Ali Abouelazm

713-834-4423 | aliazm419@gmail.com | linkedin.com/in/ali-abouelazm | aliabouelazm.com | github.com/AliAbouelazm

Education

Texas A&M University **Aug. 2023 – May 2027**

B.S. Data Engineering | Minor in Mathematics and Statistics *College Station, TX*

- Relevant Coursework: Data Structures & Algorithms, Quant Models for Machine Learning, Statistical Learning & Decisions, Optimization of Analytics, Computational Data Science, Principles of Data Science, Statistics.

Experience

Cloudflare **May 2026 – Aug. 2026**

AI Discoverability & Optimization Intern *Austin, TX*

- Working on Cloudflare's **AEO** team to understand how large language models index and surface web content, then building technical frameworks that improve AI-driven content discoverability across Cloudflare's global network.

Texas A&M AgriLife **Mar. 2026 – Present**

Machine Learning Researcher *College Station, TX*

- Building **PyTorch**-based predictive models on **50GB+** of livestock biosensor data to forecast animal health indicators across **3+** active deployments, iterating continuously on model architecture, feature design, and training pipelines.
- Architecting scalable **ML** pipelines on **AWS** (S3, Athena) to ingest and process raw biological data streams, cutting preprocessing time by **35%** and reducing time-to-insight for field researchers and enabling faster iteration.
- Implementing automated feature engineering to extract health signals from noisy multi-sensor inputs, improving model training throughput by removing manual preprocessing steps and increasing prediction reliability across all deployments.

TCG Digital Solutions **May 2025 – Aug. 2025**

Data Science Intern *Somerset, NJ*

- Architected an end-to-end **CV** pipeline using **Gemini Pro** and **OpenCV** to automate soccer highlight extraction, hitting **95%** precision across **100+** match hours and reducing manual video processing overhead by **80%**.
- Developed and tuned **XGBoost** and **CatBoost** gradient-boosting models on **1M+** rows of player tracking data, using **SHAP** values to surface tactical insights to downstream athletic analysis teams.
- Optimized real-time video inference via multi-threaded **Python** processing, reducing event extraction latency and enabling near-zero delay delivery of match-critical clips to field-based athletic analysis teams worldwide.

Projects

Sonus | *Python, FastAPI, React, scikit-learn, SQLite, WebSockets, OpenAI API* **Feb. 2026**

- Built an autonomous **LLM** assistant that reasons over calendar events, biometrics, and live device state to execute complex multi-step routines unprompted, orchestrating **10+** service integrations via a tool-calling loop.
- Trained personalized stress and sleep classifiers on biometric data from **3** wearable integrations using **scikit-learn**, with a confidence scoring system that learns from user feedback and adapts via real-time **WebSocket** loops.

Drift | *Python, FastAPI, React, DistilBERT, HuggingFace, SHAP* | drift.aliabouelazm.com **Apr. 2026**

- Fine-tuned **DistilBERT** on **40K+** labeled Reddit and Twitter posts for 3-class sentiment classification, achieving **73.1%** F1 macro on a **10,360**-sample held-out test set across positive, negative, and neutral classes.
- Built an analytics engine computing rolling sentiment averages across configurable time windows, anomaly detection with mild/moderate/severe severity scoring, **TF-IDF** keyword extraction per class, and **SHAP** token-level explanations.

Wavelength | *Python, FastAPI, React, XGBoost, PyTorch, SHAP* | wavelength.aliabouelazm.com **May 2026**

- Trained **Random Forest**, **XGBoost**, and a **PyTorch MLP** on **89K** Spotify tracks across **7** mood classes; **XGBoost** reached **99.6%** and **MLP** reached **93.5%** F1 macro, highlighting tree vs. neural boundary learning differences.
- Engineered **13** features from raw Spotify audio data including energy/valence ratios and composite mood index, applied **SHAP** **TreeExplainer** for plain-English explanations, and built a recommendation engine with cosine similarity ranking.

Leadership

Lovable.dev | TAMU Datathon **Jan. 2026 – Present**

Campus Ambassador | Challenges Organizer *Texas A&M University*

- Hosted campus-wide AI platform workshops for **200+** engineering students as Lovable Campus Ambassador; directed **ML** problem set creation for **500+** competing students at TAMU Datathon using **Python**, **SQL**, and **scikit-learn**.

Technical Skills

ML & AI: PyTorch, TensorFlow, scikit-learn, XGBoost, DistilBERT, SHAP, LLM Tool Calling (OpenAI/Gemini APIs), OpenCV.
Languages: Python, SQL, C/C++, TypeScript, JavaScript, R.
Data & Cloud: AWS (S3, Athena), PostgreSQL, Snowflake, Docker, Git, FastAPI, WebSockets.
Libraries: pandas, NumPy, HuggingFace Transformers, Matplotlib, Seaborn, Dask, SciPy.