SYLVIA BODDU

14511 Valor Circle, Tampa FL 33613 | +1 (812) 803-5227 | satyaratnab@gmail.com | LinkedIn | GitHub

EDUCATION

MASTER OF SCIENCE IN DATA SCIENCE, GPA: 3.8/4.0 — **Aug 2021 – May 2023**
Indiana University, Bloomington, Indiana

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING, GPA: 3.9/4.0 — **Jul 2012 – May 2016**
GITAM University, Visakhapatnam, India

TECHNICAL SKILLS

Programming Languages: Python (NumPy, pandas, scikit-Learn, TensorFlow, NLTK, Seaborn), R, Shell
Database and ETL Tools: MongoDB, PostgreSQL, NoSQL, T-SQL, Elasticsearch, Hive, AWS-S3, Azure, Splunk, Snowflake, Docker, Spark, Kafka, Airflow, Tableau, Logstash, Kubernetes, Hadoop, Yarn, MapReduce, Informatica, HDFS, PowerBI, Git
Data Science Skills: Data Modeling, Data Preprocessing, Exploratory Data Analysis, PCA, t-SNE, Classification, Regression, SVM, Random Forests, K-means Clustering, Lemmatizing, Stemming, GNN, GAN, LLM
Statistics: Hypothesis Testing, t-test, Z-test, A/B Testing, Bayesian Statistics, Confidence Intervals, and p-value

WORK EXPERIENCE

Indiana University, Bloomington, Indiana — **Sep 2023 – Present**
AI/ML Researcher

- Developed a memory chatbot using Datastax, AstraDB, and LangFlow to handle emergency customer care inquiries by tracking SessionIDs to have multiple conversations.
- Implemented LoRA from scratch on a single GPU to fine-tune large language models, reducing computational demand while maintaining high model performance.
- Crafted a transformation model for corneal images using OpenCV to optimize image quality for medical diagnostic tools, increasing efficiency in healthcare diagnostics.
- Designed a real-time data ingestion and processing pipeline with Kafka, Spark, and Hadoop, elevating analytics and decision-making capabilities across multiple platforms.

Kyndryl (spinoff of IBM) — **May 2022 – Dec 2022**
Data scientist

- Utilized Python's Scikit-Learn to design and implement Support Vector Machine model to classify abnormal server data, increasing uptime by 23% via anomaly detection.
- Leveraged Azure Synapse Analytics to implement Spark pools, processing and analyzing terabytes of data daily that significantly improved query performance.
- Designed monitoring and alerting mechanisms within Apache Airflow to detect failures and automatically notify stakeholders, enhancing system reliability and efficiency.
- Deployed 7+ Azure and Informatica data pipelines, facilitating seamless ETL processes and integrating both structured and unstructured data sources via Hadoop/HDFS.
- Implemented MapReduce jobs and employed Yarn within the Hadoop to optimize data sorting and aggregation, leading to significant reduction in processing times.
- Performed comprehensive tuning of PostgreSQL databases and optimized NoSQL configurations, boosting query speed and throughput in high-traffic scenarios.

IBM GTS Data & AI Services Team — **Jan 2020 – Aug 2021**
Data Scientist

- Engineered a hybrid NLP model combining LSTM and Word2Vec techniques in TensorFlow to enhance categorization of incident tickets by 35% for a global customer.
- Applied Python's Scikit-Learn for predictive modeling, and developed a time-series model to accurately forecast sales for the next quarter, based on historical data trends.
- Crafted an ITSM analytics module for a healthcare client using ELK stack, optimizing the data storage and extraction processes which increased system performance.
- Created user friendly PowerBI dashboard for the healthcare client, which allowed for enhanced interaction with ITSM analytics data, aiding in better decision-making.
- Directed a team to revamp data lake architecture, integrating Kafka, Spark, and Airflow to enhance data extraction and feature engineering processes, which improved KPI accuracy by 10% and increased revenue by 17%.

Juniper Networks Pvt. Ltd — **Sep 2016 – Oct 2019**
Data Engineer

- Built a custom Grafana plugin for dynamic dashboards with Python, facilitating time series analysis for business solutions that led to 1.5x faster troubleshooting process.
- Managed a suite of microservices-based data processing applications, deployed using Docker and Kubernetes which streamlined operations and saved 40 hours weekly.
- Automated ETL pipelines with AWS Lambda and Glue to optimize data storage and access on S3 and Athena, reducing costs through efficient data lifecycle management.
- Developed and maintained centralized data storage solutions on Amazon Redshift, ensuring data integrity and schema consistency for advanced analytical processing.
- Proposed and implemented strategies to optimize data flow within architecture using Python, reducing operations time by 27% and enhancing business insight extraction.
- Owned MX104 series, implementing testing and configuration enhancements to cut manual efforts by 48%, demonstrating strong communication, interpersonal skills.

Juniper Networks Pvt. Ltd — **May 2015 – Jul 2016**
Data Engineer Intern

- Established scalable data pipelines via Amazon Kinesis, AWS Glue, and AWS Data Pipeline, ingesting and transforming large-scale data in both real-time and batch modes.
- Employed Python scripts with Pandas and Seaborn for statistical analysis and visualization, enhancing accuracy and depth of analytical reports presented to stakeholders.
- Refined SQL queries in PostgreSQL to manage and analyze large datasets and visualized results using Tableau to streamline reporting processes across departments.
- Improved data flows with Spark for faster data processing and Kafka for reliable data streaming, maintaining scalable data infrastructure capable of handling large datasets.

Juniper Networks Pvt. Ltd — **May 2014 – Aug 2014**
Data Engineer Intern

- Applied Python and Pandas for data cleaning and exploratory data analysis (EDA), identifying correlations and dependencies in customer datasets for predictive modeling.
- Automated data pipeline processes including data cleansing and transformation using Kubernetes and Python to improve data quality and ensure timely data ingestion..

INDIVIDUAL PROJECTS

DATA MINING

- Constructed an LSTM model to accurately predict short-term stock trends, achieving 79.90% accuracy, thereby enhancing decision-making tools for investors.
- Used advanced machine learning techniques to improve predictive accuracy and support investment strategies, directly impacting financial outcomes.

DEEP LEARNING

- Created a hybrid 1D-2D CNN and GAN network to colorize black and white images, elevating user interaction by providing visually appealing content with 77% accuracy.
- Implemented cutting-edge neural networks to transform grayscale images into color, significantly improving visual content and user experience.

DATA VISUALIZATION

- Built a web-scraping and Tableau-driven visualization tool for tracking pandemic spread, offering vital insights into infection patterns and aiding in public health decisions.
- Illustrated and analyzed COVID-19 dissemination by developing dynamic visual tools, augmenting understanding and enabling better management of health resources.

GRAPH ANALYTICS

- Designed and implemented a Graph Neural Network (GNN) to refine social e-commerce recommendations, achieving a significant 18% increase in group buying revenue.
- Leveraged graph analytics to develop predictive models that boost group-purchasing by tailoring recommendations, increasing sales efficiency on e-commerce platforms.

CERTIFICATIONS

- **Microsoft Azure Fundamentals AZ-900**
- **Microsoft Azure Data Scientist DP-100 - In progress**