

## SUMMARY

Data-driven Analyst and ML specialist with 3+ years of experience in risk prevention and distributed data engineering. Expertise in building high-scale NLP and sentiment analysis pipelines using AWS (EMR, SageMaker), SQL, and SparkR. Proven ability to identify geographical and technical risk clusters, automate ETL workflows, and deliver live insights to integrity teams.

## EXPERIENCE

### **Atmos Energy / Data Analyst / Dallas, TX / Aug 2023 – Present**

- Spearheaded a cross-functional risk-mitigation initiative with Engineering and Compliance for 50+ KPIs, ensuring 100% alignment with regulatory safety standards and efficiency of operation in pipeline safety & integrity.
- Engineered SQL-queries to extract and analyze large-scale data across distributed systems, translating complex analytical insights into automated Tableau dashboards for all Pipeline Integrity teams; streamlined risk reporting and improved data-driven decision-making speed by 200%.
- Developed an R&D predictive model (XGBoost, Deep Learning) to predict unlocatable pipeline coordinates on a Large-scale GIS dataset, significantly reducing safety vulnerabilities and improving risk-prevention accuracy.
- Automated end-to-end ETL pipelines in Python to resolve data discrepancies across business units, maintaining high data integrity for data large scale digital data migration and storage. regarding system integrity and operational safety.
- Translated largescale data of 10 years of various initiatives for PHMSA annual safety reporting into executive-level Tableau dashboards, enabling leadership to make data-driven decisions. Performed large scale and distributed SQL data extraction.

### **CCC Intelligent Solutions | Data Scientist Intern | Chicago, IL | May 2022 – Aug 2022**

- Optimized a production-grade privacy-preserving ML system using SISA unlearning, reducing retraining costs by 70% while maintaining 95.6% model accuracy.
- Conducted complex trade-off analyses between accuracy and latency across varying request rates, informing scalable strategies for model updates under strict privacy and policy constraints.
- Identified non-linear performance behaviors (U-shaped trends) in high-volume data deletion scenarios to harden system reliability against edge-case vulnerabilities.

### **Visteon Corporation / Machine Learning Intern / Pune, India / Jul 2019 – Sep 2019**

- Integrated real-time object detection (PoseNet) into ADAS systems using C++, improving model efficiency and generating automated risk-based alerts for user safety.
- Collaborated cross-functionally with Engineering and Product teams to deploy ML models, ensuring technical requirements aligned with user-facing safety features.

## PROJECTS & VOLUNTEERING

### **Enterprisewide Key Performance Indicators (KPI)**

*Present*

- Engineered complex SQL queries to extract millions of rows of live data across distributed systems, identifying geographical clusters of risk and anomalies within the Texas pipeline network to drive proactive mitigation.
- Spearheaded geospatial analysis using ESRI shapefile mapping to visualize high-risk corrosion signals, providing the primary technical foundation for anomaly detection and infrastructure hardening.
- Developed and published live Tableau dashboards for 6 distinct Integrity teams (Risk, Compliance, Operations, etc.), translating 50+ KPIs into actionable insights used for daily operational planning and real-time decision-making.
- Managed the end-to-end data lifecycle for millions of records, resolving discrepancies and ensuring data integrity to support safety-critical project management and budget forecasting.

### **Data Analyst Volunteering**

*2022*

- Perform Data Mining, Data cleaning, Data collection, developing models, Validation, Visualization and Perform Gap analysis
- responsible for gathering requirements from business analysts and operational analysts.
- Work with cross functional teams and translate business requirements into SQL queries to answer business questions
- Clean and analyse data using Pandas and involved in analysis, design and documenting business requirements and data specifications.
- Utilized Power BI to create various analytical dashboards that helps business users to get quick insight of the data
- Designed and implemented multiple dashboards using Power BI - PowerPivot & Power Query tools for in house metrics.

## Cloud Based Sentiment Analysis & NLP Pipeline

2021

- Engineered a distributed data processing pipeline on AWS EMR using SparkR to analyze millions of rows of IMDb text data, optimizing cluster performance via custom Bootstrap scripts and maximized resource allocation.
- Developed and compared multiple ML models (Random Forest, Logistic Regression, SGD) for high-volume sentiment classification, achieving a peak accuracy of 84% and a 21% improvement over baseline regression models.
- Implemented NLP techniques for signal detection, cleaning and preprocessing large-scale unstructured text to extract features for model training in AWS SageMaker.
- Hardened model reliability by diagnosing and mitigating overfitting through L2 regularization and hyperparameter tuning, increasing validation accuracy from 0.63 to 0.81.
- Managed the end-to-end data lifecycle from S3-based storage and ETL to production-grade evaluation using Recall and F-score metrics to ensure model integrity.

## EDUCATION

Illinois Institute of Technology | Master's in Data Science | Jan 2021 – Dec 2022

## CERTIFICATIONS

- **Google** Data Analytics Professional Certificate
- **ESRI** essential workloads

## SKILLS

**Languages & Tools:** Python, R, SQL, C++, Github, Jupyter, Unix/Linux, ArcGIS Pro.

**Machine Learning:** Regression, Classification, Convolutional Neural Networks (CNN), Deep Learning, XGBoost, SVM, K-Nearest Neighbour (KNN), Random Forest Regression and classification, Transfer Learning.

**Libraries & Tools:** Tensorflow, Keras, PyTorch, Scikit-Learn, GeoPandas, Numpy, Pandas.

**Analytical & Data Eng:** Data Cleaning, Feature Engineering, ETL, Statistical Analysis, A/B Testing, Hypothesis testing.

**Data Visualization:** Tableau, Power BI, Matplotlib, Seaborn

**Cloud & Infrastructure:** Snowflake, BigQuery, AWS

**Soft Skills:** Stakeholder Collaboration, Cross-Functional Communication, Project Management, Strategic Thinking, Critical thinking