JITESH RAWAT

DATA ENGINEER

jitezzrawat99@gmail.com | +91 9324979351 | Linkedin | Portfolio | Github

SKILLS

- Google Cloud Platform
- Python
- SQL
- Spark
- AWS
- Big Query
- Snowflake
- ETL/ELT
- Data Science

EDUCATION

Master of Science (M.Sc)
Data Science and Computing

Sri Sathya Sai Institute Of Higher Learning

2020 - 2022
GPA: 7.8

Bachelor of Computer Applications (BCA)
Sri Sathya Sai Institute Of Higher Learning

2017-2020

COURSE

Data Engineering on Google Cloud Agilitics - Virtual Classroom

- This course offered hands-on experience in designing and implementing data processing systems on Google Cloud, utilizing services like Cloud SQL, BigQuery, DataProc, Pub/Sub, and DataFlow.
- It focused on data ingestion, processing, and analytics, as well as leveraging BigTable and AutoML for advanced solutions

PAPER PUBLICATION

- Published a paper titled "Comparative Analysis of Stacked Models and other Machine Learning Algorithms in Estimating Claim Cost for Automobile Insurance" in the International Journal of Emerging Technologies and Innovative Research.

WORK EXPERIENCE

Data Engineer

Atidiv

October 2022 - current

- Orchestrated ETL and ELT pipelines to extract and load API data into BigQuery and Snowflake for both batch and streaming analytics.
- Designed Tap-Target pipelines using ELT methodology, leveraging Google Cloud Platform's Compute Engine for virtual machine setup and Cloud Composer (Airflow) for automated scheduling.
- Led the development of Reverse ETL pipelines, proactively resolving pipeline issues, and implementing comprehensive rebuilds to ensure efficient, standardized workflows. Utilized Google Cloud tools to optimize performance, scalability, and reliability, ensuring robust data pipeline operations across multiple cloud environments.

Data science Intern

Tech Actuarial

Jan 2022 - Apr 2022

- Led a transformative initiative in Tech Actuarial Research, utilizing advanced Pyspark proficiency to implement stacked models, markedly improving precision in estimating claim costs for extensive auto insurance datasets, showcasing strong analytical and technical capabilities in driving impactful data science solutions.

PROJECTS

Amazon Product Review Analysis.

https://amazon-reviews-dashboard-11.streamlit.app/

Search for any keyword and the data for that is scraped from Amazon. Designed and executed a data engineering project focused on the systematic extraction and analysis of product reviews and ratings from the Amazon e-commerce platform, culminating in the creation of an interactive sample product review dashboard.

Estimating claims cost using Stacked models and Comparative analysis with individual Machine learning techniques.

Research based and Actuarial domain specific project where regression analysis and comparison based on individual machine learning models and stacked models in Pyspark for estimation of auto insurance claims.
Technology used : Python, Big data, Pyspark. Dataset : Crash Analysis System (New Zealand)

YouTube Data Analysis and Visualization Pipeline.

Developed a data pipeline to analyze YouTube channel performance by scraping video metrics (views, likes, comments) and transforming data with PySpark. Created an interactive dashboard using Plotly Dash to visualize key insights on channel growth and engagement, automating reporting for content creators. Tech Stack: PySpark, Python, Dash, YouTube API.