Shrikrishna Soni

+917067493944 | shrikrishnasoni111@gmail.com

LINKS

LinkedIn:// shrikrishnasoni
GitHub:// zuno10
Kaggle:// @shrikrishnasoni

SKILLS

Programming Languages
Python(More than 5000 lines), SQL(More than 5000 lines)

Statistical Expertise
Statistical analysis to solve business problems and drive business outcomes.

EDA
Conduct exploratory data analysis (EDA) to uncover patterns, trends, and anomalies in datasets.

Data Manipulation and Cleaning
Pandas, Numpy, Excel, Handling missing data, Feature Engineering, Data Transformation, Normalization

Data Visualization Tools
Matplotlib, Seaborn, PowerBi, Dash

Machine Learning Frameworks
Scikit-learn, PyTorch

Database Management
PostgreSQL, SQLite3

Cloud Platforms
Amazon SageMaker, EC2 Instance, AWS Lambda.

EDUCATION

GOVT. P.G. AUTONOMOUS COLLEGE SATNA

B.Sc. In Computer Science
Jul 2020 | Satna,MP
CGPA : 6.4

EXPERIENCE

CONTECHUB SOLUTIONS | Jr. Data Analyst

jan 2024 - June 2024 | Jaipur

- Led data scraping: Managed a team to ethically harvest marketing data from competitor websites, increasing data volume by 30% while ensuring accuracy.
- Enhanced data quality: Performed data cleaning and validation tasks, ensuring data accuracy and reliability for reporting and analysis.
- Collaborated across teams to understand data needs and provide insights to support strategy development and improve operations.

PROJECTS

YOUTUBE DATA ANALYTICS ON BIG DATA

- Built a data lake by loading multi-regional JSON and CSV files into an AWS S3 bucket. Automated data extraction, transformation, and loading (ETL) processes using AWS Glue, Lambda functions, and Apache Spark to convert raw data into optimized Parquet format for efficient analysis.
- Designed an automated pipeline using S3 event triggers to process new data uploads in real time. Implemented regional partitioning (e.g., US, Germany, France) to enhance query performance and data management.
- Leveraged AWS Athena for querying large datasets and integrated AWS QuickSight to create interactive visualizations, enabling data-driven insights into YouTube performance metrics across different regions.

DELHI HOUSE PRICE PREDICTION

- Developed a predictive model to estimate house prices in Delhi, utilizing a range of machine learning algorithms, including Linear Regression, Random Forest, and XGBoost.
- Cleaned, pre-processed, and analyzed housing data to identify key factors influencing house prices (e.g. location, property size, amenities).
- Compare Multiple Machine Learning Algorithms to find most suitable approach based on evaluation matrices r2 score, MAE, MSE.
- Packaged the model in a Docker container for seamless deployment on cloud platforms.

MOVIE RECOMMENDATION SYSTEM

- Conquered complex movie data through effective wrangling and pre-processing.
- Utilized NLP techniques Bag of Words to transform text data into numerical vectors for content analysis.
- Leveraged cosine similarity to recommend movies similar to user selections.

-