Viraj Kawade

8149976368 | virajkawadevk@gmail.com | Pune,Maharashtra

Summary

Dedicated and analytical recent graduate with a specialization in Big Data Analytics, aiming to start a career as a Data Engineer . Possess robust skills in Machine learning, Data Analysis. Experienced in using Python , SQL, and Tableau for data manipulation and visualization.Dedicated to using data-driven approaches to solve complex problems and enhance business performance.

Projects

Retail Orders ETL Pipeline

- Dataset Acquisition: Utilized Kaggle API to download the retail orders dataset, ensuring timely and efficient data retrieval.
- Data Extraction: Extracted the orders.csv file from a compressed ZIP file using Python's zipfile module.
- Data Cleaning: Handled missing values by setting specific placeholder strings (e.g., 'Not Available', 'unknown') to NaN using pandas.
- Data Transformation: Standardized column names to lowercase and replaced spaces with underscores for consistency and ease of use.
- Feature Engineering: Created new columns to calculate discount, sale price, and profit based on existing columns,enhancing data usability.
- Data Type Conversion: Converted 'order_date' from string format to datetime format to facilitate time-series analysis.
- Column Management: Dropped unnecessary columns (list_price, cost_price, discount_percent) to streamline the dataset.
- SQL Integration: Established a connection to a SQL Server database using SQLAlchemy for data storage and management.
- Data Loading: Loaded the cleaned and transformed data into a SQL Server table with an 'append' option, ensuring efficient
- data handling.
- Documentation and Code Quality: Maintained clear and organized code with comments and modular steps, making the
- process reproducible and understandable.

Build a model to predict Stock market performance using LSTM network

- Objective :To develop a predictive model using LSTM, a type of recurrent neural network (RNN), to forecast stock market performance to predict future price movements.
- The stock market is known for its unpredictability, with prices influenced by various factors such as economic indicators, investors entiment, and global events.
- After Data Collection ,Preprocessing the data was done by handling missing values, scaling features, and splitting the dataset into training and testing sets.
- Feature Engineering involved Identifying relevant features that may influence stock prices, such as historical prices, trading volumes, moving averages, and technical indicators.
- Model Development with LSTM was done using deep learning frameworks. It involed defining the input shape and architecture of the LSTM model, including the number of hidden layers,LSTM units, and activation functions.
- Training the model using the training dataset and optimize hyperparameters through techniques like grid search or random search.
- Model Evaluation was done using appropriate metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).
- Validation of the model's effectiveness was done by testing it on unseen data from the testing dataset.
- The developed LSTM model successfully predicts stock market performance with decent accuracy, providing valuable insights for investors and traders.

Predict Fare of Airlines Tickets using Machine Learning

- Objective: Predicted flight prices using features like duration, stops, airline, and departure time.
- Technologies: Python, Pandas, NumPy, Matplotlib, Seaborn, Plotly, Scikit-learn, Pickle.
- Data Handling: Cleaned and processed dataset, handled missing values, and engineered features.
- Visualization: Created static visualizations with Matplotlib/Seaborn and interactive plots with Plotly.
- Outlier Management: Detected and handled outliers using the IQR method.
- Feature Selection: Identified key features using mutual information regression.
- Model Development: Trained Random Forest and Decision Tree Regressors, achieving R^2 score of 0.808.
- Model Deployment: Saved and loaded models using Pickle for future use.
- Automation: Automated ML pipeline and conducted hyperparameter tuning with RandomizedSearchCV

Skills

Power BI, Tableau, Python, SQL, Machine Learning, Spark, Hadoop, Cloud Computing, Linux

Education

Institute for Advanced Computing & Software Development | Pune
Post Graduate Diploma in Big Data Analytics | 02/2024
64%

D Y Patil Institute of Technology | Pune
07/2019
7.44 CGPA

New English High School & Junior College | Ulhasnagar,Thane
HSC | 05/2015
74%

Adarsh Vidya Mandir | | Badlapur,Thane
SSC | 06/2013
91%

Certificates

Pwc Powerbi Certificate, Mentormind Capstone Project

Hobbies

- Keen observer of Geopolitical Events & Current Affairs.
- Watching Historical documentaries , Business case studies & Podcasts.