Operationalize Deep Learning Models for Fraud Detection with Azure Machine Learning

Francesca Lazzeri (@frlazzeri)

Jaya Mathew (@mathew_jaya)

Strata
DATA CONFERENCE

What is AI? – To sense, comprehend and act

Sense

Computer vision and audio processing, for example are able to actively perceive the world around them by acquiring and processing images, sounds and speech. The use of facial recognition at border control kiosks is one practical example of how it can improve productivity.

Comprehend

Natural language processing and inference engines can enable AI systems to analyse and understand the information collected. This technology is used to power the language translation feature of search engine results

Act

An AI system can take action through technologies such as expert systems and inference engines or undertake actions in the physical world. Auto-pilot features and assisted-braking capabilities in cars are examples of this

AI Technologies

Emerging AI technologies

- Computer Vision
- Audio Processing
- Natural Language Processing
- Knowledge Representation
- Machine Learning
- Expert Systems

Illustrative Solutions

Virtual Agents

Identity Analytics

Cognitive Robotics

Speech Analytics

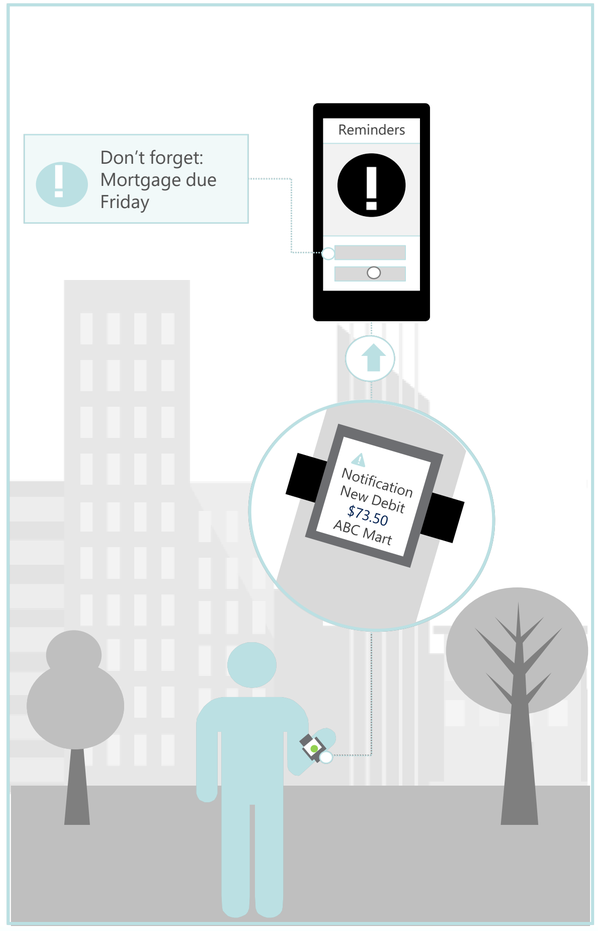
Recommendation Systems

Data Visualization

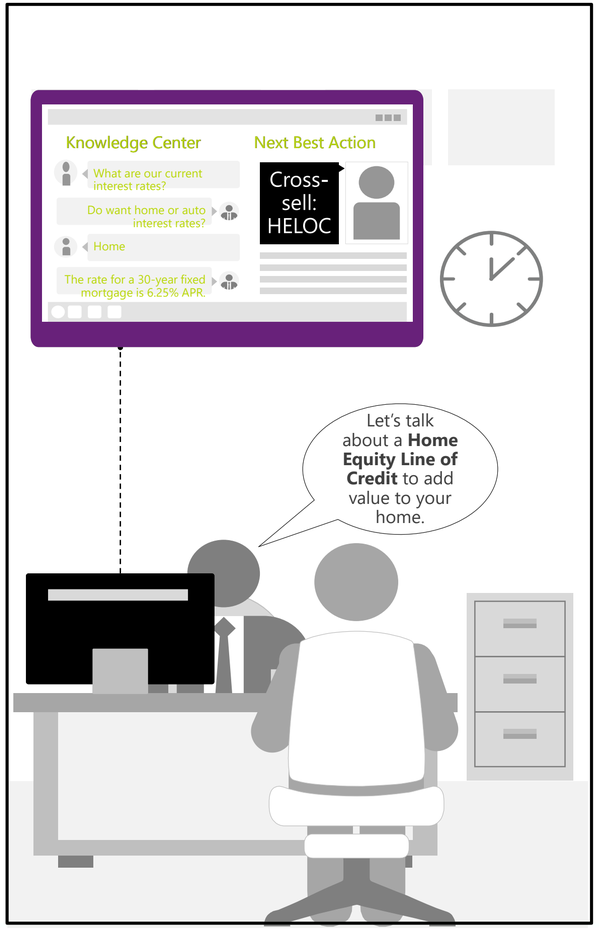
Source: Accenture: Why artificial intelligence is the future of growth, April 2016

AI enables differentiation within the Banking domain

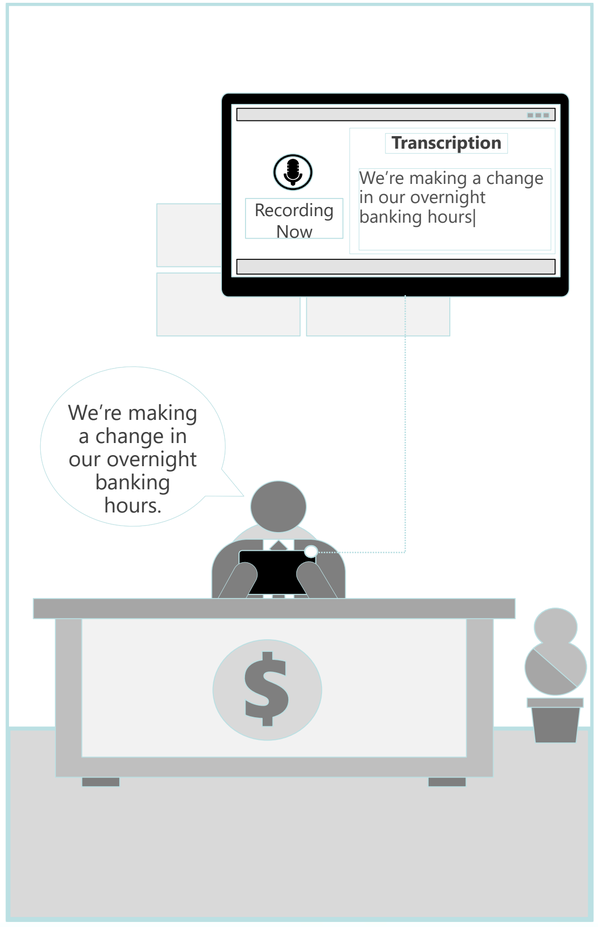
Engage your customers in new interactive ways



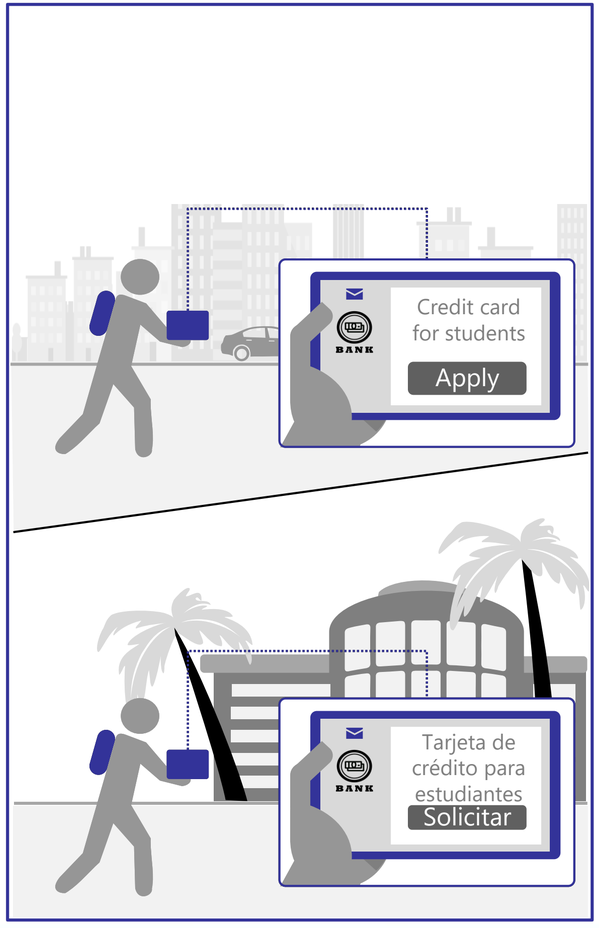
Empower your employees to innovate the customer experience



Optimize your operations to drive efficiencies across your business



Transform your products and services to become a trusted advisor



Online Fraud – when and what can be done to prevent it?

Evolving consumer purchase behavior

Customers are now looking at online shopping as a convenient option to visiting a physical store. This has led to an increasing popularity of e-commerce platforms.

What is happening with this change in consumer behavior?

This change has radically amplified the risk of online fraud for financial services companies and their customers. Failing to properly recognize and prevent fraud results in billions of dollars of loss per year for the financial industry.

What are companies doing to curb these losses?

This trend has urged companies to look into many popular artificial intelligence (AI) techniques, including deep learning for fraud detection.

Fraud detection

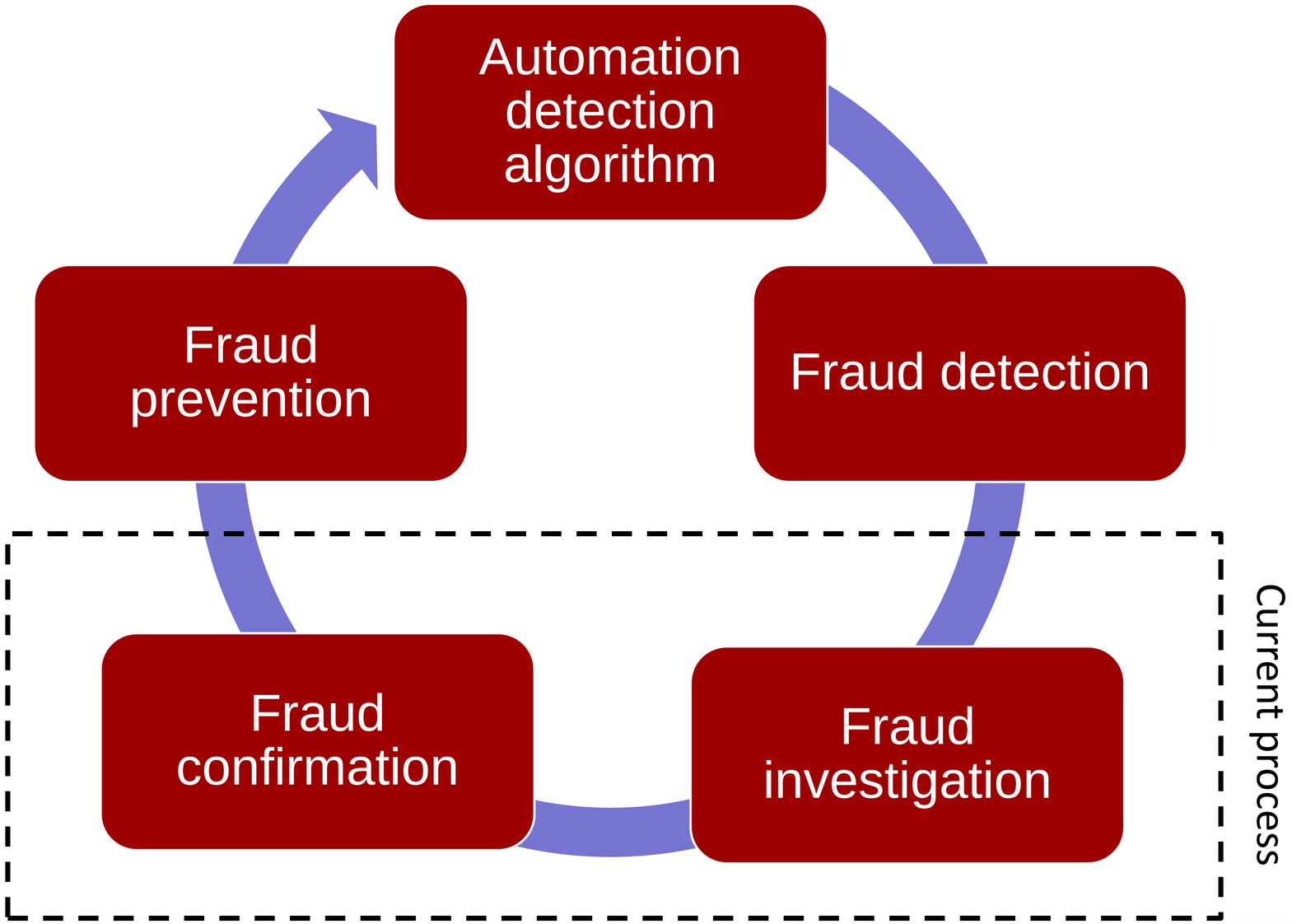
- A short definition of fraud is outlined in *Black's Law Dictionary*:

  "An act of intentional deception or dishonesty perpetrated by one or more individuals, generally for financial gain".

- This simple definition mandates a number of elements that must be addressed in order to prove fraud:

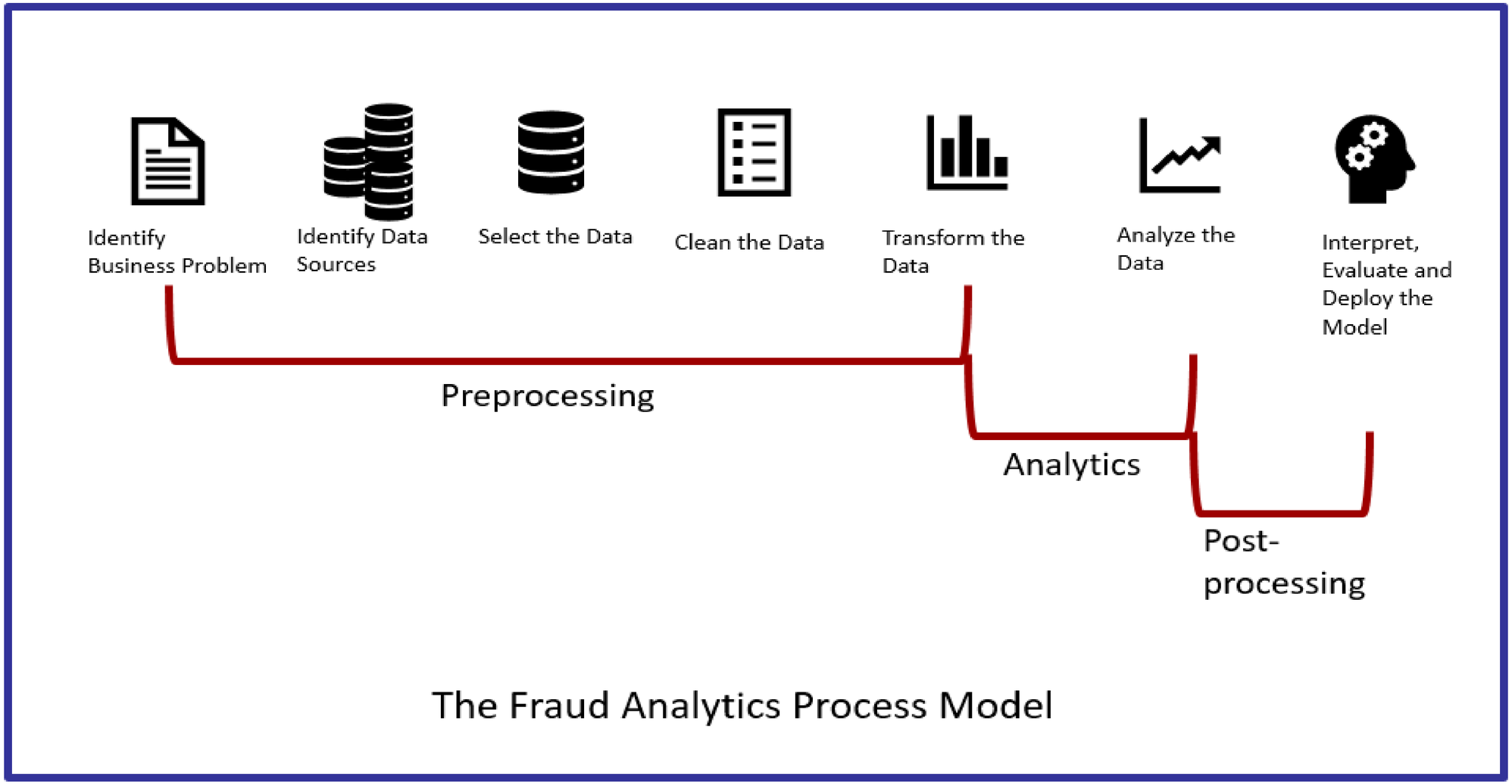
Sources: Black's Law Dictionary, "What Is FRAUD?";
Fraud and Fraud Detection: A Data Analytics Approach by Sunder Gee.

Fraud Cycle



Source: Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection by Wouter Verbeke, Veronique Van Vlasselaer, Bart Baesens.

The Fraud Analytics Process Model



Source: Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection by Wouter Verbeke, Veronique Van Vlasselaer, Bart Baesens.

Credit Card Fraud Detection

- Credit card companies “lose approximately seven cents per every hundred dollars of transactions due to fraud” (Andrew Schrage, Money Crashers Personal Finance, 2012).

- In credit card fraud there is an unauthorized taking of another's credit.

- Two subtypes can been identified, as described by Bolton and Hand (2002):
  - *Application fraud*, involving individuals obtaining new credit cards from issuing companies by using false personal information, and then spending as much as possible in a short space of time;

  - *Behavioral fraud*, where details of legitimate cards are obtained fraudulently and sales are made on a “Cardholder Not Present” basis.

Sources: Money Crashers Personal Finance by Andrew Schrage;
Statistical Fraud Detection: A Review by Richard J. Bolton, David J. Hand.

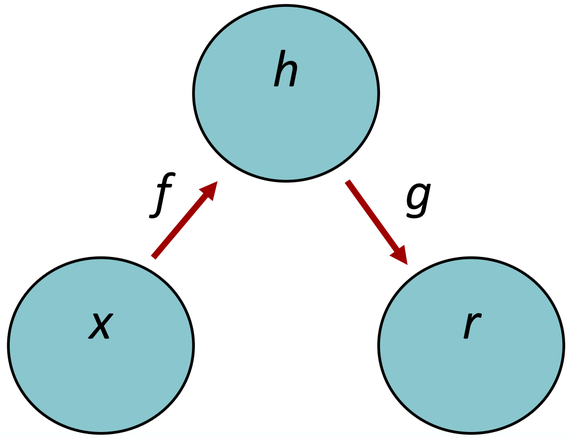
Comparison of Fraud Detection Techniques

Fraud Detection Techniques	Advantage	Disadvantage
K-nearest Neighbor Algorithm (KNN)	KNN method can be used to determine anomalies in the target instance and is easy to implement	KNN method is suitable for detecting frauds with the limitations of memory
Hidden Markov Model (HMM)	HMM can detect the fraudulent activity at the time of the transaction	HMM cannot detect fraud with a few transactions
Neural Network (NN)	NN have learned the previous behavior and can detect real-time credit card frauds	NN have many sub-techniques. So, if they pick-up this which is not suitable for credit card fraud detection, the performance of the method will decline
Decision Tree (DT)	DT can handle non-linear credit card transaction as well	DT have many type of input feature, DT can be constructed using different induction algorithm like ID3, C4.5 and CART. So, the cons are how to bring up induction algorithm to detect fraud as well. DT cannot detect fraud at the real time of transaction
Outlier Detection Method	Outlier detection detects the credit card fraud with lesser memory and computation requirements. This method works fast and well for large online datasets	Outlier detection cannot find anomalies accurately like other methods
Deep Learning (DL)	A key advantage of DL is the analysis and learning of a massive amount of unsupervised data. It can extract complex patterns	Now, deep learning is widely used in image recognition. No information to explain the other domains is available. The library of DL does not cover all algorithms

Source: Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine by Apapan Pumsirirat, Liu Yan.

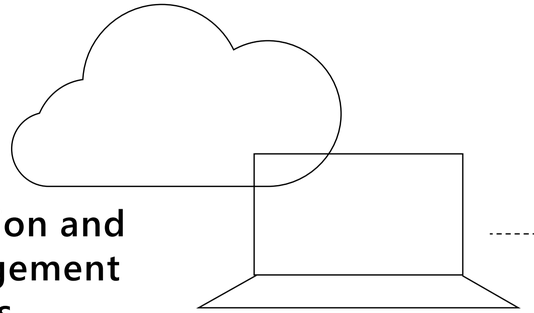
Our Approach

- An autoencoder is a neural network that is trained to attempt to copy its input to its output.
- Internally, it has a hidden layer ***h*** that describes a ***code*** used to represent the input.
- The network may be viewed as consisting of two parts:
  - an encoder function ***h = f(x)***
  - a decoder that produces a reconstruction ***r = g(h)***
- If an autoencoder succeeds in simply learning to set ***g(f(x)) = x*** everywhere, then it is not especially useful. Instead, autoencoders are designed to be unable to learn to copy perfectly:



Azure Machine Learning

AZURE MACHINE LEARNING SERVICES



Experimentation and Model Management Services

Notebooks
IDEs
Azure Machine Learning Workbench
VS Code Tools for AI

TRAIN & DEPLOY OPTIONS

AZURE

- Spark
- SQL Server
- Virtual machines
- GPUs
- Container services

ON-PREMISES

- SQL Server
- Machine Learning Server

EDGE

- Azure IoT Edge

Credit Card Fraud Detection Data Set

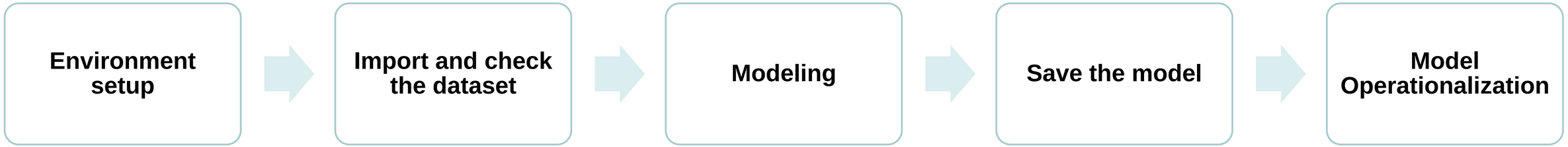
- Dataset used:
  - contains transactions made by credit cards in September 2013 by European cardholders.
  - presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions.
- Features *V1, V2, ... V28*: are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'.
- Feature *Time*: contains the seconds elapsed between each transaction and the first transaction in the dataset.
- Feature *Amount*: is the transaction Amount.
- Feature *Class*: is the response variable and it takes value 1 in case of fraud and 0 otherwise.

Time	V1	V2	V3	V4
0	-1.3598071336738	-0.0727811733098497	2.53634673796914	1.37815522427443
0	1.19185711131486	0.26615071205963	0.16648011335321	0.448154078460911
1	-1.35835406159823	-1.34016307473609	1.77320934263119	0.379779593034328
1	-0.966271711572087	-0.185226008082898	1.79299333957872	-0.863291275036453
2	-1.15823309349523	0.877736754848451	1.548717846511	0.403033933955121
2	-0.425965884412454	0.960523044882985	1.141109342232219	-0.168252079760302
4	1.22965763450793	0.141003507049326	0.045370773589944 9	1.20261273673594
7	-0.644269442348146	1.41796354547385	1.0743803763556	-0.492199018495015

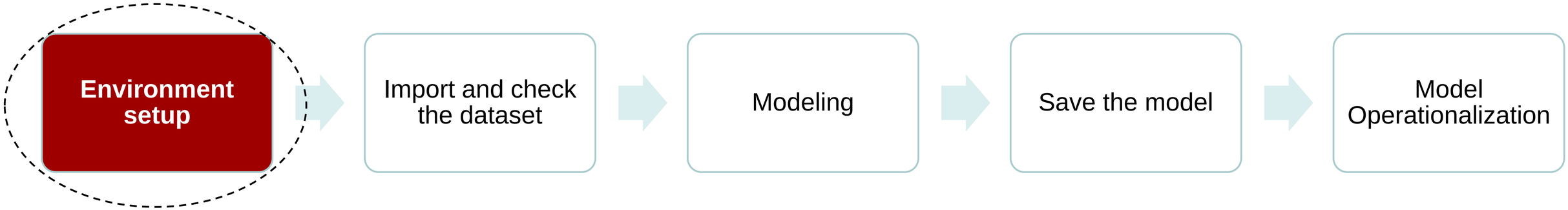
Data Source: https://www.kaggle.com/mlg-ulb/creditcardfraud

Demo

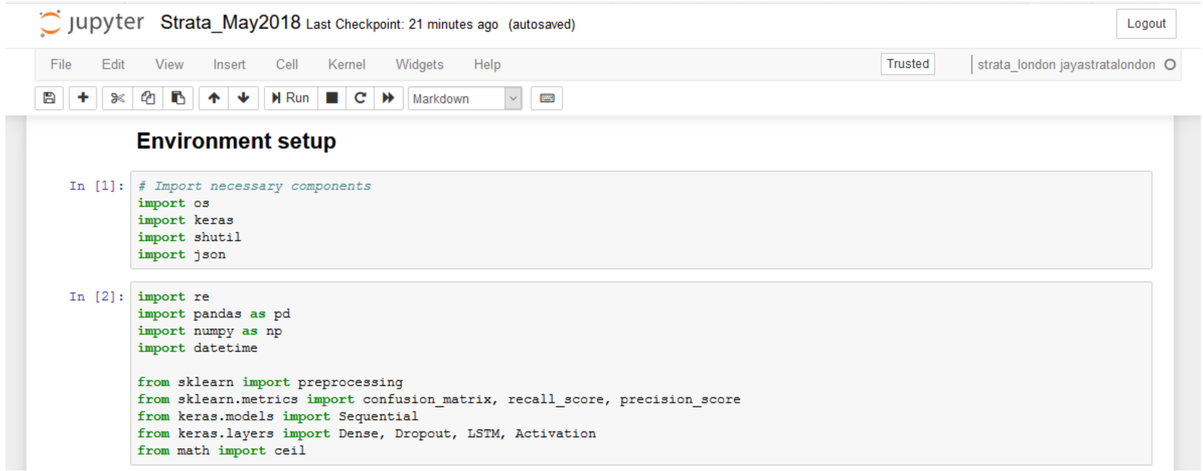
- The sample code was tested and run using the Jupyter notebook environment on a remote Azure VM (Standard F8s (8 vcpus, 16 GB memory)).
- The sample code is available at the following GitHub location: https://github.com/jayamathew/Codebase/tree/master/conferences
- The outline of the code is as follows:



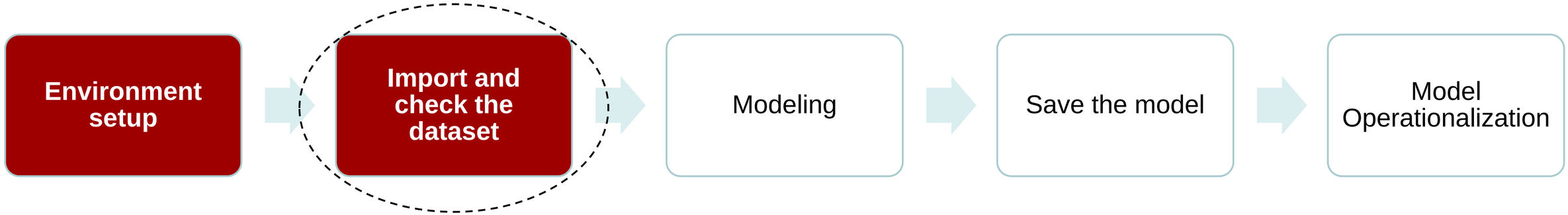
Demo



- Import the necessary libraries and provide credentials to access the data.



Demo



- Import the dataset and check the distributions of the variables.

jupyter Strata_May2018 Last Checkpoint: 25 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted strata_london jayastratalondon

Run Markdown

Import the Credit card data set

```
In [11]: # Check the path
         aml_dir
Out[11]: '/azureml-share/'

In [12]: # Ingest the dataset
         cc = pd.read_csv(aml_dir+'creditcard.csv')
```

After data ingestion from Blob, check to see the various columns and number of rows/columns of the dataset.

```
In [13]: # Check sample data
         cc.head(1)
```

Out[13]:

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V24	V25
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.066928	0.128539

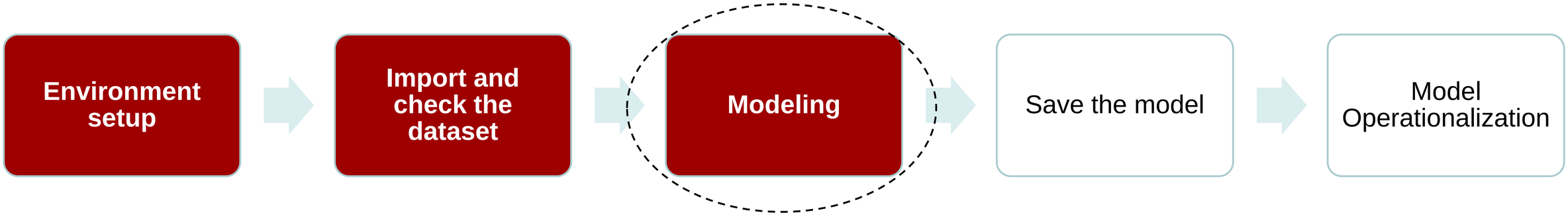
1 rows × 31 columns

```
In [14]: # Check the number of rows/columns
         cc.shape
Out[14]: (284807, 31)
```

Now that the data is properly imported, check the descriptive statistics of the columns in the dataset.

```
In [15]: # Check data statistics
         print(cc.describe())
```

Demo



- Build an autoencoder model and tune hyper parameters.

jupyter Strata_May2018 Last Checkpoint: 27 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted strata_london jayastratalondon

Run Markdown

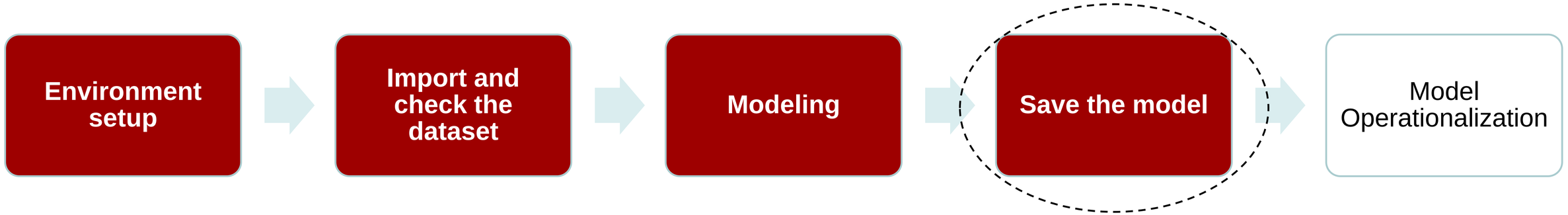
Define the framework for the autoencoder and then compile and fit using the training data.

In [24]:
```
# Define the encoded/decoder framework
input_dim = X_train.shape[1]
encoding_dim = 14

input_layer = Input(shape=(input_dim, ))
encoder = Dense(encoding_dim, activation="tanh", activity_regularizer=regularizers.l1(10e-5))(input_layer)
decoder = Dense(int(encoding_dim / 2), activation='tanh')(encoder)
decoder = Dense(input_dim, activation='relu')(decoder)
autoencoder = Model(inputs=input_layer, outputs=decoder)
```

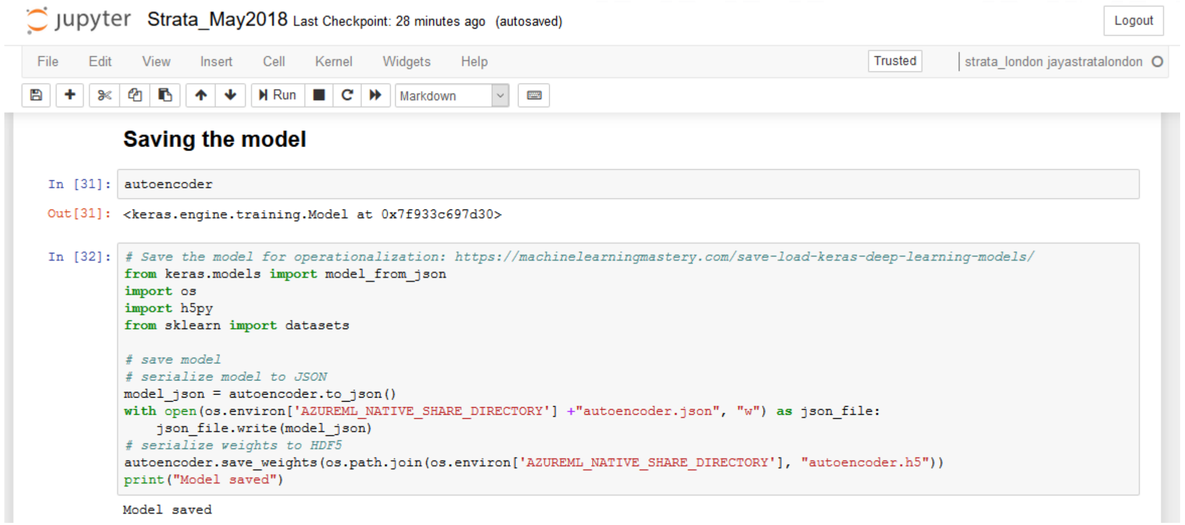
In [25]:
```
# Compile and fit the autoencoder
nb_epoch = 5
batch_size = 32
autoencoder.compile(optimizer='adam', loss='mean_squared_error', metrics=['accuracy'])

checkpointer = ModelCheckpoint(filepath="model.h5", verbose=0, save_best_only=True)

history = autoencoder.fit(X_train, X_train, epochs=nb_epoch, batch_size=batch_size, shuffle=True, validation_data=(X_test, X_t
```

```
Train on 199364 samples, validate on 85443 samples
Epoch 1/5
199364/199364 [==============================] - 15s 74us/step - loss: 0.8534 - acc: 0.5970 - val_loss: 0.7937 - val_acc: 0
.6356
```

Demo



- Save the best model for operationalization.



Demo

Environment setup → **Import and check the dataset** → **Modeling** → **Save the model** → **Model Operationalization**

- Create the necessary functions for model operationalization using Azure ML.

Links to get started with AI:

- Here are some links to get you started:
  - Azure ML: https://docs.microsoft.com/en-us/azure/machine-learning/preview/overview-what-is-azure-ml
  - Installation: https://docs.microsoft.com/en-us/azure/machine-learning/preview/quickstart-installation
  - Sample code for the Deep Learning use case: https://docs.microsoft.com/en-us/azure/machine-learning/preview/scenario-deep-learning-for-predictive-maintenance
  - Preconfigured Virtual Machines: https://azure.microsoft.com/en-us/services/virtual-machines/data-science-virtual-machines/
  - Data Source: https://www.kaggle.com/mlg-ulb/creditcardfraud
  - Blog Post by Venelin Valkov: https://medium.com/@curiousily/credit-card-fraud-detection-using-autoencoders-in-keras-tensorflow-for-hackers-part-vii-20e0c85301bd
  - GitHub location: https://github.com/jayamathew/Codebase/tree/master/conferences
  - Deep Learning Book by Ian Goodfellow, Yoshua Bengio, Aaron Courville: http://www.deeplearningbook.org/

Bibliographic References

- Black's Law Dictionary, "What Is FRAUD?"
- Fraud and Fraud Detection: A Data Analytics Approach by Sunder Gee
- Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection by Wouter Verbeke, Veronique Van Vlasselaer, Bart Baesens
- Money Crashers Personal Finance by Andrew Schrage
- Statistical Fraud Detection: A Review by Richard J. Bolton, David J. Hand
- Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine by Apapan Pumsirirat, Liu Yan

Thank you!

Francesca Lazzeri (@frlazzeri)

Jaya Mathew (@mathew_jaya)

Strata
DATA CONFERENCE