

Created: October 3, 2025

Status: **V** PRODUCTION READY



I've created a comprehensive guide for running your IDS pipeline and ensured all required Python packages are installed.

Documentation Created

1. F QUICKSTART.md

Purpose: Get the pipeline running in 5 minutes (PCAP testing)

Use when: You want to run the pipeline RIGHT NOW with PCAP files

Contains:

- TL;DR commands to start everything
- What's happening under the hood
- Quick monitoring commands
- · Common issues & fixes

2. PRODUCTION_DPDK_GUIDE.md (NEW!)

Purpose: Complete production setup with DPDK and external traffic

Use when: Running in production with real network traffic

Contains:

• DPDK interface binding

• External device setup

- Suricata DPDK mode configuration
- Flow-based ML for ALL traffic
- High-performance tuning
- · Production troubleshooting

3. RUNTIME_GUIDE.md

Purpose: Complete step-by-step execution guide

Use when: You need detailed instructions or troubleshooting

Contains:

- Prerequisites checklist
- Detailed step-by-step execution
- · Component monitoring

- Comprehensive troubleshooting
- Performance tuning
- Data flow examples

4. PACKAGES_INSTALLED.md

Purpose: Document installed packages

Use when: Checking what's available or reinstalling

Contains:

- Complete package list with versions
- Installation verification
- Reinstallation instructions

5. **\ install_missing_packages.sh**

Purpose: Automated package installation

Use when: Missing packages or fresh environment setup

Does:

- Installs LightGBM, XGBoost
- Verifies all core packages
- Installs visualization utilities
- Validates installation

What's Been Fixed

1. Missing Packages Installed

- **LightGBM** (4.6.0) For LightGBM model support
- **XGBoost** (3.0.5) For XGBoost model support
- **Matplotlib** (3.10.6) For plotting
- **Seaborn** (0.13.2) For visualization
- **V tqdm** (4.67.1) For progress bars
- **Colorama** (0.4.6) For colored output

2. Documentation Complete

- **V** Quick start guide for immediate use
- V Detailed runtime guide for troubleshooting
- Package installation documentation
- Automated installation script



Option 1: Quick Testing (PCAP Replay)

Best for: Development, testing, learning

```
cd /home/sujay/Programming/IDS/dpdk_suricata_ml_pipeline

# Read the quick guide
cat QUICKSTART.md

# Or just run:
   ./scripts/02_setup_kafka.sh # Start Kafka
sleep 30
source   ../venv/bin/activate
python src/ml_kafka_consumer.py --config config/pipeline.conf
```

Option 2: Production Mode (DPDK + External Traffic)

Best for: Real network monitoring, production deployment

```
cd /home/sujay/Programming/IDS/dpdk_suricata_ml_pipeline

# Read the production guide
cat PRODUCTION_DPDK_GUIDE.md

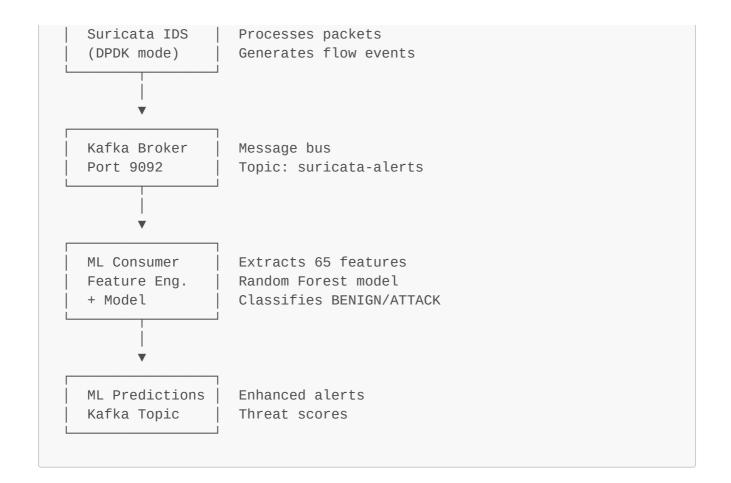
# Then follow the 6-phase setup:
# 1. Start Kafka
# 2. Bind network interface to DPDK
# 3. Start Suricata in DPDK mode
# 4. Start ML consumer
# 5. Send traffic from external device
# 6. Monitor predictions
```

Option 3: Follow the Detailed Guide

```
cat RUNTIME_GUIDE.md # Detailed step-by-step guide
```

Ш Pipeline Architecture

```
PCAP Replay or Live Traffic tcpreplay
```



@ Test Commands

Verify Everything is Ready

```
# 1. Check virtual environment
source /home/sujay/Programming/IDS/venv/bin/activate
python -c "import kafka, sklearn, lightgbm, joblib; print(' All
packages OK')"

# 2. Check ML models
ls -lh /home/sujay/Programming/IDS/ML\ Models/

# 3. Check pipeline status
cd /home/sujay/Programming/IDS/dpdk_suricata_ml_pipeline/scripts
./status_check.sh
```

Test the Pipeline

```
# 1. Start Kafka
./02_setup_kafka.sh

# 2. Start ML Consumer (new terminal)
cd /home/sujay/Programming/IDS/dpdk_suricata_ml_pipeline
source ../venv/bin/activate
```

```
python src/ml_kafka_consumer.py --config config/pipeline.conf --verbose

# 3. Generate test traffic (another terminal)
cd tests
python test_benign_traffic.py
python test_attack_generator.py
```

Q What to Expect

When running successfully, you should see:

ML Consumer Output:

Kafka Topic (ml-predictions):

```
{
  "timestamp": "2025-10-03T09:30:15.123Z",
  "flow_id": "abc123...",
  "src_ip": "192.168.1.10",
  "dst_ip": "93.184.216.34",
  "src_port": 45123,
  "dst_port": 443,
  "proto": "TCP",
  "ml_prediction": "BENIGN",
  "confidence": 0.98,
  "threat_score": 0.02,
  "features": {...}
}
```



🐛 Troubleshooting Quick Reference

Issue	Quick Fix
Kafka won't start	pkill -9 -f kafka && ./02_setup_kafka.sh
Missing packages	./install_missing_packages.sh
ML model not loading	Check path: ls -lh/ML\ Models/
No traffic flowing	Generate test: python tests/test_benign_traffic.py
Import errors	source/venv/bin/activate
Permission denied	Add sudo for DPDK/Suricata commands

Full troubleshooting: See RUNTIME_GUIDE.md section \$\lambda\$.



Project Structure

```
dpdk_suricata_ml_pipeline/

    5 - minute quick start

— QUICKSTART.md
— RUNTIME_GUIDE.md
                   Complete execution guide
├─ install_missing_packages.sh 🔧 Package installer
                Architecta.

Installation guide
 README.md
                    Architecture overview
— SETUP_GUIDE.md
 — config/
  - scripts/
   ├─ 02_setup_kafka.sh Start Kafka
   ├─ 04_start_ml_consumer.sh Start ML engine
    05_replay_traffic.sh Replay PCAP files
   - src/
   ├─ ml_kafka_consumer.py 🧠 ML inference engine
   — model_loader.py 🔖 Model loading
   └─ alert_processor.py 🚨 Alert correlation
  tests/
   test_benign_traffic.py Generate benign traffic
   test_attack_generator.py Generate attacks
```

PROF

Key Concepts

Flow-Based ML

• Processes **ALL** network flows (not just alerts)

- Extracts 65 CICIDS2017 features per flow
- Real-time classification: BENIGN or 11 attack types

Models Available

- 1. Random Forest (2017) random_forest_model_2017.joblib (2.0 MB)
- 2. LightGBM (2018) lgb_model_2018.joblib (801 KB)

Attack Types Detected

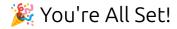
- DoS/DDoS attacks
- · Port scans
- Brute force (SSH, FTP)
- · Web attacks
- Botnet traffic
- Infiltration attempts

- Suricata Docs: https://suricata.io/
- Kafka Docs: https://kafka.apache.org/
- CICIDS2017 Dataset: https://www.unb.ca/cic/datasets/ids-2017.html
- **DPDK Docs**: https://doc.dpdk.org/

Final Checklist

Before running the pipeline:

- ✓ Virtual environment exists: /home/sujay/Programming/IDS/venv
- All Python packages installed (including LightGBM, XGBoost)
- ■ ML models available: . . /ML Models/*.joblib
- Configuration file: config/pipeline.conf
- ✓ Scripts are executable: chmod +x scripts/*.sh
- 🗹 Documentation complete
- Ready to run!



Next Steps:

- 1. **Read QUICKSTART.** md for immediate execution
- 2. Start Kafka with ./scripts/02_setup_kafka.sh
- 3. Run ML consumer with python src/ml_kafka_consumer.py
- 4. **Monitor** predictions with Kafka console consumer
- 5. Analyze results and tune as needed

Questions or Issues?

Check the **RUNTIME_GUIDE.md** for:

- Detailed step-by-step instructions
- Comprehensive troubleshooting
- Performance tuning tips
- Monitoring commands
- Example outputs

Happy Intrusion Detecting! ♥��

Created with ♥ by GitHub Copilot

Date: October 3, 2025

+8/8+