BADISAHGANDU RAVINDER

Data Scientist

919347759904 ravinder.badishagandu@gmail.com
https://www.linkedin.com/in/ravinder-badishagandu-b52783291 Hyderabad

SUMMARY

Enthusiastic data science aspirant with a Certification in Master Data Science Program and a strong foundation in statistics, machine learning, and data analysis. Proficient in Python and SQL, with coursework in data mining, predictive modeling, and data visualization. Completed a capstone project on sentiment analysis using natural language processing techniques. Eager to apply my analytical skills and passion for solving complex problems to contribute to a data-driven team's success.

EDUCATION

Electronics and Communication Engineering

Vignana Bharathi Institute of Technology - VBIT,Gahtkesar,Hyderabad.

2011 - 2016

GPA 4.0 / 4.0

Intermediate

Siddartha Junior & Degree College,Bhupalpally

2007 - 2009

GPA 4.0 / 4.0

Secondary School Certificate

AP SW RES School,Ghanpur Station

2006 - 2007

GPA 4.0 / 4.0

CERTIFICATION

Master Data Science Program

GUVIi Geek Networks,IITM Research Park

PROJECTS

YouTube Data Harvesting and Warehousing Project

https://github.com/Ravinder1729/youtube-data-project.git

Project Description:

- Designed and implemented a data harvesting and warehousing system to collect and store YouTube data.
- Utilized SQL for structured data storage, MongoDB for unstructured data, and Streamlit for the user interface.
- Gathered and processed YouTube video metadata, comments, and engagement metrics for analysis

Key Responsibilities and Achievements:

- Writing Python scripts to scrape YouTube data using the YouTube API.
- Developing a database schema using SQL to store structured data efficiently.
- Implementing a NoSQL MongoDB database to handle unstructured data like comments and video descriptions.
- Creating an interactive and user-friendly dashboard using Streamlit for data visualization and exploration.
- Performing data analysis to extract insights and trends from the YouTube data.
- Ensuring data quality, integrity, and security throughout the project.

Technologies Used:

- Technologies: Python, SQL, MongoDB, Streamlit
- API: YouTube API
- Databases: MySQL (or your specific SQL database), MongoDB
- Data Visualization: Streamlit

PROJECTS

PhonePe Pulse Data Visualization Tool

https://github.com/Ravinder1729/phonepulse.py.git

Project Description:

- Developed a user-friendly data visualization and exploration tool using Streamlit and Plotly to analyze PhonePe Pulse data.
- Created interactive dashboards and visualizations to provide insights into transaction trends, user behavior, and financial data.

Key Responsibilities and Achievements:

- Designing and implementing a Streamlit web application for data exploration.
- Using Plotly to generate dynamic and interactive visualizations, such as bar charts, line graphs, and heatmaps.
- Collecting, cleaning, and processing PhonePe Pulse data from various sources.
- Creating filters and user-friendly controls to allow users to customize their data analysis.

Technologies Used:

- Technologies: Python, Streamlit, Plotly
- Data Sources: PhonePe Pulse data
- Data Processing: Pandas, NumPy
- Data Visualization: Plotly, Matplotlib

Breast Cancer Prediction Using Ensemble Techniques

https://github.com/Ravinder1729/ensemble-techniques.git

Project Description:

- Developed a machine learning model to predict breast cancer in patients based on a dataset with key features including radius, texture, perimeter, area, smoothness, compactness, and concavity.
- Employed ensemble techniques such as Decision Trees, Random Forest, AdaBoost, and XGBoost to enhance predictive accuracy and robustness.

Key Responsibilities and Achievements:

- Data preprocessing, including cleaning and feature engineering to prepare the dataset for modeling.
- Implementation of various ensemble techniques and hyperparameter tuning for optimal model performance.
- Evaluation of model performance using metrics such as accuracy, precision, recall, and F1-score.
- Conducting feature importance analysis to identify the most influential predictors of breast cancer.

Technologies Used:

- Machine Learning Libraries: Scikit-learn, XGBoost
- Data Analysis: Python, Pandas, NumPy
- Data Visualization: Matplotlib, Seaborn

E-commerce Customer Segmentation using K-Means Clustering

https://github.com/Ravinder1729/customer_segmentation.git

Project Description:

- Conducted an in-depth analysis of existing customer data in the e-commerce domain to gain valuable insights into purchase patterns and customer behavior.
- Implemented data pre-processing techniques, including missing value treatment, to ensure data quality.
- Utilized K-Means clustering to segment customers into distinct groups based on their purchase behavior, with the optimal number of clusters determined using silhouette score.

Key Responsibilities and Achievements:

- Cleaning and preparing the customer dataset for analysis, addressing missing values and outliers.
- Applying the K-Means algorithm with varying numbers of clusters (k) and evaluating silhouette scores to determine the optimal number of clusters.
- Creating visualizations and summary statistics to communicate the results of the customer segmentation.
- Extracting meaningful insights from the segmented customer groups, such as identifying high-value customer segments or patterns in purchase behavior.

Technologies Used:

- Data Pre-processing: Python, Pandas
- Clustering: K-Means
- Data Visualization: Matplotlib, Seaborn

PROJECTS

Malaria Detection with Convolutional Neural Networks (CNN)

https://github.com/Ravinder1729/Malaria-Detection-Using-CNN.git

Project Description:

- Developed a deep learning model using Convolutional Neural Networks (CNN) for the automated detection of malaria parasites in blood smear images.
- Trained the model on a dataset containing both infected and uninfected blood smear images to achieve high accuracy in malaria diagnosis.

Key Responsibilities and Achievements:

- Data preprocessing, including image resizing, normalization, and augmentation.
- Designing and implementing the CNN architecture, including convolutional layers, pooling layers, and fully connected layers.
- Training the model with appropriate hyperparameters, optimizers, and loss functions.
- Evaluating the model's performance using metrics such as accuracy, precision, recall, and F1-score.
- Visualizing model results and creating a user-friendly interface for malaria detection.

Technologies Used:

- Deep Learning Framework: TensorFlow, Keras
- Computer Vision: OpenCV
- Data Preprocessing: Image processing libraries (e.g., PIL), NumPy
- Deployment: Flask

NLP Spam Classification and Sentiment Analysis with Heroku Deployment

https://github.com/Ravinder1729/nlpmodel_heroku.git

Project Description:

- Developed a web-based application for spam classification and sentiment analysis using Natural Language Processing (NLP) techniques.
- Trained machine learning models to classify text messages as spam or non-spam and to perform sentiment analysis on user-generated text data.
- Deployed the application on Heroku for easy accessibility and usage.

Key Responsibilities and Achievements:

- Data preprocessing and text cleaning to prepare the dataset for analysis.
- Building and fine-tuning NLP models, such as text classification and sentiment analysis models.
- Developing a user-friendly web interface using Flask for text input and displaying analysis results.

Technologies Used:

- NLP Libraries: NLTK, spaCy, Scikit-learn
- Web Framework: Flask/Django
- Deployment Platform: Heroku
- Web Development: HTML

Machine Translation: Hindi to English using Sequence-to-Sequence Models

https://github.com/Ravinder1729/machine-learning-translation.git

Project Description:

- Developed a machine translation system using sequence-to-sequence models to translate text from Hindi to English.
- Implemented encoder and decoder architectures to effectively capture and generate translated sentences.
- Leveraged deep learning techniques and trained the model on a large bilingual dataset.

Key Responsibilities and Achievements:

- Data preprocessing, including tokenization, padding, and creating vocabulary.
- Building and training the sequence-to-sequence model with attention mechanisms.
- Evaluating translation quality using metrics like BLEU score or human evaluation.
- Handling issues related to language-specific challenges and nuances in translation.
- Demonstrating the model's effectiveness by showcasing successful translations and improvements over time.

Technologies Used:

- Deep Learning Framework: TensorFlow, PyTorch
- Natural Language Processing: NLTK, spaCy (for text preprocessing)
- Data Processing: Python, NumPy
- Evaluation Metrics: BLEU score

PROJECTS

Flight Price Prediction and Web Application with Flask

https://github.com/Ravinder1729/Flight-price-prediction.git

Project Description:

- Developed a flight price prediction model using the Random Forest algorithm to forecast airfare based on historical data and relevant features such as departure location, arrival location, airline, and travel dates.
- Created a user-friendly web application using Flask to allow users to input travel details and receive price predictions in real-time.

Key Responsibilities and Achievements:

- Collecting and cleaning flight data, addressing missing values, and transforming features for modeling.
- Implementing Random Forest regression to create a predictive model.
- Evaluating the model's performance using metrics like Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE).
- Fine-tuning model hyperparameters to optimize predictive accuracy.
- Presenting the results, including insights into the factors influencing flight prices.

Technologies Used:

- Machine Learning Libraries: Scikit-learn (for Random Forest), Pandas, NumPy
- Data Preprocessing: Python, Pandas
- Data Visualization: Matplotlib, Seaborn (if applicable)

TECH STACK

Python | **MySQL** | **MongoDb** | **Machine Learning** | **Deep Learning** | **NLP** | **flask** | **Docker** | **Power BI**

STRENGTHS

Problem Solving:

- Demonstrated problem-solving skills through successful projects, such as optimizing inventory management or improving customer satisfaction.
- Ability to identify root causes of complex issues and develop effective solutions.

Adaptability:

- Quickly adapt to new technologies, tools, or methodologies in a fast-paced environment.
- Willingness to take on new challenges and continuously learn and grow.

Communication:

- Strong written and verbal communication skills, demonstrated through clear and concise reporting of findings.
- Ability to present complex technical concepts to non-technical stakeholders.

Technical Skills:

- Proficiency in programming languages such as Python, R, or SQL.
- Expertise in working with data visualization tools Power BI.

LANGUAGES

English **Hindi** **Telugu**