

# AMOL B

## DATA SCIENTIST

 8767074972

 amol.n.badhe97@gmail.com

Experienced Data Scientist with 3.5 years of expertise in Machine Learning (ML), Natural Language Processing (NLP), Deep Learning (DL), and Generative AI. Proficient in Python, TensorFlow, PyTorch, Scikit-Learn, AI frameworks. Strong background in data preprocessing, feature engineering, model development. Hands-on experience in large language models (LLMs), transformers, RNNs, CNNs, and GANs for various AI applications. Skilled in SQL, Spark, Databricks, and cloud platforms like AWS. Passionate about applying AI to real-world business problems, optimizing algorithms, and deploying scalable solutions using Docker, Kubernetes, and CI/CD pipelines. Strong analytical and problem-solving skills with expertise in predictive modeling, statistical analysis.

Coding :	Python, SQL , Numpy , Pandas
Database :	MySQL, MongoDB,MSSQL
Platforms :	Jupyter Notebook, Visual Studio, Google Colab, Workbench
ML Algorithms :	Linear Regression, Logistic Regression, Decision Tree, Random Forest, KNN, K-Means, Adaboost, SVM, PCA, Naive Bayes
Deep Learning :	ANN ,RNN , LSTM , LLM ,GRU , GAN ,NLP
Technology :	Python, Numpy, Pandas, OOPs, Machine Learning, Deep Learning, Neural Networks, Tensorflow, AWS (Cloud), Natural Language Processing, SQL, GIT, Flask, MongoDB, Postman, Hypothesis Testing, ANN , Ilm , Prompt engineering
Packages :	Scikit Learn, Matplotlib, Scipy, Seaborn, stats, nltk, tensor

## PROFESSIONAL EXPERIENCE

### NTT DATA | AI Engineer

July 2021 -Present

- **Machine Learning & Predictive Modeling:** Developed end-to-end machine learning models including data extraction, preprocessing, feature engineering, and model evaluation using algorithms like **Linear/Logistic Regression, KNN, SVM, Decision Trees, Random Forest, Naive Bayes, and K-Means**
- **Data Wrangling & Data Preparation:** Expertly managed large datasets by cleaning, transforming, and normalizing data to ensure quality input for models. Performed data manipulation and preprocessing using **Python, Pandas, and NumPy**
- I spearheaded the development of an automated financial analysis platform for PE firms using deep learning. The system reduced report generation time from days to minutes while maintaining 85% forecast accuracy.
- Advanced Machine Learning: **Applied advanced ML algorithms including Deep Learning, Natural Language Processing (NLP), and Time Series Analysis** to solve complex business problems.
- Cloud & Big Data Technologies: Experienced in deploying ML models and managing data workflows on **AWS, including EC2 and Sagemaker**. Familiar with cloud storage, computation, and orchestration.

---

**PROJECTS**

---

**Project 1: Language Translation Model Generation**

**Description:** Developed a robust Gen-AI-based document translation model for accurate and efficient translation of medical documents from English to Japanese, fostering global connectivity and collaboration.

**Roles and Responsibilities:**

- Analyzed business requirements and developed effective strategies for Large Language Models (LLMs) utilization.
- Preprocessed a dataset of approximately 65 million instances for generative AI model training.
- Tested and implemented Helsinki (HNMT) translation model, achieving a significant improvement in BLEU score from 28 to 55.
- Integrated AI models with APIs and application servers (Flask, Django) for scalable AI applications.
- Managed AI model inputs and outputs using SQL and NoSQL databases (e.g., MongoDB).
- Strong experience with AI/ML frameworks (TensorFlow, PyTorch, Transformers) and deep understanding of Generative AI models, including LLMs.
- Proficient in Vector Databases, Vector Similarity Search, and Embedding models and frameworks.
- Experienced in designing multi-lingual Generative AI applications.
- Familiar with cloud platforms (AWS) and containerization technologies (Docker) for deploying AI solutions.
- Resources used: 8xA100 NVIDIA GPU system.

**Project 2: Insurance Cost Prediction Model**

**Description:** Developed a machine learning-based health insurance prediction system to enhance cost estimation accuracy and optimize insurance pricing. With the increased focus on health insurance, the project aimed to leverage AI-driven analytics to improve risk assessment and policy customization.

**Roles and Responsibilities:**

- Analyzed business challenges and formulated effective machine learning strategies for cost prediction.
- Built and trained models using Logistic Regression, Decision Tree Classifier, KNN, and Random Forest Classifier for predictive analytics.
- Conducted statistical analysis and data exploration using Matplotlib and Seaborn for better feature understanding.
- Performed data visualization, feature engineering, extraction, and preprocessing to enhance model performance.
- Developed, trained, and evaluated machine learning models, optimizing their accuracy and efficiency.
- Deployed the model using Flask and AWS, ensuring scalability and real-time processing capabilities.

**Project 3: Building a Chatbot for Financial and Operational Risk Guidelines**

**Description:** Architected an AI-powered regulatory chatbot using LangChain and RAG (Retrieval-Augmented Generation), significantly improving compliance and risk management efficiency.

**Roles and Responsibilities:**

- Implemented RAG-based system to enhance document retrieval accuracy for financial and operational risk guidelines.
- Reduced compliance query resolution time from 2 days to 10 minutes, saving 120+ man-hours per month.
- Integrated LangChain, Pinecone, and Vector Databases for efficient information retrieval and document embeddings.
- Designed text chunking and embedding pipelines to improve chatbot response relevance and accuracy.
- Deployed scalable architecture using AWS (Amazon Bedrock, S3, Lambda), Kubernetes, and FastAPI for real-time query processing.
- Utilized Transformer models (BERT, Hugging Face) to enhance text understanding and semantic search capabilities.
- Implemented prompt engineering techniques to improve chatbot responses and generate precise regulatory insights.