SATHWICK KIRAN M S

Contact: +1 (716) 359-1432 | sathwickkiran04@gmail.com | LinkedIn | GitHub | Portfolio | Buffalo, NY

EDUCATION

University at Buffalo – State University of New York   Buffalo, New York, US
Master of Science in Data Science   *Aug 2024 – Dec 2025*

B.M.S. College of Engineering   Bengaluru, Karnataka, India
Bachelor of Technology in Computer Science and Engineering   *Jul 2018 – May 2022*

SKILLS & Certifications

Certifications: Microsoft Azure Data Science Associate (DP-100), AWS Cloud Practitioner, Snowflake Data Warehousing, Databricks Fundamentals, CompTIA, Python, Gitlab, Certification in Cybersecurity from University of Maryland, Linux CLI.
Programming: Python, R, SQL, C++, Java, Scala
CloudMLOps: AWS, Azure, Terraform, Docker, Kubernetes, CI/CD, Hadoop
Tools: PowerBI, Tableau, GitHub, GitLab, MATLAB, Databricks
Scripting & Automation: Shell, YAML, Git

WORK EXPERIENCE

Sony India Software Centre | Software Engineer| Bengaluru, India   Aug 2022 – Apr 2024
- Developed and supported scalable data pipelines using Apache Spark, Hadoop, and AWS, enabling reliable processing of structured and unstructured datasets.
- Automated ETL workflows in Python and SQL for data ingestion, cleaning, and transformation, reducing manual effort by 25% and improving pipeline efficiency.
- Deployed containerized big data solutions with Docker and Kubernetes, ensuring scalability, high availability, and fault tolerance in production.
- Designed and implemented analytical data models and performed feature engineering to support ML-driven business use cases and reporting requirements.
- Built and maintained Power BI dashboards with drill-down capabilities, translating complex datasets into actionable insights for decision-making.
- Applied statistical and predictive modeling (Python: Pandas, NumPy, Scikit-learn) to uncover patterns and trends that informed strategic business outcomes.
- Established automated quality checks and monitoring by developing Python and Shell linters, ensuring robust and consistent data pipelines.
- Led Agile stand-ups and sprint planning to coordinate across engineering and business teams, resolving blockers and aligning on deliverables.
- Recognized with 3rd place in Sony's "First Challenge" for building a document automation platform, showcasing innovation in data-driven solutions.

PROJECTS

Osteoporosis Risk Prediction using Statistical and Machine Learning Methods
- Cleaned and preprocessed over one million patient records (handling missing values, outliers, and categorical transformations) to ensure high-quality input for analysis.
- Applied ML models (Random Forest, Logistic Regression, Decision Tree) in Python and R, achieving ~83% accuracy in predicting osteoporosis risk.
- Conducted survey-style exploratory data analysis using visualizations (histograms, count plots) to identify key risk factors such as age, nutrition, and activity levels.
- Summarized findings into interpretable visuals, highlighting risk distributions and key demographic patterns for clear communication.
- Collaborated and presented findings under the guidance of Professor Jianzhen Liu.

Advanced Data Analysis and Scalable Big Data Processing with Hadoop and NLP Models
- Designed and deployed a **scalable big data pipeline** using **Hadoop and Spark** to process over **3 million Amazon book reviews**, efficiently handling unstructured text data.
- Developed **sentiment analysis models** (Logistic Regression, SVM, Random Forest, Naïve Bayes) utilizing **TF-IDF vectorization** and **hyperparameter tuning**, evaluated by **F1-score and accuracy**.
- Conducted **feature importance analysis** with Random Forest and SHAP values to identify key predictors like sentiment polarity, review length, and keywords.
- Performed **error analysis** to uncover misclassification patterns caused by linguistic nuances such as sarcasm, ambiguity, and class imbalance.
- Analyzed sentiment trends across book genres and temporal shifts, visualizing insights through **heatmaps and Power BI dashboards**.
- Leveraged sentiment analysis outputs to provide **actionable insights** for refining product offerings, improving customer strategies, and optimizing marketing campaigns.