SANTHOSH KUMAR SURESH KUMAR

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

SUMMARY

Data science expert with proficiency in Power BI, Tableau, Apache Hive, Hadoop, DBMS, SQL and MS Excel. Skilled in developing and implementing end-to-end data analytical solutions, establishing data science standards, and conducting statistical analysis. Published multiple research papers. Experienced in creating databases, data collection systems, and analytics strategies, with a focus on interpreting data and identifying business trends to support faster decision-making.

EXPERIENCE

Data Engineer Intern Feb 2024 - April 2024

Office of Commissioner of Police Chennai, Tamil Nadu, India

Overview: Developed an advanced sentiment analysis program to study Twitter data and identify customer complaints, reducing the need for manual testing by 80%. Previously, sentiment analysis was done manually, requiring significant time and effort to read and classify tweets.

- **Automated Data Extraction:** Implemented web scraping techniques combining SOAP, Selenium, and Requests to gather tweets in real-time.
- **Language Translation:** Utilized the MBART language model to translate non-English tweets into English for standardized analysis.
- **Sentiment Analysis:** Employed RoBERTa base model to evaluate sentiment polarity and intensity.
- **Complaint Prioritization:** Segregated complaints using sentiment scores, enhancing targeted response strategies.
- **Visualization:** Incorporated plotting functionalities to track complaint management metrics and trends.

Data Scientist Intern Jun 2023 - Dec 2023

Tamil Nadu Civil Supply and Corporation, Chennai, Tamil Nadu, India

Overview: Conducted a comprehensive analysis of large datasets, focusing on sales and procurement data, to uncover business trends and provide actionable insights. Leveraged statistical techniques and machine learning algorithms to identify purchase behaviour patterns and market dynamics. This analysis optimized inventory management and enhanced operational efficiency.

- **Trend Analysis:** Used statistical techniques and machine learning algorithms to identify patterns and trends within big data, particularly in purchase behaviours and market dynamics.
- **Predictive Analytics:** Developed predictive models to forecast future sales trends, aiding in proactive decision-making and inventory management.
- **Data-Driven Insights:** Analyzed sales and procurement data to uncover patterns that informed strategic decisions, such as maintaining optimal buffer stocks.
- **Reporting:** Created detailed reports and visualizations to communicate findings and recommendations to stakeholders, helping the company make better decisions and efficiently manage stock levels.

SKILLS

Software: Power BI, Tableau, Gephi, Jupiter Notebook, Looker, Pycharm, Apache Hive, Microsoft Office, PostgreSQL, Unity
Computer Languages: SQL, Python, R, HTML, CSS, JS
Competencies: Data Analytics, Data Visualization, Machine Learning, Optimisation, Business Data Analysis, Data Mining, Data Science for Application, Big Data, Deep Learning, DBMS
Soft Skills: Communication, Adaptability, Curiosity, Critical Thinking, Problem Solving, Team Working, Research

EDUCATION

Master of Science in Data Science Expected: May 2026
Lewis University, Romeoville, IL GPA: 4.0

Bachelor of Technology in Artificial Intelligence and Data Science May 2024
Panimalar Engineering College, Chennai, India GPA: 9.16/10

PROJECTS

- **AI Integration In Combat Vehicles**: Developed a security and surveillance system integrating Raspberry Pi 4B and Arduino UNO, enhancing surveillance in restricted areas 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 for precise intervention, manual control options, an 8MP Pi Camera for high-accuracy detection, servo motors for laser aiming, ultrasonic sensors and image processing for threat detection, and GPS technology for localization**.**
  Tools Utilized*: Python, Arduino UNO*

- **Uber Data Analytics**: Analyzed the "NYC 2023" dataset containing data on Uber's operations in New York City to enhance public transportation accessibility and utilization. Leveraged Google Cloud Platform (GCP) tools to uncover insights that can improve transit solutions, reduce traffic congestion, and enhance quality of life. By focusing on increasing ridership and user satisfaction, the project contributes to a more sustainable urban ecosystem.
  Tools Utilized*: Python, Google Cloud Platform (GCP), BigQuery, Looker Studio*

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 (click).
  Integrating clinical decision support (CDS) with electronic health record (EHR) systems can improve healthcare efficiency and patient outcomes. Benefits include better care coordination, increased effectiveness, and fewer errors. Challenges involve high implementation costs and the need for clinician training. Addressing these requires a clear plan and clinician involvement. Despite the difficulties, integration is recommended for its positive impact on patient care and healthcare delivery. This work was published at the PECTEAM IConIc (IEEE) - 2023 conference.

- Revolutionizing Transportation Insights using GCP Services in Uber's Data Analytics (click).
  Analyzed the "NYC 2023" dataset using Google Cloud Platform (GCP) tools to improve Uber's operations. Data was managed in Google Cloud Storage, processed with Mage, and transformed in BigQuery using SQL queries. Looker created interactive dashboards and reports. Emphasis was placed on security, compliance, and scalability to support Uber's growth and enhance decision-making and operational efficiency. This work was published at the PECTEAM IConIc (IEEE) - 2023 conference.