Dinesh Sagar

+91-8121909150 Mail GitHub LinkedIn Profile

Passionate Data Scientist with a drive to tackle real-world challenges through data analysis and machine learning. Committed to extracting value from complex data and delivering actionable insights to inform business decisions and drive growth. Eager to continuously learn and improve to deliver impactful solutions.

Experience

Capgemini

Consultant

FEB 2025- PRESENT

Amazon Development Centre India

Quality Specialist

JAN 2020 - DEC 2024

- Implemented an outlier detection program to enhance address data quality, improving logistics performance for Amazon Logistics.
- Analyzed large-scale datasets using statistical methods to optimize delivery operations and resource planning.
- Built time series models to forecast peak order volumes, enabling proactive logistics management.
- Developed machine learning models to evaluate delivery agent performance, driving operational efficiency.
- Created interactive dashboards for real-time monitoring and actionable insights.

Skills

- **Programming:** Python, SQL
- **Data Visualization:** Tableau, Power BI
- **Statistics & Math:** Statistics and Hypothesis Testing, NumPy, Pandas, Scikit-Learn
- **Machine Learning:** TensorFlow, Keras, Streamlit, Flask, FastAPI
- **Generative AI:** LangChain, LangGraph, AWS Bedrock, Azure AI Foundry, Azure Machine Learning, AWS Sagemaker

Scaler Certifications

- Python Libraries, EDA Fundamentals
- Supervised & Unsupervised Learning
- Natural Language Processing, Computer Vision

Projects

Generative AI Powered Data Assistants

- **Medical RAG agent:** Developed a medical chatbot using Pinecone DB for vector storage and Flask for real-time query handling to provide accurate, context-driven medical answers.
- **RAG Agent with ChromaDB:** Developed an intelligent recommendation system leveraging ChromaDB for reranking and content discovery based on user queries.
- **CSV Command Agent:** Built an AI-powered tool enabling natural language manipulation of CSV files and SQL query generation, simplifying data operations for non-technical users.

Named Entity Recognition (NER) in Tweets Using LSTM+CRF and BERT Models

- Developed a robust Named Entity Recognition (NER) system for tweets using LSTM-CRF and BERT models.
- Initialized LSTM-CRF with Word2Vec embeddings and fine-tuned BERT for entity detection.
- Executed data preprocessing, hyperparameter optimization, and model evaluation to boost accuracy.
- Achieved precise entity recognition on real-world tweet data, ensuring high model performance.

Ninjacart Computer Vision Project

- Developed a CNN-based computer vision system for Ninjacart to identify onions, potatoes, tomatoes, and noise in produce images.
- Leveraged transfer learning (VGG, ResNet) and data augmentation, with optimized training via callbacks and regularization techniques.
- Achieved robust and accurate vegetable classification.

Delivery Time Estimation for Porter's Delivery Service

- Performed comprehensive data cleaning, feature engineering, and categorical encoding
- Designed, trained, and optimized a neural network model using TensorFlow and Keras
- Delivered accurate delivery time predictions, enhancing service quality and operational efficiency

AdEase Wiki Page View Forecasting

- Forecasted Wikipedia page views to optimize ad placement for AdEase clients, considering regional and linguistic variations.
- Applied ARIMA, SARIMAX, and Prophet time series models, incorporating exogenous campaign data for enhanced accuracy.
- Improved ad placement strategies based on predicted page view trends, enabling better ROI and targeted advertising campaigns for diverse regions.

Education

Scaler

Data Science and Machine Learning

2024

Chaitanya Bharathi Institute of Technology

B.E(Mechanical)

2019

ZEE Recommendation System

- Implemented a movie recommender system for personalized movie suggestions.
- Utilized user-user and item-item similarity with Cosine and Pearson methods.
- Built collaborative filtering and matrix factorization models for recommendations.
- Enhanced user experience through data-driven movie suggestions

Graduate Admissions Prediction

- Analyzed graduate admissions data, constructed precise linear regression models in Python, and presented actionable insights to enhance student success for Jamboree.
- Created Linear regression model and implemented Ridge and Lasso regularizations and evaluated model metrics using sklearn stats library. Identified significant predictors (e.g., GRE scores, GPA) for graduate admissions.
- Developed accurate linear regression models with R2 score exceeding 0.8.
- Delivered insights and recommendations for improving application profiles.

LoanTap Creditworthiness Model

- Developed a Logistic Regression model for LoanTap, analyzed precision vs. recall tradeoffs, and provided actionable insights for enhanced loan decisions.
- Created several flags and indicators based on the given features and applied a logistic regression model to generate classification report, ROC AUC curve, precision-recall curve.
- Revealed key relationships between loan status and borrower attributes.
- Achieved high AUC score, indicating strong model performance.
- Provided insights into significant predictors of loan defaults, leading to improved creditworthiness assessment.