Dipanshu Kakshapati

Data Engineer | Data Analyst

Lalitpur, Nepal ■ +977-9866952349 ■ dipanshu.ksh30@gmail.com ■ LinkedIn ■ Portfolio

Data Engineer and Analyst with over 1 year of hands-on experience in real-time data processing and ETL pipeline construction using Airflow, Snowflake, and AWS. Seeking to drive data-driven strategies effectively in roles like data engineer and data analyst.

EDUCATION

Westcliff University, King's College Nepal **2022-Present**
Bachelor of Information Technology, Major in Data Science GPA 4.0, Honorary Scholarship

SKILLS

- Python
- PySpark
- SQL/MySQL
- Airflow
- Selenium/lxml
- Databricks
- Snowflake
- Docker
- FastAPI
- Data Structure & Algorithms
- GitHub/Bitbucket
- Data Analysis, Visualization
- AWS - EC2, S3, Glue, RedShift, Athena, SNS, RDS, Lambda

WORK EXPERINCE

Grow By Data **May 2024 – Present**
Data Engineer Intern Kathmandu, Nepal

- Developed a FastAPI-based API for real-time and historical NEPSE stock data scraping, enhancing data retrieval efficiency by 30%.
- Engineered an automated ETL pipeline using Airflow on AWS EC2, reducing processing time by 50%.
- Facilitated data analysis by transferring data in Snowflake and Looker Studio, boosting data accessibility for stakeholders.

Naamche **April 2024 – June 2024**
Outreach Automation Expert Kathmandu, Nepal

- Implemented marketing automation workflows that increased campaign efficiency by 40% using Zapier and Customer.io.
- Automated the collection and updating of real estate news by improving data freshness on Google Sheets using the Google Developers API.

Uber **December 2023 – February 2024**
Data Scientist **-** *Certified* India

- Refined a lasso regression model to predict Uber ride fares with an 80% accuracy rate, using a dataset of 200,000 past records.
- Analyzed and interpreted complex data, identifying key fare determinants such as distance, and time.
- Proposed a dynamic pricing strategies based on demand and temporal factors such as traffic conditions.

PROJECTS

ETL Pipeline for NEPSE | *Airflow, AWS EC2, S3, RDS, SQS, Snowflake, Looker Studio* **GitHub**

- Designed and implemented an automated ETL pipeline of NEPSE to scrape the stock data on a daily basis.
- Expanded a FastAPI-based API of NEPSE stock data scraping alongside Selenium to get the stock data from website in real time.
- Used Airflow that was hosted in AWS EC2 to daily processes stock data, upload it to S3 bucket and then transferred to Snowflake for further data analysis all within 2 minutes.

Fare Prediction Model for Uber | *Jupyter Notebook, Lasso Regression, Streamlit, Matplotlib* **Demo**

- Developed a regression model to accurately predict Uber ride fares, utilizing a dataset 200,000 historical ride records.
- Employed Lasso Regression for effective feature selection and regularization and achieved model accuracy of 80%.

Employee Performance Prediction Model for Tech for India | *Jupyter Notebook, LightGBM, Streamlit* **Demo**

- Enhanced a predictive model to predict employee performance, leveraging a a dataset of historical data on employee evaluations.
- Employed LightGBM Regression model with the Standard Scaler and achieved model accuracy of 73%.

More projects and information available at *https://github.com/DipanshuKakshapati*

CERTIFICATES

Associate Data Analyst in SQL - DataCamp, Data Scientist Associate - DataCamp, Data Engineering Essentials - IBM

REFERENCE

Sunney Sharma *susharma@growbydata.com*
Software Engineer at Grow By Data *+977 9861806230*