Dipanshu Kakshapati

Data Engineer Intern

https://www.linkedin.com/in/dipanshukakshapati/
https://dipanshukakshapati.github.io/portfolio/

(+977) 9866952349
dipanshu.ksh30@gmail.com

Work Experience

GrowByData - Data Engineer Intern – Onsite   *Kathmandu, Nepal*   May 2024 – July 2024
- Developed and deployed a FastAPI-based API for real-time and historical NEPSE stock data scraping. This API, hosted on AWS EC2, utilizes a PostgreSQL database, is containerized with Docker, and optimized with Docker Compose and Amazon RDS for efficient management.
- Developed an innovative method to scrape Google search results, distinguishing between paid and organic listings using a JSON snippet to inject an SEO-pixel into HTML. This technique groups results by consistent dimensions such as height and width, enhancing robustness against layout changes.
- Designed and implemented an automated ETL pipeline using Apache Airflow that daily aggregates and processes stock data through a FastAPI service. The data is transformed, analyzed for gains and losses, stored in PostgreSQL, and archived on Amazon S3. This system, hosted on AWS EC2, is efficiently monitored and managed via Airflow's webserver UI.

Naamche - Outreach Automation Expert – Hybrid   *Kathmandu, Nepal*   April 2024 – June 2024
- Engineered and monitored a series of marketing automation workflows utilizing platforms like Zapier, Midjourney, ChatGPT, Customer.io, and Make. These workflows seamlessly convert ideas from Google Sheets into ready-to-publish blog articles, social media content, and images in under two minutes, adhering strictly to the brand's guidelines.
- Automated the daily scraping of real estate news from Florida-based websites with RSS feeds, updating the collected data directly into Google Sheets through the Google Developers API.

Uber – Data Scientist Intern – Remote   *India*   December 2023 – February 2024
Certified by Uber
- Developed a comprehensive 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, enhancing the model's predictive accuracy.

Teach For India – Data Scientist Intern – Remote   *India*   January 2024 – March 2024
Certified by Teach For India
- Engineered a robust predictive analytics model to forecast employee performance, leveraging a dataset comprising extensive historical data on employee evaluations.
- Compiled an in-depth research report delineating the methodologies and efficacy of various employee performance prediction models, underscoring the potential for HR analytics in strategic talent management.

Certifications | Education

Certifications		Education
Data Scientist Associate – DataCamp	*NYC, USA*	***Westcliff University, King's College Nepal***
Associate Data Analyst in SQL – DataCamp	*NYC, USA*	- Bachelor's in Information Technology, Major in Data Science 2022 - 2025

Projects | Skills

Airflow ETL Pipeline for NEPSE   GitHub
- Designed and implemented an automated ETL pipeline using Apache Airflow that daily aggregates and processes stock data through a FastAPI service.

Classification Model for King's College Scholarships   GitHub
- Developed a logistic regression model with Python to identify scholarship candidates for King's College Nepal, focusing on academic excellence, financial need, and resource distribution.

Regression Model for Predicting Uber Ride Fares   GitHub
- Developed a comprehensive regression model to accurately predict Uber ride fares, utilizing a dataset 200,000 historical ride resource.

E-commerce Platforms for Pets with Django.   GitHub
- Developed an e-commerce platform for pets using Django and PostgreSQL, featuring user authentication, product listings, and secure transactions.

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

Skills

Languages
Python, PySpark, SQL, Bash

Tech Stacks
Airflow, Databricks, Snowflake, Docker, FastAPI, Anaconda, JupyterNotebook

Databases
PostgreSQL, MySQL, Microsoft SQL Server

Version Control
Git, GitHub, Bitbucket

Cloud Skills
AWS (EC2, S3, Glue, Athena, SNS, RDS, Lambda