

Chitra Lekha Sura

chitra22lekhas@gmail.com | +1 (425)-375-1786 | Seattle, WA | [linkedin](#) | [portfolio](#)

EDUCATION

Stevens Institute of Technology	Master's of Science (MS)	Computer Science	Sep 2022– Dec 2023
Osmania University	Bachelor's(BE)	Computer Science	July 2018– May 2022

EXPERIENCE

Intone Networks | Software Engineer - Data operations March 2024 – present

- Designed and optimized **ETL pipelines** using **Python, SQL**, and Apache Spark, enabling efficient data ingestion, transformation, and storage for analytics.
- Built interactive **dashboards** and reports with **Power BI and Tableau**, providing actionable insights to support business decision-making and **predictive sales** models achieving a 20% improvement in sales **forecasting**, helping the business optimize inventory.
- Implemented **data models, star schemas**, to streamline querying and enhance performance in Data Warehouses and Data Lakes
- Automated data quality checks, to work with accurate data for regulatory and operational use, using **PostgreSQL and Python**.
- Leveraged **Apache Hadoop, Apache Spark**, and **Kafka** in a production environment to process and analyze large-scale datasets, enabling real-time data streaming, batch processing, and optimized analytics for business-critical applications and decision-making.

Verzeo EdTech Pvt. Ltd. | Machine Learning Intern March 2020 – April 2020

- Analyzed 17K+ gaming app records using **Python** and **SQL**, extracting insights on genre, pricing, and ratings to support product strategy.
- Built **Decision Tree** and **Random Forest models** to predict game success and estimate billings, achieving 85% accuracy and reducing revenue prediction error by 12%.
- Identified key pricing correlations, including a 30% relationship between game cost and success, influencing pricing strategies.

PROJECTS

Data Analysis of COVID-19(2020&2021) world dataset:- (Data Analysis using various tools)

- Conducted **data exploration** and analysis on **Apache Spark** handling over 500 types of data inconsistencies, revealing a 2.5% global infection rate, and highlighting a 50% reduction in new cases post-vaccination.
- Developed interactive **Tableau** dashboards and **calculated fields** to visualize trends in infection rates, vaccination coverage.
- Identified a 0.75 correlation between mobility and infection rates using Python**, influencing lockdown policy recommendations.

Image Captioning:- (Deep Learning and partial NLP implementation)

- Integrated **CNNs** and **RNNs** in a **transformer model** for **image captioning**, achieving 92% accuracy in generating contextually-rich descriptions. Utilized an **Inception V3 module**, outperforming **ResNet-50** by 4% in **feature extraction**.
- Optimized **feature extraction** to reduce image disturbances during **CNN encoding**, resulting in a 3x reduction in model training time.
- Leveraged **TensorFlow, NLTK**, **token embeddings, attention, LSTM** and evaluated using **BLEU scores**.

Data Analysis of “Cyclistic - bike share”:- (Complete Data Analysis and forecasting using Python)

- Executed **analysis** on **normalized, standardized and preprocessed** Cyclistic bike-share app's data set of 5 million records, utilizing **Python libraries (Pandas, NumPy)** to eliminate 99% of **outliers** and ensure data integrity and feature engineering.
- Performed **geospatial analysis** by mapping 100,000+ latitude and longitude points using **Folium**, on real-time maps to visualize ride patterns.
- Generated **visualizations**, histograms and **time series graphs(SARIMA model)**, to **forecast trends** and insights that drove data-informed decisions

Predictive Analysis of Job Change Intentions:- (Machine Learning models comparative study)

- Designed and implemented **ML models (Random Forest, Gradient Boosting, SVM)**, achieving 92% accuracy in predicting job change intent among 10,000+ candidates.
- Deployed models on **Azure Machine Learning** for cloud-based inference and optimized Gradient Boosting models, improving prediction precision by 15%.

CERTIFICATIONS

Databricks GenAI fundamentals, Data Analytics & visualisation job simulation, Google Data Analytics, Cisco Cybersecurity essentials

SKILLS AND TOOLS

- Machine Learning libraries: TensorFlow, Keras, PyTorch, Scikit-Learn, pandas, PySpark, NLTK, XGBoost
- Languages & Tools: Python, Java, SQL, R || PostgreSQL, Tableau, PowerBI, Hadoop, Excel, Apache Spark, Azure ML