MANISH KUMAWAT

Data Scientist (AI/ML)

Phone: **+91** 8956012161
Email: Manish.kumawat1808@gmail.com
Address: 08-Oas appt, Gulmohar garden, Nashik - 422101
DOB: 08/18/1996

PROFILE SUMMARY:

- **Process-Oriented Data Scientist** with **5.5 years** of experience in designing, training, and deploying machine learning models.
- Expertise in **Natural Language Processing (NLP), Large Language Models (LLMs), and Applied Machine Learning** across diverse industries, including e-commerce, medical, and consulting.
- Experienced in **building conversational AI systems**, automating workflows, and fine-tuning models for real-world applications.
- Strong focus on **integrating AI into business processes** to extract actionable insights from unstructured data and deliver end-to-end solutions.
- Proficient in **cloud platforms like AWS and Azure**, optimizing scalable and cost-effective data pipelines.
- Skilled in **interpreting and analysing data** to drive growth and improve corporate performance.

PROFESIONAL EXPERIENCE:

- **2+ Years** of Working Experience at **Thinkbridge Software Pvt.Ltd**, Pune as **Data Scientist** from February 2023 to Present Date
- **3.3 Years** of Working Experience at **Bitwise Global**, Pune as **Machine Learning Engineer** from December 2019 to February 2023

KEY SKILLS:

- **Machine Learning & AI:** Model development, training, evaluation, and deployment.
- **Generative AI & LLMs:** Prompt engineering, fine-tuning, and deployment of AI models.
- **Natural Language Processing (NLP):** Text processing, sentiment analysis, topic modeling.
- **End-to-End AI Integration:** Scalable AI solutions, cost-effective ML pipelines, real-world deployment.
- **Cloud Computing:** AWS, Azure – model deployment, data pipeline optimization.
- **Data Analysis & Visualization:** Statistical analysis, data storytelling, dashboards (Power BI)
- **Databases:** SQL, NoSQL, Vector Databases (ChromaDB, Milvus, Faiss , Weaviate, Pinecone)
- **Programming & Scripting:** Python (Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch), SQL
- **Automation & Workflow Optimization:** AI-driven automation, chatbot development, process enhancement
- **Business Intelligence & Strategy:** Data-driven decision-making, performance optimization, predictive analytics

TECHNICAL SKILLS:

- **Programming & Development Tools:** Jupyter Notebook, VS Code, Google Collab, Git, GitHub, GitLab CI/CD
- **Data Processing & Workflow Automation:** Azurite (Azure Data Factory), Celery, RQ (Redis Queue), IDP
- **Machine Learning & Deep Learning Frameworks:** TensorFlow, PyTorch, Scikit-learn, XGBoost, LightGBM
- **LLM & Generative AI Tools:** Hugging Face Transformers, LangChain, OpenAI API, Anthropic, Langfuse
- **Databases:** SQL , NOSQL, Vector DB, Mongo DB
- **Cloud & MLOps Platforms:** Azure, AWS, MLflow.
- **Monitoring & Logging:** Langfuse, Flower (Celery Monitoring).
- **Model Deployment & Serving:** Docker, Kubernetes, FastAPI, Flask
- **Business Intelligence & Visualization:** Power BI, Matplotlib, Seaborn, Excel

QUALIFICATION:

- BE | Savitribai Phule Pune University | 2018 | First class
- HSC | Maharashtra State Board | 2014 | 62.5%
- SSC | Maharashtra State Board | 2012 | 77%

PROJECTS:

1.Project Name: thinkAdvisor

Domain: Strategy Consulting / AI-driven Solutions

Description:

Designed and developed thinkAdvisor, an AI-powered assistant tailored for strategy consultants, integrating proprietary consulting frameworks with Large Language Models (LLMs). Utilizing advanced techniques such as Retrieval-Augmented Generation (RAG) and agentic frameworks, thinkAdvisor delivers comprehensive company and industry insights, streamlining research efforts and significantly reducing preparation time. This solution enhances client engagement by generating key consulting artifacts, automating repetitive tasks, and improving decision-making efficiency.

Responsibilities:

- Led the end-to-end development of thinkAdvisor, architecting an AI-driven solution that transforms strategy consulting workflows.
- Integrated LLMs with consulting methodologies, leveraging RAG and agentic frameworks to generate tailored insights.
- Utilized multiple chat models from OpenAI and Anthropic (Claude) to enhance response quality and contextual relevance.
- Implemented Celery and Redis for asynchronous task execution and distributed processing, improving scalability and system responsiveness.
- Deployed Langfuse to monitor real-time performance and evaluations of LLMs, ensuring consistent quality and reliability and cost management.
- Optimized AI-generated insights to enhance decision-making for strategy consultants, reducing research time and improving client interactions.
- Designed a scalable infrastructure that supports high-volume query processing with minimal latency.

2.Project Name: Generating Monthly Highlights using Azure OpenAI

Domain: Financial Analytics / AI-driven Reporting

Description:

Developed an AI-powered automated financial summarization system that generates monthly executive highlights by processing financial data from SQL Server and leveraging Azure OpenAI'. The solution extracts, restructures, and analyzes revenue and expense data, generating high-quality summaries stored in Azure Storage. Additionally, an evaluation pipeline ensures data accuracy by verifying key financial metrics in the AI-generated reports. This system streamlines financial reporting, reduces manual effort, and enhances decision-making for stakeholders.

Responsibilities:

- Architected and implemented a serverless solution using Azure Functions for automated financial reporting.
- Developed data ingestion pipelines to extract and restructure financial data from SQL Server using pyodbc.
- Designed and deployed a Chatmodel model on Azure OpenAI Studio, optimizing prompt engineering for executive summary generation.
- Built an asynchronous execution framework using Azure Function Apps, Celery, and Redis, improving processing efficiency.
- Integrated an evaluation pipeline to measure AI output accuracy, verifying total revenue and expense values in AI-generated reports.
- Implemented a feedback collection system, enabling users to rate AI-generated summaries and provide structured insights for continuous improvement.
- Deployed and managed the entire solution on Azure, ensuring scalability, reliability, and compliance with financial data integrity standards.

3.Project Name: Intelligent Document Processing for Health Insurance Contracts

Domain: Healthcare / AI-driven Document Automation

Description:

Developed an Intelligent Document Processing (IDP) pipeline to automate the extraction and processing of key contractual details from Non-Emergency Medical Transportation (NEMT) agreements with health insurance providers. The system utilized AI-powered document parsing, NLP-based information extraction, and structured data storage to streamline contract management, ensuring compliance and operational efficiency. By automating contract analysis, the solution reduced manual effort, improved data accuracy, and enabled faster decision-making for the client.

Responsibilities:

- Designed and implemented an AI-driven IDP pipeline to process contracts from multiple health insurance providers.
- Developed document ingestion workflows to handle various contract formats (PDF, Word) and extract key terms using OCR and NLP techniques.
- Built an entity recognition and extraction model to identify critical contract details such as service terms, pricing, coverage limits, and regulatory clauses.
- Integrated LLMs and rule-based systems to ensure accurate interpretation of contract terms and conditions.
- Stored extracted data in a structured format using SQL database for easy access and reporting.
- Implemented validation mechanisms to cross-check extracted details against predefined compliance rules.
- Optimized pipeline performance to handle large volumes of contracts with minimal latency and high accuracy.
- Deployed the solution on cloud infrastructure (Azure) to ensure scalability and security in processing sensitive healthcare contracts.

4.Project Name: Customer Churn Prediction Analysis (Using ANN – Deep Learning Model)

Domain: Finance / Predictive Analytics

Description:

Developed an advanced customer churn prediction system for a major bank, leveraging deep learning (ANN) to proactively identify customers at risk of leaving. This innovative solution enabled the customer service team to focus on high-impact retention strategies and effectively reduce churn.

Responsibilities:

- Data Integration & Collection: Consolidated data from diverse sources and formats to create a comprehensive dataset for analysis.
- Data Preprocessing: Conducted extensive data wrangling, handling imbalanced classes and performing feature engineering to enhance model accuracy.
- Exploratory Data Analysis: Utilized Microsoft Excel and advanced visualization tools to identify key factors contributing to churn and uncover actionable insights.
- Model Development: Designed and built a statistical analysis model using an Artificial Neural Network (ANN), tailored to predict customer churn on large-scale datasets.
- Performance Evaluation: Interpreted model outcomes to extract key metrics, validate predictions, and fine-tune parameters for optimal performance.
- Strategic Recommendations: Proposed targeted solutions based on data insights, leading to a measurable 15% reduction in customer churn through improved operational efficiencies.

5.Proof of Concept & Small Projects

- **FWA Detection for NEMT Services:** Developed a system to identify potential Fraud, Waste, and Abuse using advanced analytics and ML techniques.
- **XML Data Migration Pipeline:** Engineered an automated pipeline to extract XML files from AWS S3, preprocess data, and load it into SQL databases.
- **Employee Churn Analysis:** Analysed internal HR data to uncover key drivers of turnover and recommended retention strategies.
- **Continuous Review Feedback System:** Built a platform for weekly feedback exchange to streamline performance reviews.
- **Trip Details Calculation Module:** Created a Python module integrating Google Maps APIs to calculate trip metrics for a NEMT cloud platform.