Jocelyn Soto

nadyasoto@icloud.com
https://www.linkedin.com/in/nadya-soto/
https://github.com/nadya-soto

Profile

AI-focused Data Scientist with expertise in predictive modeling, NLP, and operational research. Proven track record in building real-time pipelines, deploying machine learning models, and automating dashboards to support fraud detection, customer retention, and operational efficiency. Passionate about applying AI to real-world problems, including digital transformation in infrastructure, energy optimization, and ethical applications of GenAI. Currently expanding cloud-based NLP deployment and foundation model integration.

Technical Skills

- **Advanced:**
- Programming: Python, SQL, R, ETL pipelines.
- Machine Learning: Regression, XGBoost, Random Forest, SVM, Scikit-learn.
- NLP: spaCy, sentiment analysis, text classification.
- **Proficient:**
- Deep Learning: TensorFlow, PyTorch, LSTMs for time-series.
- Cloud/MLOps: AWS, Docker, GitHub Actions (CI/CD), automated retraining, scalable architecture.
- Visualisation: Tableau, Power BI, data storytelling techniques, operational dashboard creation.
- **Working Knowledge:**
- MLOps: MLflow, Azure ML, model monitoring, GDPR-compliant AI deployment.
- Foundation Models: LLaMA, BERT.
- App Development: Flask, API integration.

Soft Skills

- Problem solving
- Analytical thinking
- Team collaboration
- Strategic planning
- Agile methodologies
- Innovation
- Quick learner

Additional Details

- **Languages**: Fluent in English and Spanish
- **Visa Status:** Eligible to work in the UK and open to relocating for the role
- **Research-Driven**: Strong interest in bridging academic insight and business impact
- **Ethics and Innovation**: Enthusiastic about ethical AI and knowledge transfer

Education

MSc Optimisation and Data Science *University of Essex* **2023 - 2024**

- Grade: Distinction
- Focused coursework in Machine Learning, Reinforcement Learning, AI, and Decision Analytics.
- Dissertation: Applied Machine Learning and AI to prevent femicide (Applied NLP and classification models to analise social media and news content).

BSc Applied Mathematics *Instituto Tecnologico Autonomo de Mexico (ITAM)* **2017-2021**

- Specialised in Probability Theory, Mathematical Statistics, Operations Research, and Computational Mathematics.
- Completed advanced coursework in Time Series Analysis, A/B Testing, and Stochastic Optimisation.

Experience

Data Scientist, *DiDi Global* **Apr 2022 - Dec 2023**

Global ride-sharing, delivery and fintech platform

- Applied advanced analytics and A/B testing to assess pricing strategies and costumer incentives, accelerating product launches by 15%.
- Deployed NLP models and sentiment analysis using SpaCy and Scikit-learn on 1M+ reviews, achieving 89% accuracy. Identifying pain points that increased customer retention by 3%, showcasing collaboration and proactive problem-solving.
- Built real-time monitoring dashboards (Tableau) that improved operational decision speed by 50%, allowing proactive issue resolution.
- Developed automated ETL pipelines that reduced data processing by 60%, allowing business teams to access real-time insights and take faster action.
- Coordinated Agile-based ML projects between global and regional teams, managing planning and delivery cycles over 3-month timelines.

Data Analyst, *Banco Azteca* **Mar 2020 - Apr 2022**

Financial institution

- Engineered Machine Learning fraud detection model, cutting losses by £400K/year via real-time risk analysis, relevant for predictive systems in asset management.
- Automated model retraining pipeline using CI/CD principles to maintain high performance amid evolving fraud patterns.
- Delivered end-to-end reporting automation, saving 40+ analyst hours/month, and provided real-time leadership dashboards.
- Collaborated with legal, tech, and operations teams to turn fraud analytics into policy decisions, demonstrating cross-team knowledge transfer.

Projects

NLP Research-to-Application Project (Dissertation) **2024**

- Designed and implemented NLP pipeline to analise narrative patterns across social media and news sources related to gender-based violence.
- Applied sentiment classification models to categorize public discourse, achieving 92% accuracy.

Renewable Energy Forecasting **2024**

- Developed deep learning (RNNs/LSTMs) time-series models achieving 95% accuracy in predicting renewable energy availability.
- Outperformed classical forecasting methods by 20%, creating potential for optimised energy distribution systems.
- Applied academic research methodologies to solve practical industry challenges, demonstrating knowledge transfer capabilities.

Reinforcement Learning Algorithms **2024**

- Implemented Q-Learning algorithms for optimisation, demonstrating 12% efficiency improvement in simulations.
- Designed policy iteration framework adaptable to fintech challenges including transaction routing and fraud detection.

Certifications

- Microsoft Azure Fundamentals (AZ-900) | Certified 2025
- DataCamp Python | Certified 2024
- TensorFlow Developer Certificate | In progress (Expected July 2025)
- DataCamp Associate Data Scientist in Python | (Expected August 2025)
- Gen AI Intensive Course with Google | In progress (Expected June 2025)