

AKSHAY JOSHI

Data Scientist



+91 8827406956



akshaysjoshi24@gmail.com



<https://www.linkedin.com/in/-akshay-joshi/>

JOB OBJECTIVE

Focused on leveraging machine learning and generative AI to build scalable, data-driven solutions that address complex business problems. Committed to innovation and practical AI deployment.

EDUCATION

Post Graduate Program in Data Science and Engineering

Great Lakes Institute of Management, Bangalore — 2021

B.E., Mechanical Engineering

Acropolis Institute of Technology and Science, Indore — 2011–2015 |
CGPA: 6.8/10

CORE COMPETENCIES

Predictive Modeling and Deployment

Feature Engineering and Data Preprocessing

Large Language Models (LLMs) and Generative AI

Natural Language Processing (NLP) Applications

Machine Learning Algorithms (Tree-based, Ensemble Methods)

Model Evaluation and Validation Techniques

Exploratory Data Analysis (EDA)

Cloud-based AI Solutions and Vertex AI

AI-Powered Chatbots and RAG Integration

Document Parsing and Summarization Pipelines

Hyper parameter Tuning and Optimization

Python Programming and Data Manipulation

Containerization and Application Deployment

CERTIFICATIONS

- Generative AI with Large Language Models – Coursera
- Machine Learning Specialization – Coursera

PROFESSIONAL SUMMARY

- Data Scientist with over 3 years of experience**, currently contributing to Cybage Software by leveraging advanced machine learning and AI technologies to drive business value
- Proven expertise in designing scalable predictive models and deploying them in production environments to enhance decision-making and operational efficiency
- Strong background in applying **Large Language Models (LLMs) and generative AI** techniques to develop innovative NLP-based products and automated solutions
- Skilled in transforming complex, high-dimensional data into actionable insights through robust feature engineering and exploratory data analysis
- Adept at integrating **cutting-edge AI frameworks like LangChain and GPT** to create intelligent conversational agents and document processing tools
- Experienced in working with cloud platforms and modern deployment tools to streamline ML workflows and ensure **continuous delivery of AI-driven applications**
- Committed to stay at the forefront of emerging AI trends & technologies, with continuous learning demonstrated through relevant certifications & practical implementation

WORK EXPERIENCE

Cybage Software — Data Scientist

Pune, India | May 2023 – Feb 2025

Project: Lapsed Customer Prediction Model

- Creating a predictive model to identify lapsed customers by analyzing transactional, demographic, & **behavioral data from over 28 million records spanning 15 months**
- Engineering 115+ features through Information Value (IV) and bivariate analysis to enhance signal quality and model performance
- Achieving over 85% recall on validation datasets** while maintaining strong interpretability for business stakeholders
- Deploying the model within Redshift for automated monthly scoring, driving targeted retention campaigns
- Facilitating **proactive customer retention strategies**, contributing to a tenfold increase in sales revenue

Project: DocEvolve.AI – LLM-based Document Summary Generator

- Designing and developing DocEvolve.AI, a generative AI-powered document summarization tool utilizing LLaMA2 and GPT-based models
- Integrating a **Retrieval-Augmented Generation (RAG) chatbot** using LangChain and GPT-3.5 to enable interactive conversational Q&A functionality
- Building a comprehensive document parsing pipeline incorporating PyPDF2, pdfplumber, pytesseract, and docx2pdf for seamless data extraction and processing

Key Achievements:

- Spearheaded the end-to-end development and deployment of a **Lapsed Customer Prediction model**, empowering the organization to implement data-driven customer retention strategies.
- Engineered **critical components of a GenAI Retrieval-Augmented Generation (RAG) chatbot**, including advanced file parsing, document chunking, embedding generation, vector database integration, and efficient RAG-based information retrieval.
- Collaborated cross-functionally to architect scalable system frameworks and translated conceptual designs into robust, production-grade code modules.
- Acquired deep **operational expertise in deploying scalable machine learning** and generative AI solutions within live production environments.
- Exhibited proactive ownership, analytical problem-solving and effective teamwork in both autonomous and **collaborative project settings**.

TECHNICAL SKILLS

- **Programming & Data Tools:** Python, Pandas, NumPy, MySQL, BigQuery
- **Machine Learning & AI:** Scikit-learn, TensorFlow, LLMs (GPT, BERT, LLaMA2), RAG, GenAI
- **NLP & Deep Learning:** Text Classification, LSTM, Transformers, Summarization
- **Cloud & Deployment:** Vertex AI, Docker, Streamlit
- **Git**
- **Deployment & Frameworks:** Docker, Streamlit
- **Techniques:** EDA, Feature Engineering, Hyperparameter Tuning

SOFT SKILLS

Analytical Thinking

Problem-Solving

Effective Communication

Collaboration and Teamwork

Adaptability and Continuous Learning

PERSONAL DETAILS

Address : Pune, India

Languages Known : English, Hindi

Date of Birth : 24th April 1993

TEKsystems — Associate Data Scientist

Bangalore, India | Dec 2021 – Apr 2023

Project: Loan Default Detection

- Developing a Random Forest classifier on Vertex AI to accurately predict loan default probabilities
- Leveraging **domain expertise and statistical techniques** to identify critical risk factors such as income, loan purpose, and credit score. Enhancing risk segmentation models to improve decision-making processes for loan approvals

Project: Retail Reorder Prediction

- Analyzing reorder patterns across various product categories and aisles to understand customer purchasing behavior
- **Identifying top and bottom 20 product reorder trends** to guide inventory optimization
- Delivering actionable recommendations for aisle-level product placement and replenishment strategies to maximize operational efficiency

Key Achievements:

- Collaborated as an integral team member to develop prototype ML models on synthetic datasets leveraging Google’s cutting-edge Serverless Spark platform, **facilitating impactful client demonstrations.**
- Acquired practical expertise in orchestrating cloud-native ML pipelines, encompassing model training and deployment workflows via Vertex AI.
- Collaborated with cross-functional teams to iterate on classification algorithms and optimize **feature engineering strategies using datasets modeled** on real-world scenarios.
- Cultivated core data science competencies, including rigorous statistical analysis, hyperparameter tuning, and proficiency with GCP cloud-native tools and services.

IndiaBulls HFL — Sales Officer

Indore | Aug 2017 – Apr 2019

- Assessing loan eligibility for both salaried and self-employed clients with thorough financial analysis
- Managing and **expanding home loan portfolios through DSAs** and builder networks, actively driving increased disbursement volumes

Shadow Infra — Business Development Manager

Indore | Apr 2019 – Jan 2021

- Driving end-to-end loan sales processes, managing client acquisition through to successful disbursement
- Coordinating **closely with clients and cross-functional teams** to facilitate and finalize high-value real estate transactions