intelligence and Machine Learning from Symbiosis Institute of Technology (CGPA: 8.65/10), Kumar combines strong academic foundations with practical industry experience. His expertise spans data engineering, machine learning, computer vision, natural language processing, and generative AI, with a proven track record of delivering high-impact projects in both corporate and research settings.

Education

Symbiosis Institute of Technology (2021 – 2025)

- **Degree:** B.Tech in Artificial Intelligence and Machine Learning
- **CGPA:** 8.65/10
- **Location:** Pune, Maharashtra

St. Karen's High School (2018 – 2020)

- **Grade:** 12th (CBSE)
- **Percentage:** 95.8%
- **Location:** Patna, Bihar

Technical Skills

Programming & Scripting

- Python
- SQL

Data Engineering & Pipelines

- ETL Development
- Data Integration
- Workflow Orchestration
- Databricks

- Apache Airflow

AI/ML Domains

- Machine Learning

- Computer Vision

- Natural Language Processing

- Large Language Models (LLM)

- Generative AI

- Agentic AI Systems

Python Libraries & Frameworks

- **Machine Learning:** Scikit-learn, Pandas, NumPy

- **Visualization:** Matplotlib, Seaborn, Plotly

- **LLM & GenAI:** LangChain, LangGraph

- **API Development:** FastAPI

- **UI Development:** Streamlit

- **NLP Models:** Transformers, RoBERTa

Cloud & DevOps

- Azure DevOps

- Azure Databricks

- Azure Speech Services

- Multi-environment deployment (Dev, QAT, PPE, Prod)

Databases & Vector Stores

- MySQL

- SQL Server

- Pinecone (Vector Database)

Additional Technologies

- YOLOv8 (Object Detection)

- DaViT (Dual Attention Vision Transformers)

- HERE API

- CrewAI
- Gemini

Professional Experience

StoneX Group Inc. (January 2025 - Present)

Analyst (August 2025 - Present)

Location: Pune, Maharashtra, India (Hybrid)

- Creating comprehensive Source to Target mappings for 45 Airflow DAGs to ensure seamless migration to Databricks
- Conducting in-depth analysis of Databricks production job failure logs to identify and resolve root causes
- Developing interactive dashboards to visualize failure trends and patterns
- Delivering data-driven insights that significantly improved job reliability and system stability
- Working extensively with Microsoft Excel, Apache Airflow, and data pipeline tools

Data Engineering Intern (January 2025 - July 2025)

Key Achievements:

- Built scalable ETL pipelines on Databricks, ingesting data from SQL Server
- Implemented Medallion Architecture (Bronze, Silver, Gold layers) using Unity Catalog to improve data organization and access control
- Led critical process migration from Apache Airflow to Databricks Jobs, improving job performance significantly
- Successfully migrated an Apache Airflow DAG that delivered company and account data to Salesforce CRM, replacing it with a Databricks workflow
- Achieved 40% reduction in data latency through the migration
- Increased system reliability by eliminating daily failures
- Improved CRM data integration accuracy
- Enabled decommissioning of legacy Airflow DAGs
- Deployed 10+ production jobs using Azure DevOps pipelines
- Gained hands-on experience with multi-environment deployment workflows (Dev, QAT, PPE, Prod)
- Ensured smooth CI/CD deployment processes
- Analyzed production job failure logs and built monitoring dashboards

Technologies Used: Python, Azure Databricks, Apache Airflow, SQL Server, Azure DevOps, Unity Catalog

Symbiosis Centre for Applied Artificial Intelligence (SCAAI)

Research and Project Intern (July 2024 - January 2025) Location: Pune, Maharashtra, India (On-site)

Project: Government-funded surveillance project

- Worked on real-time aerial and naval object detection systems

- Curated an extensive dataset of 6 lakh (600,000) images

- Dataset included aircraft, helicopters, drones, and naval objects

- Developed advanced object detection model using YOLOv8

- Achieved mean average precision (mAP) of 78.5%

- Surpassed previous research benchmarks in the domain

- Focused on surveillance applications for defense and security

Technologies Used: Python, YOLOv8, Computer Vision techniques

Other Positions

Mental Health Champion

Symbiosis Centre for Emotional Wellbeing (January 2023 - June 2024) Location: Pune

- Contributed to student mental health and wellbeing initiatives

Technical Co-Head

SIT ACM Student Chapter (July 2023 - February 2024) Location: Pune, Maharashtra, India

- Led technical initiatives and activities for the ACM student chapter

- Organized coding events and technical workshops

Volunteer

SymbITech (September 2022) Location: Pune, Maharashtra, India

- Volunteered for data science visualization event at tech fest

- Collected, cleaned, and processed datasets according to requirements

- Created analytical questions based on prepared datasets

Notable Projects

Generative AI & Agentic AI Projects

AI Interview Assistant (Currently Building)

Technologies: LangChain, FastAPI, Streamlit

Description:

- Building an intelligent resume-based Q&A tool
- Leverages LangChain for conversational AI capabilities
- Implements FastAPI for efficient backend services
- Creates personalized interview preparation experience
- Provides context-aware responses based on resume content
- Streamlit-based user interface for seamless interaction

NGO Chatbot (Currently Building)

Technologies: NLP, Speech Recognition, Multimodal AI

Description:

- Developing a multimodal chatbot for NGO support
- Integrates Natural Language Processing for text understanding
- Implements speech recognition for voice interactions
- Provides accessible communication channels for diverse users
- Enhances NGO outreach and engagement capabilities

MedSure - Multimodal Medical Chatbot

Technologies: Langchain, HERE API, Streamlit, Pinecone, Gemini, Azure Speech Services, CrewAI

Achievements:

- Created advanced multimodal chatbot using Retrieval Augmented Generation (RAG)
- Integrated Pinecone vector database for intelligent medical document retrieval
- Enabled seamless voice input and output through Azure Speech Services
- Enhanced accessibility for users with different interaction preferences
- Implemented location-based features using HERE API

- Utilized CrewAI for improved agentic conversational capabilities

- Deployed comprehensive healthcare information system

- Combines text, voice, and document understanding in single platform

Resume Revealer - Intelligent Resume Parser

Technologies: RoBERTa, Plotly, NLP

Achievements:

- Built sophisticated resume parsing system using RoBERTa transformer model

- Extracts structured information from unstructured resume documents

- Implements advanced Named Entity Recognition (NER) for information extraction

- Creates interactive visualizations using Plotly for resume analysis

- Identifies skills, experience, education, and other key resume components

- Provides insights into resume quality and completeness

- Helps recruiters and job seekers analyze resume effectiveness

Computer Vision Projects

X-ray Vision - Medical Image Analysis

Technologies: DaViT (Dual Attention Vision Transformers), YOLOv8, Streamlit

Achievements:

- Engineered a comprehensive lung X-ray classification pipeline

- Utilized Dual Attention Vision Transformers for advanced image analysis

- Achieved 92% accuracy in classification tasks

- Integrated YOLOv8 for object detection capabilities

- Improved mean average precision (mAP) to 33%

- Surpassed prior benchmarks in medical image analysis

- Deployed user-friendly interface using Streamlit

Machine Learning Projects

Smart Property Valuator - Real Estate Price Prediction

Technologies: Scikit-learn, Matplotlib, Seaborn, Streamlit

Achievements:

- Developed machine learning solution for Gurgaon real estate market

- Conducted robust Exploratory Data Analysis (EDA)

- Implemented comprehensive feature engineering techniques

- Delivered 90% R^2 score using Random Forest algorithm

- Launched real-time prediction application via Streamlit

- Enabled accurate property valuation for end users

Achievements & Recognition

Research Publications

- **ICMLA 2024 Conference:** Presented research paper on Real-Time Object Detection

- **Published in Springer Conference Proceedings**

- Research focused on surveillance and object detection in aerial and naval contexts

Competition Wins

- **1st Prize - Manas Codeathon:** Organized by CDAC Bangalore
  - Recognized for excellence in product development

  - Demonstrated superior coding and problem-solving abilities

Open Source Contributions

- **Social Winter of Code (SWOC'24):** Selected as contributor

- Contributed to open-source projects in data science domain

- Collaborated with developers globally on meaningful projects

Hackathon Success

- **TIAA Hackathon:** Ranked among Top 20 teams nationwide

- Competed against leading teams across India

- Demonstrated strong technical and teamwork capabilities

Key Strengths

Data Engineering

- Expert in building scalable ETL pipelines

- Proficient in Databricks and Apache Airflow

- Strong understanding of data architecture patterns (Medallion Architecture)

- Experience with Unity Catalog for data governance

- CI/CD implementation using Azure DevOps

- Multi-environment deployment expertise

Generative AI & LLM Applications

- Building production-ready LLM applications

- Expertise in LangChain and LangGraph frameworks

- Retrieval Augmented Generation (RAG) implementation

- Multimodal AI system development

- Conversational AI and chatbot development

- Agentic AI systems with CrewAI

Machine Learning & Computer Vision

- Advanced knowledge of computer vision techniques

- Experience with state-of-the-art object detection models (YOLOv8)

- Transformer-based architectures (DaViT, RoBERTa)

- Natural Language Processing and NLP model fine-tuning

- Vision Transformers and attention mechanisms

Problem Solving & Analytics

- Root cause analysis of production failures

- Dashboard development for monitoring and insights

- Data pipeline optimization and migration

- Performance improvement initiatives

Full-Stack Development

- API development with FastAPI

- Frontend development with Streamlit

- End-to-end application deployment

- User experience optimization

Career Highlights

- Successfully transitioned from internship to full-time analyst role at StoneX Group Inc.

- Reduced data latency by 40% through strategic migration from Airflow to Databricks

- Curated and worked with large-scale datasets (600,000+ images)

- Achieved research publication in international conference (Springer)

- Won multiple hackathons and coding competitions

- Contributed to government-funded research projects

- Deployed 10+ production jobs in enterprise environments

- Building cutting-edge GenAI applications using latest LLM technologies

Professional Development

- Strong foundation in both theoretical AI/ML concepts and practical implementation

- Experience in agile development and DevOps practices

- Proven ability to work in hybrid and remote environments

- Leadership experience through technical co-head position

- Community engagement through open-source contributions

- Commitment to mental health awareness and student wellbeing

- Continuous learning in emerging AI technologies

Work Style & Approach

- Data-driven decision making

- Focus on scalability and reliability

- User-centric application development

- Continuous learning and skill development

- Collaborative team player

- Strong documentation practices

- Proactive problem identification and resolution

- Building practical AI solutions for real-world problems

Technical Interests

- Large Language Models and prompt engineering

- Retrieval Augmented Generation (RAG) architectures

- Agentic AI and autonomous systems

- Multimodal AI applications

- Computer vision for healthcare and surveillance

- Real-time data processing and streaming

- MLOps and production ML systems

This profile summary is designed to provide comprehensive information about Kumar Rohit's professional background, technical skills, and achievements for use in a digital chatbot or automated query response system.