Sowjanya Bojja

LinkedIn • GitHub • Website • LLM Portfolio • Medium
2201 Masters Way, Alpharetta, GA, 30005
+1 6692974674 | simplysowj@gmail.com

Professional Summary:

Results-driven AI/ML Engineer and Data Scientist with 2+ years of hands-on experience in developing and deploying machine learning, deep learning, and Generative AI solutions, complemented by 5 years of specialized experience in Oracle database systems and ETL processes. Proven ability to bridge advanced analytics with robust data infrastructure to deliver end-to-end solutions.

AI/ML & Data Science Expertise:
- Proficient in Python, PySpark, TensorFlow, and PyTorch for building predictive models and AI solutions
- Specialized in LLMs, NLP, and computer vision applications with production deployment experience
- Hands-on with Generative AI implementations including chatbot development and AI-powered analytics
- Experienced in full data science lifecycle from exploratory analysis to model deployment
- Skilled in developing interactive dashboards and data visualizations for business insights

Database & Data Engineering:
- 5 years of Oracle database expertise at Tata TeleServices Limited
- Designed and optimized ETL processes for large-scale datasets
- Developed complex PL/SQL procedures, functions, and packages
- Implemented data validation and quality checks for critical business systems
- Experience with data migration and system integration projects

Key Expertise:
- Generative AI & Advanced Analytics:
  - Integrated OpenAI's LLMs for geospatial intelligence and chatbot interactions.
  - Developed GAN-based data augmentation pipelines to enhance fraud detection accuracy.
  - Engineered OCR-enhanced image captioning systems using TensorFlow, improving efficiency by 50%.
- Data Engineering & ETL:
  - Proficient in pandas, SQL (PostgreSQL, MySQL, Oracle), and NoSQL (MongoDB, Neo4j) for data extraction and analysis.
  - Automated ETL workflows, reducing manual reporting efforts by 30%.
  - Collaborated with data teams to collect and analyze web-based data using advanced web scraping techniques.
- Machine Learning & Model Deployment:
  - Developed and deployed production-grade ML models using Docker and MLflow for lifecycle management.
  - Built real-time fraud detection systems with LSTM-Autoencoders, PySpark, and Kafka.
  - Designed predictive models (XGBoost, Random Forest) with 85%+ accuracy for attrition and sales forecasting.
- Cloud & Full-Stack AI Development:
  - Leveraged AWS (SageMaker, S3) and Azure (CLI, ACR, MySQL) for scalable cloud deployments.
  - Architected full-stack AI applications using Django (backend), React (frontend), and REST APIs (DRF) with JWT authentication.
  - Containerized solutions using Docker for portability and reproducibility.
- Data Visualization & Business Intelligence:
  - Created interactive dashboards (Power BI, Tableau, Looker Studio) to visualize KPIs, sales trends, and customer behavior.
  - Delivered insights that drove 20% revenue growth for clients.
- Database Optimization & ETL:
  - Led database optimization and ETL processes for large datasets at Tata TeleServices Limited, using PL/SQL, TOAD, and Forms/Reports 6i.
  - Developed service order functionality as an alternative to Metasolv Order Management System.
  - Designed CRM service requests for GSM systems, improving operational efficiency.
- Scalable Database Solutions:
  - At Electronic Arts (India), developed efficient PL/SQL procedures, functions, and packages following best practices.
  - Implemented dynamic SQL and business rules for scalable database solutions.
  - Supported users in troubleshooting data errors and system issues.
- Certified in Generative AI (DeepLearning.AI) and Data Science, with expertise in MLOps and CI/CD pipelines.

Education

Master of Science in Data Science
Chandigarh University 3.5
Bachelor of Technology in Electronics and Communications
Jawaharlal Nehru Tech University

Certifications:

- Microsoft Azure AI/ML Engineering from Coursera
- Google Data Analytics from Coursera
- Data Analysis and Visualization with Power BI – Microsoft from Coursera
- Generative AI for Data Science – DeepLearning.AI
- AWS Technical Essentials – AWS Skill Builder
- Statistics for Data Science – Coursera
- Data Visualization with Tableau, SQL, Advanced Excel – Internshala
- Multi-Agent AI Systems with CrewAI and Autogen and DeepLearning.AI
- GenAI for Data Analytics from Coursera
- Data Science, IBM Registered Business Partner (Innomatics Research Institute)

Experience

AI Developer | Realcoderz PVT Limited
Location Finder (with Generative AI)

Tech Stack: Python, Django, React, PostgreSQL, TensorFlow, OpenCV, Geopy, Generative AI, REST API, JWT Authentication

Developed a Full-Stack AI-Powered Location Finder

- **Generative AI Integration**: Integrated **Generative AI (LLMs)** to provide contextual insights based on images uploaded by users, leveraging AI models to generate descriptions, historical context, and relevant facts related to the image and its location, enhancing user interaction and experience.
- **GPS Metadata Extraction & Reverse Geocoding**: Implemented a system for **extracting GPS metadata** from images using **Exif** (Exchangeable Image File Format) data. Employed **Geopy** for **reverse geocoding**, converting GPS coordinates into readable location data (such as cities or landmarks), providing automatic location detection.
- **Deep Learning-Based Image Similarity**: Built a **deep learning model** using `ResNet50` to compare images based on visual similarity. This enables the app to identify images from the same location or with similar visual features by comparing feature vectors extracted through a pre-trained neural network, enhancing searchability.
- **Frontend Development with React.js**: Developed a **React.js frontend** with intuitive **image upload** functionality. Implemented **role-based authentication** for **SuperAdmin** and **Admin** users, ensuring secure access and management. Admins and SuperAdmins have distinct control over platform functionality, enhancing the platform's security and usability.
- **Backend Development with Django**: Created a robust **Django backend** to handle image processing, storage, and metadata management. Implemented integration with **AWS S3** to store images in the cloud, ensuring scalability and reliability. Used **PostgreSQL** for metadata storage, enabling fast retrieval of image-related location information.
- **Google Cloud Vision API Integration**: Integrated **Google Cloud Vision API** to analyze uploaded images and automatically extract key **location data** (such as identifying landmarks or text within the image). This added intelligence to the system, enabling automated location identification and contextually relevant insights.
- **Distance Calculation & Data Visualization**: Designed and implemented **distance calculation algorithms** to determine the distance between target locations based on image metadata. Developed **interactive dashboards** to visualize the distance between locations, providing users with a geographical overview of the images they upload, adding a layer of interactivity.
- **OpenAI Integration for Contextual Interaction**: Incorporated **OpenAI's GPT models** to create an intelligent **chatbot interface**, allowing users to query the application for additional insights about uploaded images or locations. The AI responds with contextually relevant information, enriching the user experience with conversational features.
- **Azure Cloud Deployment**: Optimized the deployment of the entire solution using **Azure Cloud** for **scalable AI services**. Leveraged **Azure App Services** and **Azure Functions** for serverless computing, ensuring the system could scale efficiently to handle increasing traffic and processing demands. This allowed for seamless image processing and fast response times even under heavy load.
- **Live Project Link**: https://salmon-rock-0b662ee0f.6.azurestaticapps.net/

Outcome: Automated geolocation-based image analysis, improving accuracy and efficiency in location identification and similarity detection.

Analytics ETL Project – Realcoderz PVT Limited

Technologies: Django, React, PostgreSQL, OpenAI, Azure, Docker

Developed Full-Stack Data Management and Analytics Platform

- **Django Backend & React Frontend**: Built a robust **Django backend** integrated with a **React frontend** to handle user interactions and display data visualizations. Utilized **PostgreSQL** as the database to store and manage structured data efficiently, ensuring high-performance data retrieval and storage for large datasets.
- **Data Ingestion and Management**: Designed a system to **ingest and manage data** from multiple sources, including **Excel files, custom scripts, and user-submitted forms**, streamlining the process of data collection. Incorporated **Django Admin** to facilitate easy data entry and management, allowing non-technical users to interact with the system.
- **REST APIs with Django Rest Framework (DRF)**: Developed **RESTful APIs** using **Django Rest Framework (DRF)** to handle **CRUD operations** (Create, Read, Update, Delete) on the data, allowing for smooth interaction between the frontend and backend. These APIs were used to process user data and enable seamless data analysis operations.
- **OpenAI Integration for Visual Analytics**: Integrated **OpenAI's GPT models** to generate **visual analytics and graphs** based on user queries. Users can ask for specific insights from the data (e.g., trends, anomalies) and receive automatic graphical representations (such as bar charts, line graphs, or pie charts) to better understand key metrics.
- **Automated Data Processing & ETL**: Developed **automated data processing scripts** that perform **periodic updates** to the database, ensuring that the data remains fresh and accurate over time. The **ETL (Extract, Transform, Load)** pipeline ensures efficient data flow from external sources into the database while maintaining integrity and security.
- **Data Visualization & Exploratory Data Analysis (EDA)**: Utilized **Pandas** and **NumPy** for **Exploratory Data Analysis (EDA)**, analyzing the dataset for patterns, correlations, and outliers. Created a variety of data visualizations (e.g., histograms, scatter plots) to help users understand and explore data trends effectively.
- **Business Analytics Dashboards**: Developed **interactive dashboards** to track key business metrics such as **revenue**, **profit trends**, and the **top-performing businesses**. These dashboards allow users to monitor the health of their business through clear, actionable insights, improving decision-making capabilities.
- **Data Integrity, Validation, and Security**: Ensured the integrity and security of the data within the ETL pipeline by implementing rigorous **data validation** processes. This prevents erroneous data from being uploaded, while safeguarding sensitive information from unauthorized access.
- **Azure Cloud Deployment**: Deployed the entire project on **Azure Cloud**, utilizing **CLI (Command Line Interface)** and **Azure Container Registry (ACR)** for containerized deployment using **Docker**. Employed **Azure MySQL** to manage the database in the cloud, ensuring scalability and availability for the growing user base.
- **Performance and Scalability Optimization**: Leveraged Azure's cloud resources to ensure the platform could scale efficiently with increasing data and user traffic, providing reliable performance for the backend, frontend, and data processing tasks.
  Live Project Link: [View the live application here](#)

Automated Email Agent with Gmail Auto-Reply Integration – Realcoderz PVT Limited

Technologies Used
- **Backend**: Python, Django, Gmail API, OpenAI GPT-3.5, AutoGen, OAuth 2.0
- **Frontend**: React, Vite, Axios
- **Data Storage**: PostgreSQL (for storing campaigns, email content, and replies)
- **Security**: OAuth 2.0, environment-based secrets management

Built an intelligent email automation system using Django and React to streamline personalized bulk email campaigns from CSV-uploaded recipient data. The system supports dynamic content generation, automated sending, and smart reply handling—all powered by OpenAI GPT-3.5 and Gmail API integrations.

Backend Development (Django)
- **CSV-based Campaign Processing**: Developed a robust Django backend to ingest CSV files containing recipient email addresses. Validates, parses, and stores campaign data for downstream email generation and delivery.
- **AI-Generated Email Content**: Integrated **AutoGen** with **OpenAI GPT-3.5** to dynamically generate personalized email subjects and bodies for each recipient based on campaign context or user prompts.
- **Email Sending via Gmail API**: Implemented secure email dispatch using the **Gmail API**, with **OAuth 2.0** for authentication and token management. Ensured rate-limiting and error handling for bulk email use cases.
- **Smart Auto-Reply System**: Built an intelligent reply-handling mechanism that reads inbox messages via Gmail API, identifies relevant replies, and uses **GPT-3.5** to craft contextual responses.
- **Real-Time Tracking**: Designed Django models to track email status (sent, delivered, replied), associate replies with campaigns, and store AI-generated content and responses for transparency and insights.

Key Components
- **AutoGenEmailGenerator**: Encapsulates OpenAI prompt engineering and response parsing for structured, high-quality subject/body generation tailored to campaign goals.
- **GmailService**: Manages email dispatching, inbox scanning, OAuth token refresh, and integration with Gmail API endpoints.
- **ReplyHandler**: Detects relevant user replies, extracts context, and triggers the generation of AI-based responses using GPT-3.5, ensuring conversational continuity and professionalism.

Frontend (React + Vite)
Built a modern React interface allowing users to:
- Upload recipient lists via CSV.
- Input email prompts or templates.
- Monitor campaign progress and replies in real-time.
- Preview AI-generated content before sending.

[Live app link](#)

Bank Transaction Fraud Detection (Real-time Anomaly Detection and Response System)(Banking Domain)

Technologies Used: Python, Apache Kafka, Apache Spark, Zeppelin, MySQL, Docker, LSTM-Autoencoder, KMeans Clustering, NLP, Large Language Models (LLMs).

Responsibilities:

- **Developed a real-time fraud detection system** leveraging **LSTM-Autoencoder** for anomaly detection and **KMeans clustering** for suspicious transaction classification, ensuring enhanced financial security.
- **Built a Kafka-based streaming pipeline** to ingest, process, and analyze high-velocity bank transactions in real time, enabling early fraud detection.
- **Implemented Spark-based big data processing** to handle large-scale transaction datasets efficiently, ensuring low-latency anomaly detection.
- **Designed and configured Apache Zeppelin** as an interactive environment for exploratory data analysis, machine learning model training, and fraud detection visualization.
- **Integrated MySQL** as a backend database to store and retrieve transaction data, ensuring fast and reliable access for fraud investigation teams.
- **Developed an LLM-powered chatbot** to provide real-time fraud insights and transaction validation support for banking teams, improving decision-making.
- **Deployed the solution** using Dockerized microservices, ensuring modularity, scalability, and ease of maintenance in production environments.
- **Conducted extensive feature engineering**, analyzing transaction attributes such as timestamp, account ID, merchant, transaction type, location, and monetary value to identify fraudulent patterns.
- **Implemented dynamic anomaly scoring techniques** to prioritize fraud cases based on severity and historical transaction behaviors.
- **Optimized the fraud detection pipeline**, reducing false positives by 30% and improving real-time transaction monitoring efficiency.

Data Science and Data Analyst
| Remote (for 1 year)// Jan 2021 - Nov 2022
Innomatics Research Labs

Responsibilities:

- **Demonstrated proficiency in Python problem-solving** on Hacker Rank, showcasing strong coding and algorithmic skills.
- **Developed applications** using **FLASK** and **Streamlit**, building efficient and user-friendly web applications.
- Applied **expertise in statistics** and **machine learning (ML) algorithms** with **hyperparameter tuning** to optimize model performance.
- Utilized **advanced techniques** such as **BERT vectorization** for natural language processing (NLP) tasks.
- Showcased competence in **experiment tracking** and **model management** using **MLflow** to ensure model reproducibility and version control.
- Leveraged **workflow orchestration** with **Prefect** to automate and streamline ML pipelines.
- Gained hands-on experience in **NLP** tasks and model deployment.
- Successfully **deployed models to the cloud** using platforms like **Heroku** and **GitHub** for seamless integration and accessibility.
- **Received a Letter of Recommendation** for outstanding performance, demonstrating strong technical and problem-solving abilities.

Data Science Experience
Machine Learning & Predictive Modeling

- Designed and implemented classification and regression models to solve complex business problems.
- Built a Random Forest model achieving 85% accuracy for employee attrition prediction, identifying key factors influencing employee turnover.
- Applied XGBoost and Decision Trees to predict business trends and optimize resource allocation.
  Natural Language Processing (NLP) & Text Analytics
- Leveraged NLTK, spaCy, and Python for sentiment analysis and Named Entity Recognition (NER) to extract insights from textual data.
- Developed text preprocessing pipelines (stemming, lemmatization, stop word removal) to enhance NLP model accuracy.
- Built an automated text classification system to categorize customer feedback, improving response time by 30%.

Big Data & Scalable Processing
- Conducted Exploratory Data Analysis (EDA) on large-scale datasets using PySpark and SQL, extracting meaningful insights.
- Designed a data pipeline for a food delivery business using Databricks, PySpark, and Python, enabling automated analytics and real-time reporting.

Model Evaluation & Optimization
- Performed hyperparameter tuning using Grid SearchCV and Randomized SearchCV to optimize model performance.
- Assessed model performance using ROC-AUC, confusion matrix, precision-recall curves, and F1-score to ensure robust predictions.

Recognition & Achievements

- Received a Letter of Recommendation for outstanding performance in applying machine learning models to real-world problems.

Tools & Technologies: Python (Pandas, Scikit-learn, NLTK, spaCy), PySpark, SQL, MLflow, Databricks

---

Data Analytics Experience

Sales & Business Data Analysis
- Developed an interactive dashboard in Google Looker Studio to analyze sales data for a retail company, helping management track category-wise sales trends and top-performing products.
- Processed and transformed sales transaction data using SQL and Python, extracting key insights into customer purchase behavior, order patterns, and payment preferences.
- Identified seasonal demand trends and provided actionable insights to optimize inventory management.

Financial & Revenue Analytics

- Built an interactive Power BI dashboard to track revenue performance across multiple categories, monitoring property performance and realization percentage.
- Enabled the Revenue team to boost market share and revenue by 20% within a month by providing real-time data-driven recommendations.
- Developed financial reports using MS Excel, Power BI, and Tableau, supporting executive decision-making.

Big Data Analytics & Visualization
- Analyzed large datasets using Databricks, PySpark, and SQL, creating Tableau dashboards to visualize insights for business intelligence.
- Conducted customer segmentation analysis using K-Means clustering, identifying high-value customers for targeted marketing campaigns.
- Designed interactive visual reports using Seaborn, Matplotlib, and Tableau, enabling quick data-driven decision-making.

Exploratory Data Analysis (EDA) & Reporting
- Conducted EDA on massive datasets to identify anomalies, trends, and key business drivers.
- Automated data extraction, transformation, and loading (ETL) processes to streamline reporting workflows.
- Built custom SQL queries to aggregate and filter data for deep-dive analysis on business performance metrics.

Tools & Technologies: SQL, Python (Pandas, Matplotlib, Seaborn), Power BI, Tableau, Google Looker Studio, MS Excel, Databricks, PySpark

Tata Consultancy Services:
Role: System Engineer
Duration: Jan 2008 - Jan 2014 (6 years 1 month)
Location: India (Onsite)

Project: SCDE (Supply Chain Data Exchange)
Client: HP PALM (SCDE) Bengaluru, Karnataka, India
Tools: Toad 7.6.0.11, UNIX
Skills: Oracle Database

Business Overview
SCDE (Supply Chain Data Exchange) is an Oracle Database/Application that supports PPS (Production Planning and Scheduling) supply chain activities. The main function of SCDE is to provide a central location where data can be loaded and accessed by reporting and planning applications. This environment facilitates the loading of data from many sources, applying business-required transformations, and then allowing both ad hoc and standardized access to the data.

Responsibilities

- Wrote several procedures related to the SCDE application.
- Made changes to interface procedures.
- Wrote stored procedures to migrate data from one database to another by invoking procedures via DB Link.
- Extensively involved in SQL tuning, troubleshooting, debugging, and fixing the codes/scripts.
- Prepared and executed SQL and PL/SQL scripts to fix data integrity issues.
- Developed database procedures, functions, packages, and triggers for enforcing business rules using PL/SQL and SQL in TOAD.
- Assisted users in solving system problems due to data errors or system failures.
- Prepared and executed SQL*LOADER, DATA EXPORT, and DATA IMPORT to transfer data from the production database to the development database.
- Debugged and enhanced existing applications.
- Gained experience in Dynamic SQL, PL/SQL Collections, and Exception handling.
- Created database tables, indexes, triggers, stored procedures, functions, packages, synonyms, etc., for new and existing applications.

Role: Oracle APPS Technical Consultant

Location: Bengaluru, Karnataka, India
Client: Electronic Arts (EA)
Description:
Electronic Arts, Inc. (EA) is a major American developer, marketer, publisher, and distributor of video games, one of the largest video game publishers in the world.
Responsibilities:
- Analyzed, designed, and developed software applications using Oracle 9i as a back-end and Oracle Forms and Reports 6i as a front-end.
- Made changes in the EA Order Hold Report and custom forms, and modified EA interface procedures.
- Learned and exhibited the complete Order Management (OM) flow as per project requirements.
- Created database tables, indexes, triggers, stored procedures, functions, packages, synonyms, etc. for new and existing applications.
- Tuned queries for Forms and Reports applications using TOAD.
- Wrote stored procedures to migrate data between databases via DB Link.
- Involved in SQL tuning, troubleshooting, debugging, and fixing codes/scripts.
- Prepared and executed SQL and PL/SQL scripts to fix data integrity.
- Developed database procedures, functions, packages, and triggers to enforce business rules using PL/SQL and SQL.
- Assisted users in solving system problems due to data errors or system failures.
- Prepared and executed SQL*LOADER, DATA EXPORT, and DATA IMPORT from the production database to the development database.
- Debugged and enhanced existing applications.
- Gained experience in Dynamic SQL, PL/SQL Collections, and Exception handling.

Role: Oracle APPS Technical Consultant

Location: Hyderabad, Telangana, India
Client: Tata Tele Services Limited - India
Key Functional Areas:

- Development of service order functionality (Alternative solution for Metasolv Order Management System).
- Development of CRM Service Requests for GSM system.

Tools: TOAD 7.6.0.11, PL/SQL Developer, Forms 6i, Reports 6i
Responsibilities:

- Wrote PL/SQL statements, stored procedures, packages, and functions for various service requests following best practices for QA and implementation standards.
- Used SQL*LOADER to upload SR-related information into the database tables.
- Documented developed applications for better future understanding.
- Learned a wide range of business knowledge required for projects and interacted with the client to gather requirements.
- Created and executed test cases, ensuring bugs were logged and tracked through closure.
- Developed database objects like stored procedures, functions, packages, cursors, ref cursors, and triggers.
- Responsible for sharing application knowledge with team members.
- Involved in performance tuning using advanced PL/SQL concepts like PL/SQL tables, collections, and dynamic SQL.
- Conducted end-to-end testing of applications and error causal analysis.
- Identified and fixed security holes in PL/SQL coding, packages, and control files of SQL*LOADER.
- Extensively analyzed security vulnerabilities and resolved them in packages and SQL scripts.
- Wrote complex SQL queries using joins, subqueries, and inline views to retrieve data from the database.
- Tuned SQL queries using indexes.