ML4Crypto Classification: Model Comparison Report

Overview
This report summarizes the results obtained from various machine learning models on the ML4Crypto 2024 dataset, with a focus on binary classification of bitstreams.

Dataset Description
- Features: ID, Bitstream (1024-bit length), Label (0 or 1 for random generator type).
- Preprocessing: Bitstreams are converted into sequences, reshaped for models like LSTM, and split into training and testing datasets.

Model Comparisons

Model	**Accuracy**	**Cross Validation Accuracy**
Logistic Regression	**0.54**	**0.4875**
SVM	**0.535**	**0.5215**
Random Forest	**0.52**	**0.5045**
MLP Neural Network	**0.5325**	**0.5089**
CNN	**0.4975**	**0.4981**
LSTM	**0.4925**	

Analysis and Observations
- Logistic Regression: simple one but performs well, though cross validation accuracy is low.
- SVM: Best one with cross validation accuracy.
- MLP: performs well and good choice for experimentation yet cross validation accuracy is just above the average.
- DNN: With deeper architecture accuracy drops more.
- LSTM: Tried as some literature suggests, however, it did not get good accuracy.

Conclusion

Based on cross-validation and comparative results, SVM achieved the highest performance.