SAITEJ CHOWDARY BODAPATI

New Jersey,NJ ☎862-214-0815 @ sb2793@njit.edu https://www.linkedin.com/in/tejchowdary

EDUCATION

- **M.S Data Science- NJIT-Ying Wu College of Computing** **2023-2024**
  Rel. Course: Artificial Intelligence, Machine Learning, Reinforcement Learning, Big Data, Cloud Computing, Deep Learning, Stats Methods in Data Science, Statistics, Python & Mathematics, Database Management System
- **B-tech Electrical and Communication Engineering- Gandhi Institute of Technology and Management** **2016-2020**
  Rel. Course: C, C++, JAVA, DBMS, Total Quality Mgmt., Statistics, VLSI.

PROFESSIONAL SUMMARY

Artificial Intelligence and Data Engineer with 3+ years of experience in developing scalable ML solutions, backend systems, and analytics pipelines. Skilled in Python, Java, AWS, and Power BI with a strong foundation in machine learning, cloud platforms, and business intelligence. Proven ability to translate data into actionable insights, build predictive models, and deliver end-to-end AI solutions across cross-functional teams.

EXPERIENCE

Data Integrator – Intelli data Systems Pvt Ltd | Hyderabad, India **March 2021 – March 2023.**

Key Responsibilities & Achievements:

- Led data integration initiatives across clinical and supply chain datasets, working with structured and unstructured sources including Excel, PDFs, SAP exports, and SQL databases.
- Created machine learning–ready datasets by automating ETL pipelines and designed Power BI dashboards to support predictive analytics in supply chain workflows.
- Streamlined and curated data for analytical models by building pipelines to clean, transform, and standardize data using Python (Pandas, NumPy) and SQL, improving data consistency by 98%.
- Worked cross-functionally with Product Managers, Clinical Operations, Finance, and Procurement teams to gather data requirements and translate business needs into technical solutions.
- Applied metadata tagging and ontology mapping for data governance across product development lifecycle datasets, ensuring alignment with internal compliance and regulatory reporting standards.
- Built automated validation scripts that detected data duplication, missing fields, and schema mismatches across clinical supply records.
- Maintained version control and documentation for data transformation logic using Git, and authored process documentation to support knowledge transfer across teams.
- Conducted training sessions for non-technical teams on data access, SAP exports, and data quality checks to support self-service analytics.
- Designed Excel-based data summaries and transitioned them to Power BI for automated stakeholder dashboards, improving decision turnaround by 40%.
- Collaborated with IT and Data Engineering to enhance data discovery and sharing platforms using internal tools and dashboards.

WIPRO TECHNOLOGIES | Bangalore, India **April 2020 - August 2021.**

- Led the **backend development** of an internal **interview scheduling platform** using **Java (Spring Boot)**, **MySQL**, and **AWS EC2/S3**, resulting in a **50% improvement in scheduling speed** and reducing manual coordination overhead.
- Developed and integrated **REST APIs** to streamline data communication between services; optimized backend services for scalability and latency using efficient data structures and caching strategies.
- **Collaborated with cross-functional teams** (HR Tech, Cloud Ops, and Delivery) to gather business requirements, turning them into scalable backend features while ensuring **data integrity and security**.
- Developed a prototype AI-based recommendation engine for employee onboarding using collaborative filtering and Python.
- **Integrated SAP modules** to retrieve employee onboarding data and map with internal HR applications; built automation scripts using **Python** and **SAP BAPI/RFC APIs** to reduce data pull latency and human dependency.
- Participated in the **cloud infrastructure design** for internal projects; provisioned AWS resources (S3, EC2, Lambda) and contributed to **CI/CD pipelines** using GitHub and AWS CodePipeline.

- Worked on **data flow pipelines**, preparing datasets from HR systems, **SAP exports**, and third-party scheduling tools, using **Python** and **SQL** to clean and load data into dashboards for internal analytics.
- Created internal Power BI dashboards for HR analytics by connecting SAP exports and MySQL views. Delivered insights into onboarding, attrition, and project allocation metrics used by delivery teams.
- Supported a team of junior engineers, conducted **code reviews**, and maintained detailed project documentation using **Confluence and JIRA**.
- Developed an internal **recommendation engine** prototype using **collaborative filtering** in Python for learning suggestions during employee onboarding.

Data Science Intern – Nexus AI Labs **Sep2019 – Jan 2020**

- Built a PySpark-based data pipeline on AWS EMR to process over 8 million e-commerce records, optimizing data ingestion into AWS S3 and PostgreSQL for scalable storage and analysis.
- Developed XGBoost and Neural Network models to predict customer churn (**Recall = 87%, AUC = 0.91**).
- Automated model retraining with Airflow and containerized the stack with Docker; tracked experiments using MLflow. Created real-time monitoring dashboards in Streamlit. monitoring dashboards in Streamlit, deployed on EC2, and used SHAP for model interpretability.
- Validated model outcomes with A/B testing and statistical analysis to support marketing retention campaigns.

AARKIC Technologies – INTERNSHIP **June 2018 - August 2019.**

- Internship focused on Static RAM Design and Synthesis.
- Highlighted the importance of system reliability, especially in critical applications such as medical, imaging defense communications, and computing sciences.
- Emphasized the role of SRAM subsystems in storing and retrieving information.
- Explored techniques to enhance the fault-tolerant capability of memory subsystems.
- Utilized Electronic Design Automation (EDA) tools such as DSCH and Micro wind effectively for the study

Cloud-Based Sensor Analytics – Freelance / Academic Project **(2024 – Present)**

- Simulated edge-node and sensor data collection for industrial environments using IoT device logs and streaming data ingestion tools.
- Designed scalable ETL pipelines using PySpark and Hadoop; deployed on AWS EMR clusters.
- Built dashboards and insights from structured and unstructured data for anomaly detection and root-cause analysis.
- Practiced end-to-end data engineering lifecycle from ingestion to visualization, aligned with Micron's real-world data transformation strategies.

Technical Skills

Big Data & Pipelines:	Spark (PySpark), Hadoop, AWS EMR, ETL workflows, Airflow
Cloud & DevOps:	AWS (EC2, S3, Lambda, EMR), Git, Terraform, GitHub Actions, CodePipeline, Docker
Programming & Scripting:	Python, Pandas, NumPy, Java, SQL, Bash, C, C++
Databases & Storage:	PostgreSQL, MySQL, Redshift, Snowflake, SAP HANA (integration), SQL Optimization
Reporting & Visualization:	Power BI, Tableau (basic), Streamlit, Microsoft Excel
ML & Data Science:	Scikit-learn, XGBoost, SHAP, TensorFlow (basic), MLflow, Neural Networks, SVM, Logistic Regression, Random Forest
Tools & Platforms:	JIRA, Confluence, Parquet, JSON, CSV, SAP (BAPI/RFC)

PROJECTS

Exploration of Advanced Deep Learning Models

- Conducted a comparative analysis of methodologies for fruit image classification using CNN, VGG16, and InceptionV3 architectures. Leveraged transfer learning techniques to enhance model generalization and performance.

Achievement of High Classification Accuracy.

- Optimized InceptionV3 to achieve a classification accuracy of 94%, outperforming VGG16 (85%) and a custom CNN (70%), demonstrating the effectiveness of advanced transfer learning strategies.

Developed and Implemented Logistic Regression Model

- Designed a logistic regression model to classify wine quality, involving meticulous data preparation, including feature standardization with StandardScaler, and dataset splitting into training and testing subsets to ensure robust evaluation.

Analyzed and Preprocessed Large-Scale Data

- Preprocessed the wine quality dataset using advanced Python libraries such as Pandas, NumPy, and Scikit-learn, ensuring clean and standardized input for machine learning models while addressing real-world data variability challenges.

Predictive Modeling for Marketing Optimization

- Applied ML to optimize term deposit campaigns, simulating real-world financial marketing use cases — aligning closely with Xtramile's business-focused AI solutions.
- Designed and implemented machine learning models,in XGBoost, Logistic Regression, Random Forest, SVM,Neural Networks, to predict term deposit subscriptions with a focus on imbalanced datasets. Achieved AUC-ROC scores above 85% and highest recall of 82% using XGBoost.

Data-Driven Insights

- Conducted exploratory data analysis on a dataset of 45,211 client records, leveraging 16 features such as client demographics, call details, and campaign aggregates to optimize model input and improve interpretability using SHAP values and permutation importance.

Model Evaluation and Hyperparameter Tuning

- Enhanced model performance through hyperparameter optimization (grid/randomized search) and evaluated results using precision, recall, F1-score, and ROC metrics, achieving significant improvements in resource allocation for marketing campaigns.
- Aimed to help correct the vision of those that are affected by various eye diseases like age-related macular degeneration and retinitis pigmentosa.

Cloud-Based Distributed Machine Learning Solution

- Implemented a distributed machine learning pipeline on AWS using EMR, Spark, and S3 for scalable data processing and model training.
- Developed and containerized a machine learning application using Docker, enabling scalable deployment across multiple EC2 instances.