Anomaly Detection Models Report (IEEE-CIS Dataset)

Evaluation of Isolation Forest and Autoencoder models using 10-fold cross-validation.

Parameter Tuning Methodology

- **Optuna framework** for hyperparameter optimization
- **Common evaluation protocol**:

  ```
  # Metric aggregation template
  results = []
  for fold in kf.split(df_normal):
      # Train-test split with injected anomalies
      # Model training with trial parameters
      # Metric calculation (F1 + Precision + Accuracy)
  return np.mean(metrics)
  ```

---

1. Isolation Forest

Baseline vs Optimized Performance

Configuration	F1 Score	Precision	Accuracy	AUC
Default (contam=0.2)	0.504	0.486	0.726	-
Optuna Tuned	0.480	0.590	0.770	0.649

Optimal Parameters:

```
{'contamination': 0.1, 'n_estimators': 150, 'max_features': 0.4}
```

---

2. Autoencoder

Architecture Evolution

Configuration	F1 Score	Precision	Accuracy
Default	0.422	0.553	0.751
Optuna Optimized (128-64 layers)	0.457	0.577	0.761

Optimal Parameters:

```
{
  'hidden_layers': [128,64], 'contamination': 0.1, 'lr': 0.00216,
  'epochs': 40, 'batch_size': 2048, 'activation': 'relu', 'dropout': 0.0,
}
```

Performance Profile:

```
Normal Transaction Recall: 90%
Fraud Detection Precision: 57.7% (1.7pp improvement from baseline)
```

---

3. One-Class SVM

Status: Failed operational viability test

- Single fold training exceeded 2-3 hours vs 20min for full 10-fold IF training
- Immediate disqualification for production use

Conclusion & Recommendations

Performance Benchmark

Model	F1 Score	Precision	Training Time
Isolation Forest	0.480	0.590	20 min
Autoencoder	0.457	0.577	4-6 hrs
One-Class SVM	N/A	N/A	>24 hrs

Implementation Guidance:

1. **Isolation Forest** - Default choice for real-time fraud detection
    - 77% accuracy with 40% fraud recall at 0.1 contamination
    - 150 estimators provide optimal speed/performance balance
2. **Autoencoder** - Specialized use cases only
    - Justifiable when 1.7% precision boost outweighs 15x longer training
    - Requires GPU acceleration for practical deployment

Critical Insight: The 0.1 contamination level (10% assumed fraud rate) outperformed the dataset's actual 27% fraud prevalence, suggesting label noise or detection threshold calibration benefits from conservative anomaly estimates.