KIRAN ANDOLU

Frisco, TX | +1 (801)414-6894 | kiranand638@gmail.com

PROFESSIONAL SUMMARY

- 12+ years of experience designing, developing, and deploying scalable AI/ML applications.
- Expert in end-to-end ML pipelines: data ingestion, feature engineering, model training, and deployment using Python, Java, PySpark, TensorFlow, PyTorch, and Keras.
- Proficient in predictive modeling, fraud detection, credit risk scoring, and customer behavior analytics using XGBoost, LightGBM, Random Forest, and ensemble techniques.
- Skilled in deploying ML models on AWS, Docker, and Kubernetes with CI/CD pipelines for continuous integration and delivery.
- Experienced in Big Data and real-time analytics using Spark, Hadoop, Hive, Kafka, and Delta Lake to process high-volume datasets efficiently.

TECHNICAL SKILLS

Programming & Scripting: Python, Java, TypeScript, SQL, R, C++, Scala, Julia.
Machine Learning & AI: Supervised Learning, Unsupervised Learning, Reinforcement Learning, Deep Learning, Neural Networks (CNN, RNN, LSTM,GAN,Transformers), Autoencoders, Transfer Learning, Feature Engineering, Dimensionality Reduction, Ensemble Learning, Hyperparameter Tuning, Cross-Validation, Predictive Modeling, Statistical Modeling.
ML & Data Science Frameworks: TensorFlow, Keras, PyTorch, scikit-learn, XGBoost, LightGBM, CatBoost, Hugging Face Transformers, NLTK, spaCy, Gensim, OpenCV, Pandas, NumPy, Matplotlib, Seaborn, Plotly, SHAP, Optuna.
Big Data & Analytics: Apache Spark, PySpark, Hadoop, Hive, HDFS, Databricks, Delta Lake, Airflow, Apache Kafka
Cloud & DevOps for ML: AWS (EC2, ECS, Lambda, S3, DynamoDB, OpenSearch, Glue, EMR, CloudWatch, IAM, Step Functions, SageMaker), Docker, Kubernetes, Jenkins, GitHub Actions, MLflow, DVC
Web & Deployment: REST APIs, Flask, FastAPI, Spring Boot, Node.js, Angular, Microservices, gRPC, Streamlit, Gradio.
Databases & Storage: Oracle, MongoDB, PostgreSQL, Cassandra, DynamoDB, Snowflake.
Data Engineering & ETL: SQL, PySpark, Airflow, Luigi, Kafka Streams, Spark Structured Streaming, Delta Lake, Glue.

PROFESSIONAL EXPERIENCE

ML Engineer — Mar 2020 -Present

CapitalOne | Plano, TX

- Engineered end-to-end PySpark ETL pipelines to process 2M+ records per batch, improving data accessibility for machine learning workflows by 40% and enabling robust feature extraction with Python libraries.
- Built and fine-tuned fraud detection and credit risk models using XGBoost, LightGBM, Random Forest, and Logistic Regression, achieving 92% accuracy and significantly enhancing detection of high-risk transactions.
- Leveraged PyTorch and TensorFlow to develop predictive models for customer engagement, applying advanced feature engineering and cross-validation to reduce churn by 15% and increase targeted intervention success.
- Migrated millions of records from Cassandra to DynamoDB by implementing automated Python scripts, resolving inconsistencies and improving query performance by 35%, accelerating downstream ML model training.
- Developed NLP solutions using fuzzy matching and semantic similarity to extract insights from unstructured financial data, enabling faster and more accurate fraud and AML investigations.
- Deployed OpenSearch to perform scalable semantic queries on large historical datasets, reducing investigative reporting time for compliance and fraud teams by 50% and enhancing audit efficiency.
- Designed Kafka-based real-time streaming pipelines for transaction data, reducing alert latency by 35% and supporting immediate identification of anomalous behavior for ML applications.
- Automated deployment and retraining of machine learning models using AWS Lambda, ECS, Step Functions, and CloudWatch, cutting manual effort by 60% while maintaining continuous production availability.
- Optimized Spark transformations and Cassandra UDT handling for POS system data, increasing ETL throughput by 25% and improving accuracy and reliability of downstream machine learning models.

Big Data Engineer — Oct 2017 - Jan 2018

Heartlands Payment Systems | Charlotte, NC

- Transformed raw transactional data into structured datasets with PySpark and MapReduce, enhancing predictive fraud model inputs by 40% and enabling more precise anomaly detection.
- Automated backup and archival processes using Python and Shell scripts, ensuring complete regulatory compliance while cutting manual processing time by half.
- Reduced false positives by 30% by engineering advanced features and statistical models for anomalous transaction detection, improving overall model accuracy.
- Conducted comprehensive exploratory data analysis and preprocessing with Pandas, NumPy, and Matplotlib, generating clean datasets that accelerated ML training efficiency by 25%.
- Monitored system performance through CloudWatch and Splunk dashboards, identifying bottlenecks and achieving a 20% reduction in downtime.
- Enhanced data retrieval efficiency from MongoDB and Cassandra by optimizing queries and aggregations, accelerating ML pipeline execution by 35% and enabling faster model predictions.
- Improved ETL and batch processing with Hive and Hadoop, producing scalable AI-ready datasets while reducing preprocessing time for machine learning training by 40%.
- Integrated Redos caching to reduce repetitive data fetches, lower query latency, and boost overall system and ML pipeline performance significantly.

Predictive Analytics Engineer — Feb 2013 - Aug 2017

AT&T | Richardson, TX

- Migrated legacy monolithic systems to Spring Boot microservices, enabling scalable ML model deployment and reducing service latency by 35%.

- Developed Angular 2.0 front-end components linked to ML REST APIs, delivering real-time analytics and improving dashboard engagement by 25%.
- Integrated Netflix OSS stack (Eureka, Ribbon, Hystrix, Feign) to ensure service discovery and fault tolerance in ML pipelines, achieving 99.5% uptime
- Built predictive models using Decision Trees, Random Forest, and Gradient Boosting, increasing accuracy of customer churn predictions by 28%.
- Created feature engineering and preprocessing pipelines with Python and PySpark, reducing training time on large-scale datasets by 40%.
- Exposed ML models via SOAP and REST endpoints, enabling automated production predictions for over 100,000 daily transactions.
- Implemented CI/CD pipelines using Jenkins and Shell scripts for ML services, cutting deployment errors by 50% and accelerating release cycles.
- Optimized database queries and caching with Hibernate and Redis, increasing throughput for ML-driven applications by 25% and lowering query latency.

Java Developer **Oct 2012 - Feb 2013**

PBGC | Washington, DC

- Implemented MemCache and AWS ElastiCache to accelerate pension claim queries, reducing database access latency by 30% and improving availability of ML-ready datasets.
- Developed modular Spring 3.0 components with dependency injection, enabling seamless integration with AWS RDS and improving maintainability of cloud-based claim pipelines by 25%.
- Built interactive JSP, HTML, CSS, and JavaScript dashboards with jQuery and Ajax, enabling analysts to monitor AWS-hosted ETL workflows efficiently and reducing manual review time by 20%.
- Designed and deployed SOAP-based JAX-WS services on AWS EC2 instances, automating structured pension data ingestion for downstream predictive analytics and increasing dataset availability by 40%.
- Engineered Spring Batch ETL pipelines with AWS S3 integration, processing high-volume claims data and boosting ETL throughput by 35% for cloud-based analytics.
- Implemented data validation and preprocessing routines in Java using AWS Lambda and S3, producing ML-ready structured datasets and reducing downstream errors by 30%.
- Optimized Hibernate ORM queries for AWS RDS to accelerate large-scale data retrieval by 28%, enabling faster feature extraction for predictive modeling.
- Developed JUnit test suites and integrated logs with AWS CloudWatch, achieving 95% code coverage and improving reliability of automated claim data pipelines.

Jr Java Developer **Jul 2012 - Oct 2012**

US Patent & Trademark | Alexandria, VA

- Migrated the legacy Work Credit application to Spring IOC modules and integrated with Hibernate ORM, reducing system errors by 30% and improving claims data processing speed.
- Built dynamic JSP and JSF front-end components with HTML, CSS, and JavaScript, enhancing user interactions and accelerating claim submission workflows by 25%.
- Implemented Spring Batch jobs and automated ETL pipelines for claims, enabling scalable ML-ready datasets and reducing batch processing time by 40%.
- Developed unit and integration tests using JUnit, ensuring 95% code coverage and improving reliability of production-deployed services.
- Streamlined data preprocessing and validation routines for structured claims datasets, improving ML model input quality by 30% and enabling predictive analytics.
- Configured AWS S3 storage and ElastiCache caching for processed datasets, improving data retrieval efficiency by 28% for ML training pipelines.
- Deployed application in Docker containers on AWS ECS, achieving 99% uptime and supporting continuous ML-ready data availability for enterprise processes.
- Designed REST endpoints and SOAP services to expose claims data for analytics consumption, reducing reporting latency by 35% and enabling automated downstream ML workflows.

Jr Developer **May 2012 - Jul 2012**

PNC Bank | Strongsville, OH

- Built Spring MVC microservices for processing financial transactions, reducing ETL runtime by 30% and enabling near real-time analytics.
- Designed interactive JSP and Spring tab library dashboards, improving analyst workflow efficiency by 25% and providing actionable insights.
- Optimized MyBatis SQL queries and Oracle stored procedures, accelerating feature extraction for predictive models by 35%.
- Automated batch processing with Spring Quartz scheduling, cutting manual oversight by 40% and ensuring timely data availability.
  • Streamlined CRUD operations and ETL pipelines, reducing preprocessing time for ML models by 30% and improving dataset quality.
  • Exposed RESTful endpoints to integrate downstream ML pipelines, accelerating fraud detection model deployment and evaluation.
- Implemented JUnit testing for service and DAO layers, achieving 95% coverage and enhancing reliability of production workflows.
- Enhanced data validation and error-handling mechanisms, improving transaction accuracy by 28% and reducing false positives in anomaly detection.

Java Developer Intern **Jan 2010 - Jul 2010**

Kalyan Enterprises | Hyderabad, India

- Developed Spring MVC microservices to preprocess financial transaction data, reducing ETL runtime by 30% and enabling real-time analytics.
- Implemented MyBatis queries with complex joins and stored procedures on Oracle 10g, increasing feature extraction efficiency by 25% for ML pipelines.

- Built interactive JSP/JSF dashboards with JavaScript and jQuery, improving analyst productivity and decreasing verification time by 20%.
- Automated batch processing with Spring Quartz scheduler, ensuring timely availability of structured data for predictive modeling, increasing throughput by 35%.
- Optimized front-end rendering and database integration, enhancing real-time transaction monitoring and reducing latency by 15%.
- Designed reusable service and DAO layers with unit testing using JUnit, achieving 90% code coverage and ensuring reliability of MLready pipelines.
- Created structured ETL workflows, improving data consolidation and accuracy for downstream ML model training by 28%.
- Integrated dynamic HTML/CSS components, enhancing reporting clarity and enabling faster decision-making for financial analysts.

PROJECTS

Vintage Platform

- Engineered PySpark ETL pipelines to ingest and clean 2M+ transactional records per batch, improving ML feature availability by 40% for fraud and credit risk models.
- Developed XGBoost, LightGBM, and Random Forest models for fraud detection, achieving 92% accuracy and reducing high-risk transaction processing time by 30%.
- Implemented OpenSearch with NLP-based semantic search for historical datasets, cutting investigative report generation time for AML and fraud teams by 50%.

Predictive Analytics for Card Transactions

- Built ML-ready pipelines using PySpark and Hadoop to process large-scale transaction data, improving anomaly detection efficiency by 35%.
- Applied feature engineering and statistical modeling to identify transaction fraud, reducing false positives by 30% using Python and scikit-learn.
- Delivered REST APIs via Spring Boot to expose predictive analytics results, enabling downstream integration and accelerating ML pipeline usage by 25%.

Customer Behavior Prediction

- Migrated monolithic systems to Spring Boot microservices for real-time ML model deployment, reducing system latency by 35%.
- Built predictive models for churn and engagement using Decision Trees, Random Forest, and Gradient Boosting, improving targeted intervention success by 28%.
- Implemented PySpark-based preprocessing and feature engineering pipelines, reducing model training time on large-scale datasets by 40%.

Pension Claims Analytics

- Designed caching and batch processing solutions using MemCache and Spring, reducing ETL runtime by 30% and enabling MLready pension claim datasets.
- Developed interactive dashboards and automated reporting pipelines using JSP, JavaScript, and jQuery, accelerating claim review and feature extraction by 20%.
- Optimized Hibernate ORM for large-scale claims datasets, improving retrieval performance by 28% for downstream ML models.

Work Credit Application Modernization

- Refactored legacy applications into Spring IOC modules, creating scalable ML-ready pipelines and reducing data errors by 30%.
- Built automated preprocessing and validation routines, producing high-quality datasets and reducing ML downstream errors by 35%.
- Developed REST and SOAP endpoints to expose structured patent data, improving analytics throughput for predictive modeling by 40%.

Financial Transaction Analytics

- Designed Spring MVC microservices and MyBatis queries to preprocess financial data, reducing ETL runtime by 30% for real-time ML pipelines.
- Automated batch data pipelines using Quartz scheduler, improving ML model readiness and increasing throughput by 35%.
- Built interactive JSP and JSF dashboards for transaction monitoring, improving analyst productivity by 20% and enabling faster model feedback.

Enterprise Data Management

- Implemented Struts MVC-based ETL workflows to preprocess large enterprise datasets, accelerating ML-ready data availability by 20%.
- Created Java-based data cleaning and normalization routines, improving dataset quality by 35% for predictive analytics.
- Developed modular front-end components with JSP and JavaScript validations, reducing data entry errors by 25% and enabling faster ML experimentation.

EDUCATION

Masters of Science in Computer Science **Dec 2011**

Texas A&M University | Kingsville, TX

CERTIFICATIONS

- IBM AI Engineering Professional Certificate - **Coursera**
- Microsoft AI & ML Engineering Professional Certificate - **Coursera**
- Google Cloud Machine Learning Engineer Professional Certificate - **Coursera**
- IBM Generative AI Engineering Professional Certificate - **Coursera**
- Associate AI Engineer for Developers - **DataCamp**
- AI Engineer Development Plan - **Coursera**