MANDIRA SAWKAR

Rochester, NY, USA 14623 | mandirasawkar@gmail.com | +1 (585) 303 2973 | Linkedin.com/in/mandira-sawkar

EDUCATION

Rochester Institute of Technology, New York, USA　　　　　　　　　　　　**May 2026 (Expected)**
Master of Science in Artificial Intelligence　　　　　　　　　　　　　　　　**GPA: 3.70 / 4.0**

IIIT – Hyderabad, Hyderabad, India　　　　　　　　　　　　　　　　　　　**Jun 2023**
Post Graduate Certification in Software Engineering for Data Science with *Distinction*　　**GPA: 3.84 / 4.0**

Indira Gandhi National Open University, Delhi, India　　　　　　　　　　　**Dec 2020**
Bachelor of Computer Applications with First Class　　　　　　　　　　　　　**GPA: 3.07 / 4.0**

PROFESSIONAL EXPERIENCE

Machine Learning Engineer | **Kloud9.nyc, Bengaluru, India**　　　　　　　**Apr 2023 - Jul 2024**
- Led Feature Engineering: Completed feature engineering and validation phases for 50+ features of a machine learning model for ghost inventory prediction at a major variety store chain, achieving 92% accuracy, contributing to 10% reduction in stock discrepancies.
- Data Analysis & Model Expansion: Analyzed ~10 datasets with multimillion data points using PySpark and Pandas to identify key statistical patterns, trends, and anomalies. Enhanced the existing ML model's generalization, resulting in a 15% improvement in prediction stability across diverse store environments.
- Automation & Efficiency: Implemented a data analysis and insights pipeline to optimize the insights workflow, decreasing processing time from 3 hours to 20 minutes per batch.
- Cross-Functional Collaboration: Collaborated with data engineers and business stakeholders to align model objectives with business goals, streamlining the analysis pipeline and reducing deployment time by ~80%.
- Training & Mentorship: Trained and mentored 5+ interns in feature engineering, PySpark, and ML best practices, leading to 3 successful full-time conversions.

Software Engineer | **WSP Global, Bengaluru, India**　　　　　　　　　　　**Oct 2020 – Apr 2023**
- Developed CLI tool using PowerShell and ProjectWise to automate creation and management of employee records. Utilized by Digital Operations team for induction of ~60K new employees, saving ~2.5 human-months per year.
- Automated pipeline to detect project archival status using PowerShell, PowerAutomate and SharePoint. Developed a monthly workflow to notify project-owners and automatically archive projects. Led to 71% reduction in Digital Operations team's maintenance overhead.
- Designed and implemented a Python application to maintain integrity in employee and project databases. Employed Pandas DataFrames to detect duplicates and inconsistencies across ~80,000 employee records per month.

Visa Official | **Consulate General of France, Bengaluru, India**　　　　　**Mar 2020 – Sep 2020**
- Authored 5 analysis reports on Covid-19 (w.r.t economy, education, transport), distributed to the French community of 3 states (1000+ persons); Also processed ~50 visas per day.

Management Intern | **Campus France India, Bengaluru, India**　　　　　　**May 2018 – Oct 2018**
- Counselled and interviewed ~100 students for higher education. Organized 6 education fairs at 10 institutions while also training interns.

TECHNICAL SKILLS

- Programming Languages - Python, R, C, C++, PowerShell, Java, HTML, CSS, JavaScript
- Software Development Tools – Google Cloud Platform, Docker, Flask, Swagger, MongoDB, Git
- Data Science and Machine Learning Tools - SparkML, SciKit-learn, SparkSQL, MySQL, PySpark, Dask, Pandas, NumPy, seaborn, MatPlotLib, MLFlow
- Big Data Processing and Mining, ML Production, Exploratory Data Analysis (EDA), Data Visualization, Regression and Classification Modelling, Forecasting, Data Structures and Algorithms, Workflow Automation

PROJECTS

LLM-as-a-SupremeCourt-Judge | RIT
Tools: GPT API, Fine-Tuning, Chain-of-Thought Reasoning, Prompt Engineering.
- Developed an AI-driven judicial simulation framework using large language models (LLMs) to analyze 200 Supreme Court cases from the JUSTICE dataset, evaluating legal reasoning through chain of thought reasoning and persona-based modeling.
- Benchmarked multiple LLMs across 4+ experimental setups, assessing the impact of panel size, prompting, and ideology-based personas on verdict alignment with human judges.

Interpretable Parking Ticket Location Classifier | IIIT-H
Tools: PySpark, Spark Streaming, SparkML, SparkSQL, Docker, Flask.
- Built a SparkML RandomForest classifier to predict police-precinct of ticketed vehicles in NYC parking-ticket dataset.
- Employed Spark Streaming and PySpark for parallel computation. Achieved 98% accuracy and identified police precincts with anomalous ticketing practices.

Product Shipping Status Classifier | IIIT-H
Tools: SciKit-learn, Pandas, NumPy, MatPlotLib, seaborn.
- Developed a RandomForest model to predict issues in the product supply-chain which might delay product shipment to customers.
- Employed PCA and LDA optimization for visualization and dimension reduction. RandomForest model achieved accuracy of 97% (with PCA) and 92% (with LDA).