Sonu Yadav

sonu.yadav19997@gmail.com · (682) 408-4395 · Austin, TX ·
https://www.linkedin.com/in/sonu-yadav-a61046245/ ·
https://github.com/sy22478

Professional Summary

Aspiring AI engineer, and data scientist with hands-on experience designing, developing, and deploying intelligent systems to address real-world business challenges. Proficient in building Retrieval-Augmented Generation (RAG) pipelines, conversational agents, and predictive models using OpenAI Agent SDK, LangChain, Streamlit, and Scikit-Learn. Combines technical expertise in machine learning and NLP with a product-focused mindset to deliver scalable, user-centric AI solutions that drive measurable business impact.

Professional Experience

Laboratory Associate · Natera Inc.

Austin, TX · Feb 2023 - Nov 2024

- Optimized workflows for processing 240–336 blood samples daily for non-invasive prenatal testing, achieving a 99% success rate in automated plasma isolation using liquid-handling automation while ensuring compliance with CLIA and GCP standards.
- Redesigned standard operating procedures (SOPs) for DNA extraction and automated plasma isolation, reducing error rates by 90% and increasing team efficiency by 80%, enabling processing of up to 336 samples per shift.
- Collaborated with cross-functional teams during monthly meetings to identify procedural gaps, optimize workflows, and enhance productivity by 80%.

Student Technological Assistant · The University of Texas at Austin

Austin, TX · Jan 2022 - Dec 2022

- Collaborated with a team of 5+ professionals to design and distribute 20+ educational and promotional materials, ensuring accuracy and alignment with institutional guidelines.
- Managed and maintained an online Learning Management System (LMS) supporting 500+ users and 30+ courses, ensuring data integrity and system uptime.
- Processed datasets containing 10,000+ records, prepared 10+ detailed reports monthly, and conducted system reviews to identify inefficiencies.

Projects

Resume tailor: AI agent to tailor resume based on job description

https://github.com/sy22478/resume-tailoring-ai-agent

OpenAI Assistant API LangChain

Retrieval-Augmented Generation (RAG) Streamlit

- Engineered a resume tailoring AI agent using Retrieval-Augmented Generation (RAG), leveraging OpenAI's Assistant API, Cursor, and LangChain.
- Implemented document retrieval from user-uploaded resumes and generated customized versions aligned with specific job descriptions.
- Built a web UI with Streamlit to enable seamless interaction between users and the AI agent.

Restaurant Turnover Prediction

https://github.com/sy22478/Restaurant-Turnover-Prediction

Skills

Tools & Technologies

Jupyter Notebook
Git/GitHub
VS Code
Google Colab
n8n
V0
OpenAI SDK
Streamlit

AI Agents & LLMs

OpenAI Agent API
LangChain
Retrieval-Augmented Generation (RAG)
Prompt Engineering
Function Calling
Memory Management
Agent Workflows
Model Context Protocol

Data Science & Machine Learning

Python
Scikit-Learn
LightGBM
XGBoost
CatBoost
Optuna
Decision Trees
Ensemble Learning
Hyperparameter Tuning
Predictive Modeling
Classification
Regression
Model Interpretation
Feature Engineering
Handling Class Imbalance

Data Analysis & Visualization

NumPy
Pandas
Matplotlib
Seaborn
Exploratory Data Analysis (EDA)

Python LightGBM XGBoost CatBoost Optuna Ensemble Learning

Feature Engineering Predictive Modeling

- Participated in a hackathon to predict restaurant turnover using historical data, with RMSE as the evaluation metric.
- Built a high-performing predictive model by leveraging ensemble learning techniques to outperform individual models.
- Developed a weighted ensemble of three gradient boosting models—LightGBM, XGBoost, and CatBoost—after individually tuning each using Optuna for hyperparameter optimization.

Personal Loan Campaign

https://github.com/sy22478/Personal_Loan_Campaign

Python Scikit-learn Decision Trees Pruning

Class Imbalance Handling Feature Engineering Business Analytics

- Built a classification model to help AllLife Bank identify liability customers likely to accept personal loans and improve future campaign targeting.
- Achieved high recall to minimize missed opportunities while maintaining acceptable precision on a highly imbalanced dataset (only 9.6% positive class).
- Implemented a post-pruned Decision Tree with class_weight='balanced', engineered meaningful features, and extracted simple decision rules for stakeholder alignment.
- Achieved 84.56% recall and 92.65% precision on test data, delivering an interpretable model with actionable segmentation strategies for marketing.

FoodHub: Enhancing Restaurant Experience with Data Science

https://github.com/sy22478/Food-delivery-analysis

Python NumPy Pandas Matplotlib Seaborn

Exploratory Data Analysis (EDA) Multivariate Analysis

Business Insights Data Visualization Data Cleaning

- Analyzed 1,898 food delivery orders to uncover customer preferences, restaurant performance trends, and operational inefficiencies for data-driven business decisions.
- Identified popular cuisines, peak ordering times, high-performing restaurants, and revenue drivers using exploratory data analysis (EDA).
- Conducted multivariate analysis using Pandas, visualized patterns in cuisine popularity and order timing with Seaborn/Matplotlib, and derived actionable metrics like platform revenue and restaurant eligibility for promotions.
- Revealed key insights including weekend surge in orders, top-performing cuisines (Indian & Italian), and opportunities to streamline preparation time for high-demand but slow cuisines.

Education

Computer Science · MS

Westcliff University

May 2025 - Present ·

AI/ML · Post-graduate certificate

The University of Texas at Austin

Mar 2025 - Present ·

Neuroscience · BS

The University of Texas at Austin

2021 - 2022 · 3.63

Multivariate Analysis
Data Cleaning
Business Analytics
Customer Segmentation
Revenue Estimation

Soft Skills

Effective Communication
Attention to Detail
Problem-Solving
Time Management
Adaptability
Critical Thinking
Independent and Self-Motivated
Active Listening
Team Collaboration and Coordination
GenAI Literacy
Workflow Optimization
Cross-Departmental Collaboration
Technical Aptitude
Mentoring Peers

Achievements

Maverick Academic Scholarship

University of Texas at Arlington

Demonstrated history of strong academic performance.

Carmen Trujillo Nunez Pre-med Scholarship in Science

University of Texas at Arlington

Demonstrated academic achievement in the field of biology and premedical sciences

William L. and Martha Hughes Award for the Study of Biology

University of Texas at Arlington

Showed potential for outstanding achievement in their work as a Department of Biology major.

Honors College General Scholarship

University of Texas at Arlington

Recognized for being in good standing and making progress toward their Honors degrees

General ISSS scholarship

University of Texas at Austin

Recognized for being outstanding international student in the Department of Neuroscience

Certificates

Agent Engineering

Break Into Data · Present

Biology · BS
The University of Texas at Arlington
Aug 2018 - Dec 2020 · 3.63

Machine Learning
Great Learning · 2025

Python Foundations
Great Learning · 2025

Signal Processing (Python) for Neuroscience Practical Course
Udemy · 2025

Computational Neuroscience
Neuromatch Academy · 2022