Session 9: Emerging Technologies and Use Cases
PART B: AI/ML

Session Contents

Artificial Intelligence (AI) and Machine Learning (ML)

Advanced Analytics

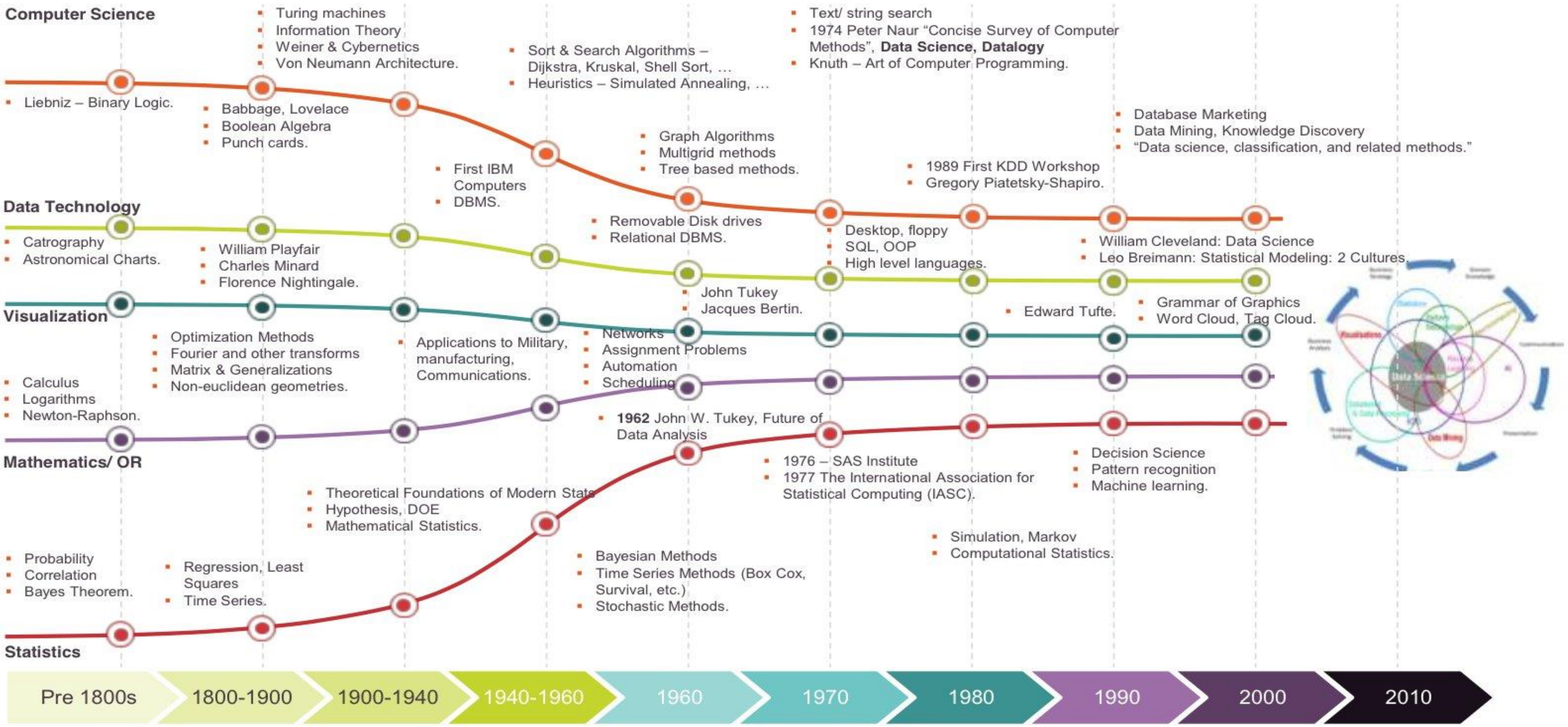
Applications of AI and ML in Electric Utilities

Speaker

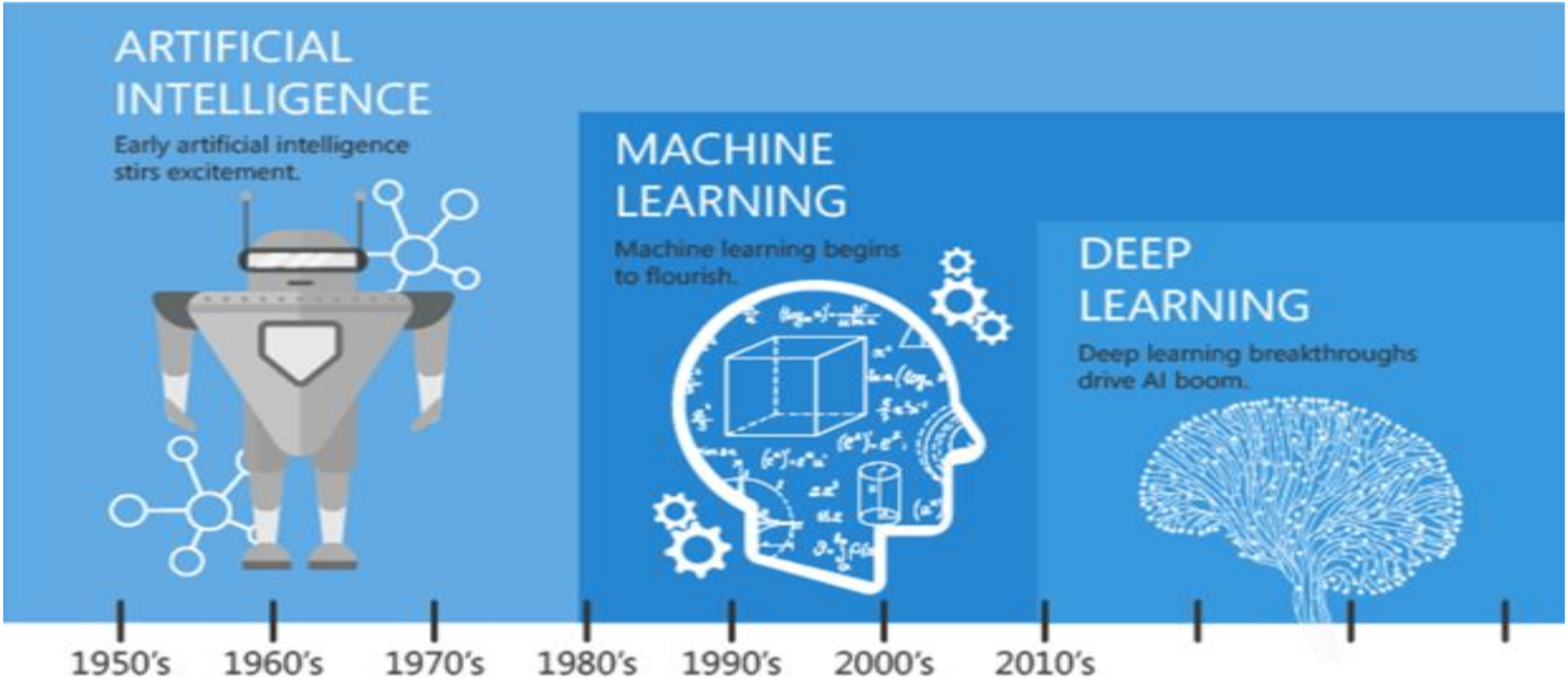
Reji Kumaar Pillai

President - India Smart Grid Forum
Chairman – Global Smart Energy Federation

History of Computing and AI Development



Introduction to AI, ML and Advanced Analytics



AI Workflow

Data Preparation → AI Modeling → System Design → Deployment

Power System Studies — Grid Analytics — Digitization/Real-Time Simulation

Data Preparation

- Data cleansing and preparation
- Human insight
- Simulation-generated data

AI Modeling

- Model design and tuning
- Hardware accelerated training
- Interoperability

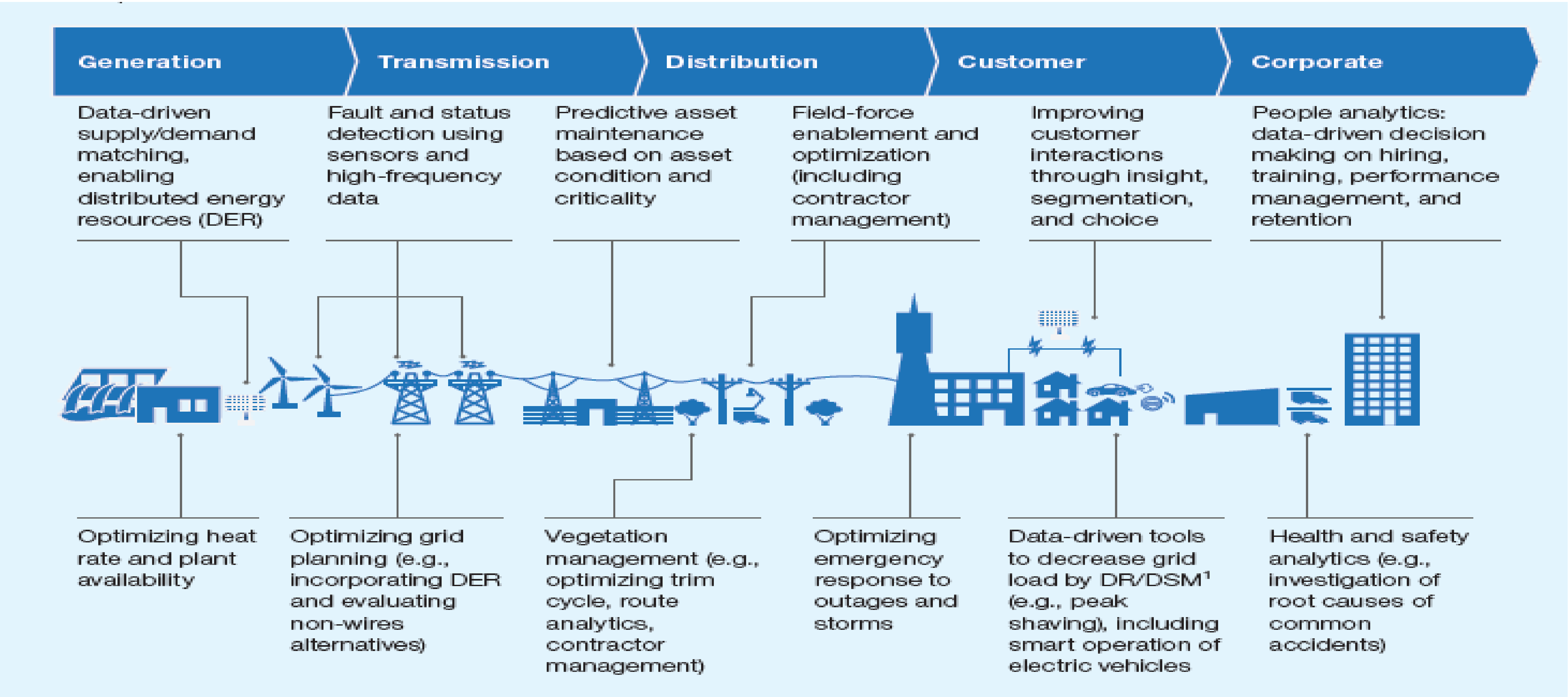
Simulation & Test

- Integration with complex systems
- System simulation
- System verification and validation

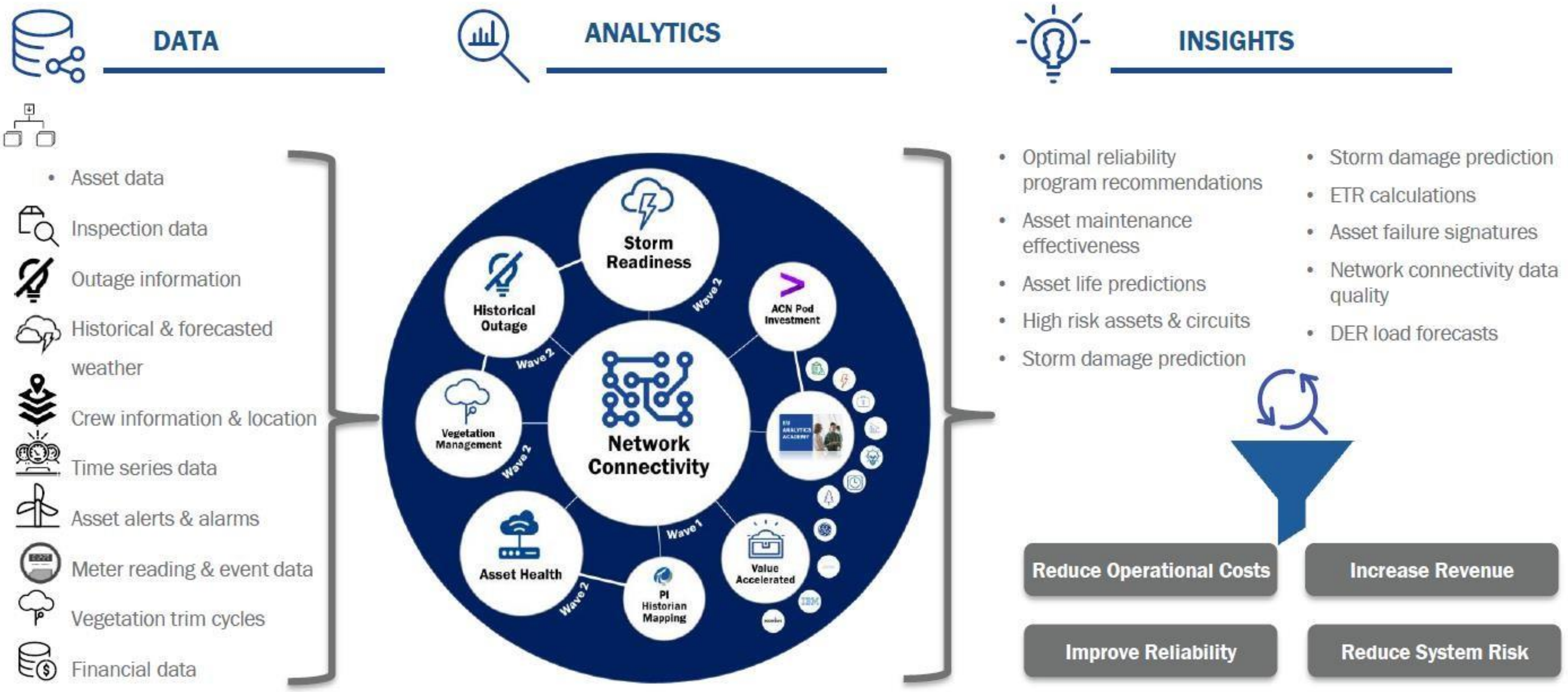
Deployment

- Embedded devices
- Enterprise systems
- Edge, cloud, desktop

Applications of AI, ML in the Power Utility Value Chain



AI for Grid Analytics



Key Domains for AI and Advanced Analytics

Customer Operations

Use Cases across...

1. Customer Strategy
2. Customer Operations
3. Revenue Cycle
4. Products & Services

...that will:

- Enhance cust. experience
- Automate low value interactions

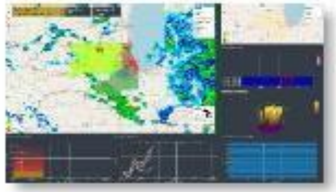
Grid

23 Use Cases across...

1. Asset Management
2. Grid Operations
3. Extended Systems

...that will:

1. Improve Reliability
2. Improve Customer Sat.
3. Reduce O&M Expenses
4. Capture new Revenue

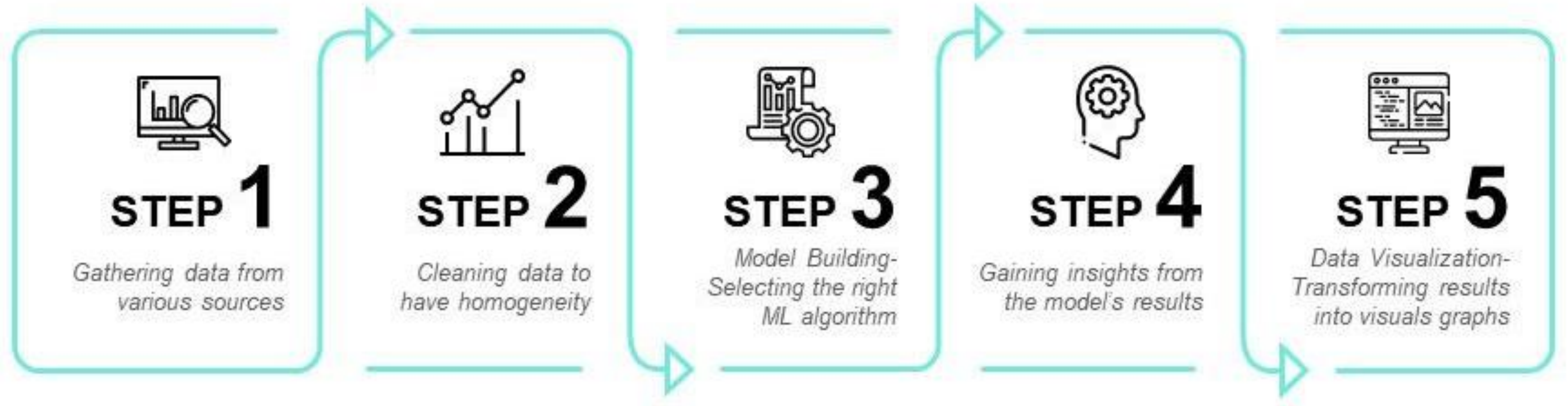


Advanced Metering Infrastructure

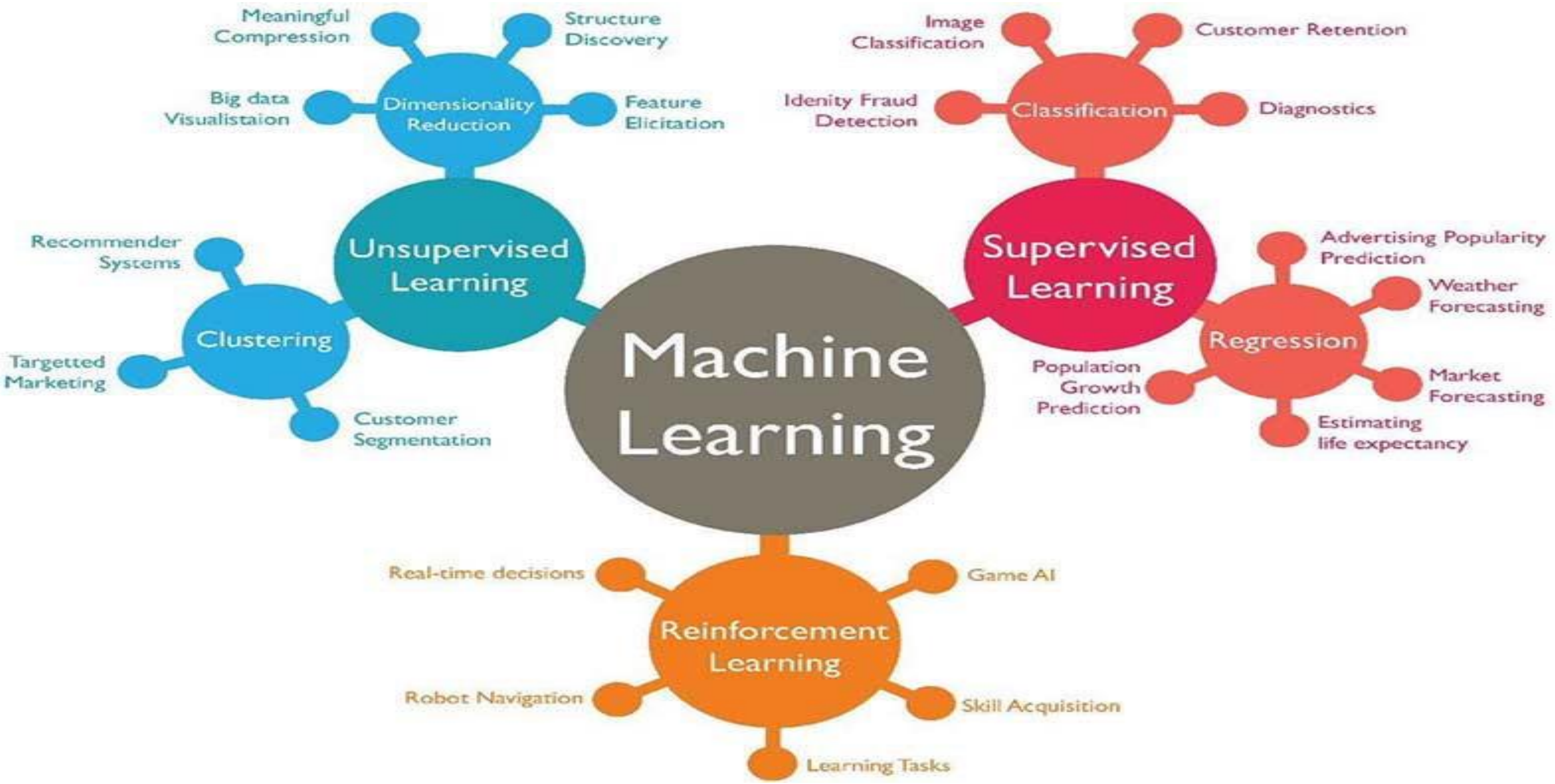
33 Use Cases across:

1. Meter Operations
2. Network Operations
3. Theft Detection
4. Inactive Meters

Machine Learning - Process



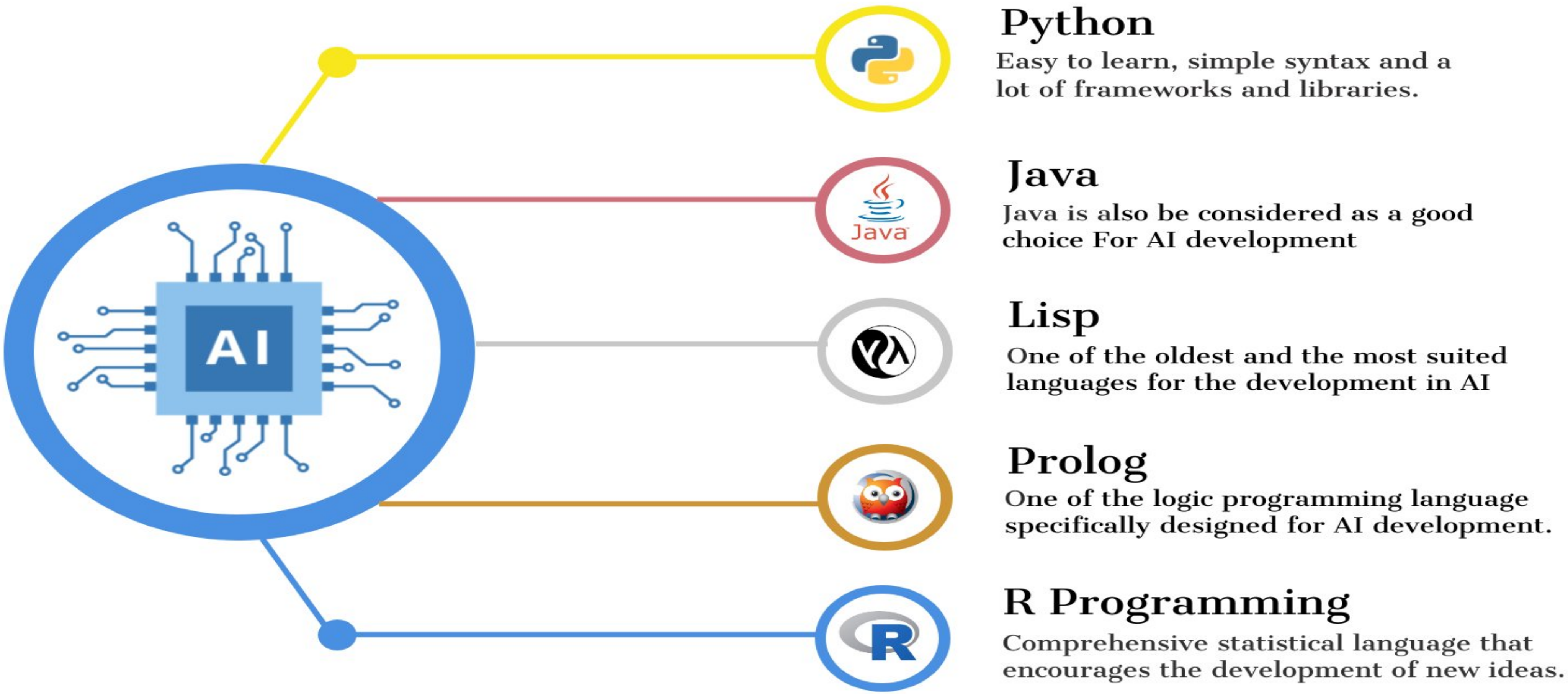
Machine Learning - Types



Machine Learning - Techniques

	Supervised Learning	Unsupervised Learning	Deep Learning	Ensemble Learning
Techniques	• Classification • Logistic or Linear Regression • Multivariate regression	• Clustering • Natural Language Processing (NLP)	• Neural Networks	• Random Forest
Potential customer problem	• Classifying product segments • Predicting machine problems to avoid downtime • Identifying linear relationships in machine performance	• Identifying similarities and characteristics e.g. if it is shaped like a 'car', it could be grouped as a 'car'	• Solving complex problem patterned towards intelligent beings using multiple sources of dissimilar inputs. e.g "smart" plant facility	• Identifying linear and probabilistic relationships between an outcome and its components using Bootstrap Aggregation
Data/Input format	• Structured imagery, numeric, strings/characters, sensory data with well identified labels	• Unstructured imagery, numeric, strings/characters, sensory data with semi-identified labels	• Multiple structured data types e.g., imagery, audio, numeric, strings/characters, sensory, etc., from multiple sources	• Structured imagery, numeric, strings/characters, sensory data with well identified labels
Applications	• Predictive and prescriptive maintenance • Predictive decision support • Production optimization • Product segmentation • Predictive root cause analysis • Asset management and quality control	• Pattern processing • Anomaly or defect detection for quality control • Asset performance management	• Complex production optimization • Quality control • Speech and pattern recognition • Complex anomaly detection • Autonomous processing • Asset management	• Powerful and accurate Predictive and prescriptive maintenance • Predictive decision support • Product segmentation • Predictive root cause analysis • Asset performance analysis and quality control

Programming Languages for AI



AI and ML Libraries

Artificial Intelligence Basic Libraries

1. Numpy
2. Pandas

Machine Learning Libraries

1. Scikit-Learn
2. Spark

Deep Learning Libraries

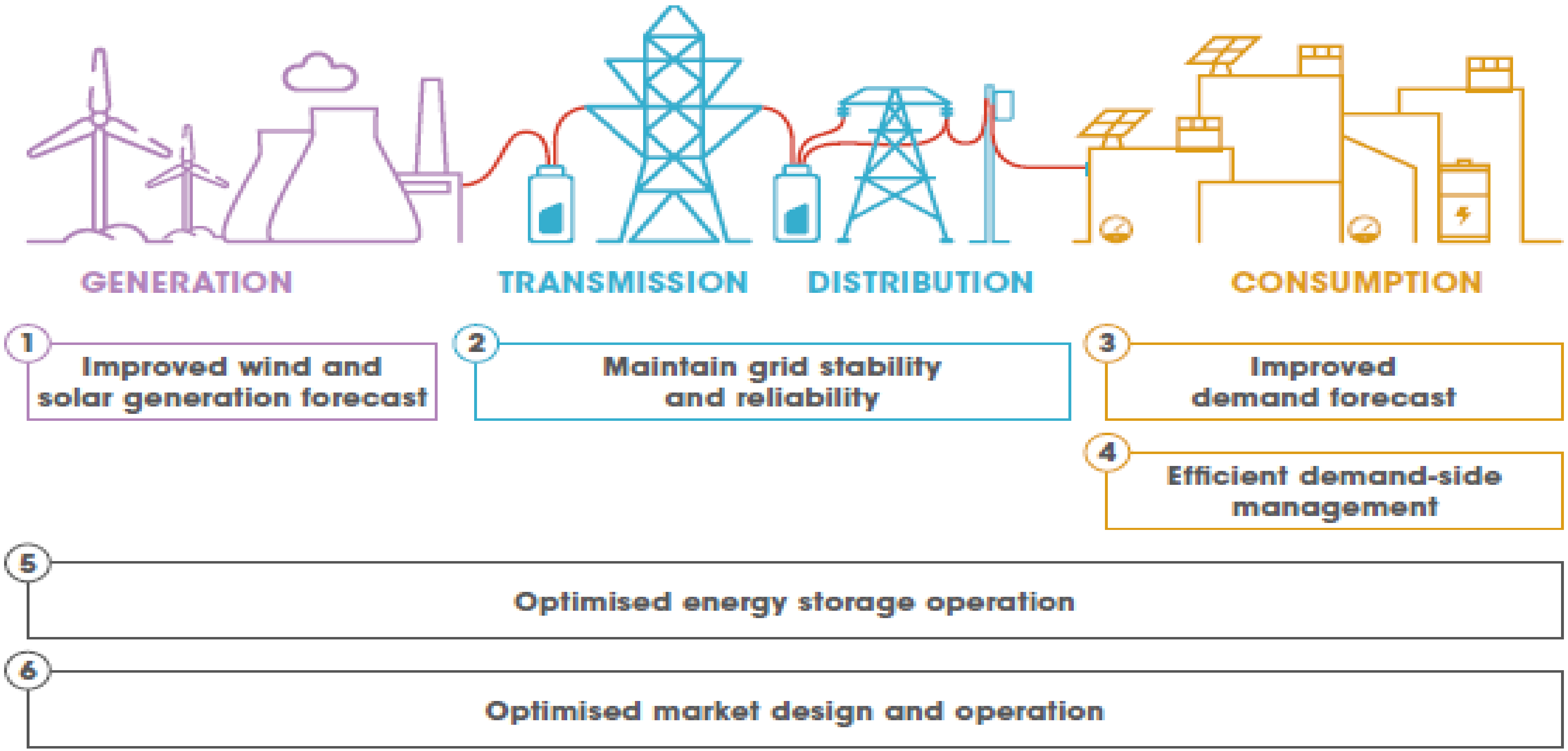
1. Keras
2. TensorFlow

Platform for Python Programming

1. NLTK - The Natural Language Toolkit
2. MXNet
3. Jupyter Notebook

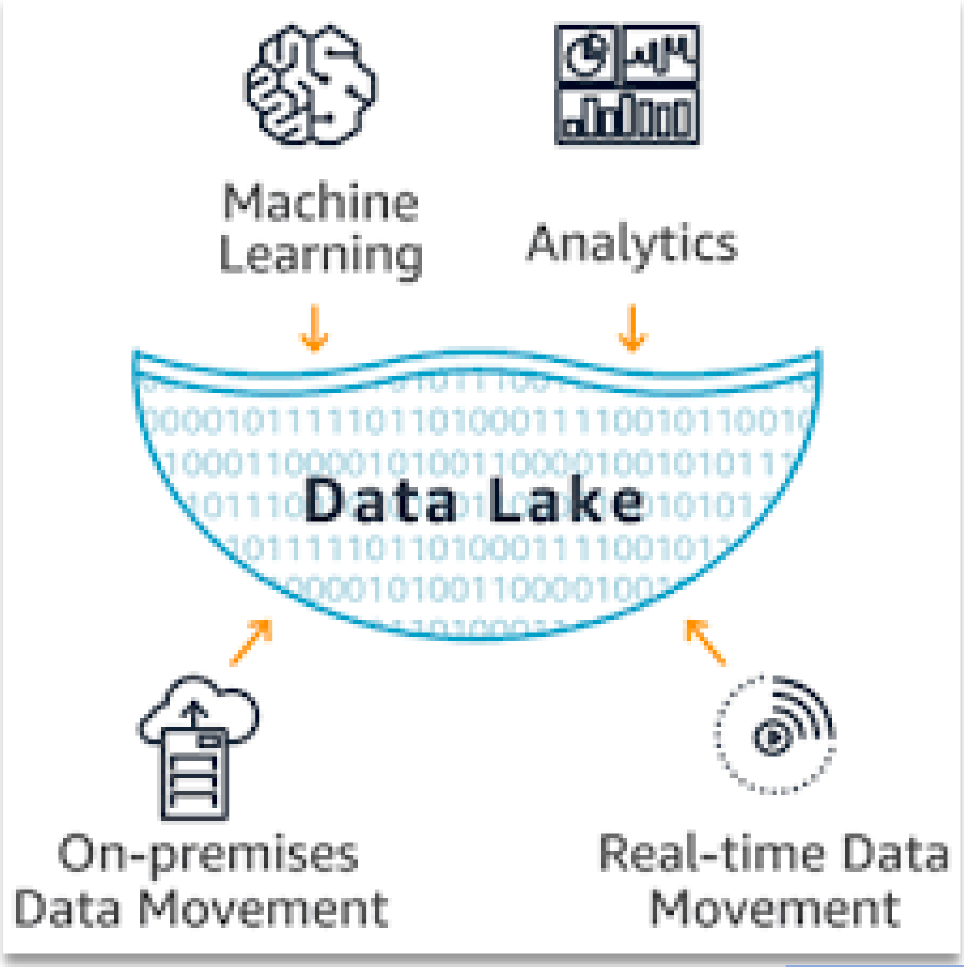
SELECT USE CASES

Use Case 1: Applications of AI for RE Integration



Source: IRENA 2019 Report

Use Case 2: Revenue Maximization – Early Revenue Recognition (1/2)



Business Challenge: Helping Utility to Realize the Revenue early by leveraging data from different sources

Solution

- Data pipeline from different sources
- Raw data and process data was sent to Data Lake
- Data Lake has been created using Big data platform
- Applying AI/ML methodology to get
  - Payment Behavior Analytics
  - Customer 360 Degree Profiling
  - Customer Risk Scoring Model
  - Customer Defaulter
- Right Campaign Strategy to enhance revenue

Business Benefits

- Data Driven Decision Application
- Better Decision Making capability
- Data Lake is future ready solution which has established further deep dive analytics and decision science methodology
- 20-30% more revenue recognition
- Prevention of Revenue Leakage
- Better Revenue Management

Results/Solution Methodology

Actual_Default_Oct	31554 (From Data)
Predicted_Default_Oct	23175 (from the model)
Accuracy	~73%
Total Value(Actual)	~11 Cr
Total Value (Predicted)	~9 Cr (~81%)
Actual_Default_Nov	**33915 (from Data)**
Predicted_Default_Nov	20280 (from the model)
Accuracy	~60%
Total Value(Actual)	~12 Cr
Total Value (Predicted)	~9 Cr (~75 %)

Use Case 2: Revenue Maximization – Early Revenue Recognition (2/2)

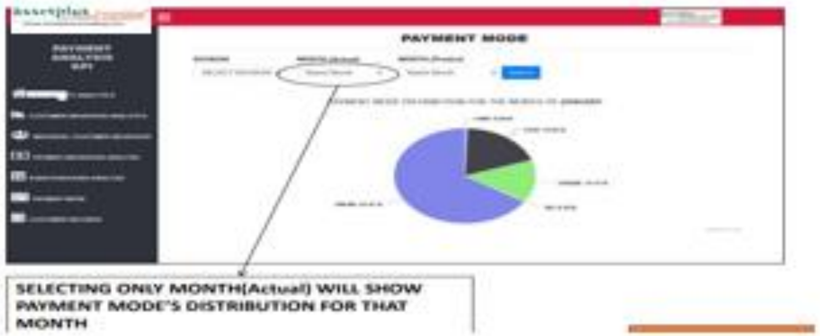
CUSTOMER BEHAVIOUR ANALYSIS

INDIVIDUAL CUSTOMER BEHAVIOUR

SEASON EVENT ANALYSIS

SEASON/EVENT ANSLYSIS

PAYMENT MODE

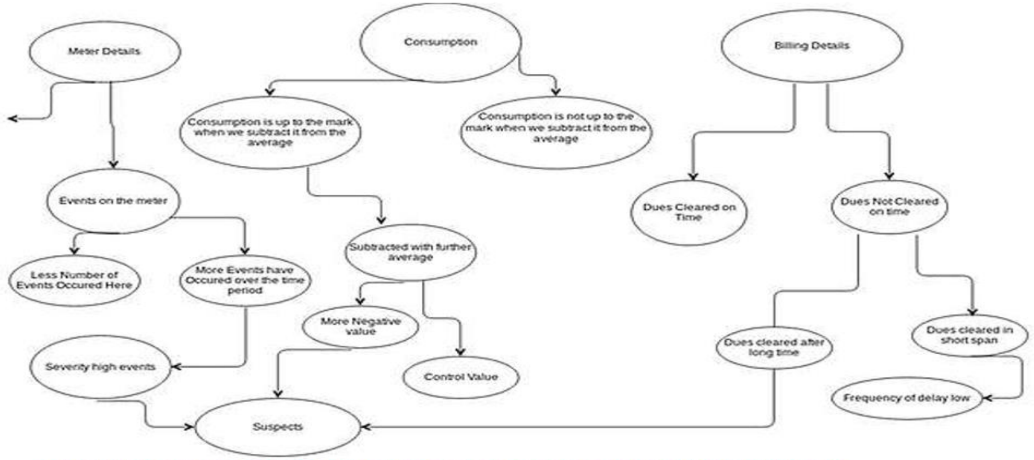


Use Case 3: Loss Reduction – Finding Pilferage Points in Customer Category (1/2)

Narrowing the Parameter

- Customer Master – Rate Category , Region , Location, Sanctioned load
- Meter Consumption Data - MDI, Average Hourly Consumption, Load Factor, TOD Rates
- Meter Event Data -HV, Tamper, Power-off, Network Event
- Billing Data- Amount, Date, Arrears
- Payment Data – Time, Delay
- MRD File-Technical Parameters – Current, Voltage Levels Phase-wise
- Customer Complaints – Frequency, Type

APPROACH: RANDOM FOREST



PINNING THE PILFERAGE POINTS WITH 99% ACCURACY

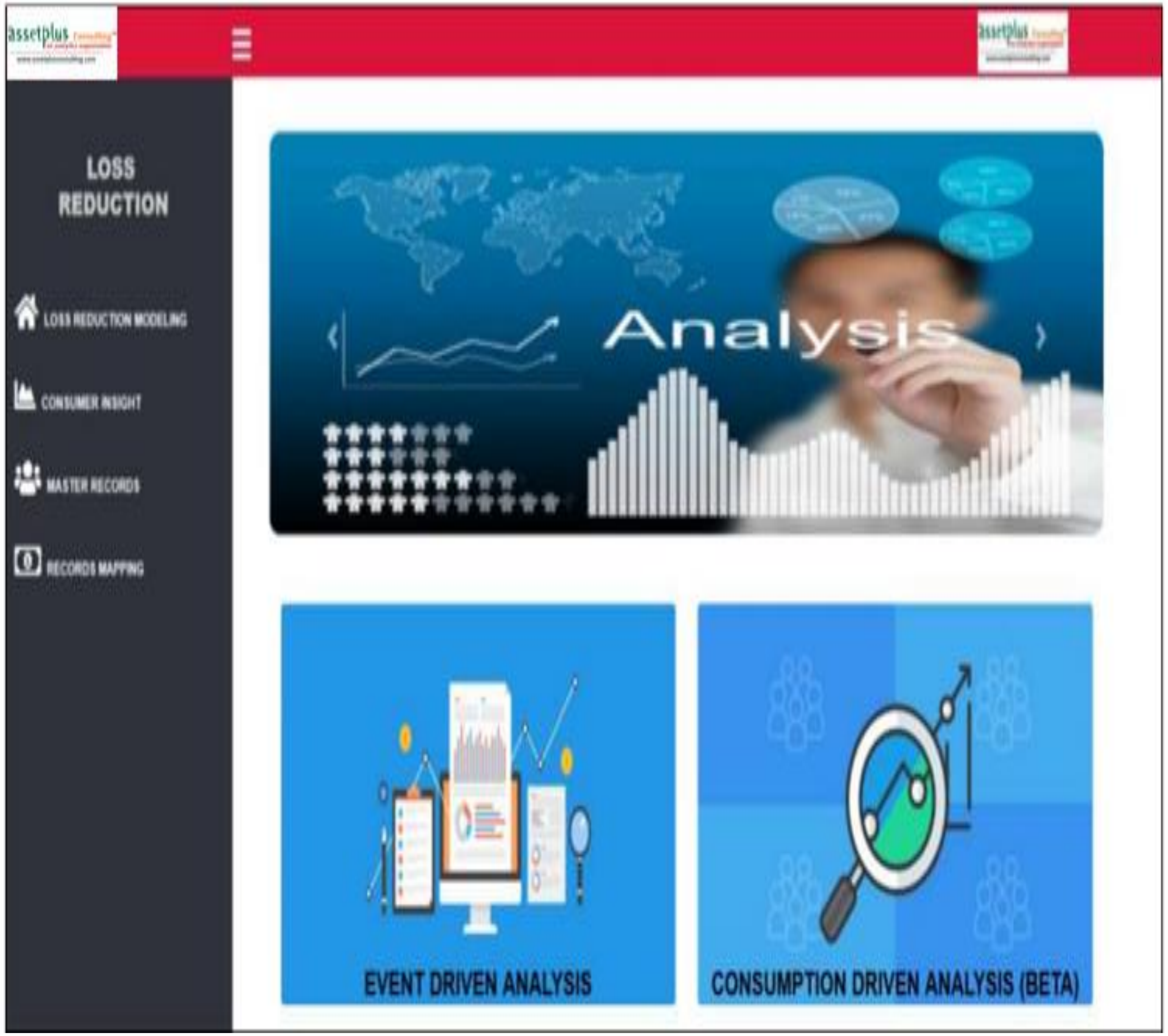
Use Case 3: Loss Reduction – Finding Pilferage Points in Customer Category (2/2)

Methodologies

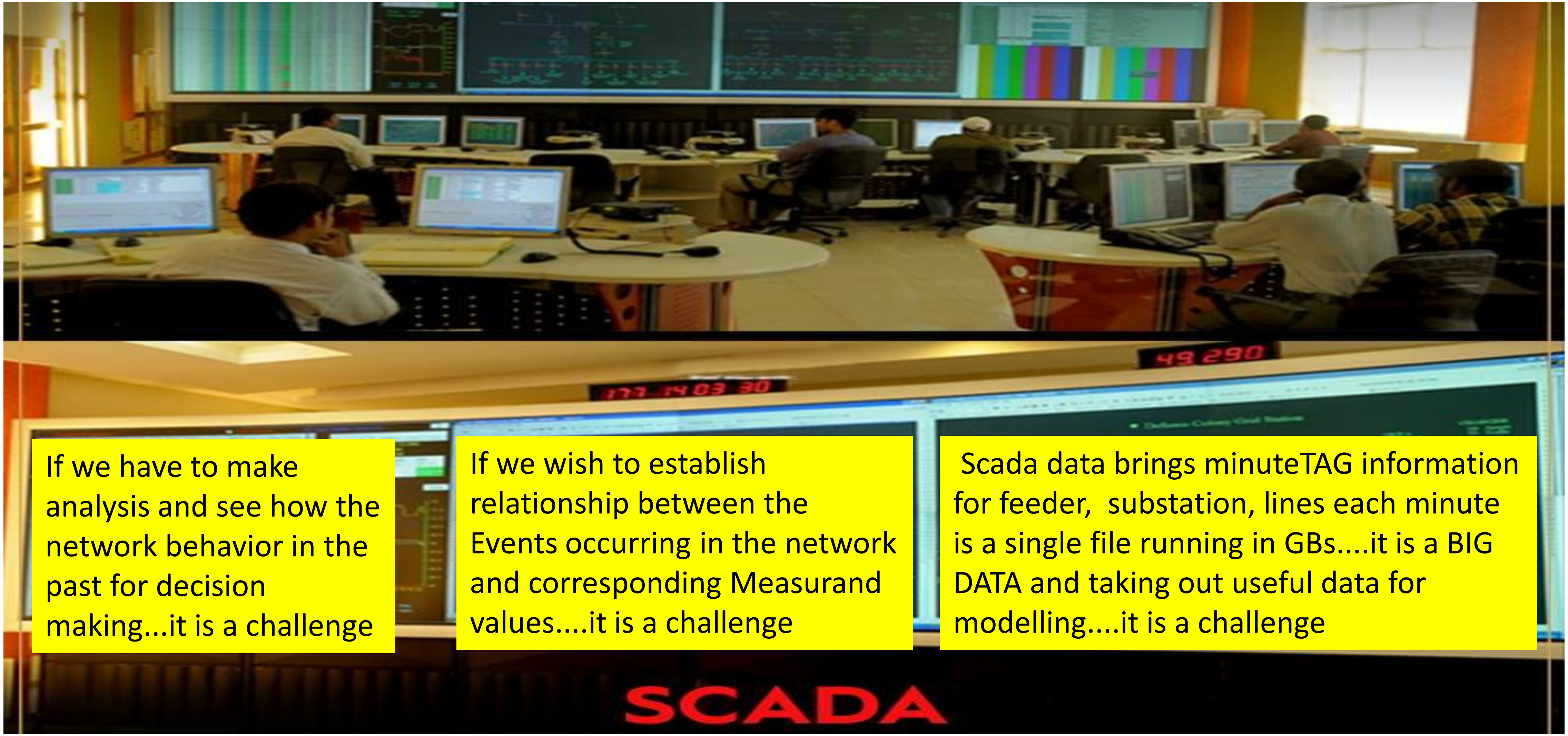
- Algorithms used – Random Forest, SVM, KNN, ANN
- Main classification done through – Multiple techniques of random forest and finding out the best group of possible suspects
- Deciding Factor are Consumption of Group, Consumption of individual, Meter Events (CT,PT,VT), Power Quality
- Statistical Analysis done using Python, Matplotlib, Basic Mathematics, Scikit-Learn Libraries.

Outline

- At first used unsupervised technique i.e. Clustering
- To get better results switched to KNN as the domain further expanded towards supervised learning
- SVM was the next step followed by the analytical study using Decision Tree which further expanded to Random Forest
- Prolonged Study of these patterns will gave more grip using Deep Neural Networks – CNN, RNN



Use Case 12: SCADA ANALYTICS-RECOMMENDER 11KV to 400 KV network modelling



If we have to make analysis and see how the network behavior in the past for decision making...it is a challenge

If we wish to establish relationship between the Events occurring in the network and corresponding Measurand values....it is a challenge

Scada data brings minuteTAG information for feeder, substation, lines each minute is a single file running in GBs....it is a BIG DATA and taking out useful data for modelling....it is a challenge

Use Case 12: SCADA ANALYTICS-RECOMMENDER 11kV to 400 kV network modelling

ANALYTICS AS A SERVICE –AaaS Platform

Different data aspect

Pdf

XML/JASON

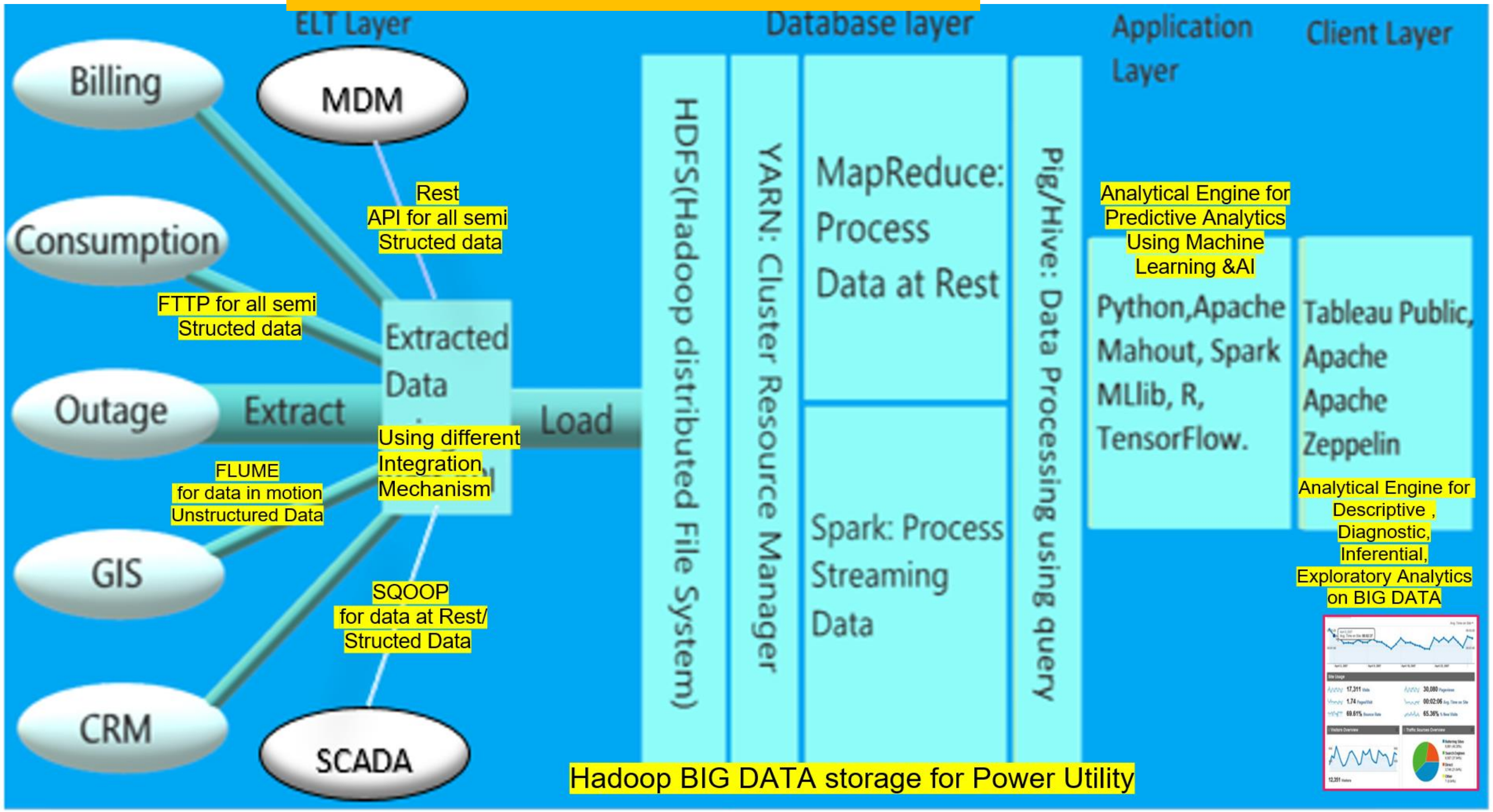
EXCEL/CSV

Images

Video

txt

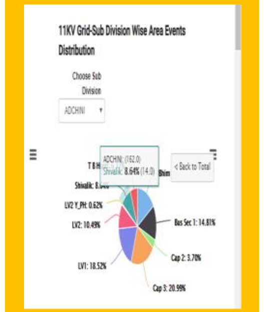
Streaming



Outcomes Driven on

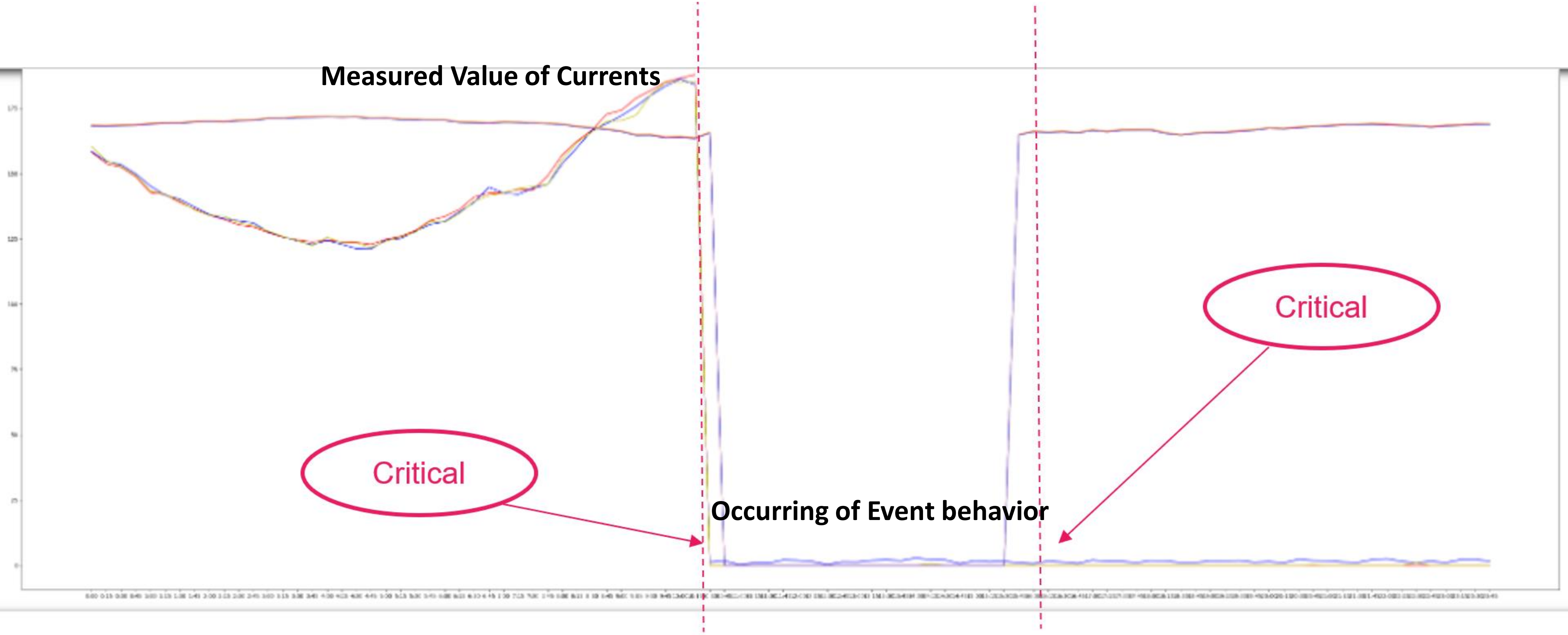
Mobile

Portal



Use Case 12: SCADA ANALYTICS-RECOMMENDER 11kV to 400 kV network modelling

MODELLING behavior of NETWORK TRIPPING EVENTS and Subsequent MEASUREND VALUES for Machine Learning modelling



Use Case 12: SCADA ANALYTICS-RECOMMENDER 11kV to 400 kV network modelling

Total feeders covering high voltage network and 11KV network for Karnataka state is around 25000, with 11 KV at 15000 feeder

1500 stations covering the feeder network

(a) 11KV first and then modelling SCADA environment for

(b) Higher voltage levels from 33KV to 400KV.

In India this is one of the largest feeder data aggregation and modelling in any of the states in the country

AWS

The SCADA analytical modelling platform handling 1kv network to 400kv network

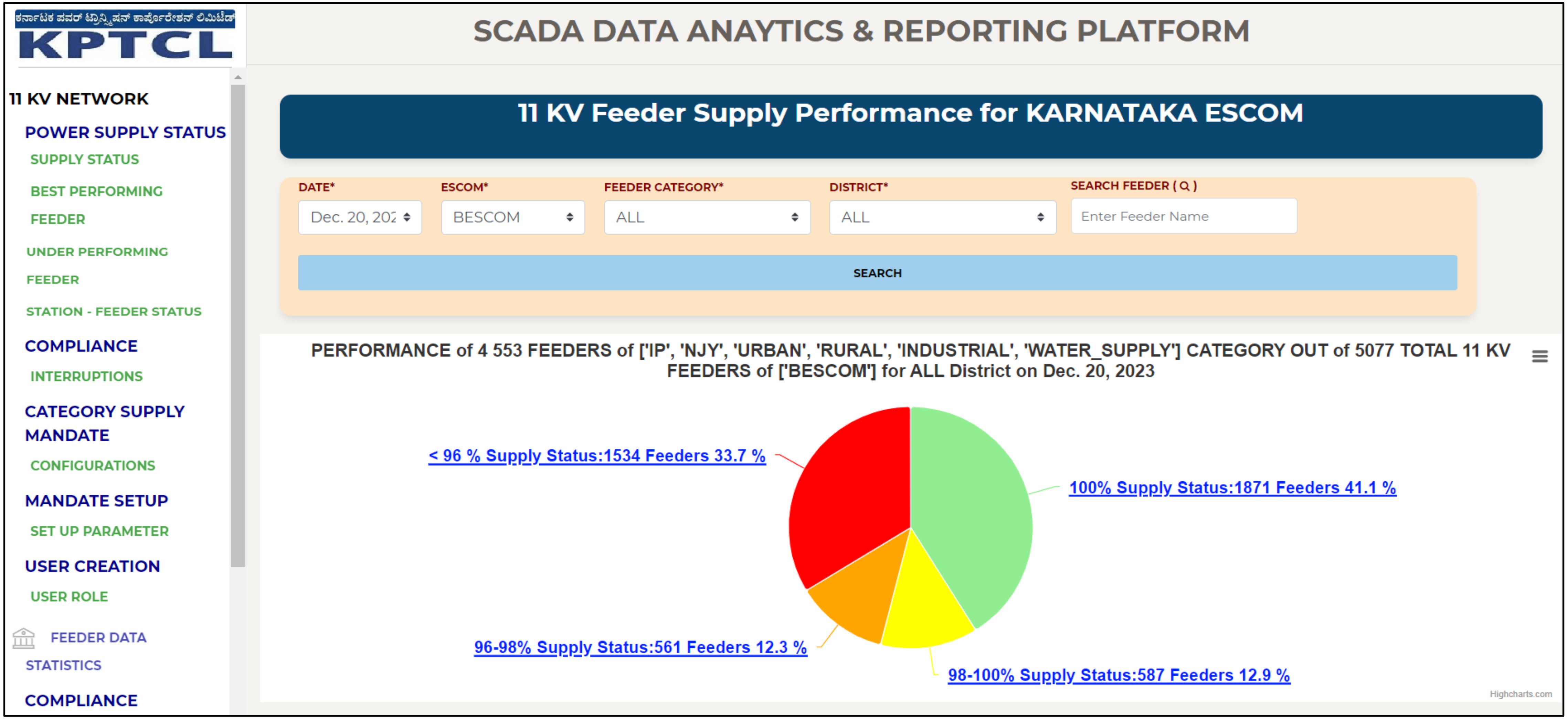
Instant deployable on the cloud

Readily accessible to other transmission network in the country for adoption

Analysis time N+2 days has been reduced to hours and minutes

Sustainability - Hardware and systems used previously are being reduced to one third

Use Case 12: SCADA ANALYTICS-RECOMMENDER 11kV to 400 kV network modelling



Thank You

Reji Kumaar Pillai

reji@rejikumar.com

@rejipillai

Machine Learning Algorithms (1/5)

ML Algorithm	Applications in the Power Utility Domain
Supervised Learning	Revenue Maximization and Customer Risk Scoring
Unsupervised Learning	Customer Profiling and Segmentation
Ensemble Learning	Loss Reduction – Finding Pilferage Points in Customer Category Power Theft/Loss – Technical or Commercial Loss
Deep Learning	• Solar Module Image Analytics • Meter Reading Image Analytics

Machine Learning Algorithms (2/5)

ML Algorithm	Applications in the Power Utility Domain
A) Regression Techniques	
i. Linear Regression	• SCADA Analytics • Alarm Analytics • Transformer Monitoring Analytics
ii. Polynomial Regression	
iii. Advanced Regression	
B) Classifications	
i. Naïve Bayes	• Loss Reduction – Finding Pilferage Points in Customer Category • Power Theft/Loss – Technical OR Commercial Loss
ii. Decision Tree	
iii. Random Forest	

Machine Learning Algorithms (3/5)

ML Algorithm	Applications in the Power Utility Domain
iv. Support Vector Machine (SVM)	• Revenue Maximization and Customer Risk Scoring • Loss Reduction – Finding Pilferage Points in Customer Category • Power Theft/Loss – Technical OR Commercial Loss
v. Logistical Regression	
vi. K Nearest Neighbour (KNN)	
vii. Gradient Boosting Algorithm (GBA)	
viii. Adaptive Boosting (AdaBoost)	
ix. Extreme Gradient Boost (XGBoost)	
C. Clustering	
i. K-Means Clustering	Customer Profiling and Segmentation
ii. Latent Dirichlet Allocation (LDA)	• Customer complaint analytics • Insights form call logs

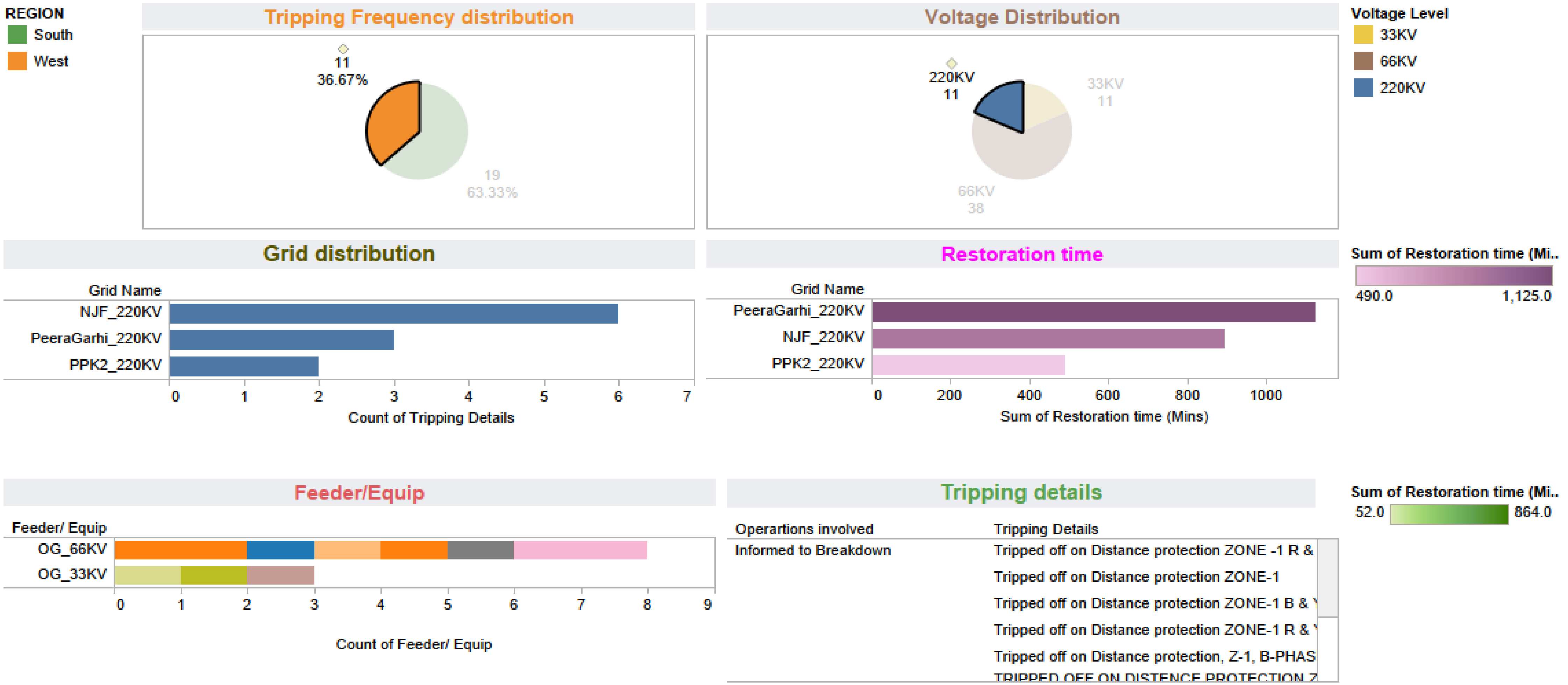
Machine Learning Algorithms (4/5)

ML Algorithm	Applications in the Power Utility Domain
D) Collaborative Filtering	
i. Alternating Least Square (ALS)	Recommendation of Energy Algorithm (Balancing of Power Procurement Strategy with impact of Distributed Generation)
E) Dimensionality Reduction	
i. Principal Component Analysis (PCA)	• Network Reliability (SAIFI, SAIDI, CAIFI etc.) • Transformer Ageing • MV Transformer and LV feeder Alarm Analysis
F) Deep Learning	
i. Convoluted Neural Network (CNN)	Electricity Meter Reading
ii. You Look Only Once (YOLO)	Electricity Meter Reading

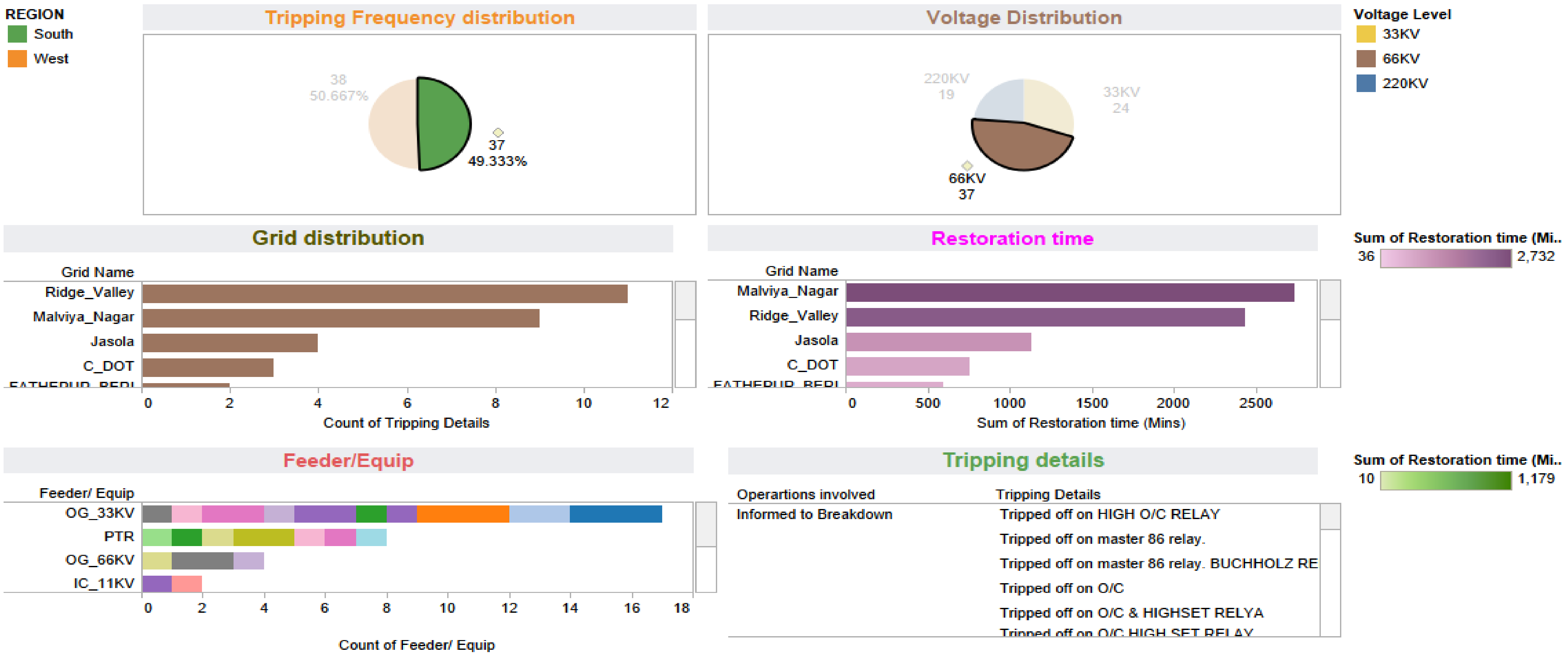
Machine Learning Algorithms (5/5)

ML Algorithm	Applications in the Power Utility Domain
iii. Recurrent Neural Network (RNN)	• Customer Complaint Analytics • Insights form call logs • Energy Market Prediction
iv. Automated Neural Network (ANN)	• Network Reliability (SAIFI, SAIDI, CAIFI etc) • Transformer Ageing • MV Transformer and LV Feeder Alarm Analysis

Use Case 4: EHV Tripping Analysis at GRID/Substation Level (1/2)



Use Case 4: EHV Tripping Analysis at GRID/Substation Level (2/2)



Use Case 5: Distribution Transformer Monitoring (1/7)

Variables

- Age of the transformer in days
- Average Maximum Temperature of transformer
- Temperature of Oil
- Temperature of various gases in Oil
- Average Maximum Load of transformer
- Average KVA rating of transformer
- Number of times the transformer was normal
- Number of times the transformer was overloaded
- Indicator (Residential – Commercial)
- Type of User(Mix-Industrial-Home-commercial)
- Harmonics Data
- Phase (Red-Yellow-Blue)
- Maintenance Cycle

Oil /submerged Gases

- Hydrogen (H_2),
- Methane (CH_4),
- Acetylene (C_2H_2),
- Ethylene (C_2H_4)
- Ethane (C_2H_6)
- Carbon monoxide (CO)
- Carbon dioxide (CO_2)

Devices/Sensors used

1. Buchholz (Gas) Relay
2. Pressure Relay
3. Oil Level Monitor Device
4. Winding Thermometer
5. Fiber optic sensor used to measure temperature
6. RTD used to measure temperature

Important Results

- Transformer Life Time Prediction and Analysis
- Predictive Maintenance of Transformer

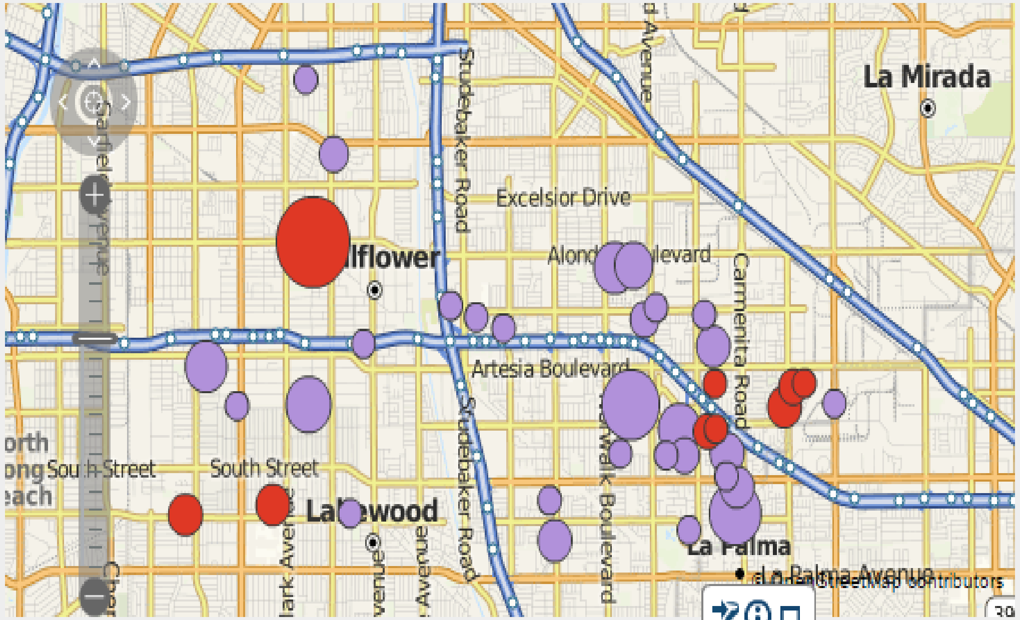
Use Case 5: Distribution Transformer Monitoring (2/7)

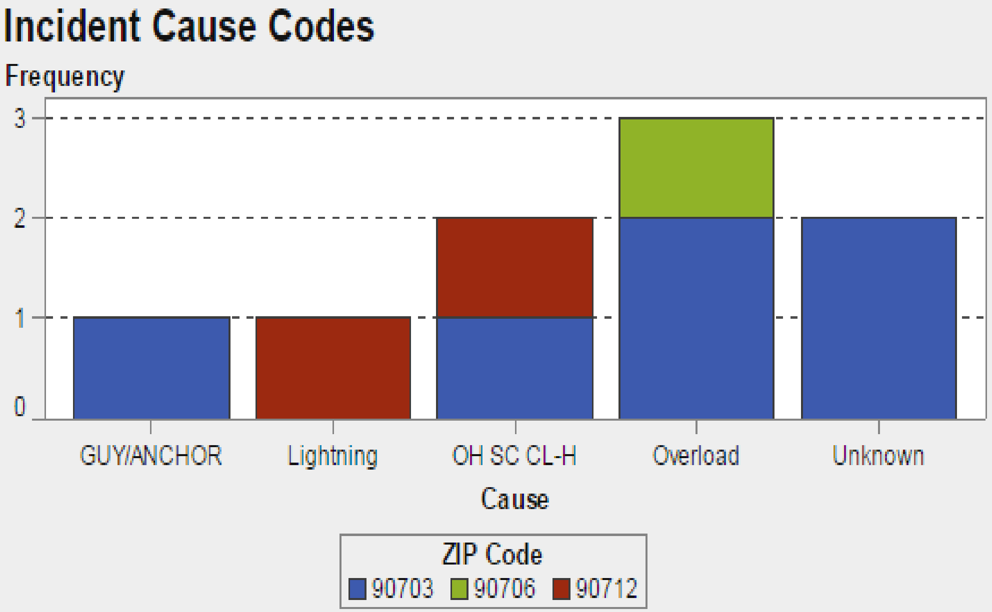
Use Case 5: Distribution Transformer Monitoring (3/7)

Transformer Analysis | Transformer Load | Transformer Ratings

Zip Code

☑	90703	44
☑	90706	14
☑	90712	12
☑	90713	11





Configuration

☐	(missing values)	30
☐	CDT	1
☐	ODT	10

Under/Overhead

☐	O	43
☑	U	38

Transformer Details

Transformer	ZIP Code	Highsideconfig...	DTMType	Ratedkva	Probability... ▼	MeterCount	Probability of Clas...	Cause
5253055899	90712	ODT	U	25	100%	5	100%	Lightni
5252602815	90712	SNG	U	25	100%	5	100%	OH SC
5069511541	90703	SNG	U	50	99%	3	99%	Overloa
4969609696	90703	SNG	U	50	98%	4	98%	Unknov
5071052658	90703	SNG	U	50	97%	4	97%	Overloa

Meters

Transformer	Meter Num.. ▲
5555850981	56819256
5255620927	60115356
5255620927	60115489
5555850981	60119561
5555850981	60119562

Use Case 5: Distribution Transformer Monitoring (4/7)

Transformer Analysis | Transformer Load | Transformer Ratings

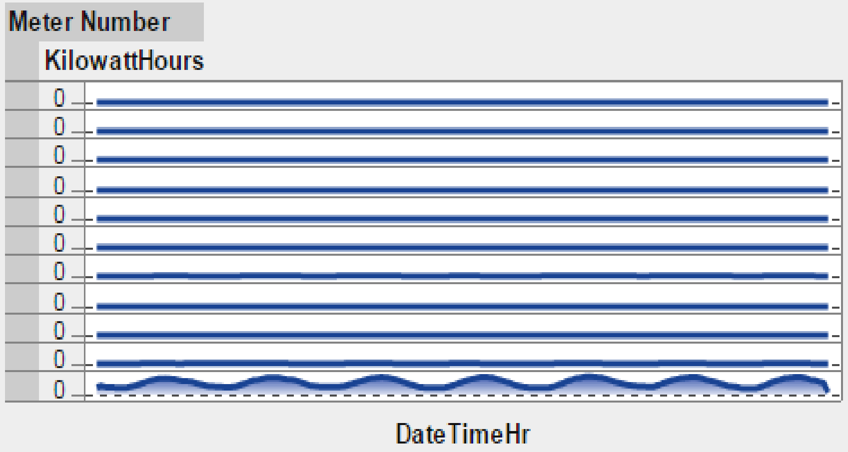
Transformer Ratings

Transformer	Ratedkva	Highsidecon...	Probability for level 1 of failed	MeterCount	Probability of Classification	PowerOutage6	PowerOutage3	PowerOutage12
4959931079	50	SNG	10%	7	90%	24	11	41
4962900368	75	ODT	86%	8	86%	30	30	30
5060458770	50	SNG	47%	3	53%	65	59	65
5062634014	25	SNG	80%	5	80%	48	24	89
5155952919	100	SNG	60%	19	60%	226	68	226

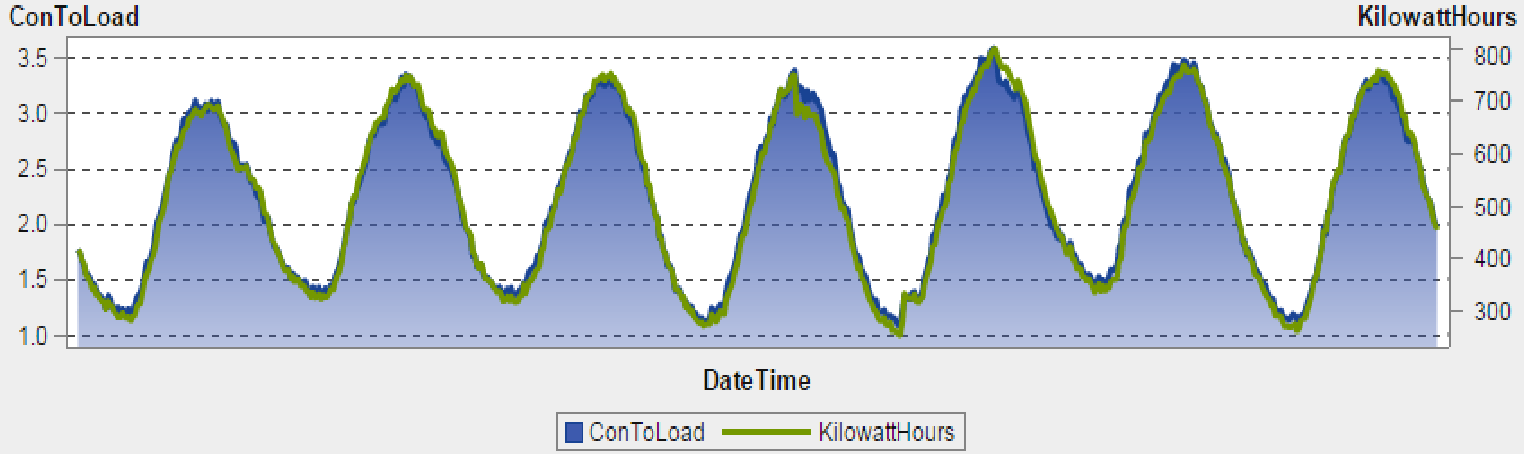
Meters

Meter Num..	KilowattHours ...	ConToLoad
56819256	0.5630	0.786875
60115356	2.2570	1.45288
60115489	3.4910	5.157385
60119561	0.0000	0
60119562	0.1170	0.117625
60119563	0.9820	1.41389

Meter Load

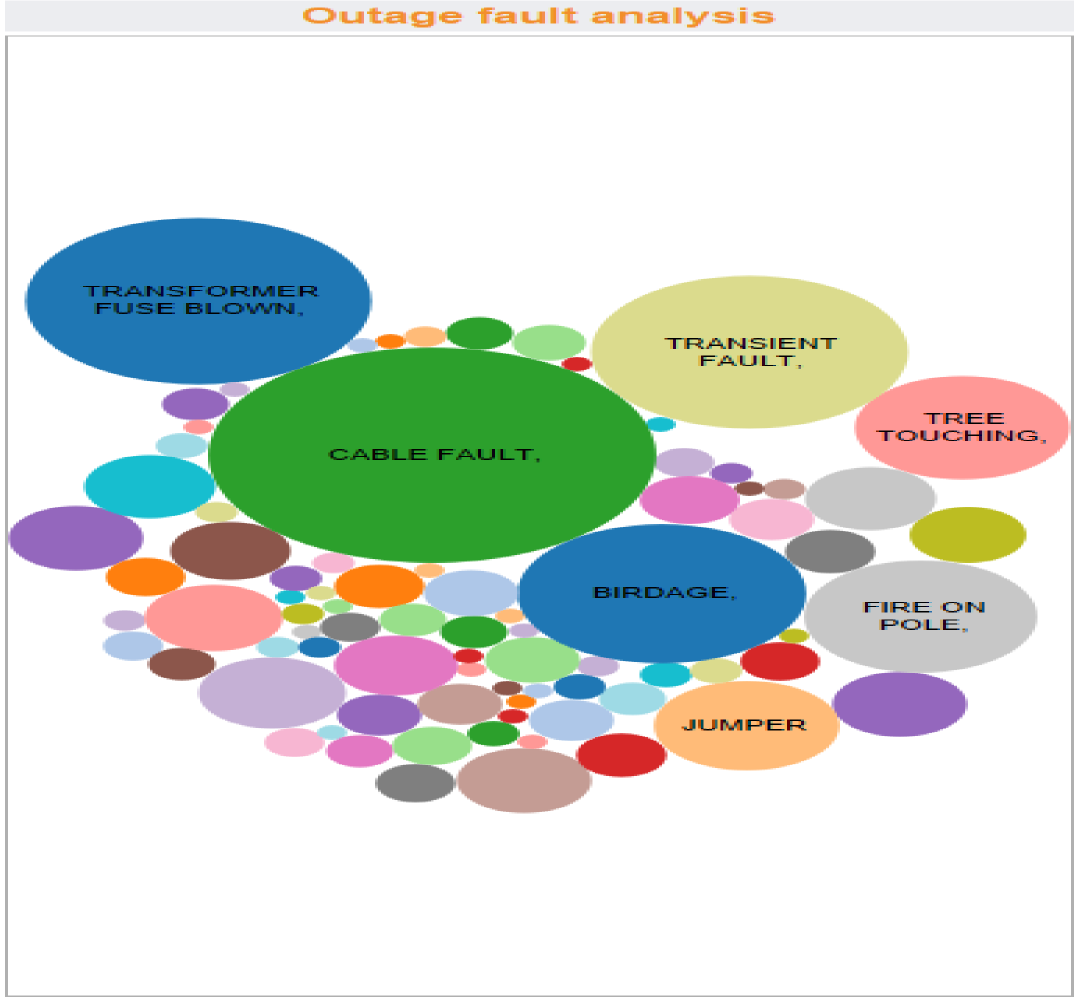


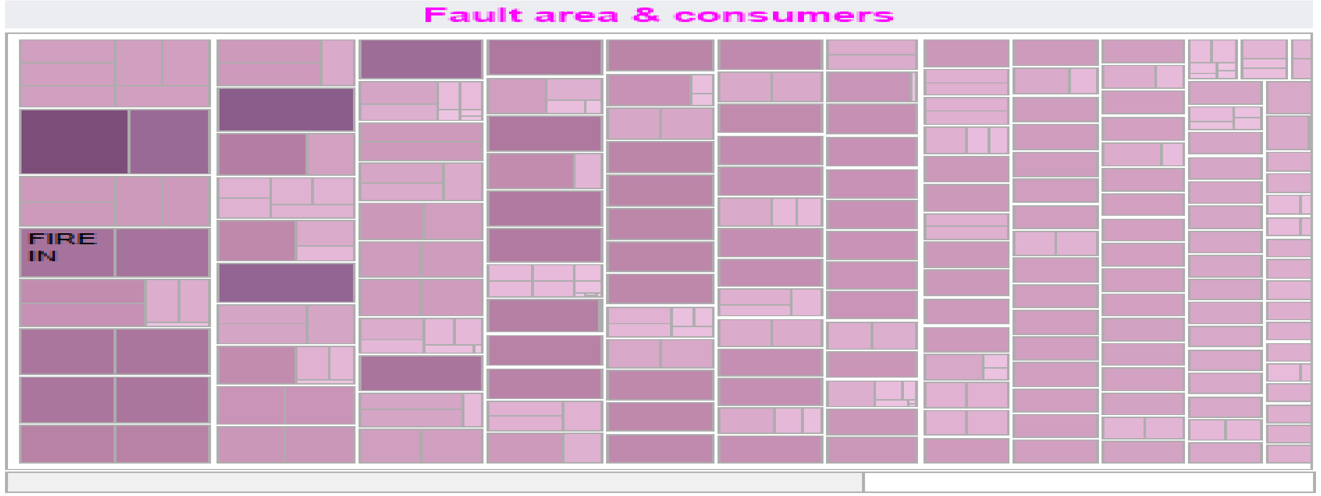
Load Analysis

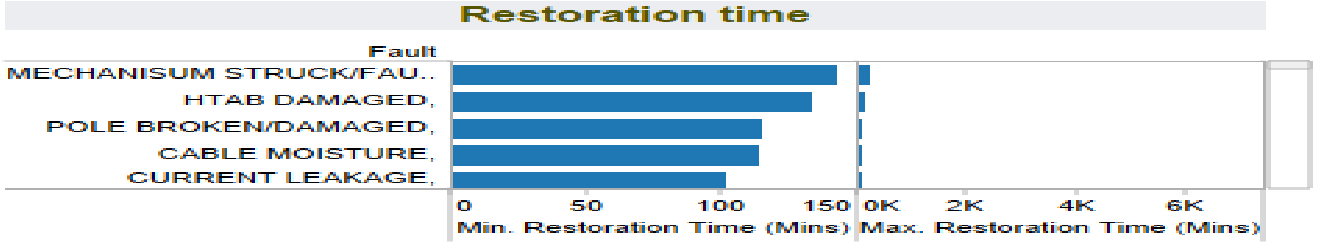


Use Case 5: Distribution Transformer Monitoring (5/7)

Breakdown Analysis





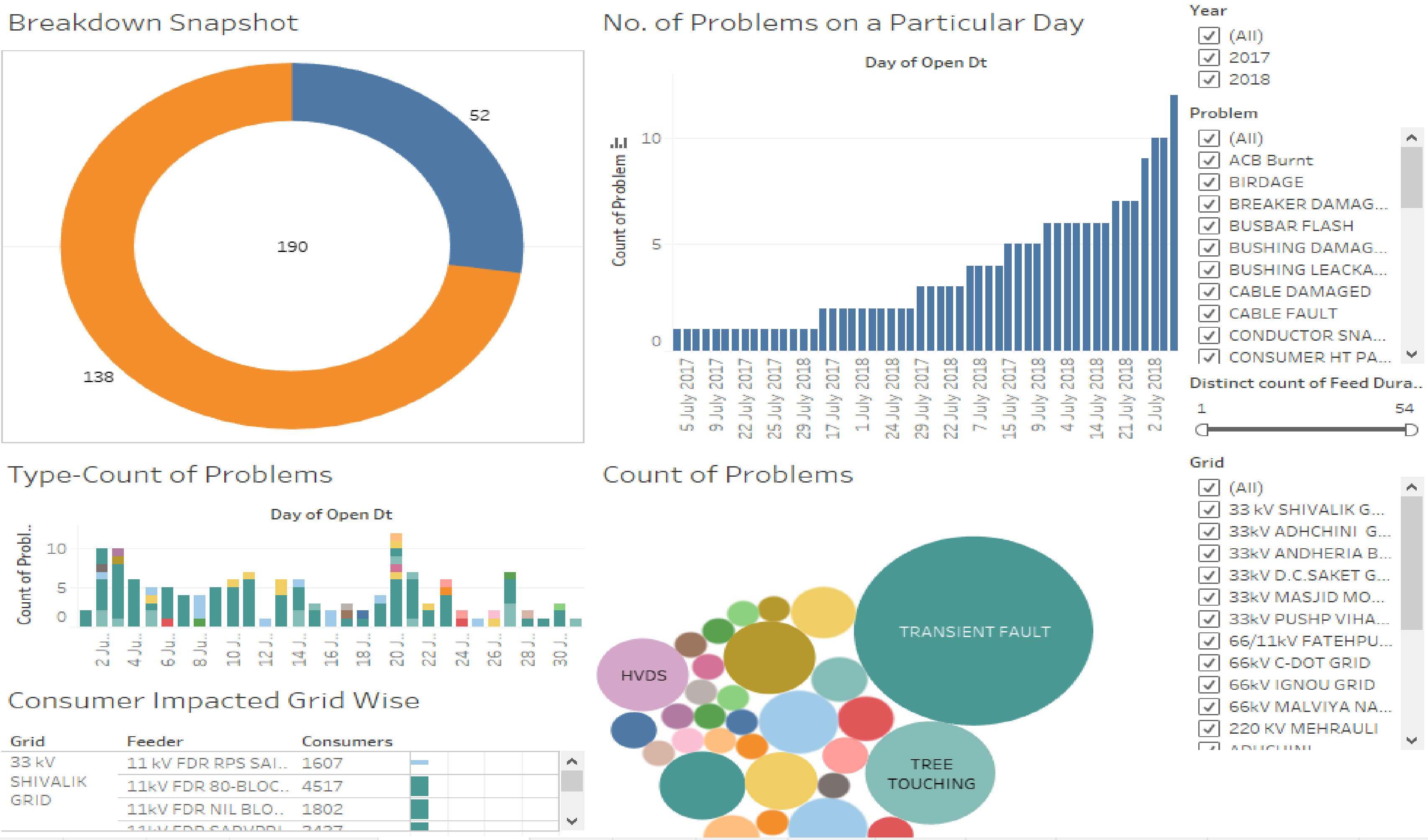


Problem description

Fault	Count of Fault	Restoration Tin
ABNORMAL SOUND FROM THE TRANSFO..	10	
BIRDAGE,	92	
BUS BAR BURNT,	1	
BUSBAR FLASH,	5	
BUSHING DAMAGED,	5	
CABLE BURNT,	9	

Use Case 5: Distribution Transformer Monitoring (6/7)

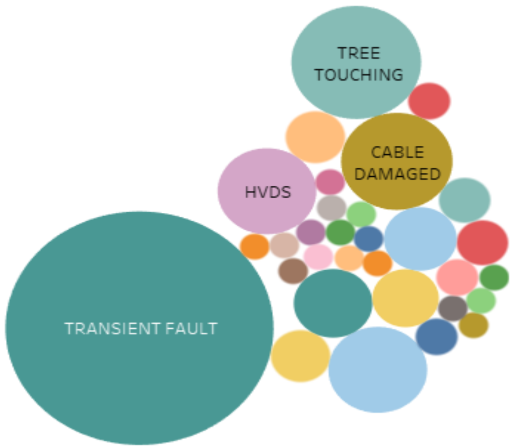
Breakdown Analysis



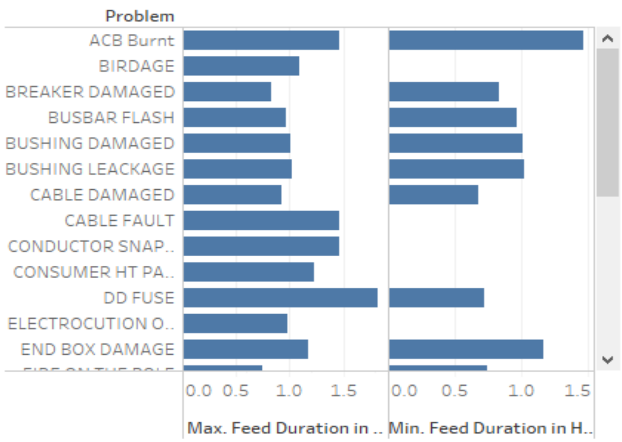
Use Case 5: Distribution Transformer Monitoring (7/7)

Outage Analysis

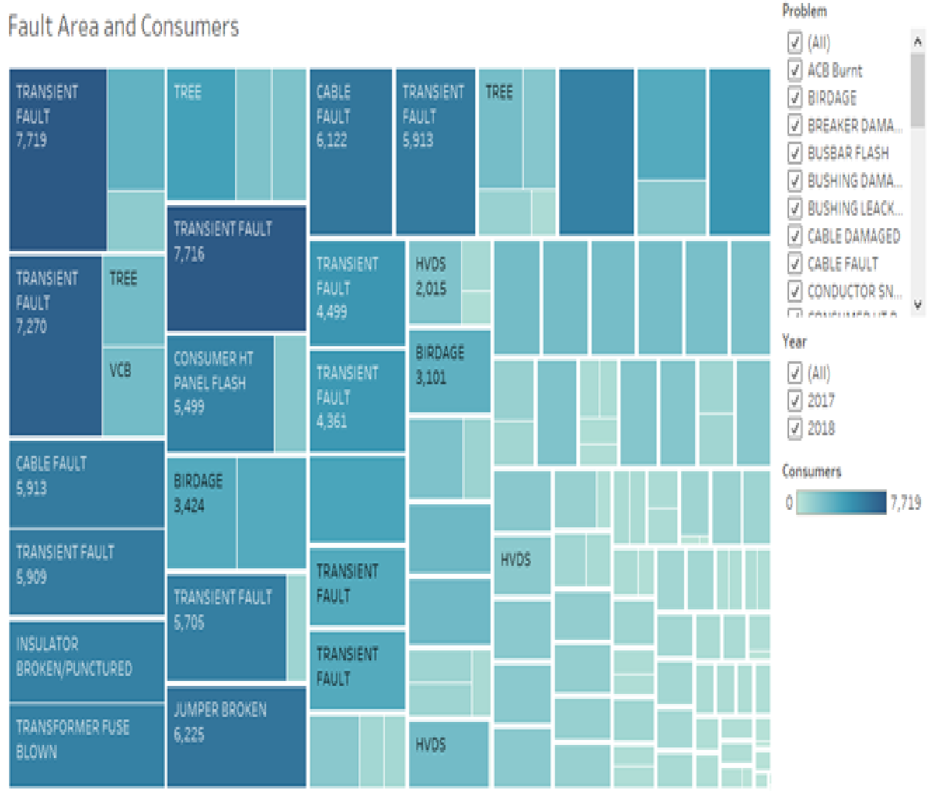
Outage Fault Analysis



Problem Description



Fault Area and Consumers

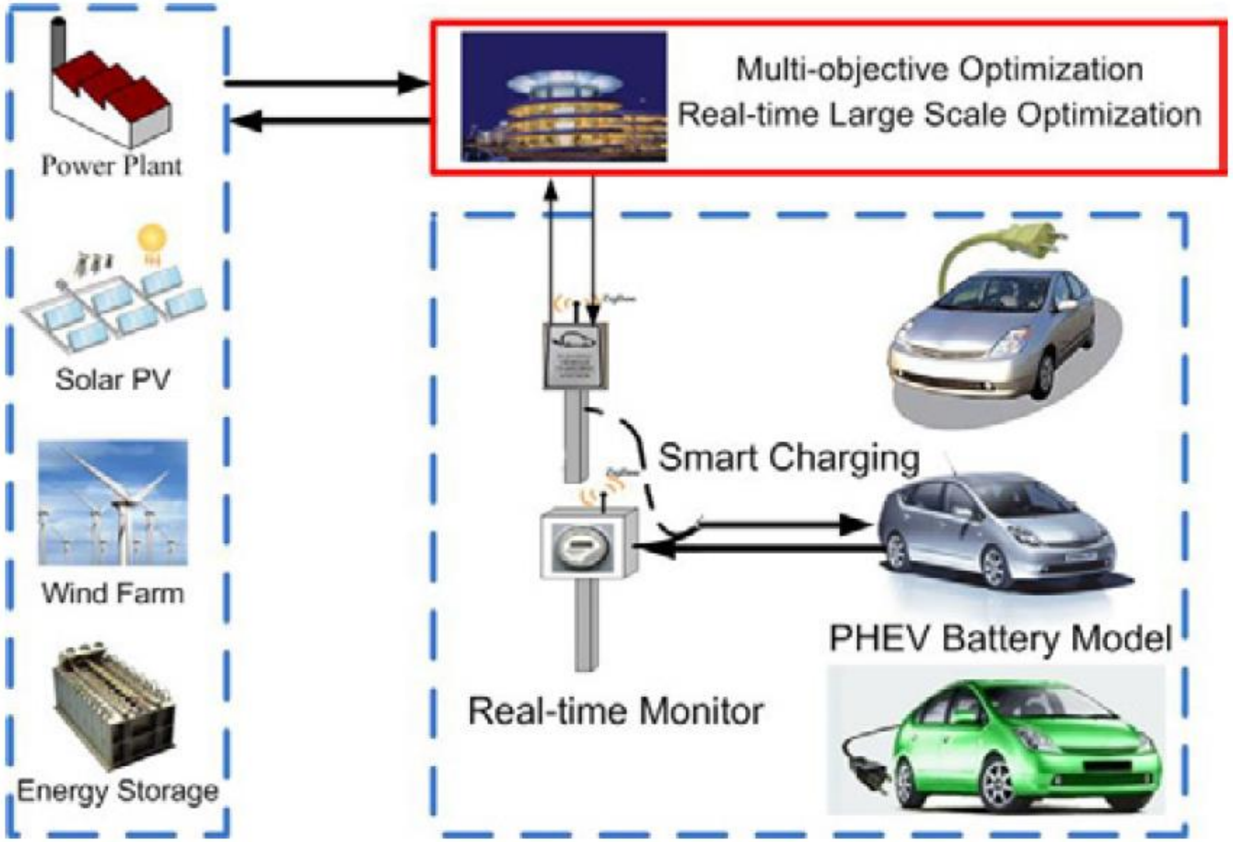


Feed Restored Duration

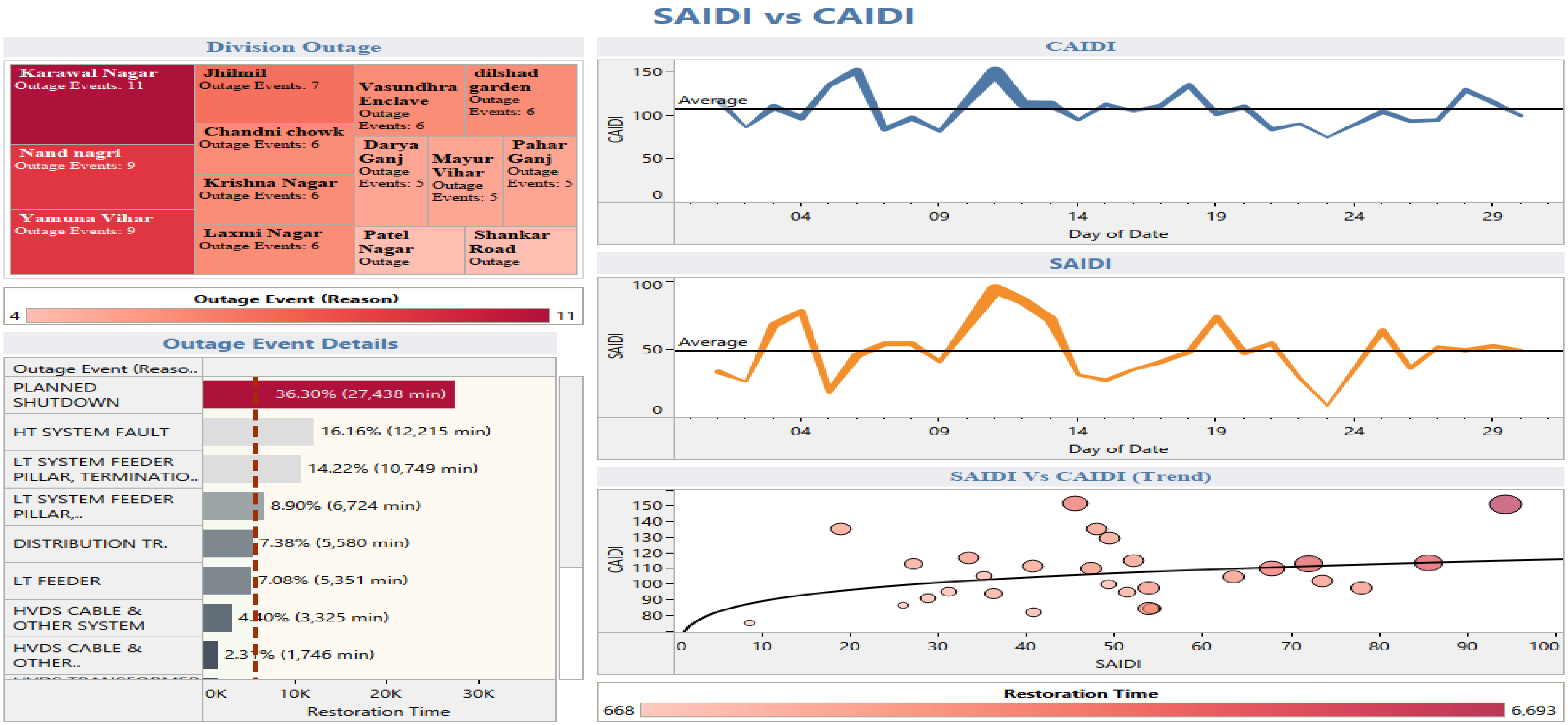
Problem	Consumers	Feed Duration	Count of Problem
ACB Burnt	722	1	1
BIRDAGE	11,516	4	6
BREAKER DAMAGED	619	1	1
BUSBAR FLASH	349	1	1
BUSHING DAMAGED	2,220	1	1
BUSHING LEACKAGE	14	1	1
CABLE DAMAGED	10,555	11	14
CABLE FAULT	15,133	3	5
CONDUCTOR SNAPPED	5,695	6	7
CONSUMER HT PANEL FLA..	5,499	2	3
DD FUSE	2,366	3	3
ELECTROCUTION OF EXTE..	0	1	2
END BOX DAMAGE	87	1	1
FIRE ON THE POLE	41	1	1

Use Case 6: Electric Vehicles

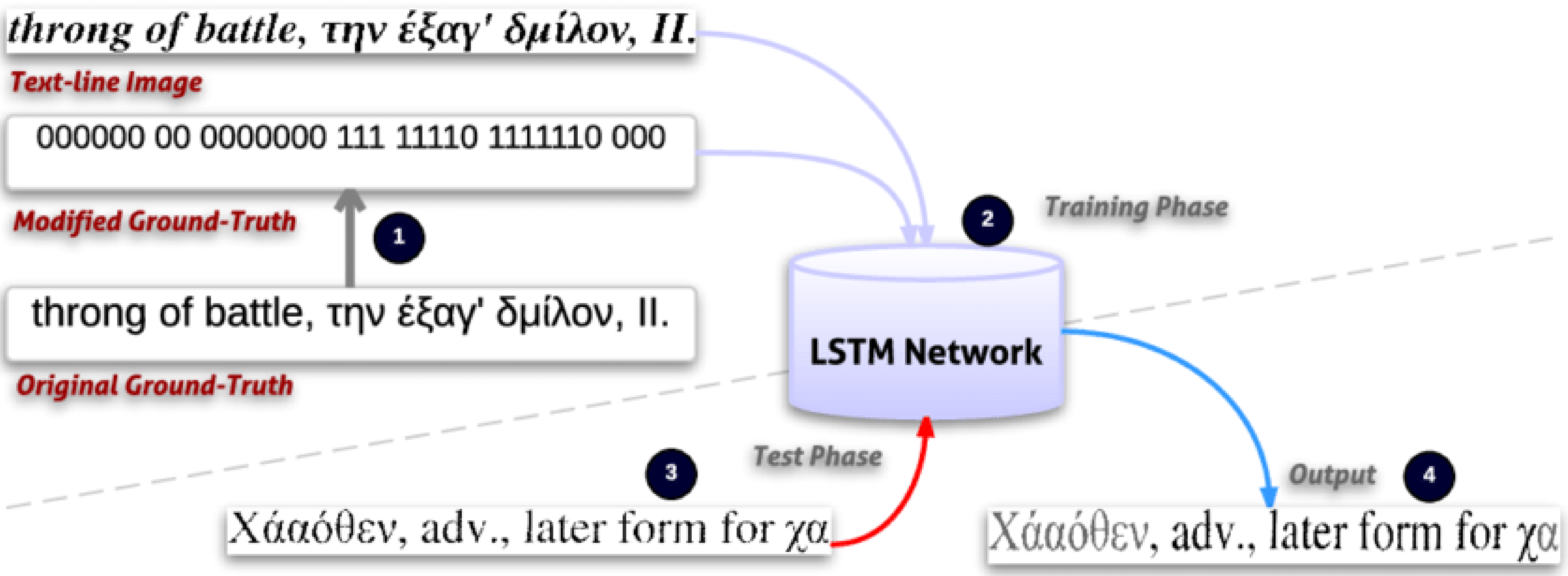
- Electric adoption is fast increasing
- BYD, Tesla, GM, Ford and Nissan are the market leaders in this segment
- EV integration with Smart Grid can optimize the charging time to reduce the charging cost
- Integration with Smart Grid would also enable an EV to trade energy with utilities - Vehicle to Grid (V2G)
- This enables the load balancing on the grid – large number of EVs can be aggregated as Virtual Power Plants (VPPs) with AI Algorithms



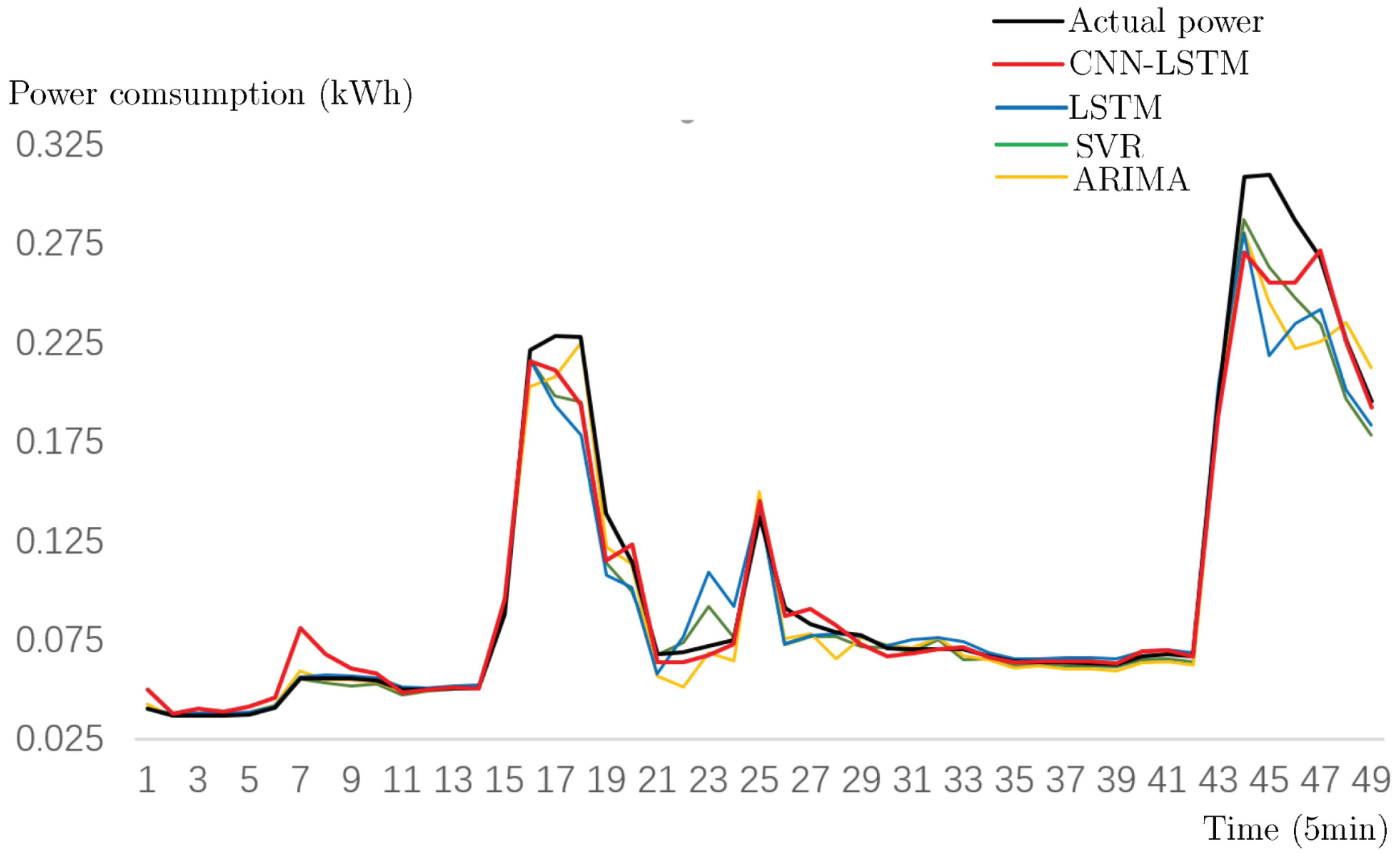
Use Case 7: Network Reliability Analysis



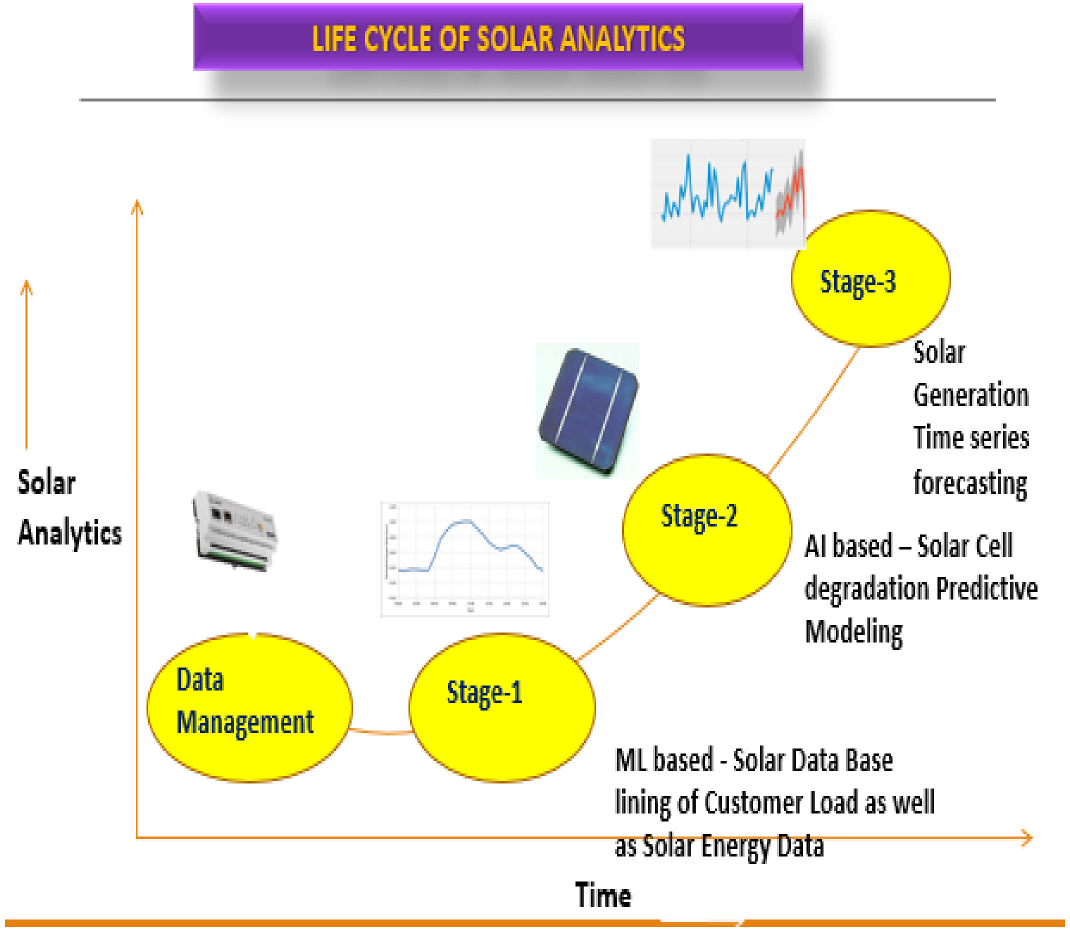
Use Case 8: Textual Data – Insights Form Call Logs



Use Case 9: Time Series Data – Energy Market Prediction



Use Case 10: Balancing of Power Procurement Strategy with Impact of Distributed Generation



SOLAR Analytics

Data KPI Analysis

- Show Table
- Energy Generation Analysis
- Temperature Analysis
- Energy & Temperature Relation
- Power Curve Analysis (Active, Reactive & Apperant Power)
- Grid Phase, Voltage and Current Analysis
- PV Curve Analysis
- Residual Energy Analysis

Advance Analytics

- Inverter Efficiency Cluster Analysis
- String Regression Modelling
- Module/Environment Variable Correlation
- Solar cell/module Degradation Analysis
- Plant-Output Forecasting
- Plant Optimization with Log Book Analysis

Use Case 11: Power Generation

Renewable Management

- Short-term enhancement of renewable energy forecasting and optimizing equipment's can be done through AI
- AI is used to predict sunlight intensity to improve the wind turbine data and solar panel sensor data.
- Energy storage and estimating lifetime of a battery are also few of the areas where AI can be used to optimize the operation.

Demand Management

- AI can help strike a balance between renewable energy and conventional energy like fossil fuel to ensure a steady energy supply.
- Individuals with renewable energy equipment's like solar panels and battery storage system can trade energy with utilities through AI. This can help in load balancing in the grid.

Infrastructure Management

- AI techniques like machine learning algorithms, intelligent systems can analyze the risk and optimize opportunities by collating, comparing, analyzing and highlighting the data
- AI is also used to suggest actions and impact by modelling of the data set
- AI helps network operators to inform decision making and judge better situational awareness
- Usage of AI tools can improve the lifetime of the grid equipment
- Grid management companies like Siemens and GE are deploying AI applications to increase the output of traditional assets.

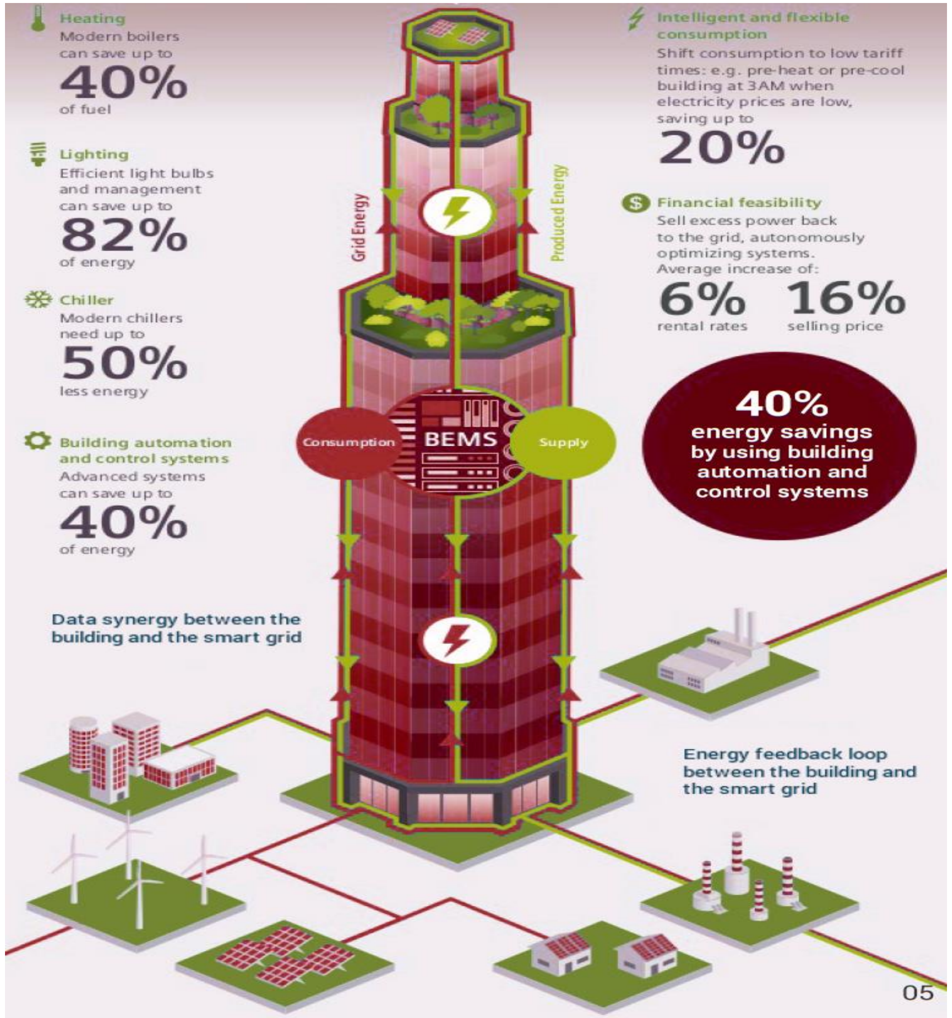
Use Case 13: Building Energy Management System

Building Energy Management System (BEMS)

- Reduce energy consumption and carbon footprint
- Increase tenant satisfaction and loyalty
- Protect tenants with non-intrusive security methods
- Lower operating costs
- Maintain buildings and comply with regulations
- Manage modern buildings' technology and systems, and
- Reap maximum ROI

Solar and Wind energy

Financial benefits of smart buildings:



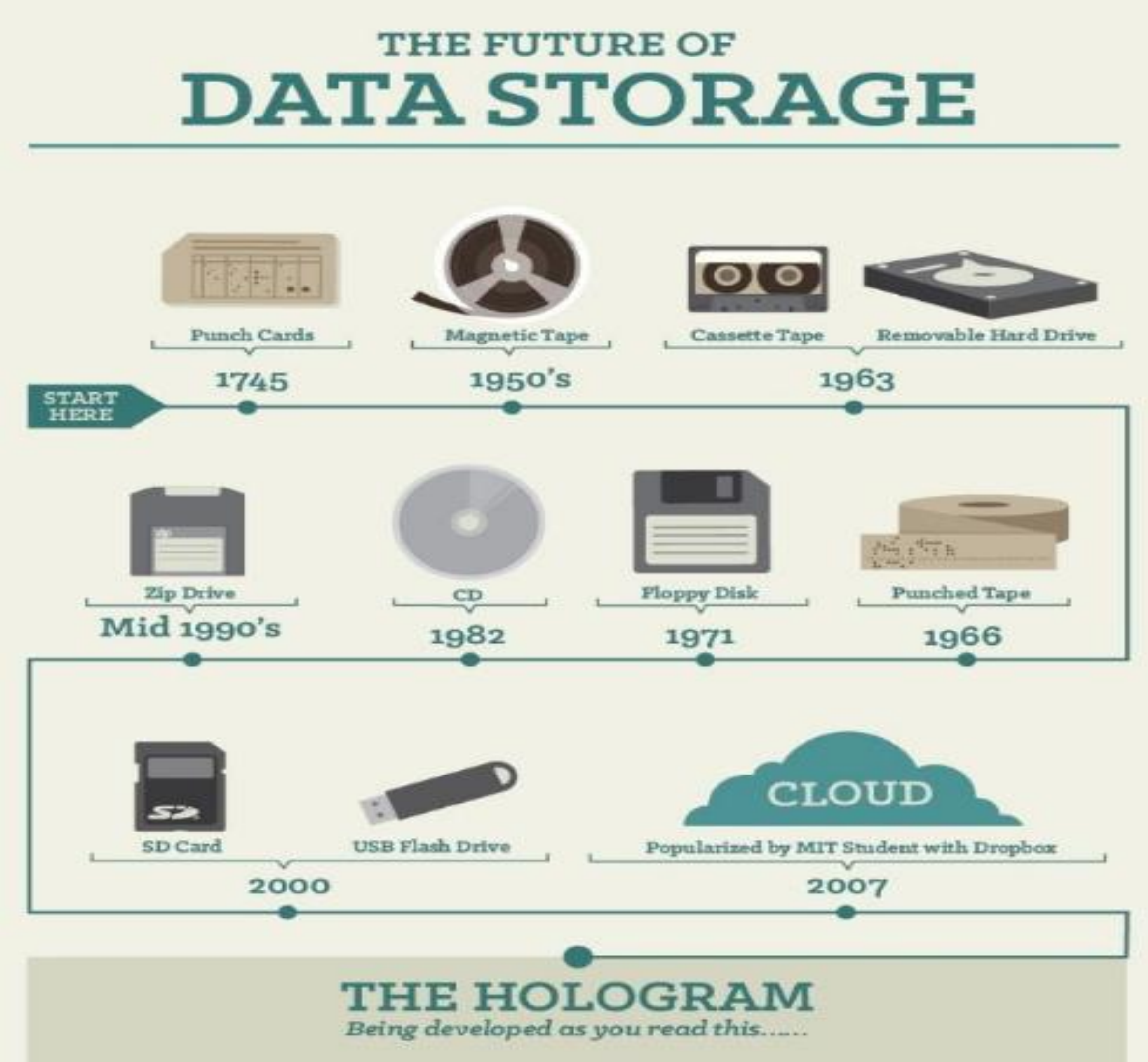
Advanced Analytics (1/2)

Type of Analytics	What it Does
Descriptive Analytics – Business Intelligence and Data Mining	▪ Condense big data in to smaller more useful nuggets of information ▪ Analyze the real-time data and historical data for insights on how to plan for future
Predictive Analytics – Forecasting	▪ Forecast what might happen in future ▪ All predictive analytics are probabilities ▪ Give answers to questions that cannot be answered by Business Intelligence ▪ Data Reduction ○ What will happen next if <condition> : Predictive Modelling ○ Why this happened: Root Cause Analysis ○ Identify correlated data: Data Mining ○ What if same trend continues: Forecasting ○ What could happen in an unknown scenario: Montecarlo Simulation ○ When should an action be invoked: Pattern Identification and Alerts

Advanced Analytics (2/2)

Type of Analytics	What it Does
Prescriptive Analytics – Simulation and Optimization	▪ Stochastic optimization to understand how to achieve best outcomes ▪ Identify data uncertainties to make better decisions ▪ Reduce duplications and readmissions o Used in calculating credit score of customers who are likely to pay loan installments on time

Evolution of Data Storage (1/2)



Evolution of Data Storage (2/2)

Byte
(8 bit)
1 byte=1 single character "A"
10 bytes= 1 word "APPLE"
100 bytes= Telegram or Punch card

Kilobyte
(1000) Bytes
1kb= A very short story
200 kb= A box of punch cards

Megabyte
(10^{6})
1MB=small Novel
500 MB= A CD ROM or hard disk of PC

Gigabyte
(10^{9})
1 GB= A pickup truck filled with paper
500 GB= The biggest FTP Site

Terabyte
(10^{12})
1TB= 50,000 trees made in to paper
400 TB= National Climatic Data Center Database

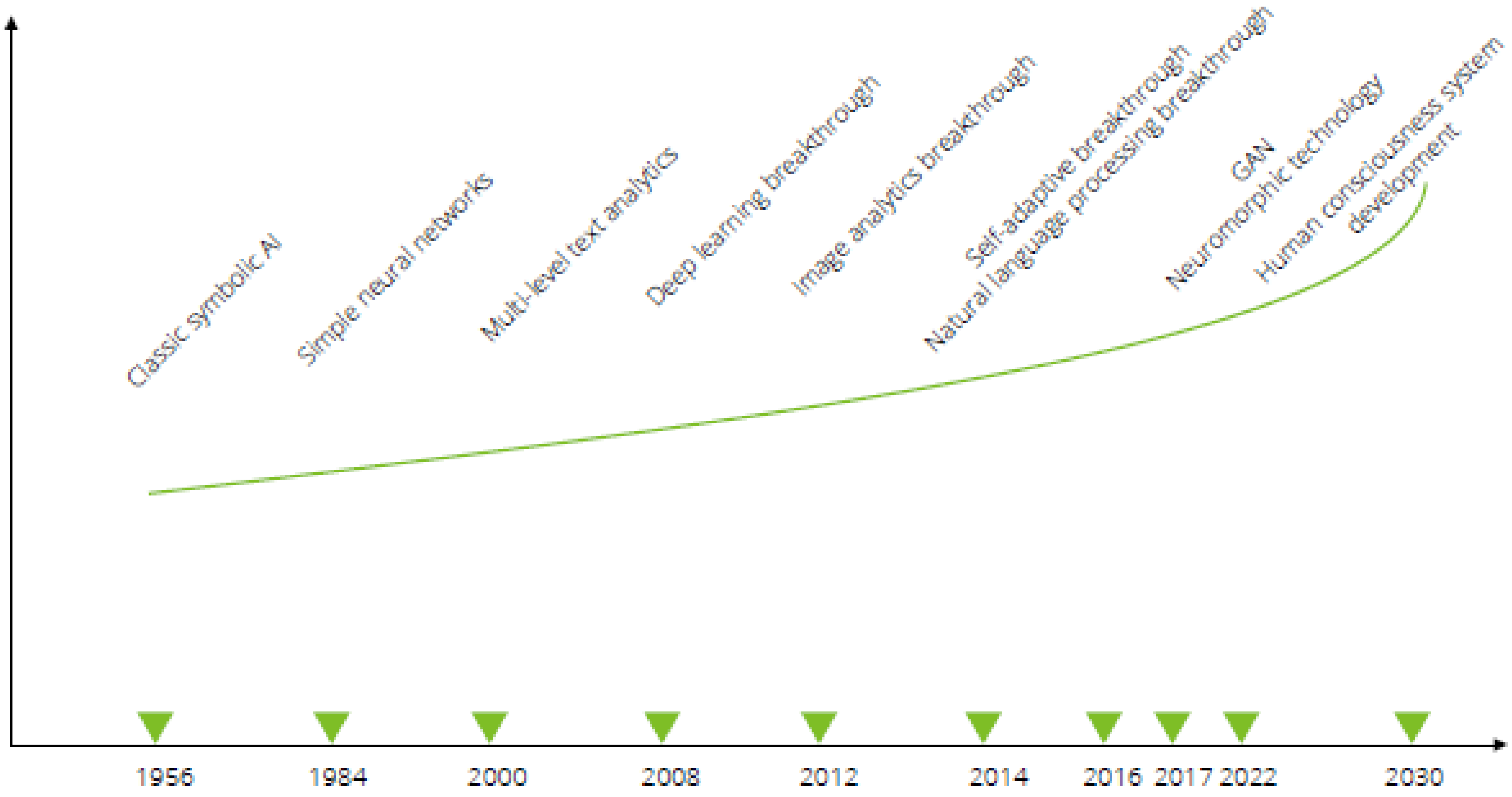
Petabyte
(10^{15})
1PB=3 Years of EOS Data
200 PB= All printed material or production digital tape in 1995

Exabyte
(10^{18})
2 EB= Total volume of information generated annually
5 EB= All words ever spoken by human beings

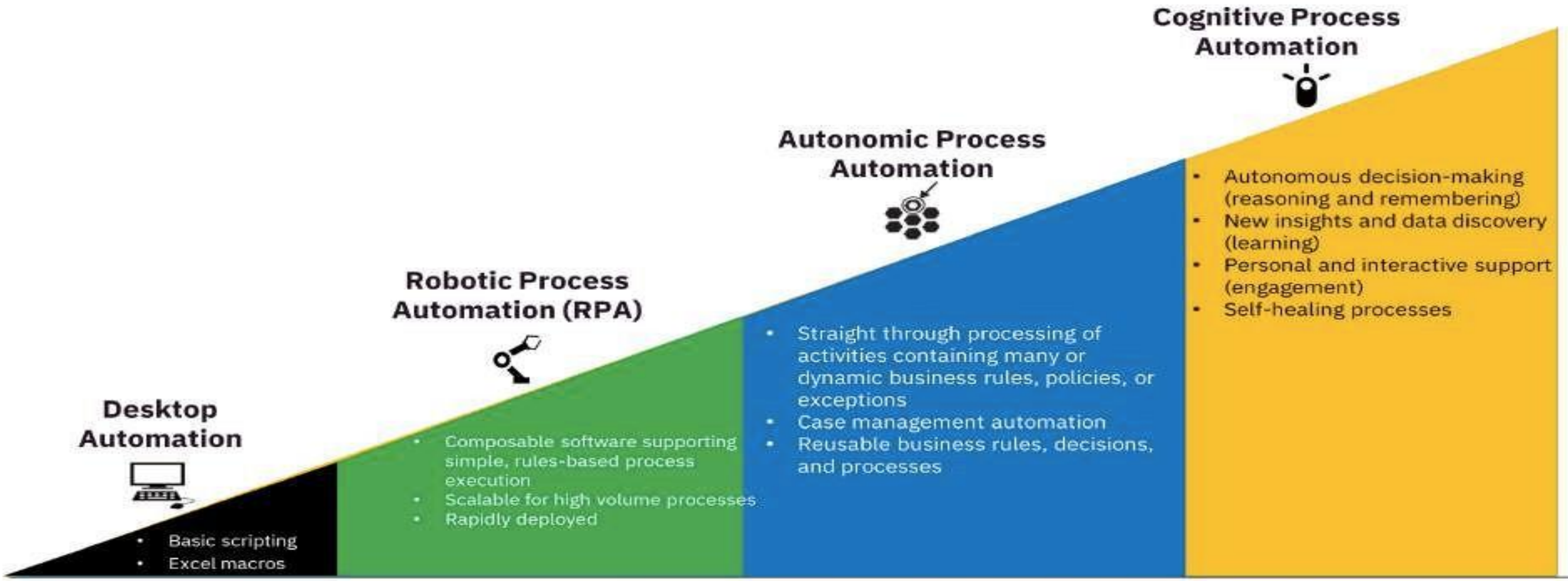
Zettabyte
(10^{21})
Too big for any application (as of now)

Yottabyte
(10^{24})
Too big for any application (as of now)

History of AI Development



Paths to Automation



What it automates...	Simple data migration	Rules-based processing when structured data exists	Processes with moderate to high exception rate when bundled with RPA	Analytics, interaction or transformation when data is unstructured
When it's not appropriate...	Virtually everything else	Unstructured data Decision making Managing workflow	Predictive reasoning, automated analytics, Natural Language Processing, etc.	Simple process automation