

Anomaly Detection Models Report (SECOM Dataset)

This report summarizes experiments with three anomaly detection models: **Isolation Forest**, **One-Class SVM**, and **Auto Encoder**, evaluated through 10-fold cross-validation.

Parameter Tuning Methodology

- Selective parameter tuning focused on high-impact hyperparameters
- **Common framework:**

```
# KFold cross-validation template
kf = KFold(n_splits=10, shuffle=True, random_state=42)
results = []
for train_index, test_index in kf.split(df_normal):
    # Data preparation with anomaly injection
    X_train = df_normal.iloc[train_index]
    X_test = pd.concat([df_normal.iloc[test_index], df_anomalous])
    # Model training & evaluation
```

1. Isolation Forest

Key Parameters Tuned

- **contamination** (0-0.4 range)
- **n_estimators** (100-1000 trees)
- **max_features** (0.0-1.0 ratio)

Optimization Breakthrough

- **Best combination:** **contamination=0.1** + **n_estimators=900** + **max_features=0.1**
- Performance progression with all other parameters fixed at default values:

Contamination	F1 Score	Precision	Accuracy
0.1	0.280	0.559	0.600
0.2	0.374	0.503	0.585
0.3	0.447	0.490	0.577

Final Metrics:

F1 Score: 0.702 | Precision: 0.624 | Accuracy: 0.602 | Recall: 0.802

2. One-Class SVM

Optuna Hyperparameter Search

- 50 trials exploring kernels (rbf/linear/poly), nu (0.01-0.5), and gamma strategies
- Best parameters:**

```
{'kernel': 'rbf', 'nu': 0.181, 'gamma_type': 'scale'}
```

Performance Summary

1 is anomaly, 0 is normal

F1 Score: 0.421 | Precision: 0.562 | Accuracy: 0.613

Classification Report:

	precision	recall	f1-score
0.0	0.63	0.81	0.71
1.0	0.56	0.34	0.42

3. Auto Encoder

Architecture Experiments

Configuration	F1 Score	Precision	Accuracy
[256,128,64], 75 epochs	0.503	0.477	0.563
[256,128,64], 50 epochs	0.504	0.500	0.583
[512,256,128], 0.2 contam	0.479	0.500	0.583
[128,64,32], batch_size=8	0.449	0.543	0.607

Key Findings:

- Shallower networks ([128,64,32]) outperformed deeper architectures in accuracy
- Lower batch sizes (8 vs 16) slightly improved precision (0.543 vs 0.539)
- Optimal contamination: 0.2 across most configurations

Conclusion & Recommendations

Performance Comparison

Model	F1 Score	Precision	Inference Speed
Isolation Forest	0.702	0.624	Fastest

Model	F1 Score	Precision	Inference Speed
Autoencoder	0.504	0.500	Moderate
One-Class SVM	0.421	0.562	Slowest

Operational Guidance:

- 1. **Isolation Forest** - Best balance of speed and performance for real-time systems
- 2. **Autoencoder** - Use when interpretability of deep features is valuable
- 3. **One-Class SVM** - Not recommended due to high compute cost and lower F1

Critical Insight: Anomaly-free training data with carefully tuned contamination (0.1-0.2) proved most effective across all models, contradicting initial assumptions about needing representative anomaly samples in training.