NEERAJ KUMAR LOKIREDDY

Kansas City, Missouri | +1 816 299 6467| neerajkumarlokireddy@gmail.com |Hackerrank| LinkedIn

SUMMARY

Data Science graduate with a focus on machine learning and hands-on experience in analyzing and modeling data to derive actionable insights. Proficient in Python, data manipulation, and visualization, utilizing tools like Pandas, NumPy, Matplotlib, and Seaborn for data-driven decision-making.

Data Science & Machine Learning: Experienced in building regression, classification, and clustering models using Scikit-learn and other machine learning libraries. Strong understanding of model evaluation, tuning, and validation techniques.

Programming Knowledge: Solid foundation in algorithms, data structures, and OOP, with a focus on applying these principles to real-world data science problems. Passionate about transforming data into valuable insights to drive business strategies.

Data Wrangling & Analysis: Skilled in data cleaning, preprocessing, and feature engineering to prepare raw data for analysis. Proficient in working with large datasets, ensuring data quality and integrity throughout the analysis process.

Tools and IDEs: Experienced with Jupyter Notebook, VS Code, and Git for version control, enabling efficient data exploration, analysis, and collaboration. Comfortable working with cloud platforms like AWS and Azure for deploying data science models.

Data Visualization: Proficient in creating intuitive and informative visualizations using Matplotlib, Seaborn, and Plotly to communicate findings and insights effectively to stakeholders.

TECHNICAL SKILLS

Programming Languages: Python (3.7-3.10)

Frameworks and Libraries: FastAPI, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn

Version Control: Git (Git 2.35)

IDE: VS Code, Eclipse, Jupyter Notebook

Databases: SQL, MySQL 8.0

Concepts: Data Structures, Algorithms, Object-Oriented Programming (OOP), Machine Learning, Regression, Classification, Clustering, Data Wrangling, Model Evaluation

Tools and Platforms: AWS, GitHub, CI/CD, Power BI – Data Visualization, Dashboard Creation, DAX, Data Modeling, Microsoft office(Outlook, PowerPoint, Word and Excel).

Cloud: Amazon AWS

WORK EXPERIENCE

Machine learning intern **Chennai, India**

Virtusa **05/2023 – 11/2023**

- Developed and implemented machine learning models for predicting student success, achieving an accuracy rate of 85% using Scikit-learn.
- Performed data cleaning and preprocessing on large datasets with over 1 million records, improving model performance by 20%.
- Conducted exploratory data analysis (EDA) using Pandas, NumPy, and Matplotlib, uncovering key insights that helped optimize marketing strategies.
- Built interactive data visualizations using Seaborn and Plotly to effectively communicate findings to stakeholders, driving a 15% increase in actionable business decisions.
- Deployed machine learning models on AWS, integrating them into the existing web application for real-time predictions, improving prediction speed by 40%.
- Collaborated with cross-functional teams to identify key business problems and provided data-driven solutions using statistical methods and machine learning techniques.

Machine Learning Consultant (Freelance)

Self-Initiated Collaboration

- Assisted **industry professionals** in solving **real-world machine learning problems**, including **data preprocessing, model selection, and performance optimization**.
- Provided insights on **data cleaning, feature engineering, and hyperparameter tuning** to improve model accuracy.
- Helped troubleshoot **machine learning pipelines** and optimize workflows for efficiency.
- Assisted in integrating **Power BI and Python-based visualizations** for data-driven decision-making.

PROJECTS

Laptop Price Prediction using Machine Learning

- **Regression:** Built predictive models (Linear Regression, Decision Trees, Random Forests) to estimate laptop prices based on features like brand, RAM, storage, and processor type.
- **Classification:** Used Logistic Regression and Random Forest to classify students into "Likely to Succeed" or "At Risk" based on educational and socioeconomic factors.
- **Clustering:** Applied KMeans and Hierarchical Clustering to group laptops into different price-performance categories.
- Evaluated models using R^2 score, RMSE, accuracy, and silhouette scores to select the best-performing techniques.

Technologies: *Python, Scikit-learn, Pandas, NumPy, Matplotlib, Regression, Classification, Clustering*

Obesity Data Analysis and Clustering using Scikit-learn and Pandas

Developed a machine learning pipeline to analyze and classify obesity levels from a dataset.

- Cleaned and preprocessed obesity data using Pandas, handling missing values with statistical imputation.
- Implemented Logistic Regression and K-Nearest Neighbors (KNN) classifiers to predict obesity levels.
- Applied K-Means and Agglomerative Clustering to group data based on obesity-related features.
- Visualized data and model performance using Matplotlib and Seaborn.
- Evaluated clustering quality using silhouette scores to provide insights into obesity trends.

Technologies: *Python, Scikit-learn, Pandas, Clustering, Logistic Regression, KNN, Data Visualization*

COVID-19 Case Fatality Ratio Analysis using PySpark and Plotly

Developed a data analysis pipeline to evaluate COVID-19 case fatality ratios across U.S. states using PySpark for data manipulation and Plotly for geospatial visualization.

- Utilized PySpark to load, clean, and aggregate COVID-19 data from CSV files, focusing on the Province_state and Case_Fatality_Ratio columns.
- Implemented a loop to consolidate data by matching state names and summing case fatality ratios across multiple files.
- Performed sorting and extracted the top 10 U.S. states based on case fatality ratios.
- Leveraged Plotly to visualize the data on a choropleth map, displaying case fatality ratios for each state.
- Demonstrated proficiency in big data handling, visualization, and geographic data representation.

Technologies: *PySpark, Data Aggregation, Plotly, Data Visualization, Geospatial Mapping, Python*

EDUCATION

Master's in computer science with Data science emphasis

UNIVERSITY OF MISSOURI-KANSAS CITY, JAN 2024 – MAY 2025

CERTIFICATIONS

Machine Learning Using Python – APSSDC

Source Code Management Using Git & GitHub – APSS DC

Elite + Silver Badge – NPTEL Programming in Java, by Prof. Debasis Samanta, IIT Kharagpur.

Certificate – C Programming (Spoken Tutorials, IIT Bombay)