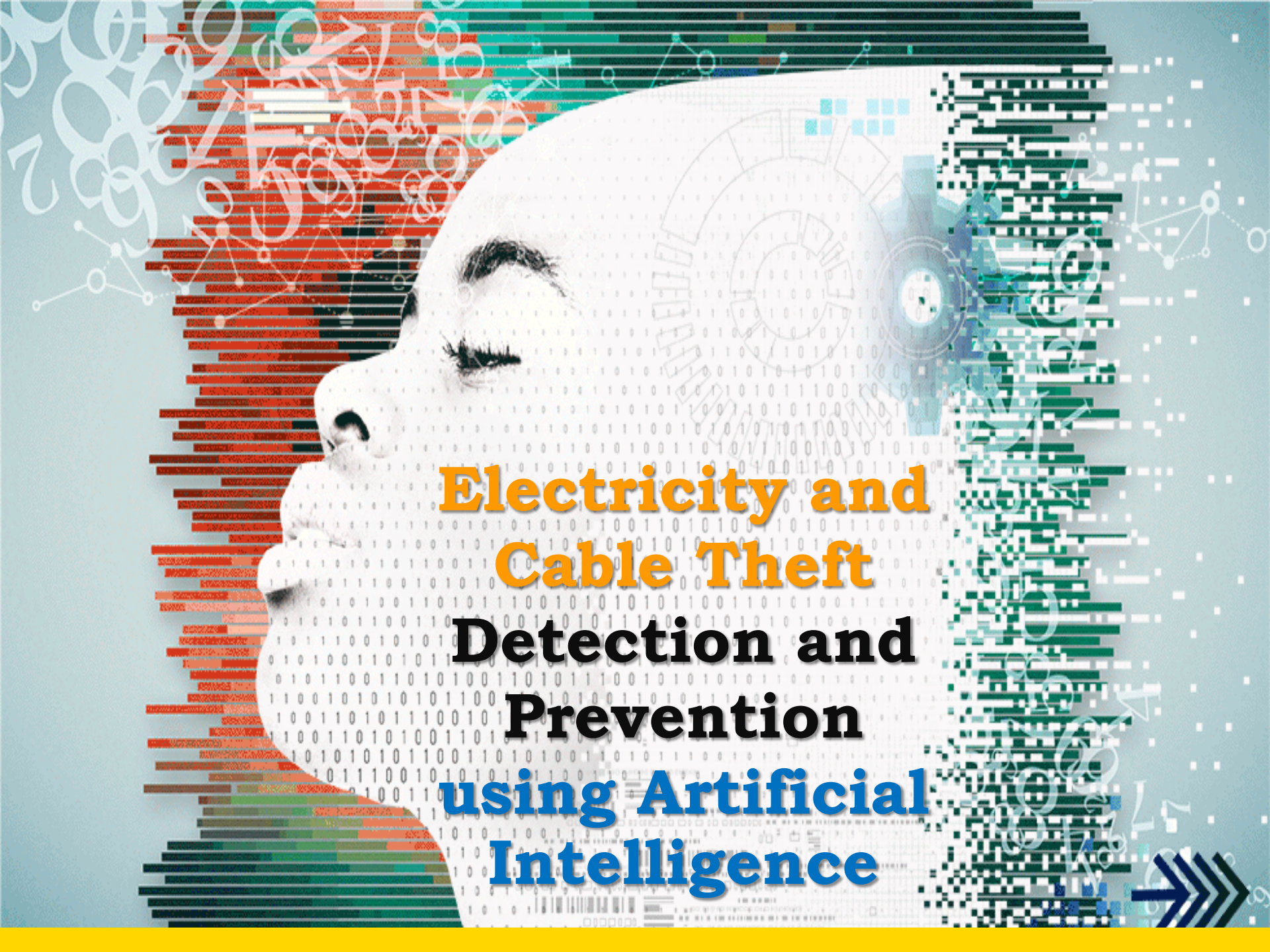
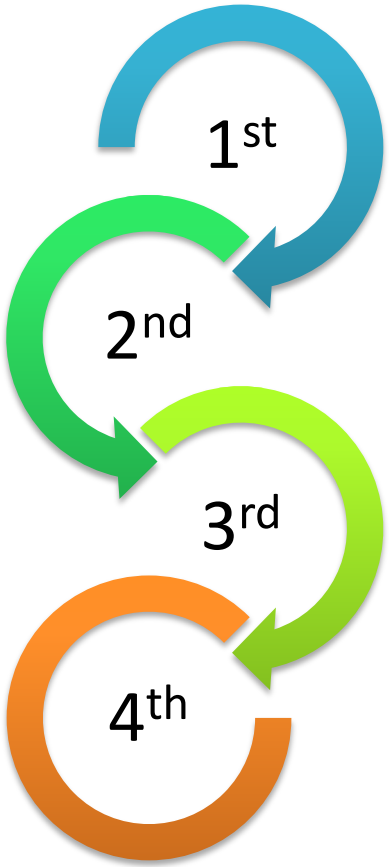

Electricity and
Cable Theft
Detection and
Prevention
using Artificial
Intelligence

# Table of Contents



# INDUSTRIAL 4.0

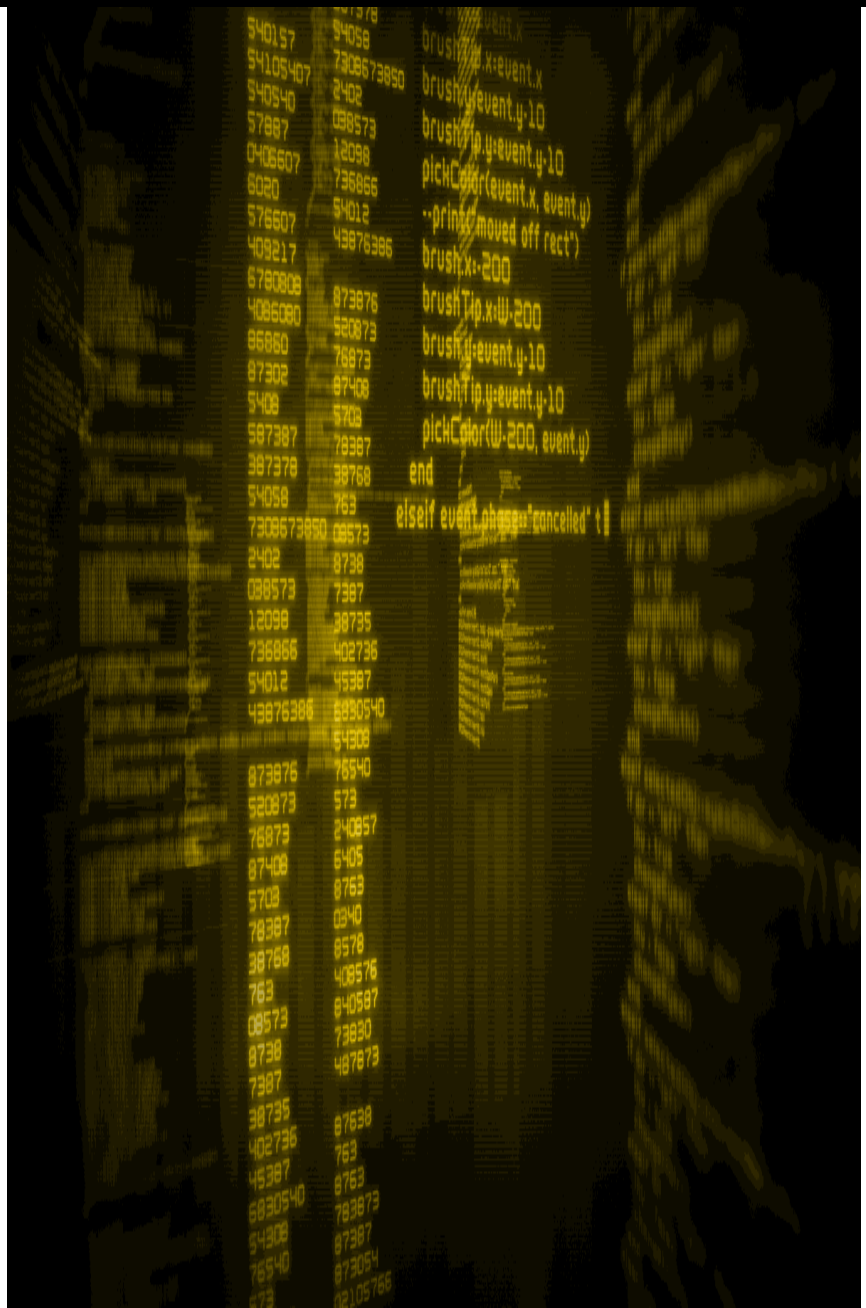



**Mechanization, water power, steam power**

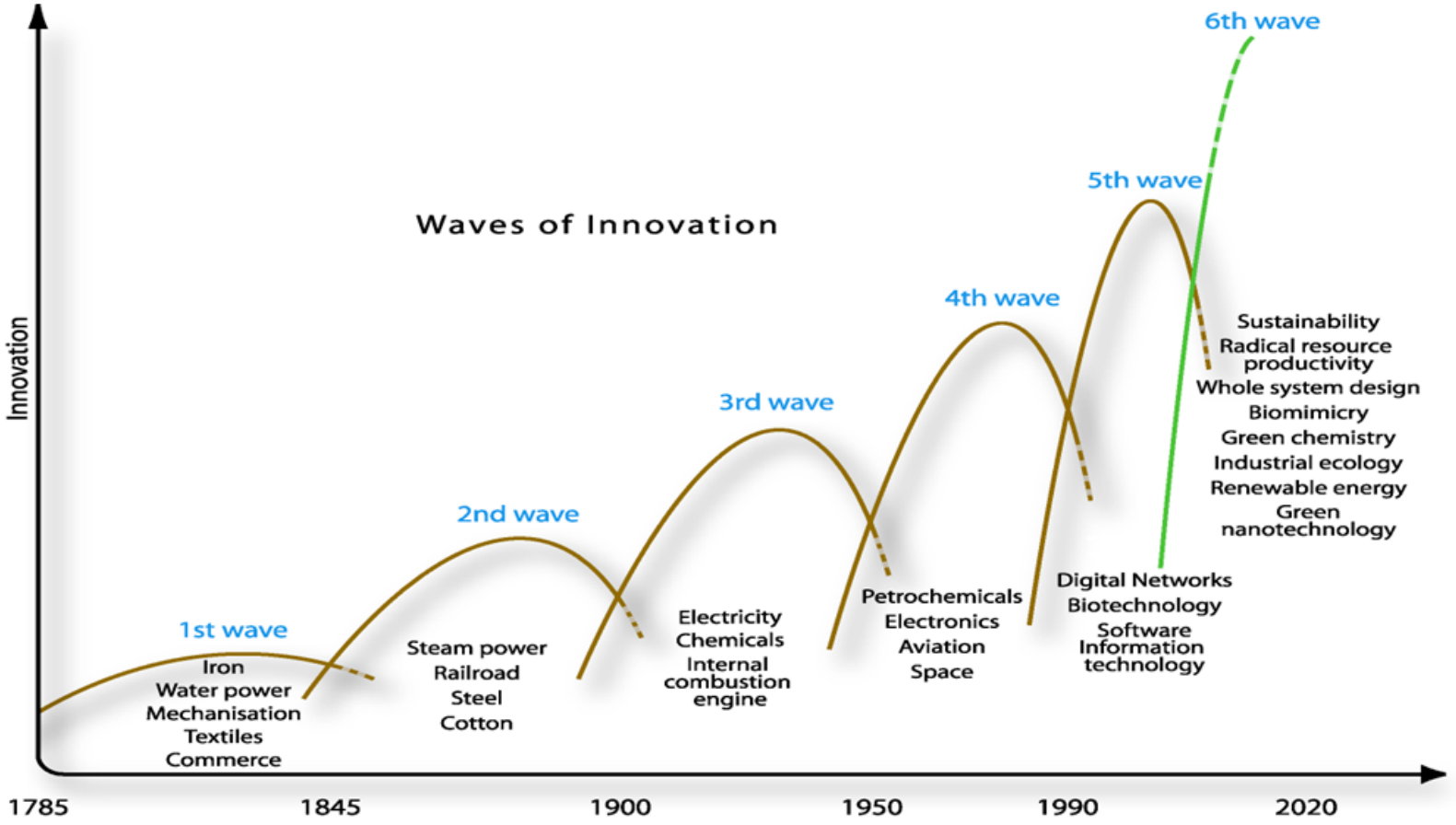
**Mass production, assembly line, electricity**

**Computer and automation**

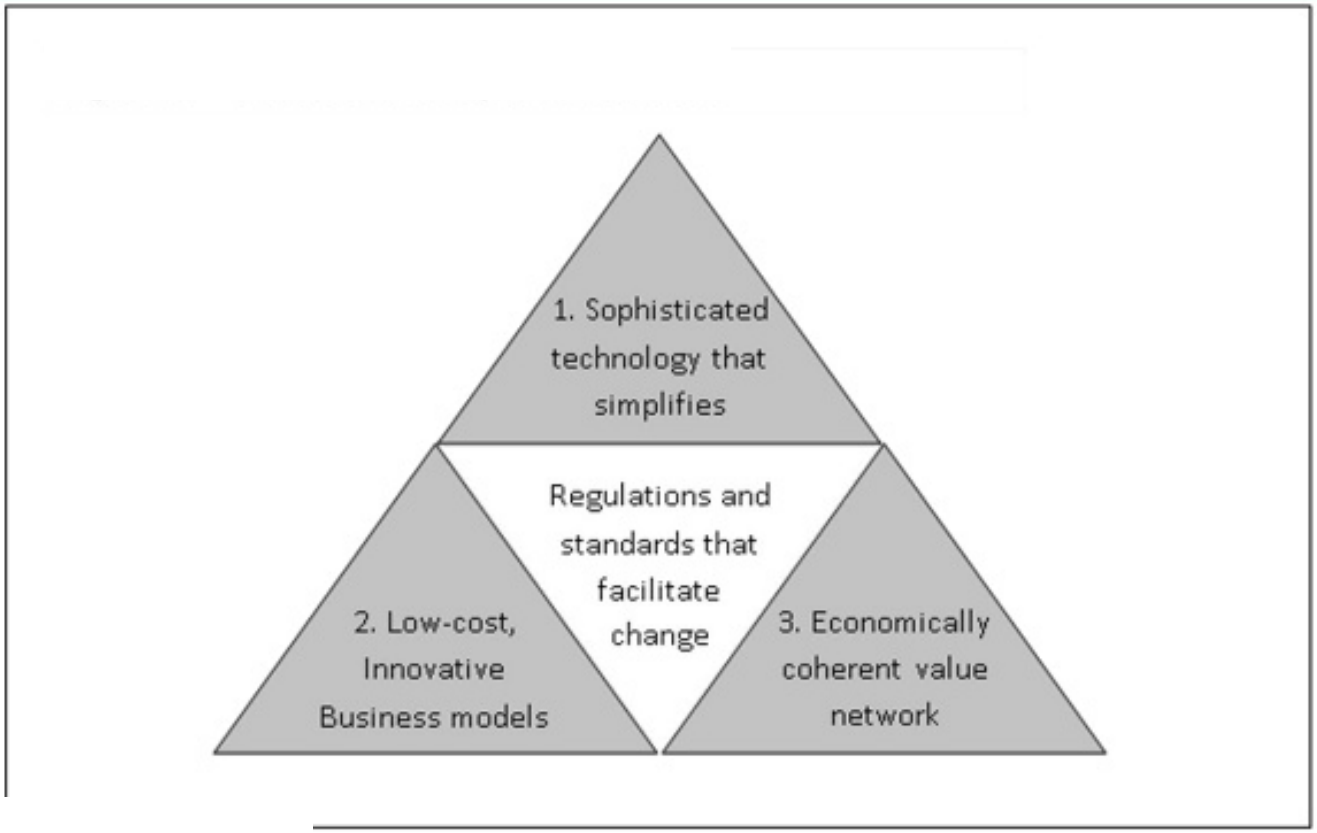
**Cyber physical systems**



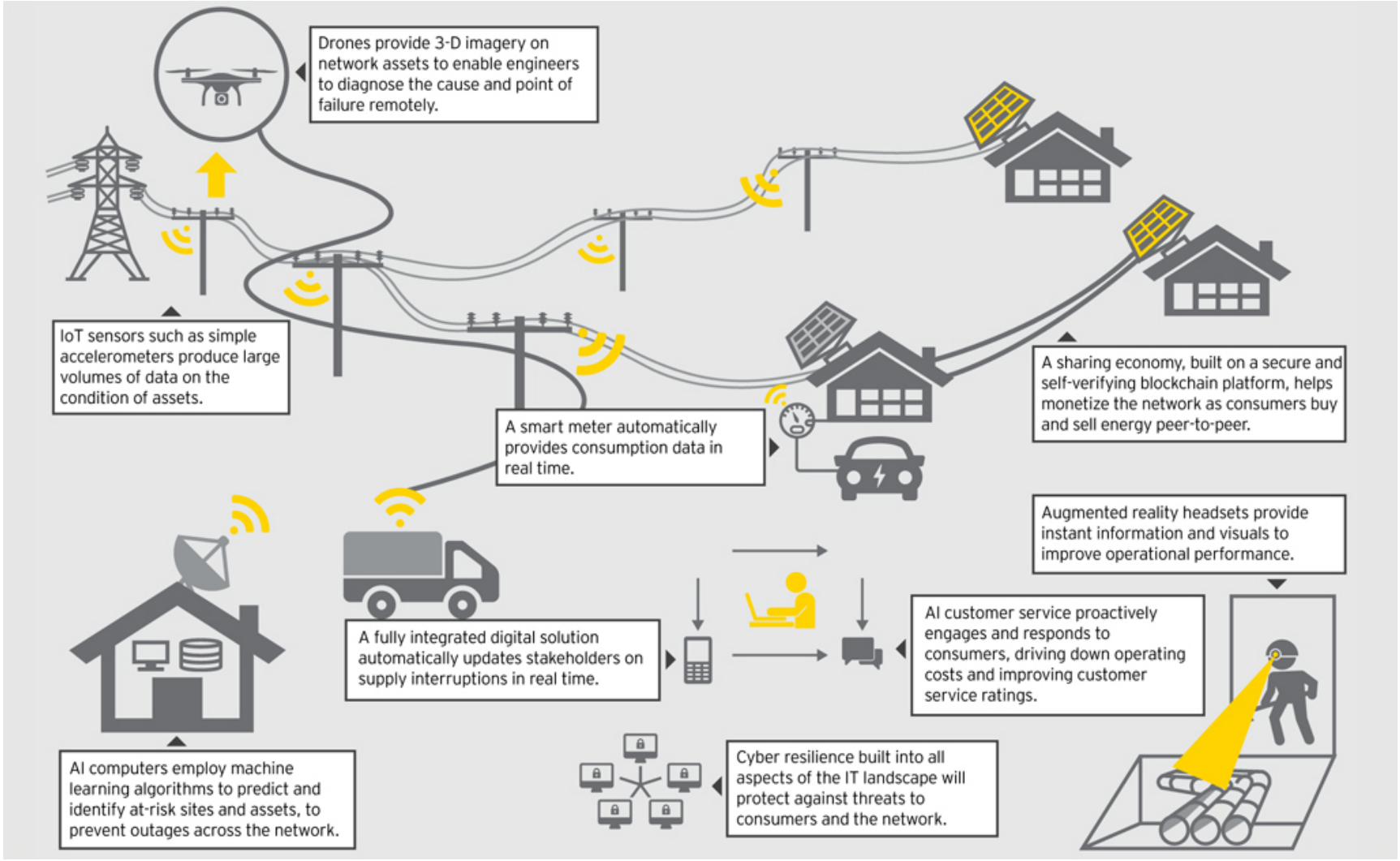
*Innovation is the central issue in economic prosperity.*
**Michael Porter, Harvard Business School**



# ELEMENTS OF DISRUPTIVE INNOVATION




*Christensen et al (2009) p. xx*


Drones provide 3-D imagery on network assets to enable engineers to diagnose the cause and point of failure remotely.
IoT sensors such as simple accelerometers produce large volumes of data on the condition of assets.
A smart meter automatically provides consumption data in real time.
A sharing economy, built on a secure and self-verifying blockchain platform, helps monetize the network as consumers buy and sell energy peer-to-peer.
Augmented reality headsets provide instant information and visuals to improve operational performance.
A fully integrated digital solution automatically updates stakeholders on supply interruptions in real time.
AI customer service proactively engages and responds to consumers, driving down operating costs and improving customer service ratings.
AI computers employ machine learning algorithms to predict and identify at-risk sites and assets, to prevent outages across the network.
Cyber resilience built into all aspects of the IT landscape will protect against threats to consumers and the network.


Disruptors vs Industry: Theft Detection

Municipal

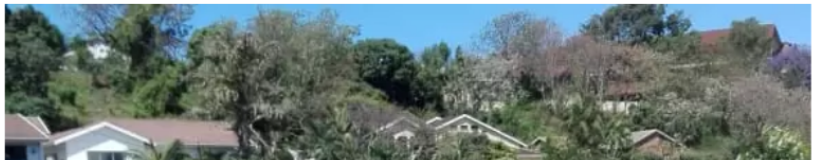
## Electricity theft a worry for eMba residents

Some residents are making a living from electricity theft

October 13, 2019

## Durban family left with no power for 26 days after cable theft

KWAZULU-NATAL / 27 SEPTEMBER 2019, 10:35AM / ZAINUL DAWOOD



## Cable theft cripples public library in Cape Town

2019-10-23 15:19

Jenna Etheridge

news24

Cable thieves have effectively crippled a library in Lwandle, 50km outside central Cape Town, leaving its users in the dark and unable to use the computers.

South Africa

## Soweto residents to take Eskom to court

## Stop illegal connections to stop outages, Eskom pleads

14

*ANA*

nt township from 'sliding furth
ections.

an / 23 October 2019 10:10

## Snakebites, theft and assault short-circuit green power development

Sub-Saharan Africa is a key market for new electricity, but hurdles are making project development on the continent challenging

# Over 2,500 held in last 18 months for power theft in Delhi

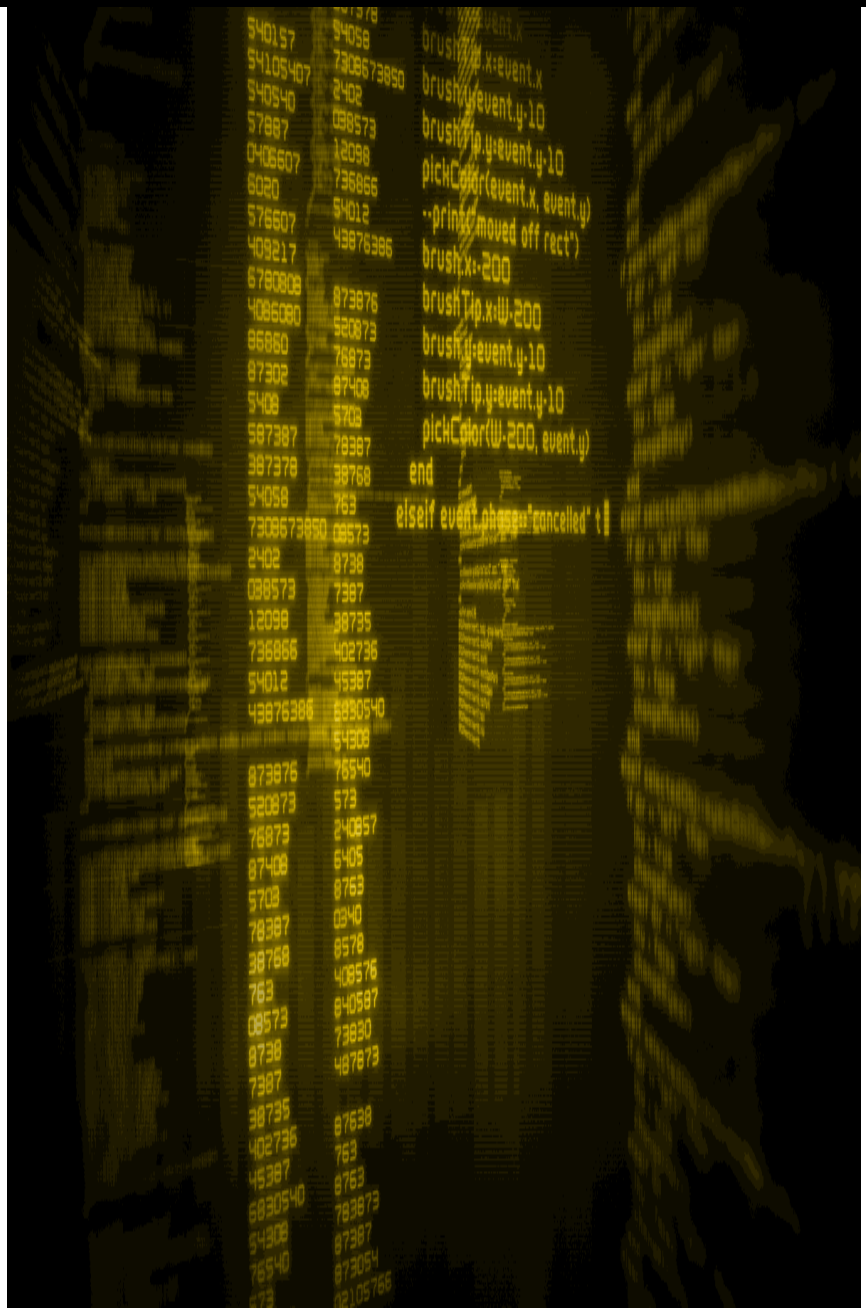
***As part of a crackdown on electricity theft that causes annual losses running into hundreds of crores of rupees, the power distribution companies in Delhi have filed over 5,500 complaints in the last one and half years leading to arrest of more than 2,500 violators.***

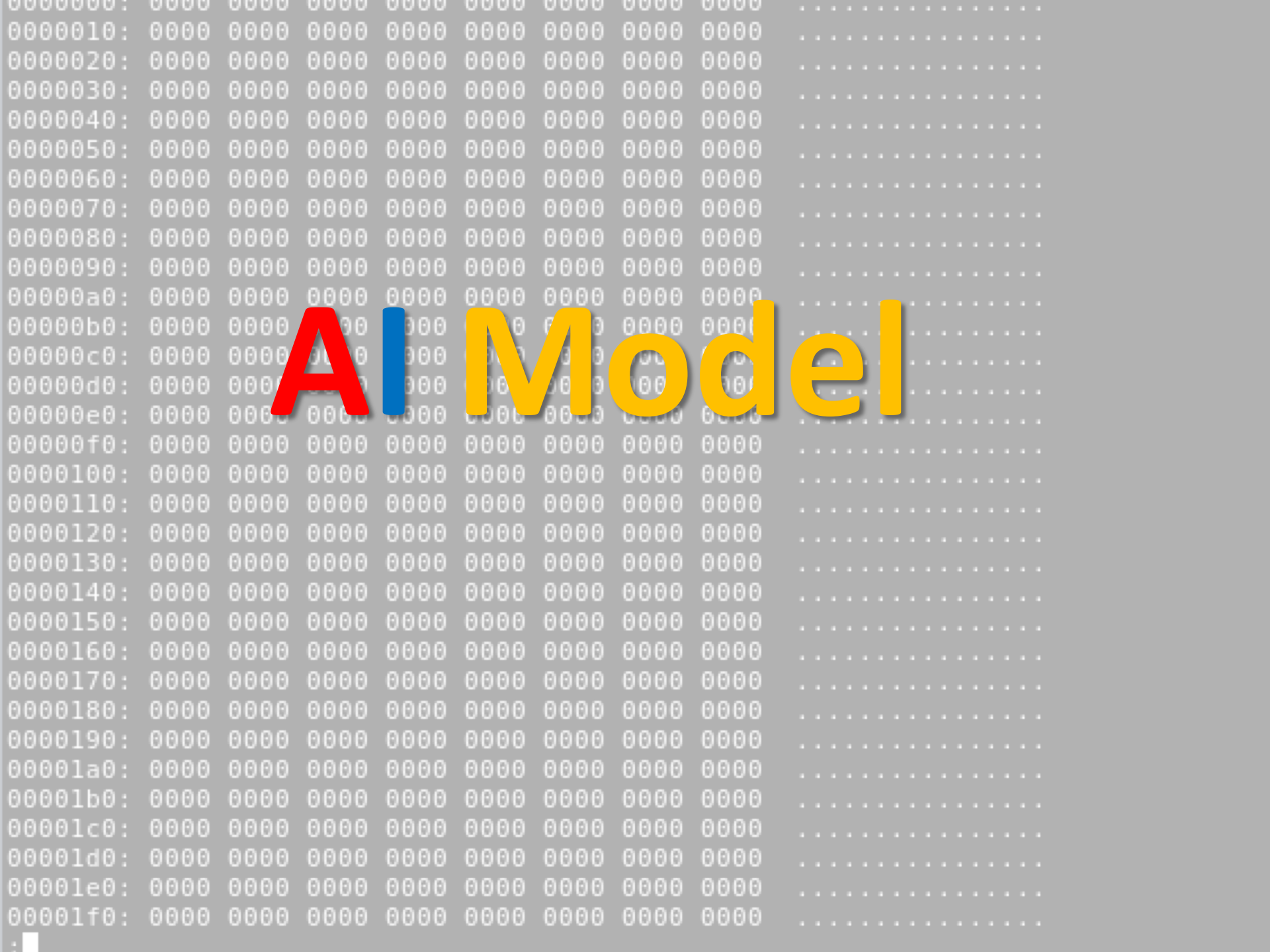
PTI | October 13, 2019, 17:52 IST

# THEFT DETECTION

Modern smart grids rely on advanced metering infrastructure (AMI) networks for monitoring and billing purposes. However, such an approach suffers from electricity theft.

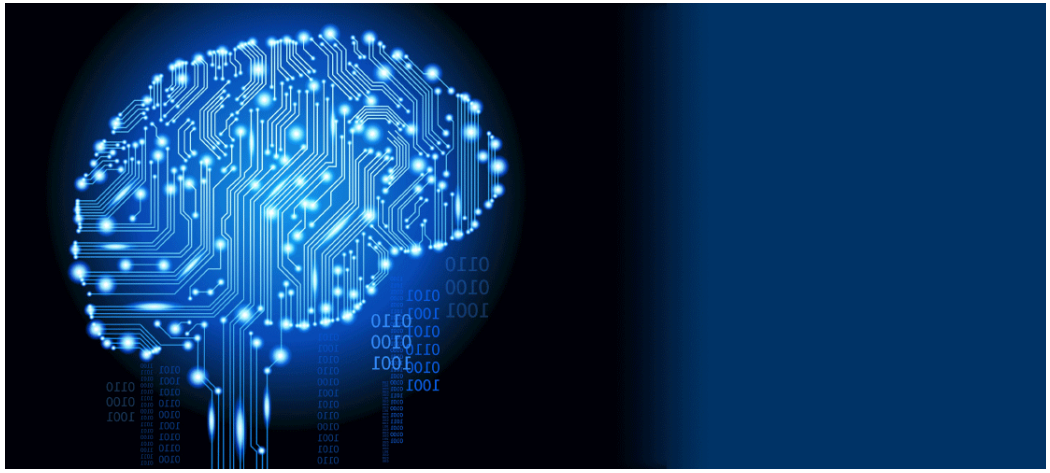
The AMI networks rely on smart meters located in the customer’s premises to regularly report their energy consumption. This approach has the potential to hinder traditional physical electricity theft attacks including line hooking or meter tampering.




0000000: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000010: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000020: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000030: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000040: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000050: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000060: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000070: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000080: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000090: 0000 0000 0000 0000 0000 0000 0000 0000 ................
00000a0: 0000 0000 0000 0000 0000 0000 0000 0000 ................
00000b0: 0000 0000 0000 0000 0000 0000 0000 0000 ................
00000c0: 0000 0000 0000 0000 0000 0000 0000 0000 ................
00000d0: 0000 0000 0000 0000 0000 0000 0000 0000 ................
00000e0: 0000 0000 0000 0000 0000 0000 0000 0000 ................
00000f0: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000100: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000110: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000120: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000130: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000140: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000150: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000160: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000170: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000180: 0000 0000 0000 0000 0000 0000 0000 0000 ................
0000190: 0000 0000 0000 0000 0000 0000 0000 0000 ................
00001a0: 0000 0000 0000 0000 0000 0000 0000 0000 ................
00001b0: 0000 0000 0000 0000 0000 0000 0000 0000 ................
00001c0: 0000 0000 0000 0000 0000 0000 0000 0000 ................
00001d0: 0000 0000 0000 0000 0000 0000 0000 0000 ................
00001e0: 0000 0000 0000 0000 0000 0000 0000 0000 ................
00001f0: 0000 0000 0000 0000 0000 0000 0000 0000 ................
AI Model

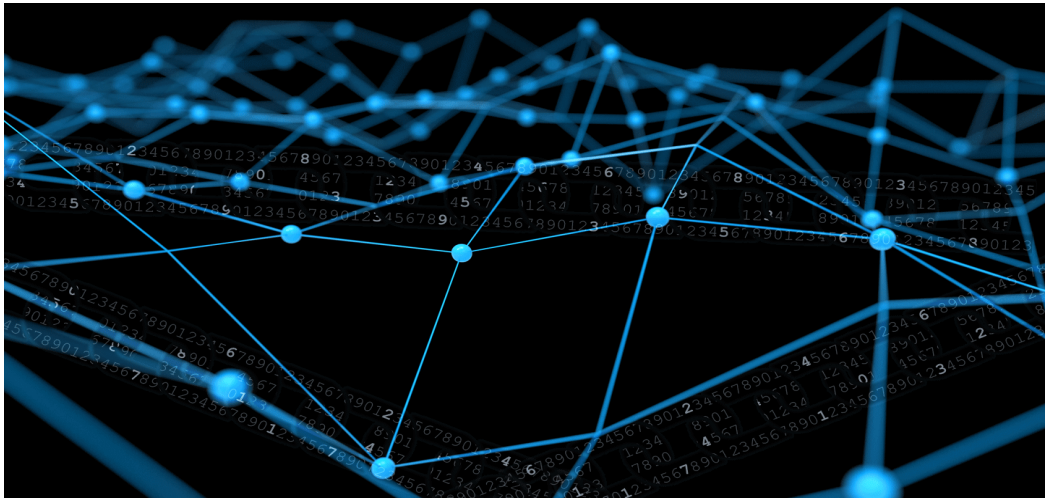
# MACHINE LEARNING

The first nation to successfully digitize its built infrastructure, and thereby generate the data suitable for AI/ML, will reap huge benefits in improved infrastructure provision, in better public services, and in generating whole new areas of economic activity and enterprise.



# MACHINE LEARNING

An application of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. The process of learning begins with observations or data such as examples, direct experience or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide.



# PROPOSED SOLUTION

- The machine learning algorithm employed trains on a customer's historical energy consumption to be able to predict the future energy usage. An irregularity between the predicted energy usage and current energy usage will then be flagged as potential electricity theft, allowing rapid alarm detection. The model developed uses a type of a **decision tree machine learning algorithm to train and predict customer energy consumption.**

# PROPOSED SOLUTION

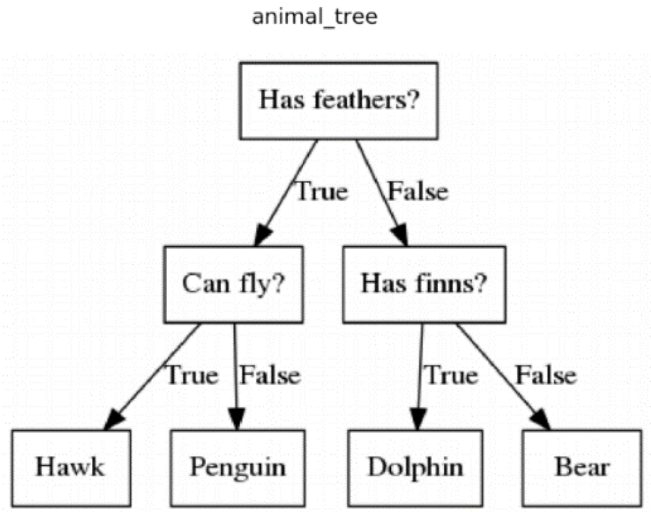
- The machine learning model is trained and tested on a Smart Meter dataset provided by **Irish Social Science Data Archives (ISSDA), Ireland**. This data comprises of both residential and commercial premises. However, for this research the residential smart meter data set considered. The data is accumulated from over **5000 houses in Ireland**. The data contains the information about the **customer id, code for date/time, electricity consumption for every 30 minutes (in kWh)**. This data was recorded over two years. The daily profile for every customer comprises then of **48 power consumption readings**. The customer ID is crucial since the utilities can use this to get the location of the different customers. The dataset also comes with post-trial survey data which gives information on the number of inhabitants at each house

# ML Algorithm

The model designed used a supervised machine learning algorithm called **M5P decision tree** to detect electricity theft. Decision trees are generated by algorithms that split a dataset into multiple branching segments based on decision rules.

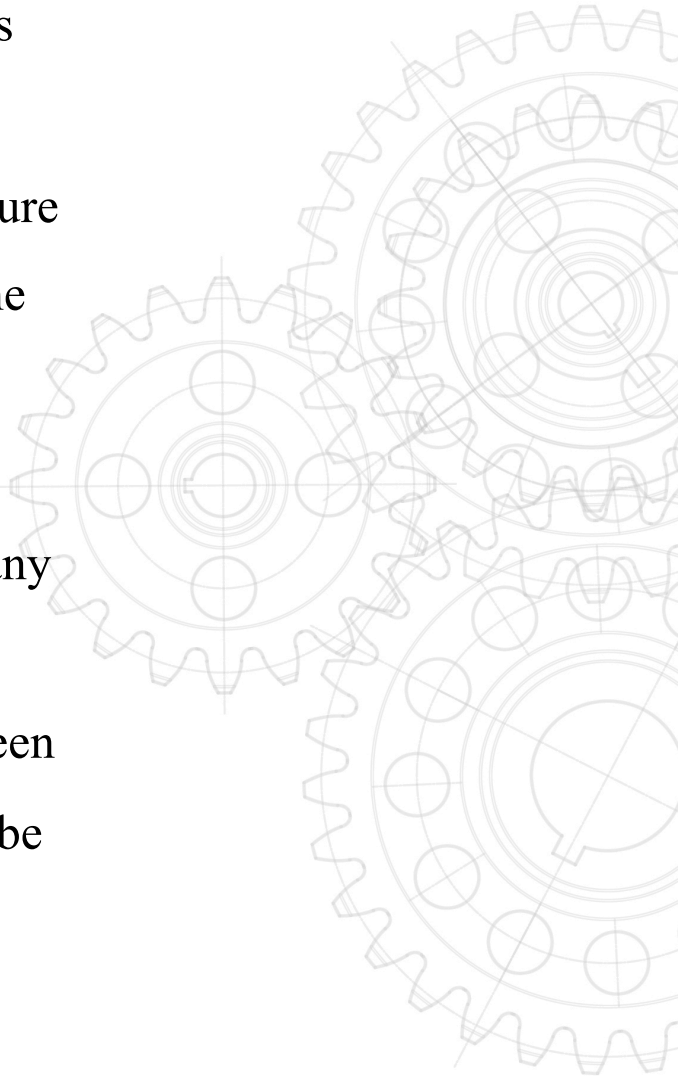
These decision rules are determined by identifying a relationship between input attributes and the outputs. Decision trees empower predictive modelling with higher accuracy, better stability and provide ease of interpretation. Decision trees are well suited for this task since there is a classification algorithm.

***Fig. Animal decision***



# ML Algorithm

The M5P decision tree algorithm used in this task combines **decision tree and linear regression** algorithms. This combination is then used to build a model that can predict future values based on past data. We created a model that learns the **individual behavior per customer to create an energy consumption model** with the capabilities to predict future energy consumption. The predicted energy consumption on any day would then be compared with the present energy consumption. By measuring the root mean square error between the predicted and actual energy usage, energy theft can then be flagged.



# ML Algorithm: Generation of Training & Validation Datasets

The training and validation dataset was created for each of the different customer data. The training data was obtained over a certain season for a certain customer in one year and the validation data was then chosen for the same season and same customer but for a different year. Ireland has 4 seasons, but these can be divided into Summer and Winter due to the similarity of Autumn with Summer and Spring with Winter. Summer was then assumed to run from May to October and Winter from November to April.

# ML Algorithm: Training of ML Algorithm

The training dataset, is used generate the decision rules representative of normal energy consumption for each of the customers in the training dataset. Decision trees are more prone to overfitting and hence careful consideration is made to prevent overfitting. Overfitting is a phenomenon in machine learning when the algorithm is too closely fit to the training data that it cannot be able to reliably predict future observations that were not in the training data.

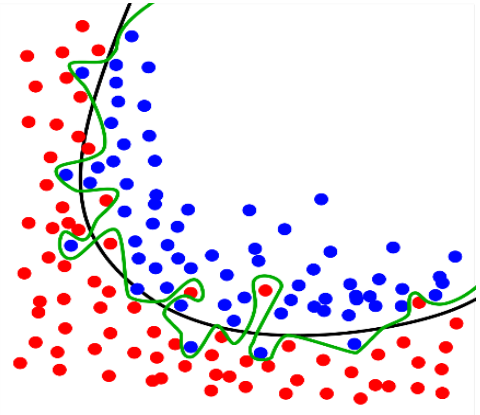


*Fig. Overfitting*

# ML Algorithm: Predicting of Future Energy Use

The generated decision rules are then used to predict the energy which each consumer is expected to use based on the on the features: day, month, time, energy used and people staying at the house. This prediction is done and tested on the validation dataset.

# ML Algorithm: Energy Theft Modelling

Two types of energy theft were simulated with the first being when the consumer's smart meter reports **less energy consumption than actual consumed energy.** This would normally occur either because of smart meter tampering to slow down its reading or by passing of the smart meter.

The second energy theft type is one more common in third world countries, **unauthorized tapping of the electricity line**. This would result in a higher than normal smart meter reading. These 2 types of energy theft were simulated by adding/subtracting a random value deviating from 0 to 1 kWh to every measurement of the energy consumption data in the validation dataset. The original data was used to learn the consumption model and simulated energy theft data was used in validating the model for an energy theft case.

# ML Algorithm: Energy Theft Detection

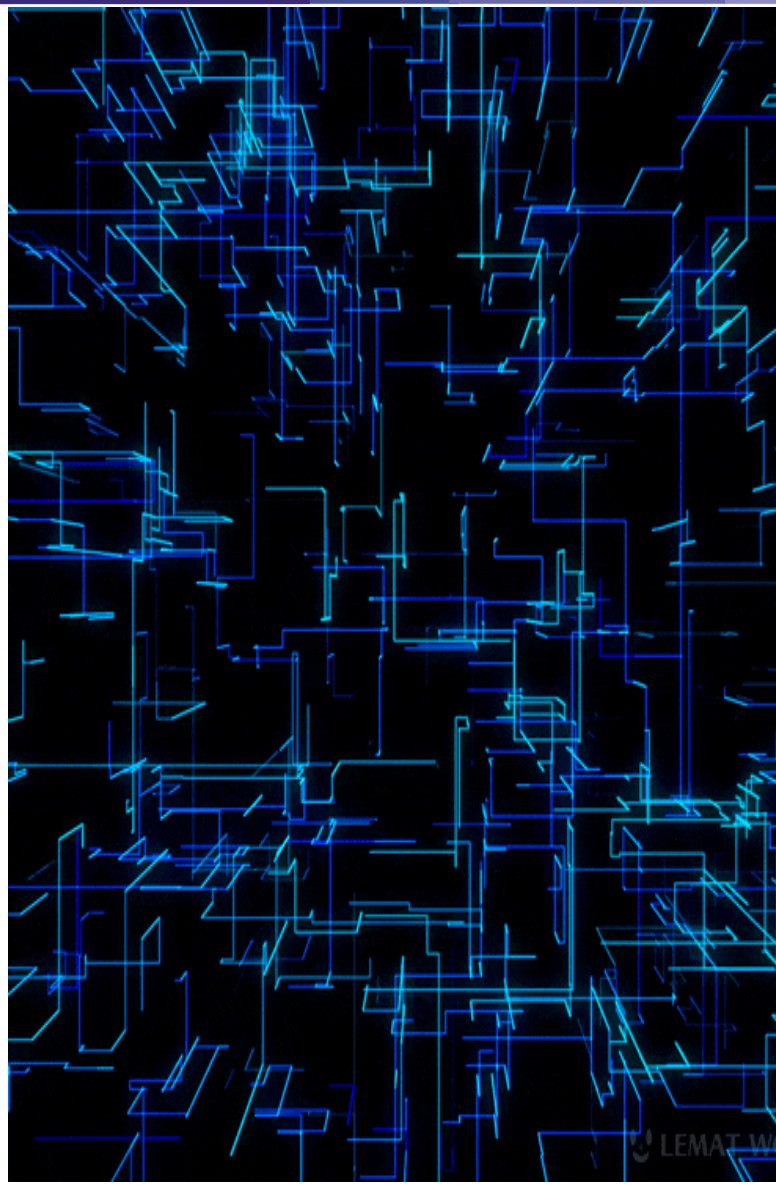
To measure the difference between the predicted energy consumption values and the actual energy consumption as in the validation dataset the **Root Mean Square Error (RMSE)** is used. The RMSE is a statistical measure of the deviation between the predicted and actual values. It is calculated for each day for each customer as:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N}(P_t - A_t)^2}{N}}$$

*where:*

*Pt is a predicted energy value from the machine learning algorithm;*

*At is the actual energy consumption value in the validation set, and N is the total number of energy values taken per day for each customer. N = 48 since 30-minute readings of the energy consumed are taken by the smart meter.*



# TECHNOLOGY STACK

*Python 3.6*

*Libraries :-*

*Tensorflow – For Deep Learning Layers*

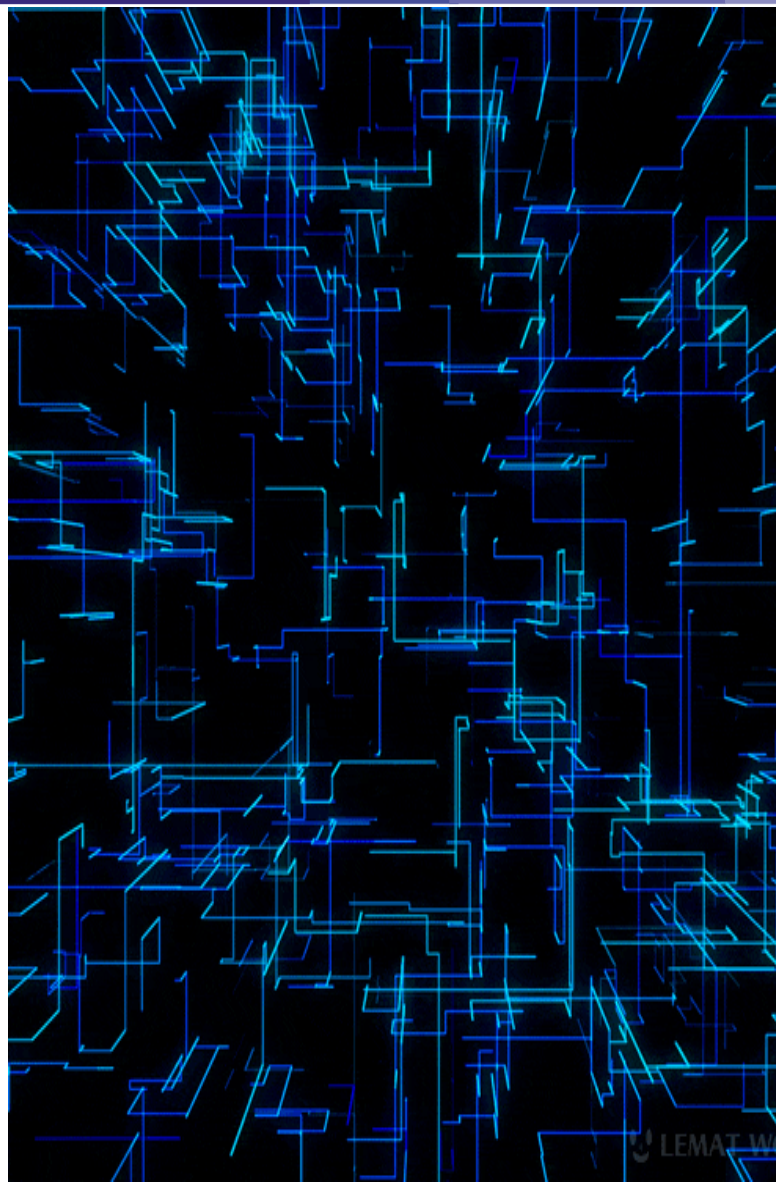
*Keras – Implementing Tensorflow*

*Numpy – Basic Numerical Operations*

*Matplotlib – To Visualize the data plot*

- *sklearn - For using k-means algorithm*

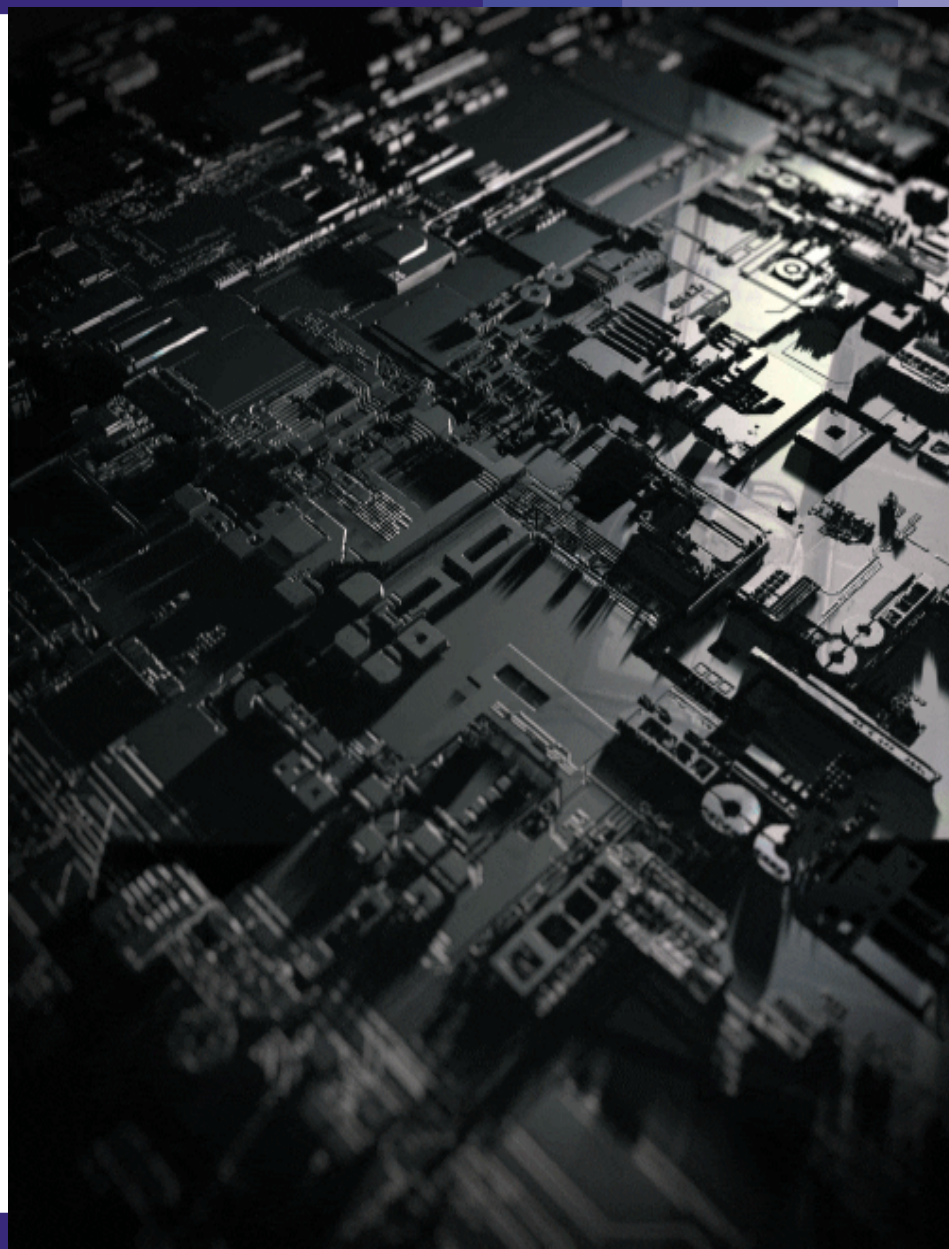
*Jupyter Notebook*



# EXPERIMENT RESULTS

A set of experiments were then conducted to evaluate effectiveness of the employed algorithm to first effectively learn the consumer model and after that to be able to detect any energy theft.

Two experiments were set up and are discussed below:



# Experimental 1 Results

The first experiment was designed to evaluate the model accuracy in predicting energy consumption for the same month a year after the training set. In experiment 1, the decision tree was trained using energy measurements from June 2009 and validating the results by using energy consumption values with and without the simulated energy theft from June 2010.

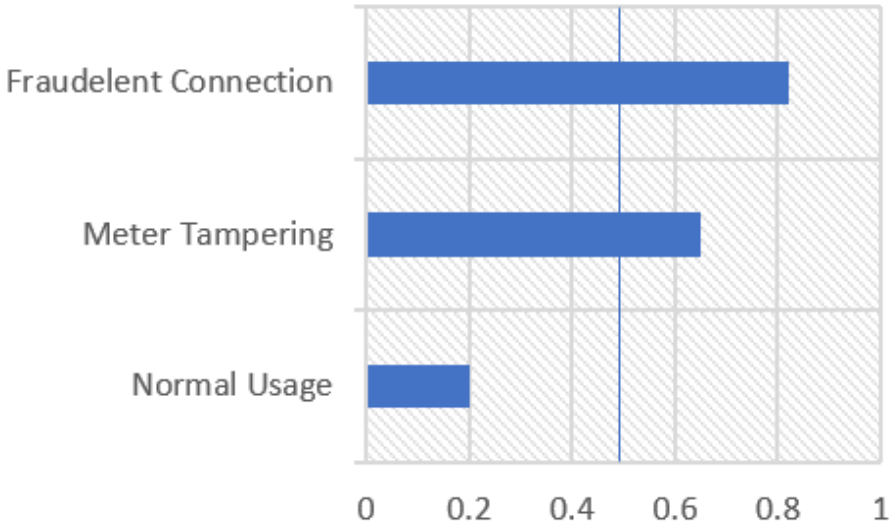



***Fig. shows the resulting RMSE values using a validation set consisting of both normal and energy theft activities (monthly values)***

# Experimental 2 Results

The second experiment revealed the ability of the machine learning algorithm to predict the values for the week following several weeks in the training set. In experiment 1, the machine learning algorithm was trained using energy measurements from November 8 - 28, 2009 and validated using energy measurements with and without energy theft activities from November 29 – December 19, 2010. .

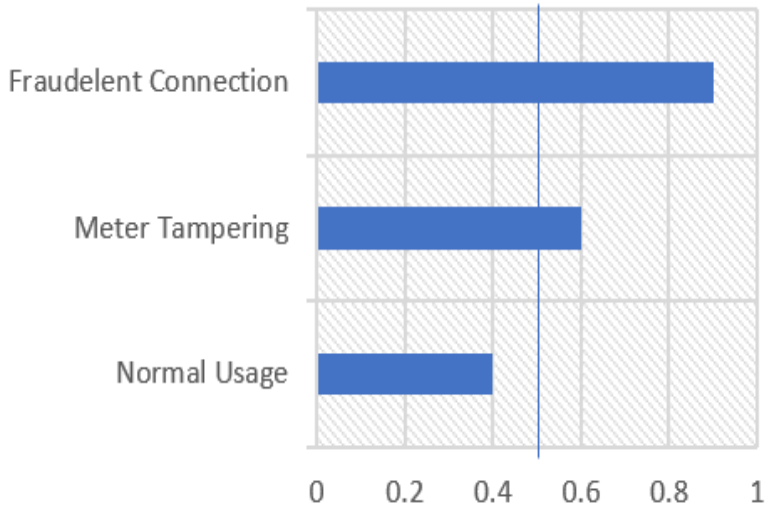



***Fig. shows the resulting RMSE values using a validation set consisting of both normal and energy theft activities (weekly values)***

# Electricity Theft Dash Board for City X

**Area: City X**
**Location: GPS Coordinates**
**Month and Year: August 2018**


Regular Usage

Irregular Usage

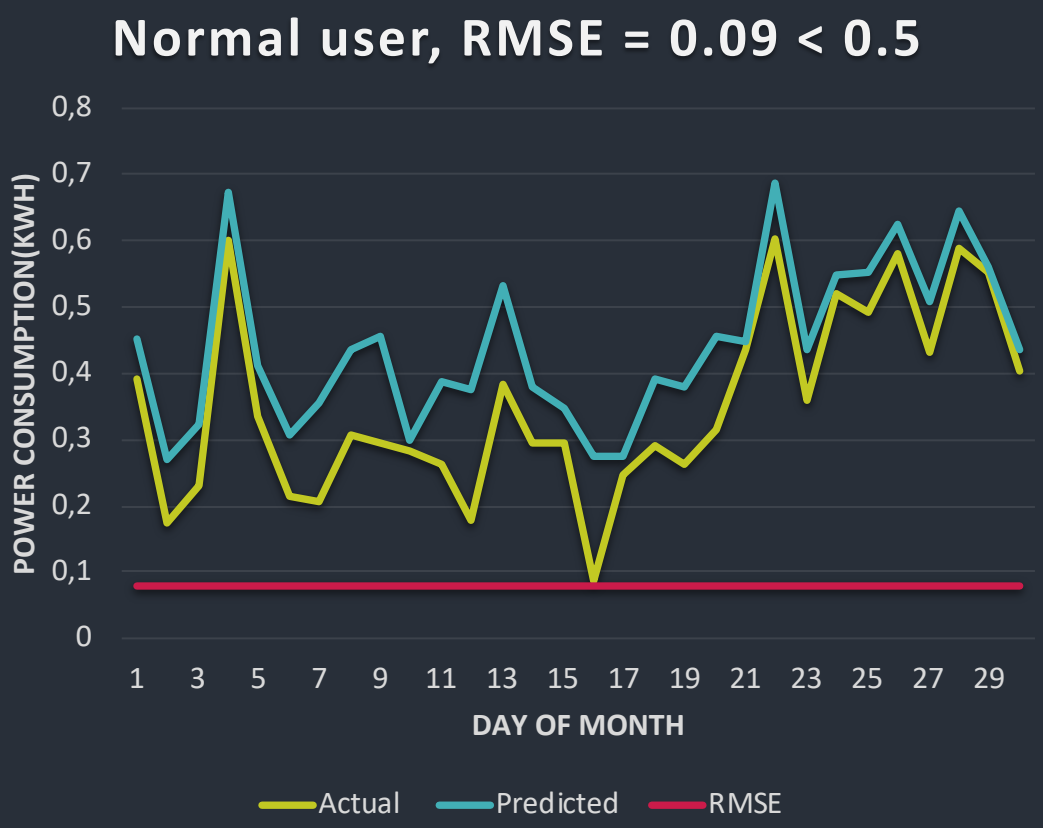
View/Print Report

Help

# Potentially good customer

Address: 22 Northumberland close
Customer Name: John Dory
Month: August

Previous Month



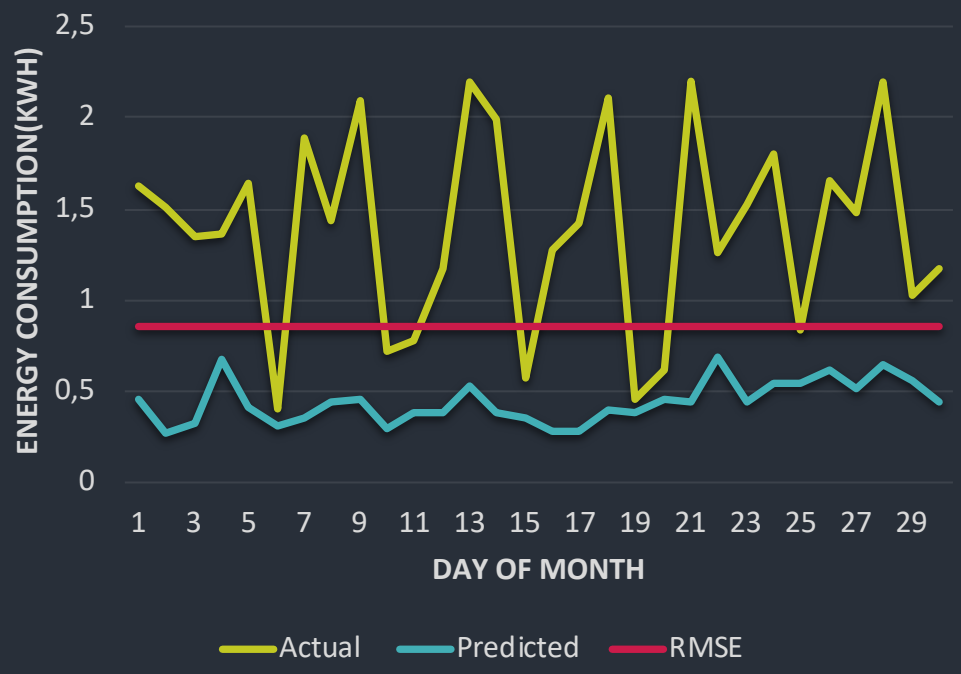

Next Month

1. Predicted Energy Consumption/month — 13.5 kWh
2. Actual Energy Consumed/month — 10.6 kWh
3. Potential energy stolen — 2.9 kWh
4. Potential Money lost by utility — R85

# Potentially bad customer

## Potential energy theft, RMSE = 0.86 > 0.5

Previous Month




Next Month

Suspected Causes:
Illegal tampering(Common in the area)

**Address: 170 Athens Road**
**Customer Name: Raymond Harris**
**Month: August**

1. Predicted Energy Consumption/month — **39.5 kWh**
2. Actual Energy Consumed/month — **13.2 kWh**
3. Potential energy stolen — **26.3 kWh**
4. Potential Money lost by utility — **R790**


09:48:25
Friday, November 23, 2018
zhengzhou
27°C
Cloudy to clear
Order transaction (2018/5/25)
538,859
Total order
Rising by the ring 3.5%
8,765
Make up amount
147,097
Paid amount
18,621
Amount owed
Time income (2018/5/25)
200
150
100
13:00
Income : 145,096,654
9:00 10:00 11:00 12:00 13:00 14:00 15:00
payment method (2018/5/25)
38%
WeChat payment
42%
Pay with Ali-Pay
20%
cash payment
65,876
Total order
87,842
Total parking lot
8,987,842
Total parking space
Zhengzhou Parking Lot
Arrival : 46,323
Appearance : 40,807
238,809
Parking space
38,809
Empty parking
263
Abnormal
98
Alarm
City income (2018/5/25)
44%
Shanghai
1
16%
Guangzhou
4
28%
Beijing
2
12%
Chendu
5
18%
Suzhou
3
8%
Wuhan
6
Revenue share (2018/5/25)
34%
Community
29%
Market
24%
Office
13%
Other
Access statistics
Appearance
Admission
2000
1500
1000
500
9:00 10:00 11:00 12:00 13:00 14:00 15:00

# FUTURE SCOPE

- With the model built and currently testing we aim to built a software which will be able to display this data and give accurate reporting
- For a larger dataset with more number of features, a time-series based model can be trained( using LSTM and GRU) for fine tuning of hyperparameters to recognize highly complex, time variant power usage patterns and determine fraudulent customers.
- After analysis of Power usage patterns of different localities with more features, the system can be used by the authorities at higher levels in hierarchy.
- With the knowledge of power usage patterns, specific power demand for the future can be predicted which can help reduce transmission losses and installation of power storage infrastructure at specific locations.

# CONCLUSION

This paper demonstrated successful application of decision tree learning for detecting energy theft. The conducted experiments unveiled the ability of the machine learning model to accurately predict energy consumption values from the same month of a year, subsequent weeks, and within the same weather season.

Furthermore, the historical data were used in these experiments to generate the machine learning model and predict future energy consumption. Using the RMSE it was shown that the machine learning algorithm was able to accurately predict the future values and hence detect electricity theft. In smart grid data analytics system, it is necessary to know the real time electricity consumption data to forecast the exact future demand of electricity and plan accordingly. The identification of power theft will also extend its support for load forecasting that permits the utilities to exactly predict the power demand for future specific to individual customer.

# ENGINEER OF THE FUTURE

"Change is inevitable, and the disruption it causes often brings both inconvenience and opportunity." - **Robert Scoble**

# Thank you

**Keith Tinashe Katyora** BEng (Electrical) , MPhil (Energy Studies)
Mail & Guardian Top 200 Young South Africans 2019
Chairperson of CESA YPF
2019 Top 100 Young Africans Finalist
**T** +27 12 427 2000 **M** +27 73 107 0483
**katyorakeith@gmail.com**