# Aarav Jain

803-833-9682 | aaravj@vt.edu | linkedin.com/in/aarav-jain-11b672382
github.com/aaravjj2 | aaravjain.dev

## Education

**Virginia Tech** Blacksburg, VA
*B.S. Computer Engineering* *Aug 2025 – May 2029*

- Technical Focus: Financial Technology, Machine Learning, Full-Stack Development
- Relevant Coursework: Data Structures, Algorithms, Computer Architecture, Linear Algebra

## Technical Skills

**Languages**: Python, TypeScript, JavaScript, SQL, Bash
**Frameworks & Libraries**: React, Next.js, FastAPI, Streamlit, Pandas, NumPy, XGBoost, LightGBM
**FinTech & Trading**: Market Data APIs (Finnhub, Alpaca, Yahoo Finance), Technical Indicators, Backtesting Systems, Real-time Streaming, WebSocket
**Data & Infrastructure**: PostgreSQL, SQLite, Redis, Celery, SQLAlchemy, Docker, Git, Pytest, Playwright
**AI/ML Tools**: GPT-4 API, PaddleOCR, Scikit-learn, Ensemble Methods, Feature Engineering

## Projects

**TradingView-Style Market Workstation** | *React, TypeScript, Python, FastAPI, PostgreSQL* GitHub

- Built production-grade trading platform with real-time WebSocket data streaming from 5+ providers (Finnhub, Alpaca, Yahoo Finance)
- Implemented 35 technical indicators (SMA, EMA, RSI, MACD, Bollinger Bands, VWAP) across trend, momentum, and volatility categories
- Developed backtesting engine with deterministic replay (0.5x-10x speed), paper trading system with P&L tracking, and 14 Bloomberg-style dashboard tiles
- **Metrics**: 50,000+ lines of code, 275 automated tests (Pytest + Playwright), multi-timeframe support (1m-1W)

**ML Stock Prediction Research Pipeline** | *Python, XGBoost, LightGBM, Pandas, Jupyter* GitHub

- Developed ensemble forecasting system using XGBoost, LightGBM, and sentiment analysis models with meta-gating for high-confidence predictions
- Implemented walk-forward backtesting to prevent look-ahead bias, modular feature engineering pipeline with 50+ technical and sentiment features
- Integrated multiple data sources (Finnhub, News API, Tiingo) with reproducible workflow compatible with Google Colab
- **Architecture**: Abstention-based trading (only signals when confident), 6 sequential notebooks for data → features → training → backtesting

**Intelligent Document Processing Pipeline** | *Python, FastAPI, PaddleOCR, GPT-4, Celery, Redis* GitHub

- Built production document ingestion service with PaddleOCR for multi-format support (PDF, PNG, JPG, TIFF) achieving <5% CER for printed text
- Implemented async processing architecture using Celery workers and Redis queue for scalable document processing (50+ concurrent uploads)
- Integrated GPT-4 and local Llama models for intelligent document classification and structured field extraction
- Developed real-time Streamlit monitoring dashboard with PostgreSQL full-text search and processing analytics

## Leadership & Community Service

**Podcast for Change Fundraiser** – Lead Organizer: Raised $1,000 for rural sanitation facilities in India

**Sonu Sood Foundation** – Remote Volunteer: Coordinated medical aid for 9 families during COVID-19

**Mahatma Gandhi School** – Summer Tutor: Taught math and science to underprivileged children