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 - 05/2025

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   Git   Github   Flask   Power BI   Tableau

Advanced Excel   Salesforce

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%

Rice Leaf Disease Detection

A deep learning project aimed at identifying and classifying common rice leaf diseases through image-based recognition techniques.

- Collected and organized rice leaf images infected with various diseases, splitting the dataset into structured training and testing sets.
- Developed a Convolutional Neural Network (CNN) architecture incorporating layers such as convolution, max pooling, flatten, and dense layers, and evaluated model performance using accuracy and confusion matrix.
- Enhanced model accuracy by applying data augmentation techniques and transfer learning to address the limitations of a small dataset.
- Integrated EfficientNetB0 as a feature extractor and combined it with an XGBoost classifier, further improving classification accuracy through hyperparameter tuning.
  Tools Used: VS code, Jupyter Notebook
  Accuracy Achieved: 97%

Skin Disorder Prediction

A machine learning project focused on predicting the type of skin disorder based on clinical and histopathological features.

- Imported the skin disorder dataset into a Jupyter Notebook environment and performed exploratory data analysis (EDA) using data profiling techniques to understand feature distributions and correlations.
- Observed that all input features, except for age, contained categorical values ranging from 0 to 3, as they represented medical indicators from clinical and histopathological examinations.
- The objective was to classify skin disorders into 6 distinct categories: *Psoriasis*, *Seborrheic Dermatitis*, *Lichen Planus*, *Pityriasis Rosea*, *Chronic Dermatitis*, and *Pityriasis Rubra Pilaris*.
- Built and evaluated classification models using Logistic Regression and Support Vector Machines (SVM), achieving strong accuracy scores in multiclass prediction tasks.
  Tool Used: VS code, Jupyter Notebook, MS Excel
  Accuracy Achieved: 99%

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%

CERTIFICATIONS

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

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

Career Essentials in Generative AI – Microsoft and LinkedIn
August 2024
Career Essentials in Generative AI - Microsoft & Linkedin

Azure Fundamentals (AZ-900) – Microsoft
July 2022
Azure Fundamentals - Microsoft