VENUDATTA KARUMURI

Aspiring Data Scientist

+91-9490676066 venukarumuri986@gmail.com linkedin.com/venudatta-karumuri github.com/venudatta67
Bengaluru

SUMMARY

Enthusiastic and results-driven data science aspirant with experience in building machine learning models, data preprocessing, data modeling, and visualization. Skilled in Python, SQL, Power BI, Tableau, and statistical techniques, with a good understanding of database management and ETL processes. Strong communicator with a passion for solving business problems using data-driven insights.

EXPERIENCE

09/2024 - present

Bengaluru

Data Science Consultant

Rubixe AI Solutions

- Designed and deployed supervised and unsupervised machine learning models, applying statistical techniques to address complex business problems and support data-driven decision-making.
- Performed comprehensive exploratory data analysis (EDA), feature extraction, and data preprocessing on large-scale structured and unstructured datasets.
- Optimized model performance through advanced feature engineering techniques, hyperparameter tuning (using GridSearchCV/RandomizedSearchCV), and model evaluation metrics (AUC-ROC, F1-score, etc.)
- Built dynamic Power BI dashboards and reports to visualize sales trends, customer segmentation, and key performance indicators (KPIs) for a pizza retail client, enabling actionable insights.
- Developed and maintained automated ETL pipelines using Python and SQL to streamline data extraction, transformation, and loading processes, reducing manual reporting effort by over 30%.

07/2022 - 03/2024

Hyderabad

Software Engineer

Persistent Systems Ltd

- Contributed to the development of an automated loan-processing application on the Salesforce platform by integrating third-party APIs for real-time data exchange and validation.
- Built custom Salesforce Lightning Components and Apex Triggers to implement complex business logic and enhance user experience across multiple workflows.
- Developed robust Apex unit tests with over 80% code coverage to ensure production readiness and maintain application reliability.
- Managed deployment workflows by migrating components and test classes across environments (Developer, Sandbox, Production) using Inbound and Outbound Change Sets.
- Optimized backend automation scripts and data processes, resulting in a 25% improvement in overall system performance and processing speed.

EDUCATION

05/2018 - 06/2022

CGPA: 8.22

B. Tech in Electronics & Communication Engineering

Vishnu Institute of Technology

TECHNICAL SKILLS

Python | SQL | Machine Learning | Deep Learning | Flask | Power BI | Tableau | Git/Github | Open AI API

Advanced Excel | unit Testing | Apex Triggers | Salesforce LWC | Langchain

SOFT SKILLS

Adaptability | Analytical Thinking | Data Storytelling | Effective Communication | Time Management

PROJECTS

Sales Effectiveness for FicZon Inc

A data science project focused on enhancing sales performance by classifying customer leads based on quality and conversion potential.

- Extracted and queried sales data from a MySQL database, followed by comprehensive data preprocessing to handle missing values, outliers, and inconsistent formats.
- Performed feature analysis on key attributes such as customer location, delivery mode, lead source, lead status, and sales agent activity to evaluate their impact on lead conversion.
- Built predictive classification models using Decision Tree and Random Forest ensemble techniques to categorize leads into high-potential and low-potential segments.
- Achieved strong model performance with high accuracy and F1-scores, contributing to improved targeting strategies and sales team efficiency.

Tools Used: MySQL, Jupyter Notebook, MS Excel, VS Code
Accuracy Achieved: 98%

Employee Performance Prediction – Inx Future Inc

A machine learning project focused on predicting employee performance ratings based on structured HR data, supporting proactive talent management and decision-making.

- Processed and explored a rich HR dataset containing attributes such as demographics, job role, experience, and satisfaction metrics, with the goal of predicting the performance rating of employees.
- Trained and evaluated multiple machine learning algorithms including Logistic Regression, Decision Tree Classifier, Random Forest, optimizing performance through hyperparameter tuning and model comparison
- Conducted detailed exploratory data analysis (EDA), generating distribution plots, boxplots, and a correlation heatmap to identify key factors influencing employee performance.
- Implemented modular Jupyter Notebooks structured into data processing, model training, predictions, and visualization stages for better project organization and clarity.

Tools Used: Jupyter Notebook, Pandas, Scikit-learn, Seaborn, Matplotlib
Accuracy Achieved: 91%

Employee Salary Prediction

A regression-based machine learning project aimed at estimating employee compensation in the IT industry based on various demographic and professional attributes.

- Gathered and analyzed employee data across different IT domains, focusing on features such as gender, age, experience, domain, and technical skills.
- Developed a Random Forest Regressor to predict salaries, improving model performance through hyperparameter tuning and cross-validation techniques.
- Assessed model effectiveness using regression evaluation metrics including R2_score, Mean Absolute Error (MAE), and Mean Squared Error (MSE), given the continuous nature of the target variable.
- Deployed the trained model into a Flask web application, enabling users to input custom feature values and receive real-time salary predictions as output.

Tools Used: VS Code, Jupyter Notebook, Flask
R2_score Achieved: 88%

Conversational AI Prototype (Internal POC)

Designed and developed a proof-of-concept conversational AI assistant aimed at automating document-based query resolution for internal finance operations

- Built a document-based QA assistant by integrating Langchain with the OpenAI API, enabling context-aware query responses from internal financial documentation.
- Utilized LlamaIndex and Pinecone vector database to implement semantic search across multiple PDF sources, ensuring relevant context retrieval for user queries.
- Designed and tested conversational flow logic using the RASA framework, simulating human-like interactions for internal support and ticketing use cases.
- Leveraged SQL to fetch contextual data from internal databases for dynamic response generation during RAG (RetrievalAugmented Generation) simulations

Tool Used: Langchain, OpenAI API, LlamaIndex, Pinecone, RASA, Python
Accuracy Achieved: Delivered 90–95% contextually accurate responses in prototype testing.

CERTIFICATIONS

Pynum AI Developer Certification
July 2025
Pynum AI Developer Certification

Certified Data Scientist Foundation -- IABAC
June 2025
Data Scientist Foundation - IABAC

Certified Data Scientist – NASSCOM Future Prime Skills
April 2025
Certified Data Scientist - Nasscom

Machine Learning Foundation – Internshala
October 2024
Machine Learning Foundation - Internshala