

SANTHOSH KUMAR SURESH KUMAR

santhoshkumarsures@lewisu.edu | +1 (779) 390-8922 | [LinkedIn](#) | [GitHub](#) | [Website](#)

SUMMARY

Aspiring Data Engineer with experience as a Data Engineer Intern, skilled in developing automation scripts and advanced sentiment analysis programs. Successfully reduced manual preprocessing time by 50% and improved data quality through innovative solutions. Eager to leverage data engineering expertise to drive efficient data integration and support strategic decision-making.

EXPERIENCE

Data Engineer Intern

Nov 2024 - Present

Novitium Pharma, USA

- Developed an automation script using Selenium to retrieve pharmaceutical sales and marketing data from sources like Clarivate and the Orange Book. The script performed data preprocessing, removing missing values and encoding categorical data using OneHotEncoder, significantly reducing manual preprocessing time and effort by 50%, enhancing team productivity
- Implemented web scraping with SOAP, Selenium, and Requests for real-time data gathering. Improved data quality and analysis efficiency through preprocessing. Created a growth and sales dashboard offering comprehensive insights into product performance and sales over several years.

Data Engineer Intern

Feb 2024 - April 2024

Office of Commissioner of Police Chennai, India

- Created a sentiment analysis program to study Twitter data and identify customer complaints, reducing manual testing by 80%. Previously, sentiment analysis required significant time and effort to read and classify tweets.
- Implemented web scraping with SOAP, Selenium, and Requests for real-time data extraction. Used the MBART language model for translating non-English tweets and employed the RoBERTa base model for sentiment analysis. Segregated complaints using sentiment scores, enhancing targeted response strategies, and incorporated plotting functionalities to track complaint management metrics and trends.

Data Scientist Intern

Jun 2023 - Dec 2023

Tamil Nadu Civil Supply and Corporation, India

- Conducted comprehensive analysis of sales and procurement data to uncover business trends and provide actionable insights. Utilized statistical techniques and machine learning algorithms to identify purchase behaviour patterns and market dynamics, optimizing inventory management and enhancing operational efficiency.
- Used statistical methods and machine learning for trend analysis, developing predictive models for forecasting sales trends and aiding inventory management. Analyzed data to inform strategic decisions, such as maintaining optimal buffer stocks. Created detailed reports and visualizations to communicate findings and recommendations, which supported decision-making and stock management.

SKILLS

Programming: Python, R, HTML5, CSS, JS

Data Base: MySQL, PostgreSQL, MongoDB, PostgreSQL, Apache Hive, Hadoop, DBMS, ETL

Analytics: Power BI, Tableau, Gephi, Looker, Microsoft Office, Machine Learning

EDUCATION

Master of Science in Data Science

Lewis University, Romeoville, IL

Expected: May 2026

GPA: 4.0

Bachelor of Technology in Artificial Intelligence and Data Science

Panimalar Engineering College, Chennai, India

May 2024

GPA: 9.16/10

PROJECTS

- AI Integration In Combat Vehicles:** Developed a security and surveillance system with autonomous robotic capabilities, real-time threat detection, and intelligent decision-making algorithms. The system includes an autonomous robot car for patrolling, face detection for identifying intruders, laser targeting, manual control options, image processing for threat detection, and GPS technology for localization.
- Uber Data Analytics:** Analyzed the "NYC 2023" dataset on Uber's New York City operations to enhance public transportation accessibility. Leveraged Google Cloud Platform (GCP) tools to uncover insights for improving transit solutions, reducing traffic congestion, and enhancing quality of life. Focused on increasing ridership and user satisfaction for a sustainable urban ecosystem.

HONORS AND CERTIFICATIONS

Certifications: Microsoft Power Platform fundamentals, Microsoft Azure AI fundamentals, Microsoft Azure fundamentals, Preparing Data for Analysis with Microsoft Excel, Harnessing the Power of Data with Power BI, Data Analysis and Visualization with Power BI

Publications:

- Optimizing Healthcare Delivery with Seamless Integration of Clinical Decision Support EHR Systems (doi.org/10.1063/5.0255988)
- Revolutionizing Transportation Insights using GCP Services in Uber's Data Analytics ([click](#)).