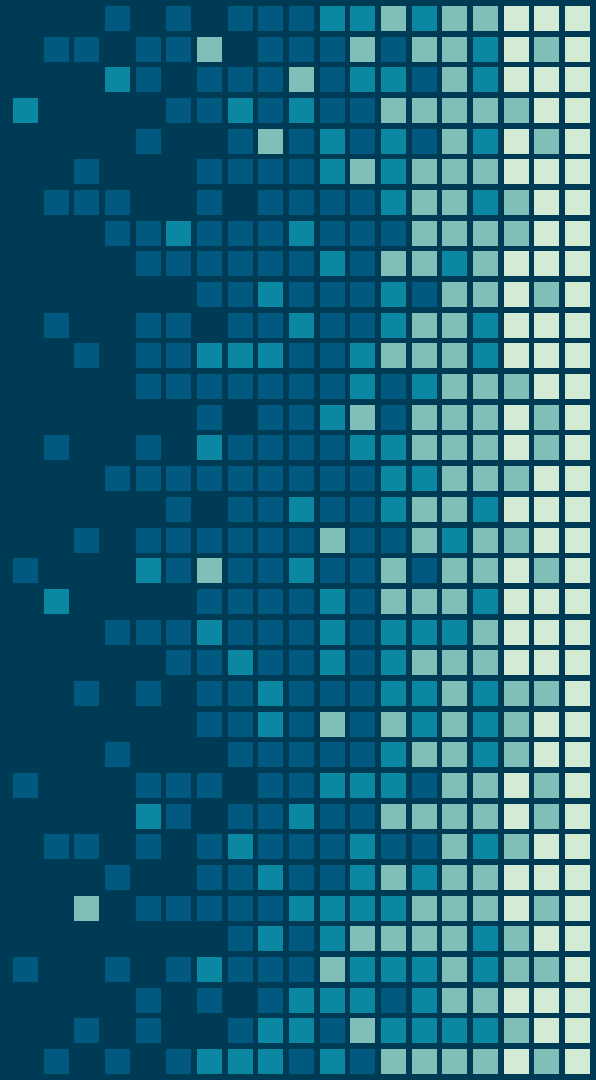


Detecting Economic Crime using Deep Autoencoder Neural Network

A capstone project for the Data Science Immersive Course
Conducted by General Assembly, Singapore

Amir Yunus



Economic Crime

Understanding the prevalence of fraud

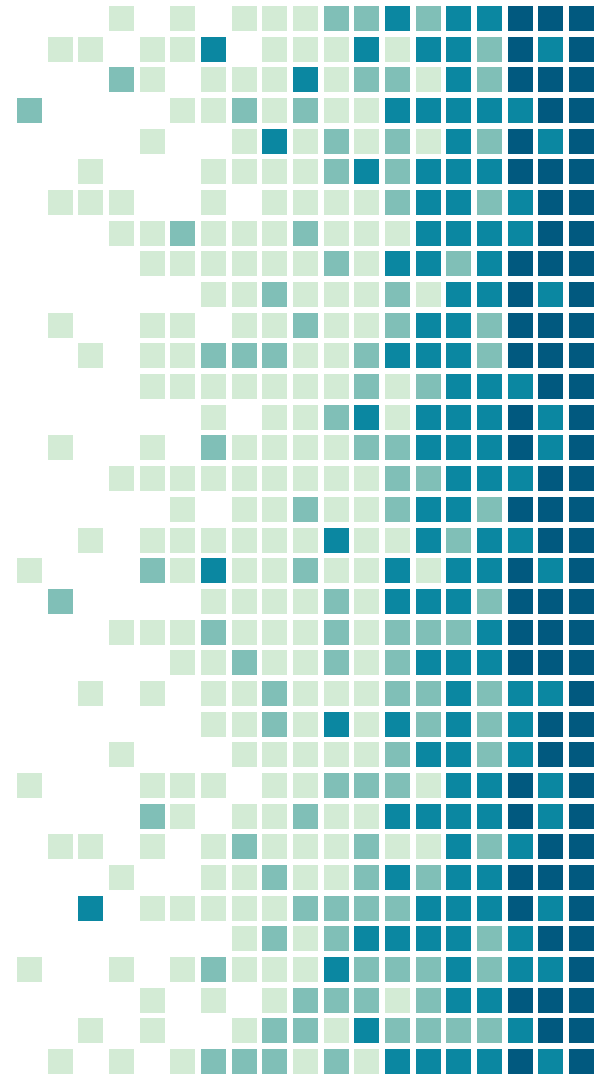


49% of respondents said that their
organisation has been a victim of fraud
or economic crime in the past 24
months



The median loss of a single financial statement fraud case is USD 150,000.

The duration from the fraud
perpetration till its detection was
18 months.



Ex-Woodbridge Group CEO Gets 25 Years in \$ 1.3b Fraud

16 Oct 2019

Singaporean Faces 30 Years for \$5m Cryptocurrency Fraud

10 Oct 2019

Atlantic City Mayor Resigned for Stealing \$87k

04 Oct 2019

Ex-Nissan CEO Charged with Fraud of \$140m

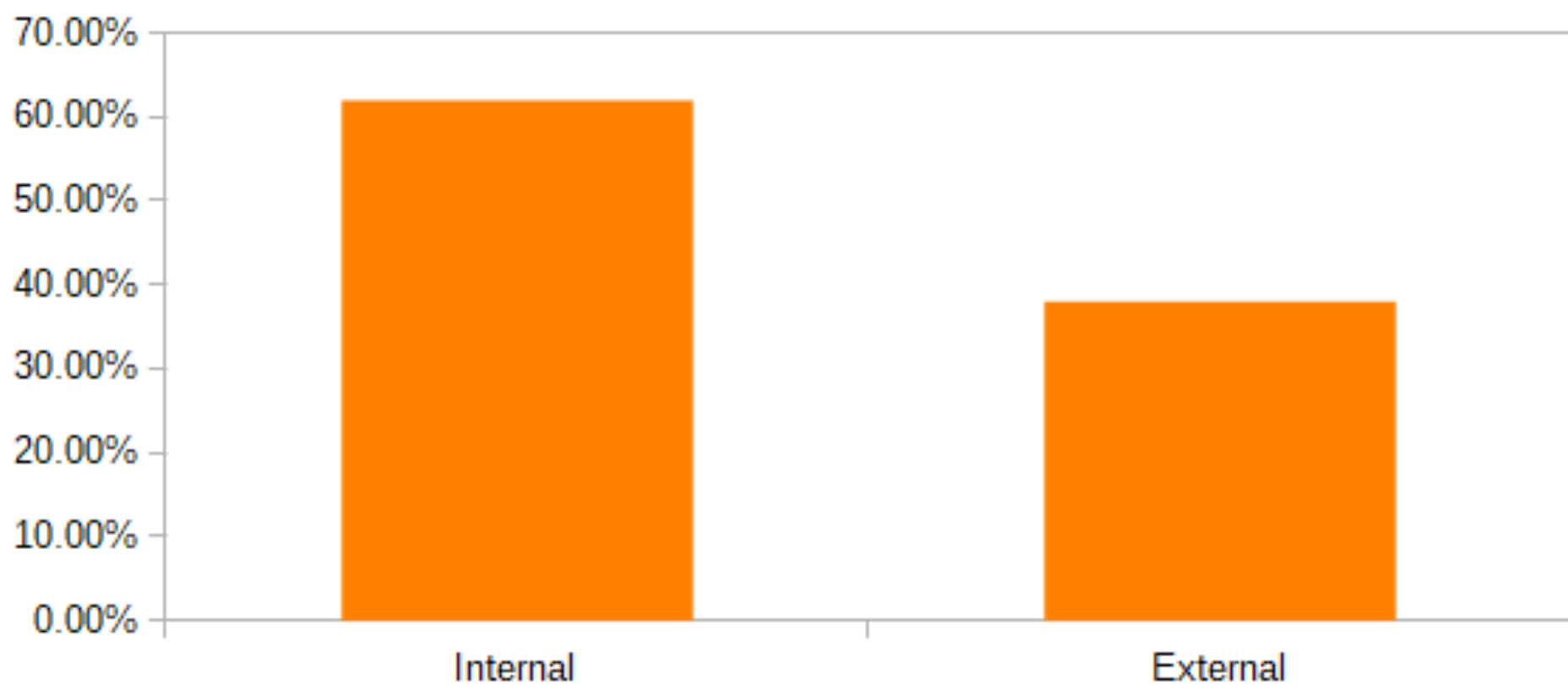
23 Sept 2019

Telemedicine CEO Pleads Guilty to \$424m Fraud Scheme

09 Sept 2019

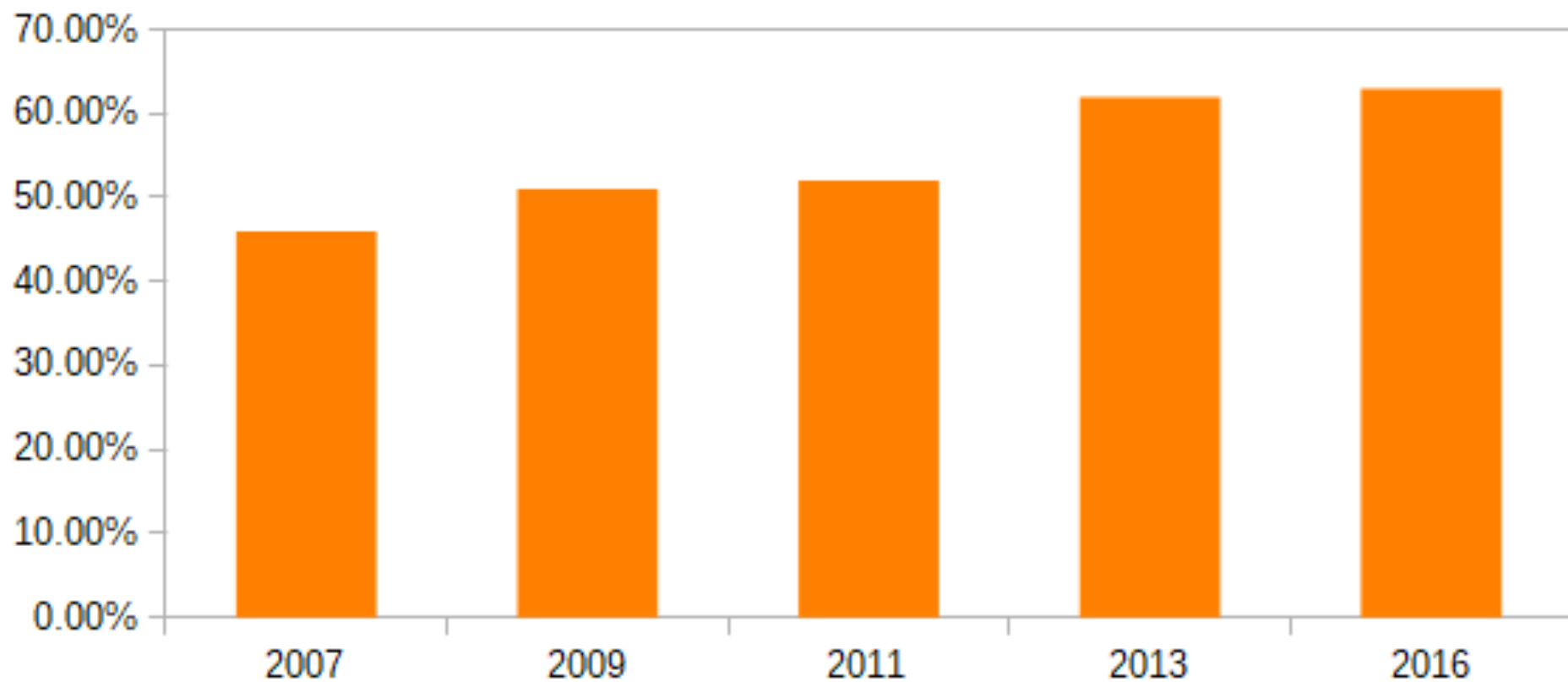
Relationship of Actor and Victimised Organisation

Internal actors are the main perpetrators of fraud



Fraction of Internal Actors Conducting Economic Crime

Internal actors committing fraud are becoming more common



Traditional Approach

Anomaly detection



Naïve Bayes

Bay, et al. & McGlohon, et al.

User Profiling

Kahn, et al. & Islam, et al.

Benford's Law

Debreceeny and Gray & Poh-Sun, et al.

Univariate and Multivariate Clustering

Jans, et al.

Euclidean Distance and Extreme Value Theory

Argyrou, et al.



Deep Learning Approach

Anomaly detection



Replicator Neural Network

Hawkins, et al. & Williams, et al.

Shallow Autoencoder Neural Network

Zhou and Paffenroth

Deep Autoencoder for Image & Video Forgery

Cozzolino and Verdoliva

Deep Autoencoder for Money Laundering

Paula, et al.

Adversarial Autoencoder for Accounting

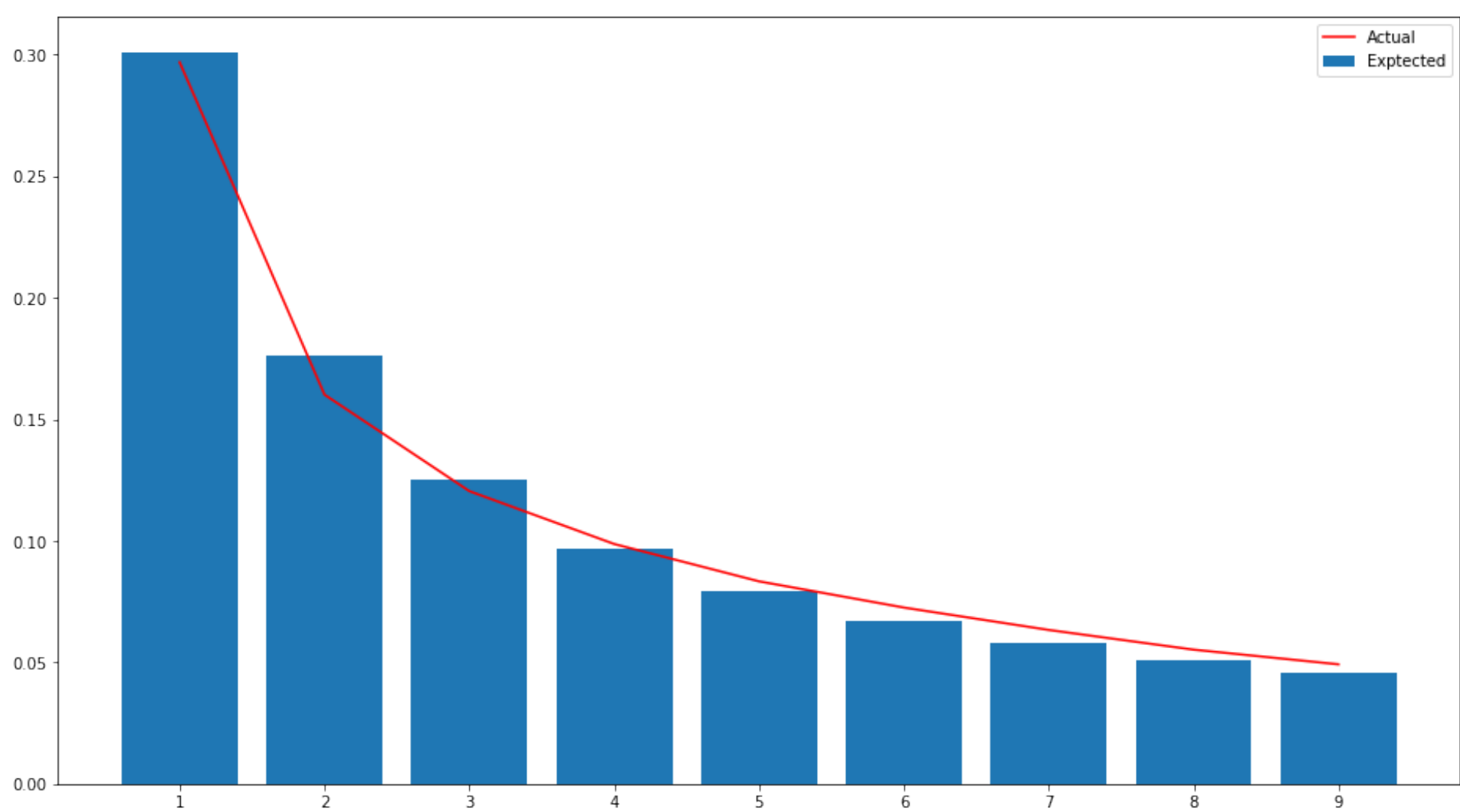
Schreyer, et al.

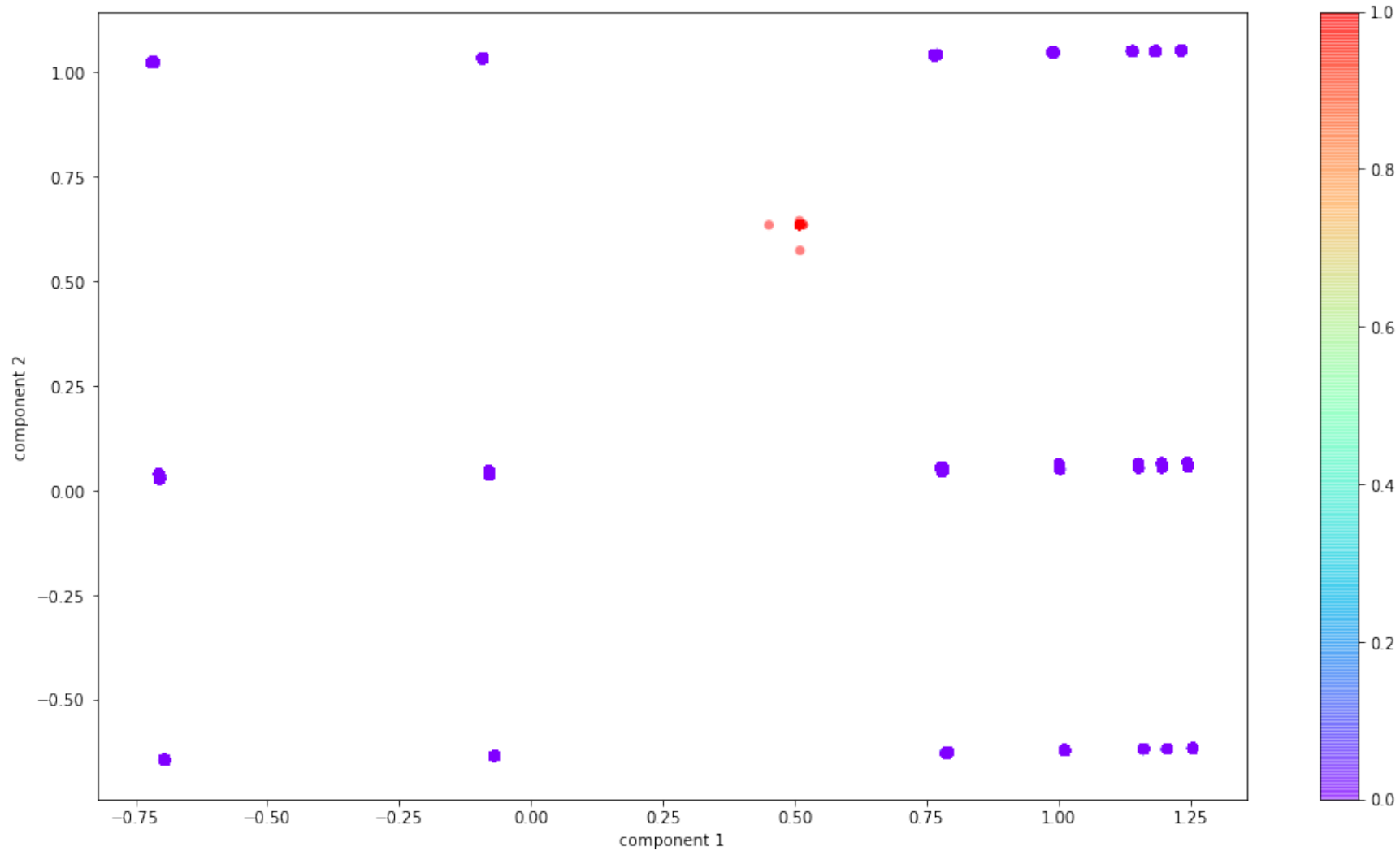


Evaluate Dataset

Detecting anomalies



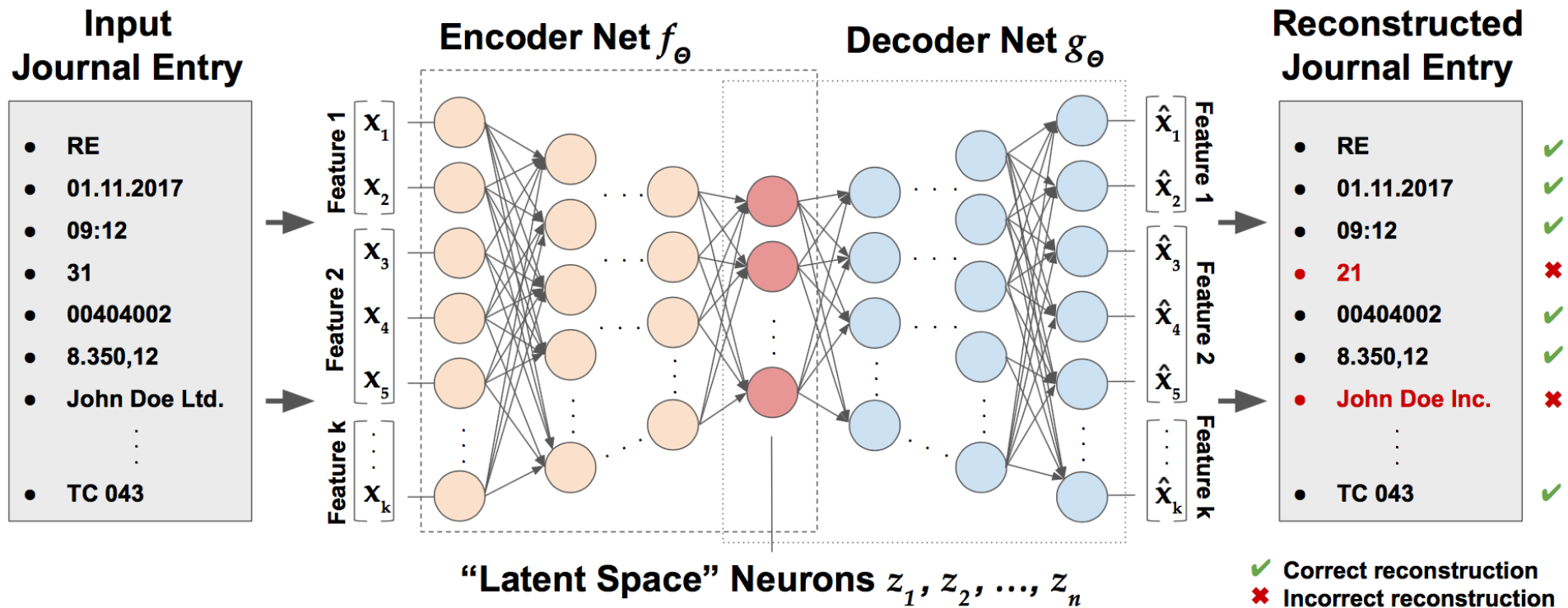




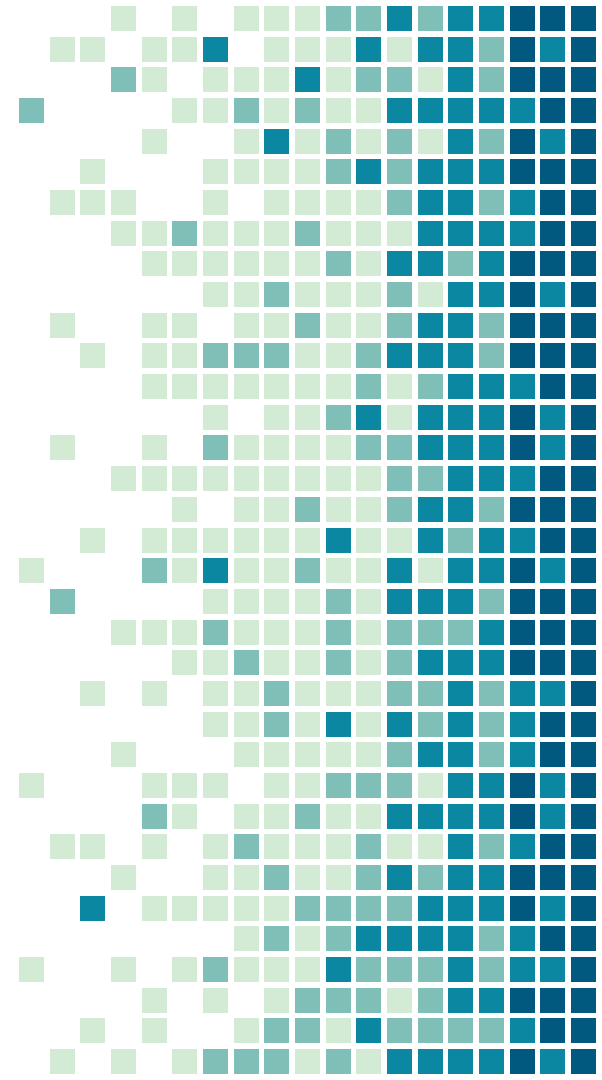
Architecture

Deep Autoencoder Neural Network

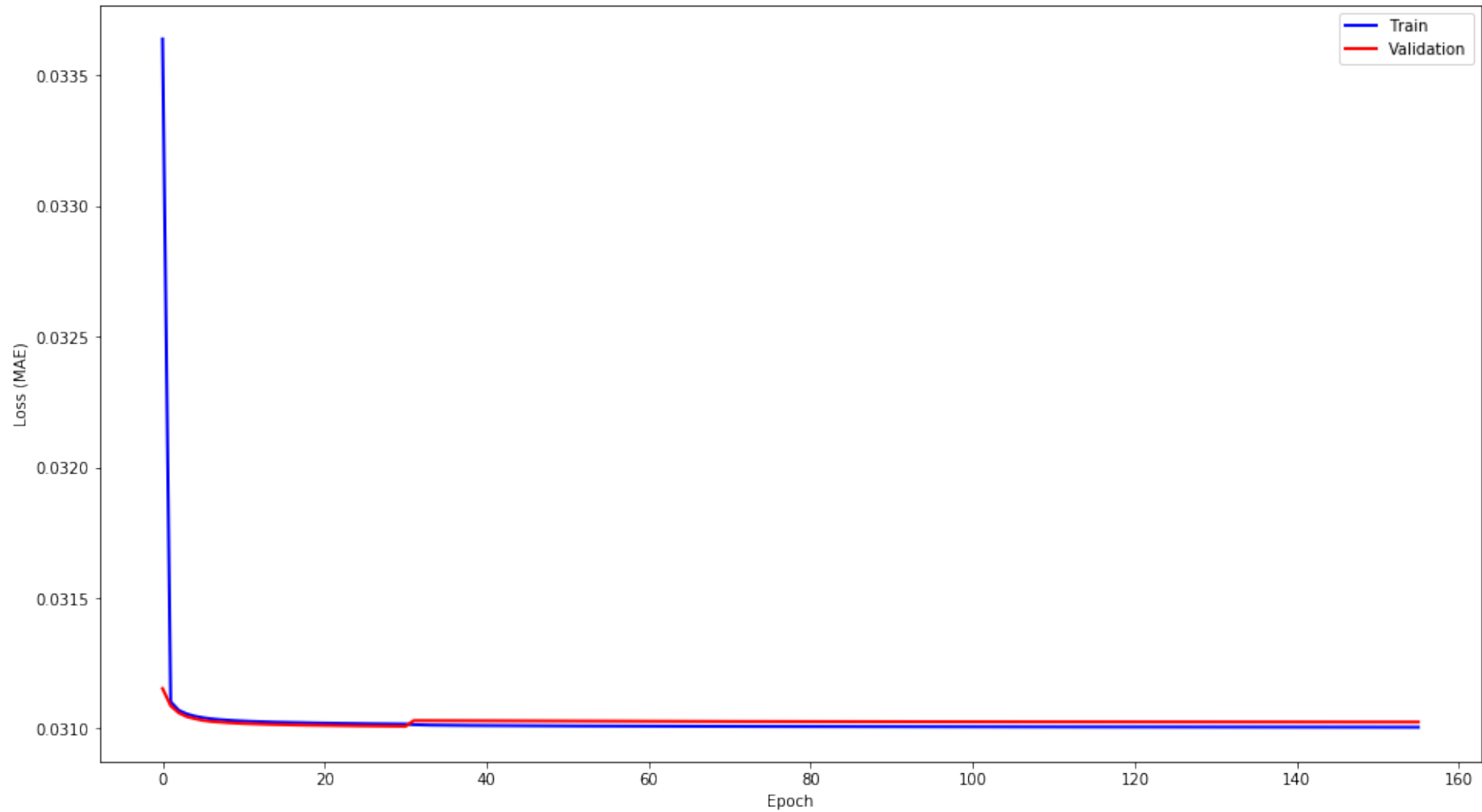




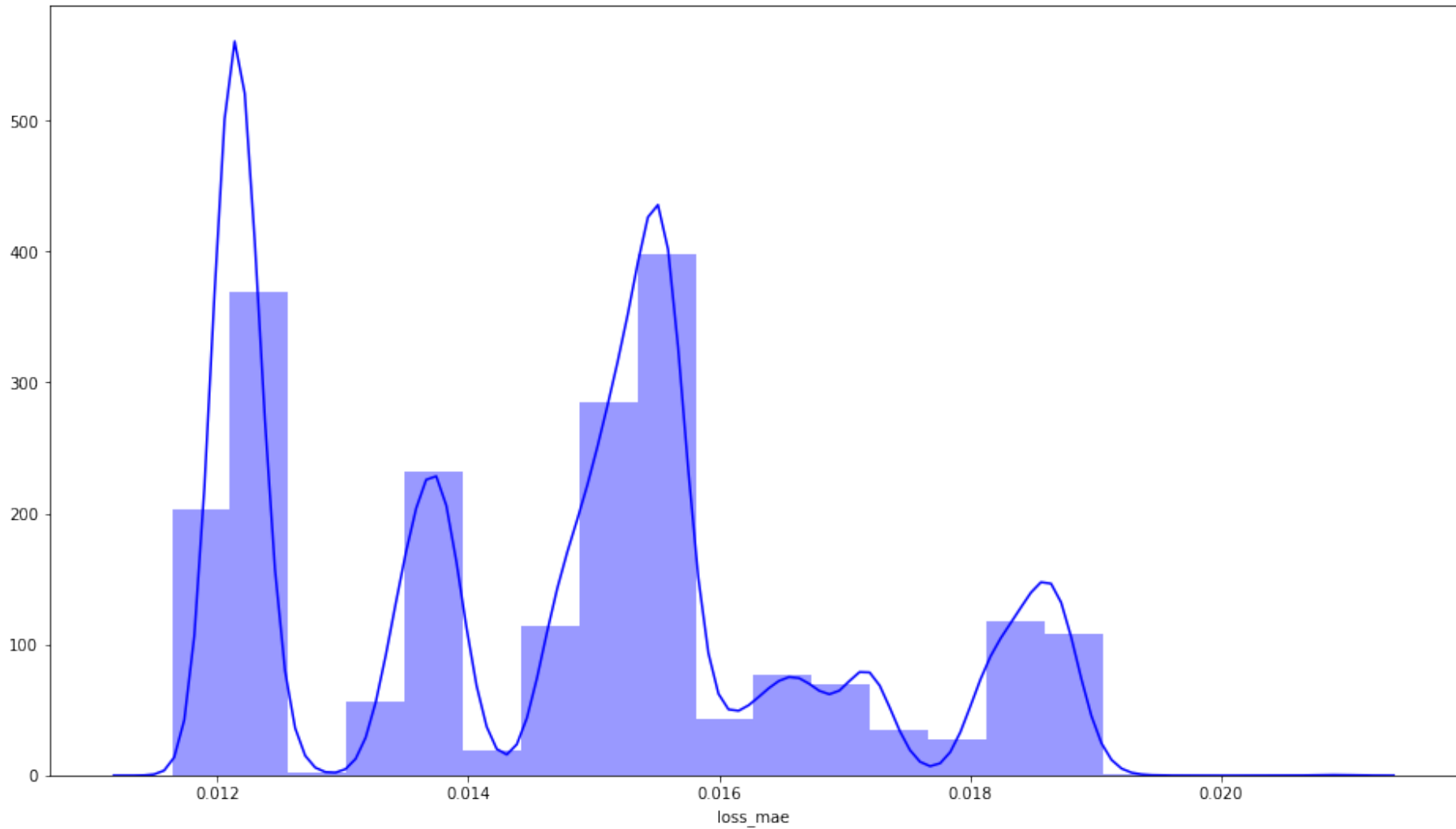
Layer (type)	Output Shape	Param #
=====		
input_1 (InputLayer)	(None, 618)	0
dense_1 (Dense)	(None, 512)	316928
dense_2 (Dense)	(None, 256)	131328
dense_3 (Dense)	(None, 128)	32896
dense_4 (Dense)	(None, 64)	8256
dense_5 (Dense)	(None, 32)	2080
dense_6 (Dense)	(None, 16)	528
dense_7 (Dense)	(None, 4)	68
dense_8 (Dense)	(None, 2)	10
dense_9 (Dense)	(None, 4)	12
dense_10 (Dense)	(None, 16)	80
dense_11 (Dense)	(None, 32)	544
dense_12 (Dense)	(None, 64)	2112
dense_13 (Dense)	(None, 128)	8320
dense_14 (Dense)	(None, 256)	33024
dense_15 (Dense)	(None, 512)	131584
dense_16 (Dense)	(None, 618)	317034
=====		
Total params: 984,804		
Trainable params: 984,804		
Non-trainable params: 0		



Model Loss



Reconstruction Error



Results for Deep Autoencoder

	Train Dataset	Test Dataset	Clean Dataset
True Positive	100	6	0
False Positive	6	0	6
False Negative	0	0	0
True Negative	532,903	33,308	33,301
Recall	1.00	1.00	N/A
Precision	0.943	1.00	N/A
F₁-Score	0.971	1.00	N/A

Baseline Evaluation

Comparing with Machine Learning



Deep Autoencoder vs. Unsupervised Machine Learning

	Deep Autoencoder	kMeans	Local Outlier Factor
True Positive	100	0	15
False Positive	6	330,243	92
False Negative	0	100	85
True Negative	532,903	202,666	532,817
Recall	1.00	0.00	0.150
Precision	0.943	0.00	0.140
F₁-Score	0.971	0.00	0.145



GUARD

Anti-Fraud System



Anti-Fraud System



Secure



Scan more datasets



../data/train.csv



../data/test.csv



Ready for a fraud scan?

Quickly scan your dataset, run a full scan, or
customise it the way you want

Quick Scan

Deep Scan

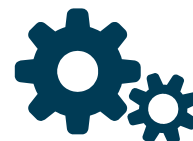
Custom Scan



My Account



Updates



Settings



Anti-Fraud System



Secure



Scan more datasets



../data/train.csv



../data/test.csv



Upload dataset

Supported formats (csv, txt)



Anti-Fraud System



Secure



Scan more datasets



../data/train.csv



../data/test.csv



Scanning for fraud

Estimated time left
10 minutes



Anti-Fraud System



Secure



Scan more datasets



../data/train.csv



../data/test.csv



Scan completed

Review the following 6 journal entries

#	BELNR	WAERS	BUKRS	KTOSL	PRCTR	BSCHL	HKONT	DMBTR	WRBTR
21912	19163	C8	C83	C3	C33	A1	B2	53,035	0
9139	29752	C8	C80	C3	C33	A1	B2	815,939	0
14395	377133	C8	C88	C3	C33	A1	B2	5,689,666	0
2997	241104	C8	C81	C3	C32	A1	B2	28,312	0
22545	90486	C8	C84	C3	C33	A1	B2	1,059,803	50,640
13157	515582	C8	C87	C3	C32	A1	B2	860,034	56,881

Export Results

Done



Anti-Fraud System



Secure



Scan more datasets



../data/train.csv



../data/test.csv

Custom Scan

Anomaly probability

0.0002

Threshold score

0.019

Patience level

100

Restore Default

Apply and Scan

Thank You!

Let's Connect

Amir Yunus

[linkedin.com/in/Amir-Yunus](https://www.linkedin.com/in/Amir-Yunus)

github.com/AmirYunus

Download the whitepaper at:
bit.ly/amir-yunus-capstone

