

SATAR SHAMSI GOUSHKI

ML / AI Engineer

Kerman, Iran | satar.shams2@gmail.com | +98 991 335 8171 | satar-shams.github.io/satar-portfolio
[satar-shamsi](#) | [satar-shams](#)

Professional Summary

ML / AI Engineer with hands-on experience in machine learning, data engineering, MLOps, and applied AI. I build practical systems that cover the full workflow, from data pipelines and model development to testing, deployment, and API-based inference. Recent work includes an end-to-end LSTM forecasting system, a PostgreSQL/Airflow market data pipeline, and a document intelligence pipeline. My academic background includes graduate-level research in machine learning, mathematical modeling, image processing, and high-performance computing.

Technical Skills

Programming: Python, SQL, C++

Machine Learning: Machine Learning, Deep Learning, Neural Networks, Time Series Forecasting, Classification, Natural Language Processing, Computer Vision

ML Engineering / MLOps: MLflow, Optuna, Model Tracking, Model Registry, Model Versioning, FastAPI, Model Serving, Docker, Docker Compose, GitHub Actions, CI/CD

Data Engineering: ETL / ELT, Pandas, PostgreSQL, SQLAlchemy, Apache Airflow, Data Validation, Incremental Loading, Idempotent Loading

Deep Learning: TensorFlow, Keras, LSTM, CNN, Autoencoders, Transfer Learning

Data Science: NumPy, scikit-learn, Statistical Analysis, Mathematical Modeling, Time Series Analysis

NLP / Document AI: Text Processing, OCR, Embeddings, Vector Databases, Document Processing, LangChain

Development & Tools: Git, GitHub, Linux, Jupyter Notebook, Pytest, Structured Logging, HashiCorp Vault

Professional Experience

ML / AI Engineer Intern, GoalEarn – Iran

Oct 2025 – present

Worked across machine learning, data engineering, and ML engineering projects, with an emphasis on modular, testable, and reproducible systems.

- Built a modular ETL pipeline for acquiring, validating, transforming, and loading data into PostgreSQL.
- Implemented incremental loading, structured logging, automated tests, and secure credential management using HashiCorp Vault.
- Developed a machine learning inference API for serving model predictions.
- Applied software engineering practices including modular project structure, documentation, Git, testing, and reproducible workflows.

Selected Projects

LSTM Time Series Forecasting — End-to-End MLOps

Oct 2025 – present

Built an end-to-end machine learning system for AAPL stock-price forecasting, covering data preprocessing, model development, evaluation, inference, and deployment.

- Developed a stacked LSTM model with TensorFlow/Keras and evaluated it on a held-out test set.
- Used Optuna for hyperparameter optimization and MLflow for experiment tracking and model management.

- Identified and fixed a data-leakage issue in the preprocessing pipeline that was affecting validation results.
- Built a FastAPI inference service and used Docker for the training and inference components.
- Added automated tests and GitHub Actions for CI/CD and deployed the inference API to Render.

Market Data Pipeline — ETL

June 2025 – present

Built a production-oriented Python ETL pipeline that acquires market data from yfinance, validates and transforms it, and stores it in PostgreSQL.

- Implemented idempotent database loading and a pipeline-run audit table.
- Added data validation and 11 unit tests for the main pipeline components.
- Integrating Apache Airflow for scheduling and pipeline orchestration.
- Developing a FastAPI query layer for accessing stored market data.
- Used Docker to run the project services.

Intelligent Document Analysis

July 2025 – present

Building a document intelligence pipeline for processing documents, extracting and preparing text, and supporting semantic search.

- Added PDF and DOCX processing using PyMuPDF and python-docx.
- Added OCR support for scanned documents using Tesseract.
- Working on text processing and preparation for semantic search.
- Adding embeddings and vector database support with Sentence Transformers and ChromaDB.
- Added automated tests for document ingestion and processing.

Retail Orders ETL Pipeline

Nov 2024 – Dec 2024

Built an end-to-end ETL pipeline that processes retail order data from CSV files and loads validated results into PostgreSQL.

- Separated the pipeline into extraction, transformation, and loading stages.
- Added checks for missing values, negative amounts, duplicates, and required fields.
- Implemented incremental and idempotent database loading.
- Added structured logging and automated tests with Pytest.
- Used HashiCorp Vault for database credential management.
- Containerized the pipeline with Docker.

ML Inference API

Jan 2025 – Feb 2025

Built a REST API for serving machine learning model predictions.

- Developed prediction endpoints with FastAPI and request and response validation.
- Added environment-based configuration and structured logging.
- Added automated tests with Pytest for the main API components.
- Containerized the service with Docker for local and deployment use.
- Organized the code into separate components for the API, model handling, configuration, and supporting utilities.

Academic Research & Projects

Graduate Researcher — Big Data and Machine Learning, ITMO University – Saint Petersburg, Russia

Aug 2018 – May 2020

Graduate research and university projects covering machine learning, mathematical modeling, data analysis, image processing, and high-performance computing.

- Researched machine-learning-based music generation and submitted a ~20-page research paper.
- Researched and implemented an SEIR-based mathematical model for epidemic dynamics, supported by literature research and a ~20-page paper.
- Worked on image processing projects including image classification and image segmentation.
- Worked on parallel algorithms and high-performance computing, including a Traveling Salesman Problem project using the nearest-neighbor algorithm and analysis of its limitations and parallel execution time.
- Completed projects in multivariate data analysis, statistical methods, MATLAB simulation, data warehousing, and SQL.

Education

ITMO University, Graduate Studies — Master's Program (degree not completed) in Big Data and Machine Learning – Saint Petersburg, Russia Sept 2018 – May 2020

Completed graduate coursework and research through the third semester of the Master's program.

- GPA: 3.91/4.00 in Big Data and Extreme Computing; 3.79/4.00 in Machine Learning and Data Analysis.
- Research in machine-learning-based music generation and SEIR-based mathematical modeling of epidemic dynamics.
- Completed projects in machine learning, image processing, high-performance computing, parallel algorithms, multivariate data analysis, and data warehousing.

Shahid Bahonar University of Kerman, BSc in Electrical Engineering — Electronics – Kerman, Iran Feb 2007 – Sept 2012

- GPA: 14.47/20.00 · Top 5 students

Awards

- Russian Open Doors Olympiad — Full Scholarship, HSE University (2024)

Languages

English: Professional proficiency (TOEFL iBT 85; graduate studies conducted in English)

Persian: Native

Russian: Basic

German: Basic