Ivan Vargas

SENIOR AI/ML ENGINEER

El Paso, TX | +1 (478) 200 2770 | Gmail | LinkedIn

Summary

Passionate AI/ML Engineer with 7+ years of hands-on experience transforming complex data into intelligent solutions that drive real business impact. Spanning computer vision systems, NLP applications, and generative AI solutions, I thrive on tackling challenging problems at the intersection of cutting-edge technology and practical applications

Experience

Velocity Data Solutions, Austin, TX | Senior Software Engineer   **Mar 2022 – Present**

- Architected end-to-end machine learning platform serving 50,000+ users with predictive analytics models, achieving 94% accuracy in customer behavior prediction and reducing churn by 28% through proactive intervention strategies.

- Implemented scalable MLOps pipelines using Python, TensorFlow, and AWS SageMaker with automated model retraining, feature engineering, and A/B testing frameworks that improved deployment efficiency by 60%.

- Designed sophisticated ensemble methods combining XGBoost, Random Forest, and neural networks for multi-class prediction tasks, resulting in 25% improvement in model performance over baseline approaches.

- Built real-time inference APIs using FastAPI and Redis caching that handle 500+ requests per second with sub-100ms latency, implementing comprehensive monitoring and alerting systems

- Developed advanced feature engineering pipelines processing 15+ million daily records using Apache Spark and Kafka, enabling real-time model updates and personalized recommendations

- Mentored team of 6 junior ML engineers, establishing best practices for model versioning, experiment tracking with MLflow, and production deployment strategies

- Implemented comprehensive model validation frameworks including statistical testing, fairness assessments, and drift detection using Evidently AI and custom monitoring solutions

- Collaborated with product and business teams to translate complex ML concepts into actionable insights, delivering executive-level reporting dashboards that influenced strategic decision-making

- Deployed containerized ML services on AWS ECS with auto-scaling capabilities, ensuring 99.95% uptime while optimizing infrastructure costs by 35%

- **Technologies Used:** Python, TensorFlow, PyTorch, AWS (SageMaker, ECS, Lambda, S3, RDS), Apache Spark, Kafka, FastAPI, Redis, Docker, MLflow, XGBoost, PostgreSQL

BorderTech Analytics, El Paso, TX | Machine Learning Engineer   **Jun 2020 – Feb 2022**

- Developed computer vision models using PyTorch and OpenCV for automated supply chain optimization, processing 50,000+ daily shipment images with 91% accuracy in package classification and damage detection

- Built end-to-end MLOps workflows with Docker, Kubernetes, and Jenkins CI/CD pipelines, reducing model deployment time by 40% and enabling continuous integration of ML improvements

- Implemented deep learning solutions for time series forecasting using LSTM networks and Transformer architectures, achieving 30% improvement in delivery time predictions for cross-border logistics

- Created real-time data processing pipelines using Apache Kafka and Apache Airflow to handle streaming logistics data from 200+ sensors and IoT devices

- Designed A/B testing frameworks for model validation in production environments, implementing statistical significance testing and automated rollback mechanisms for underperforming models

- Optimized model inference performance through TensorRT optimization and GPU acceleration, reducing processing time by 40% while maintaining prediction accuracy

- Developed RESTful APIs using Flask and FastAPI for model serving, implementing comprehensive authentication, rate limiting, and monitoring capabilities

- Collaborated with logistics teams to identify key performance indicators and translate business requirements into ML problem formulations

- Implemented data quality monitoring and anomaly detection systems using statistical process control and isolation forest algorithms

- **Technologies Used**: Python, PyTorch, OpenCV, Docker, Kubernetes, Apache Kafka, Apache Airflow, Flask, FastAPI, TensorRT, Jenkins, PostgreSQL, Redis, Git

Southwest Credit Union, El Paso, TX | Junior Data Scientist   **Aug 2018 – May 2020**

- Built predictive models for loan approval and risk assessment using scikit-learn, XGBoost, and ensemble methods, improving approval accuracy by 22% while reducing default rates by 15%

- Developed automated data preprocessing pipelines using Python and Pandas to clean and transform financial datasets containing 2+ million customer records

- Implemented feature engineering techniques including polynomial features, interaction terms, and dimensionality reduction using PCA to improve model performance

- Created interactive dashboards using Tableau and Python Plotly for risk assessment visualization, enabling loan officers to make data-driven decisions with 30% faster processing times

- Performed statistical analysis and hypothesis testing to validate model assumptions and ensure regulatory compliance with fair lending practices

- Built customer segmentation models using K-means clustering and hierarchical clustering algorithms to identify high-value customer segments for targeted marketing campaigns

- Implemented cross-validation strategies and hyperparameter tuning using GridSearchCV and Bayesian optimization to optimize model performance

- Collaborated with business analysts and compliance teams to ensure model interpretability and regulatory adherence in financial decision-making processes

- Developed automated reporting systems using Python and SQL that generated weekly model performance metrics and business impact assessments

- **Technologies Used**: Python, scikit-learn, XGBoost, Pandas, NumPy, Tableau, Plotly, SQL, PostgreSQL, Jupyter Notebooks, Git

Education

University of Texas at Austin | Bachelor of Computer Science   **Aug 2018 – May 2022**

Major: Software Engineering | Minor: Artificial Intelligence

Certifications

- AWS Certified Machine Learning Specialty
- Deep Learning Specialization
- Tensorflow Developer Certificate
- Generative AI with LLMs