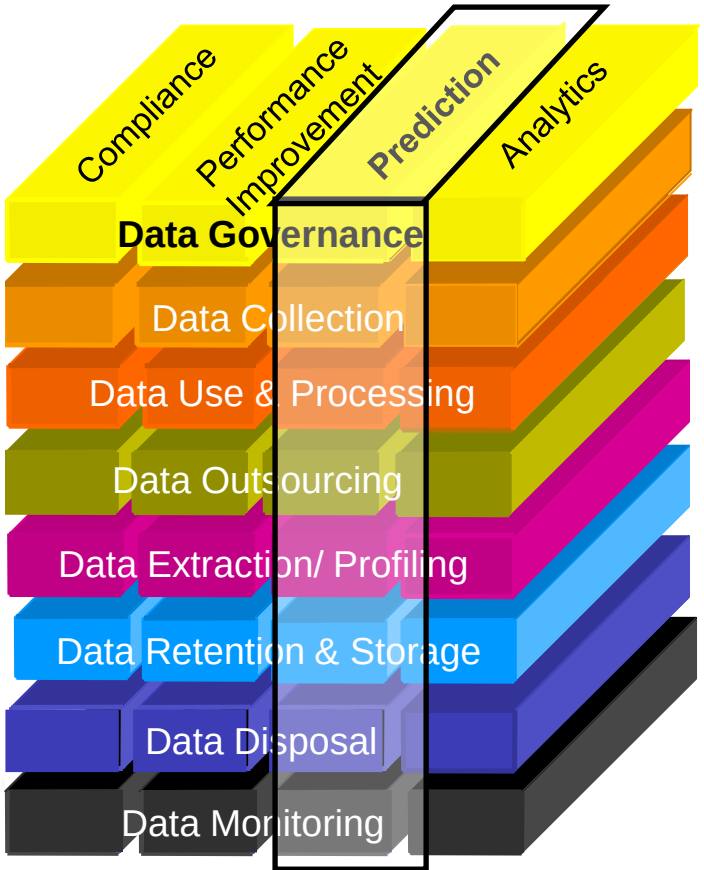
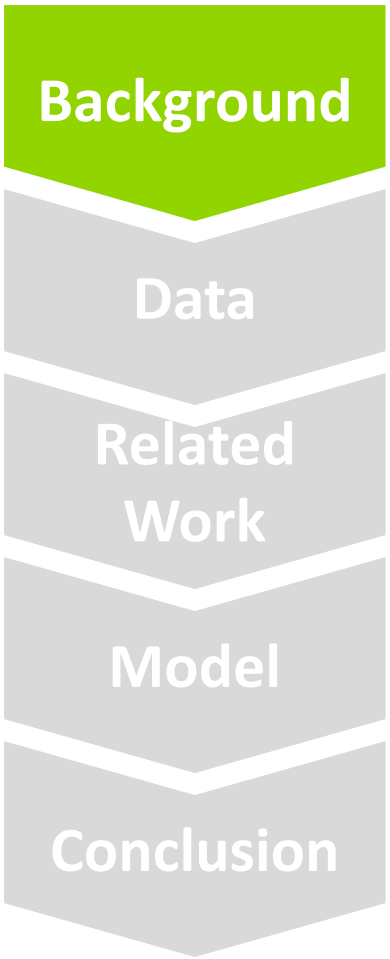
Learning System Security Compliance for Banking



SECTIONS



- Provide background information.

Motivation

Jan 2020 –

A data breach resulting from a cyber-attack on the customer management system for P&N Bank.

August 2019 –

A malware has been used by hackers stealing the PayID of customers across Australian big four banks.

March 2019 –

Bank of Queensland experienced a personal data breach by a third-party provider.

Why Machine Learning

1. Prediction of information security levels (ISLs) in a real time mode

2. Generation of compliance reports to analyze ISLs by thirteen information security controls (ISCs)

3. Meeting the CPG 234 for the security compliance purpose benefiting the banking industry

Problem Statement

Australian Prudential Regulation Authority provided a new guideline in 2019 - “Prudential Practice Guide CPG 234 Information Security” (CPG 234)*. Financial institutions are required to comply with the CPG 234.

- Many banking systems attacked → data breach incidents
- Banks manage hundred systems and monitor the information security of their hundred contractual parties handling their customer data
- Cases of failure in preventing System Risks

Resolution of the Problem

- Machine learning to automate the compliance process

* There is a range of information security controls (ISCs) influencing information security levels (ISLs) over systems, networks & information assets (collectively named “Systems”)

Paper Focus

Develop a machine learning model to train neural networks (NN) for automating the compliance process:

1. Recurrent Neural Networks (RNNs);
2. Long Short-Term Memory (LSTM) RNNs;
3. Divergent LSTM RNNs (Feedforward and Backward); and
4. Attention mechanism (ATT) applied to LSTM RNNs.

Train multiple LSTM RNNs to predict ISLs and generate compliance reports.

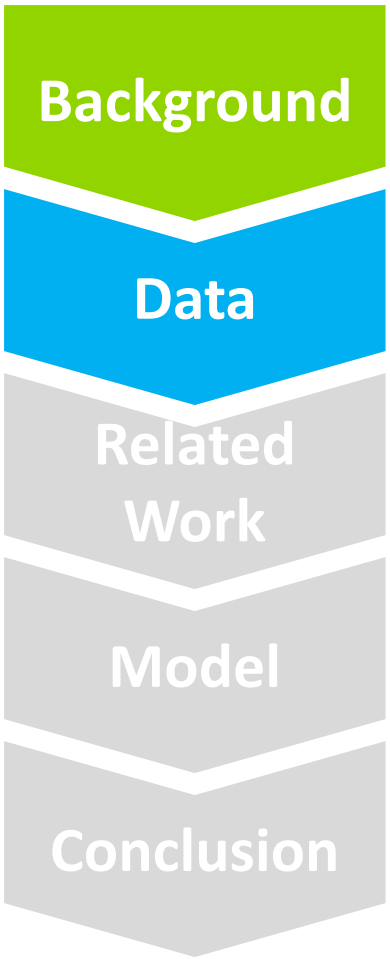
Objectives

To automate compliance process and predict ISLs using machine learning:

1. For improving financial institutions compliance readiness for the CPG 234; and

2. For assisting financial institutions in System Risks through reporting them by analytics.

SECTIONS



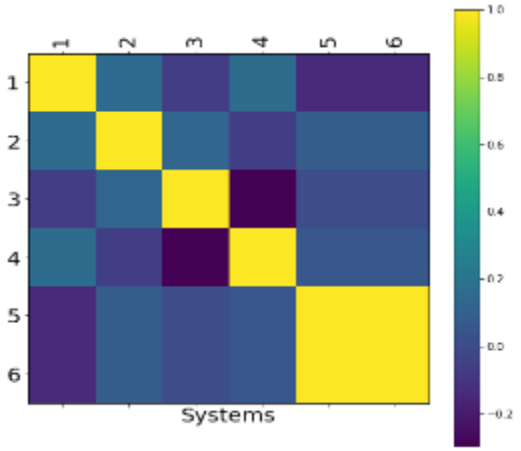
- Dataset generation and relevant features.

Dataset Characteristics

- Data Features → 6 Systems, 40 Instances, 3 data inputs (question number, question and score)

Data Features and Correlations

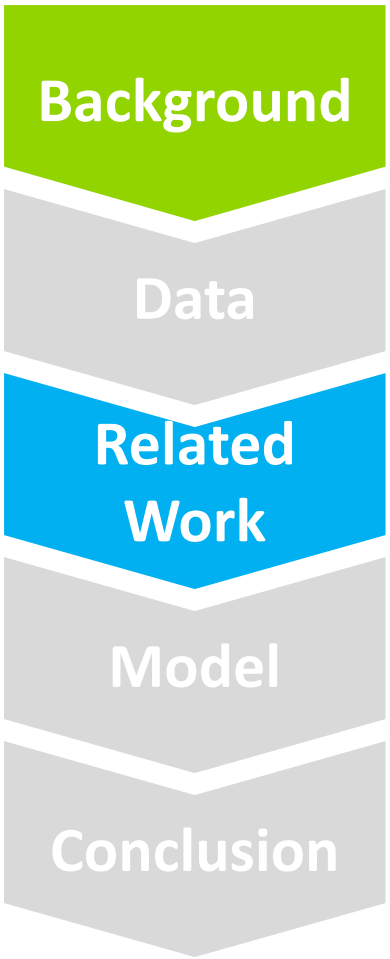
ISR No	ISR	Question No	Questions	System 1	System 2	System 3	System 4	System 5	System 6
1	1	1.1	.	5	1	5	3	1	1
2	1	1.2	.	1	3	2	4	3	3
3	1	1.3	.	1	3	5	2	1	1
4	2	2.1	.	4	5	5	1	5	5
5	2	2.2	.	4	5	5	4	3	3
6	2	2.3	.	1	5	4	3	1	1
7	3	3.1	.	1	2	5	1	4	4
8	3	3.2	.	4	4	2	5	4	4
9	3	3.3	.	1	5	4	3	5	5
10	4	3.4	.	4	5	5	2	1	1



Data Statistics

Scores	System 1	System 2	System 3	System 4	System 5	System 6
Mean	2.8250	3.5500	3.3500	3.0750	2.9500	2.9500
S.D.	1.5340	1.2598	1.5115	1.3085	1.4841	1.4841

SECTIONS



- Prior research efforts.

Machine Learning Compliance Process - Sequential Learning with Attention

No	Aspect	Title	Prediction Relevance
1	Image contents	Show, attend and tell: Neural image caption generation with visual attention	1. Two attention- based image caption generators to learn dependencies of a problem 2. Investigation of models that can attend to salient part of an image while generating its caption
2	Machine translation	Attention is all you need	1. Computation of response at a position in a sequence by attending to all positions in a self-attention network 2. Use of a transformer to train translation tasks faster than architectures based on recurrent or convolutional layers
3	Image generation	Self-attention generative adversarial networks	1. Improvement in the balance between the ability to model dependencies 2. Experimentation with a generative adversarial network with a self-attention
4	Sales prediction	TADA: Trend Alignment with Dual-Attention Multi-task Recurrent Neural Networks for Sales Prediction	1. Dual attention multi-task RNNs for sales prediction 2. Modelling of internal & external features within influential factors in sales time series in a multi-task fashion

1 – from ICML 2015
2 – from NIPS 2017
3 – from *CoRR* 2018
4 – from ICDM 2018

AI for GDPR Compliance

No	Aspect	Title	Relevance
1	AI Regulatory Compliance	Using artificial intelligence to support compliance with the general data protection regulation	Recognized the importance of AI technology in terms of compliance checklists, risk assessments, an automatic profiling and the reporting of breaches
2	Regulatory Evaluation Automation	Claudette meets GDPR Automating the Evaluation of Privacy Policies using Artificial Intelligence	Automated the legal evaluation of privacy policies under the GDPR with the Support Vector Machine, convolutional neural networks and LSTM network
3	Regulatory Compliance	Queryable Provenance Metadata For GDPR Compliance	1. Provided SPARQ queries and tested metadata to populate the GDPR readiness checklist automatically 2. Documented GDPA compliance, visualized information flows and provided knowledge graphs

1 – from *Artificial Intelligence Law* 2017

2 – from http://dx.doi.org/10.2139/ssrn.3208596 2018

3 – from *International Conference on Semantic Systems* 2018

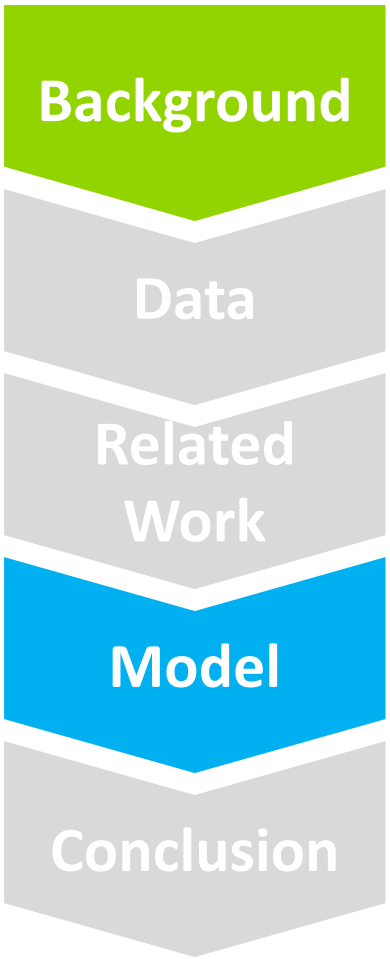
Machine Learning Model In Financial Services Industry

No	Aspect	Title	Relevance
1	Stock price prediction	Stock Priced Prediction Using Attention-based Multi-Input LSTM	1. Use of a multi-input LSTM model extracting valuable information from low-correlated factors and discarding their harmful noise 2. Introduction of new factors including the prices of other related stocks to improve the prediction accuracy
2	Liquidity risk assessment	An Artificial Neural Network and Bayesian Network model for Liquidity Risk Assessment in Banking	1. Measurement of liquidity risks with the artificial neural networks and Bayesian networks 2. Implementation of various algorithms to validate the model

1 – from *Machine Learning Research* 2018

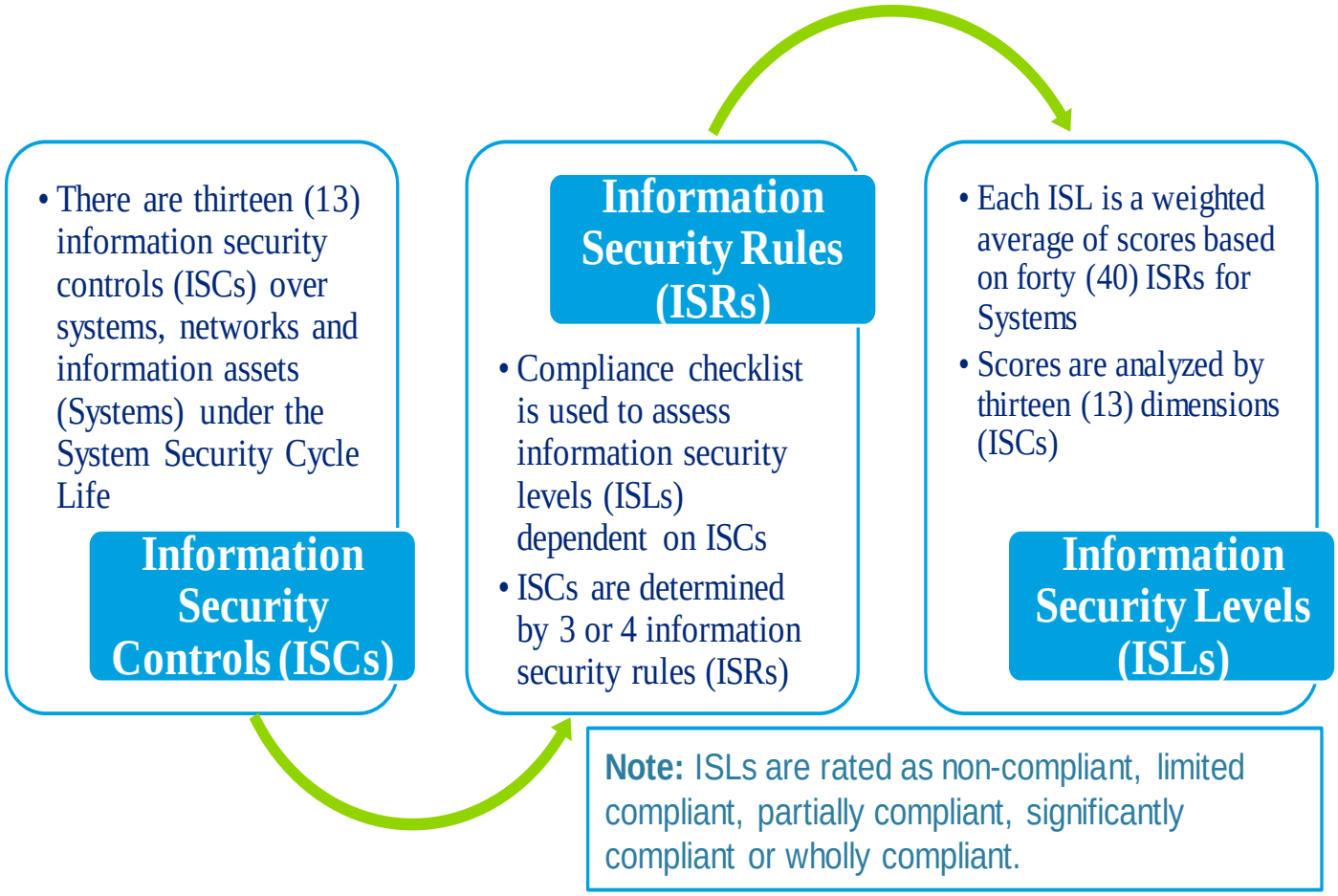
2 – from *Journal of Neurocomputing* 2018

SECTIONS



- Machine learning model.

Baseline Learning Approach



Baseline Information Security Rules (ISRs 1/3)

	ISR_1 to ISR_3 Under Process 1 – Change Management
1.1	Are changes to systems reviewed for the changes in risk profiles? (Jassal and Sehgal 2013)
1.2	Are data kept confidential and private in new e-banking technologies? (Abukhzam and Lee 2010)
1.3	Is the data model adjusted to the changed system? (Bajaber et al. 2016)
	ISR_4 to ISR_6 Under Process 2 – Configuration Management
2.1	How well is system configured to protect against vulnerability? (Farn et al. 2004)
2.2	Are uses restricted from accessing server configuration files to avoid directory traversal attacks? (Katkar and Kulkarni 2012)
2.3	Is virtualization focused configuration management tool deployed? (Brooks et al. 2012)
	ISR_7 to ISR_9 Under Process 3 – Deployment and Environment Management
3.1	Is banking software development segregated from software testing? (Grüttner et al. 2010)
3.2	Is cloud computing in bank systems segregated by logical storage? (Nedelcu et al. 2015)
3.3	Is a regular review performed to confirm who manages and administers data, and controls to detect and react to security breaches? (Caroll et al. 2011)
	ISR_{10} to ISR_{12} Under Process 4 – Access Management Controls
4.1	Is system access assigned based on user roles which are constantly updated? (Schaad et al. 2001)
4.2	Are access controls implemented in banking biometrics systems? (Venkatraman 2008)
4.3	Are access controls for outsourced vendors defined in SLA? (Baldwin et al. 2001)
	ISR_{13} to ISR_{15} Under Process 5 – Hardware & Software Asset Controls
5.1	Is an external machine authenticated and authorized based on cyber banking security protocols and standards? (Mbelli and Dwolatzky 2016)
5.2	Does IP packet filtering protect networks against intruder attacks? (Zhu 2002)
5.3	Are firewalls configured to protect against unauthorized access? (Zhu 2002)

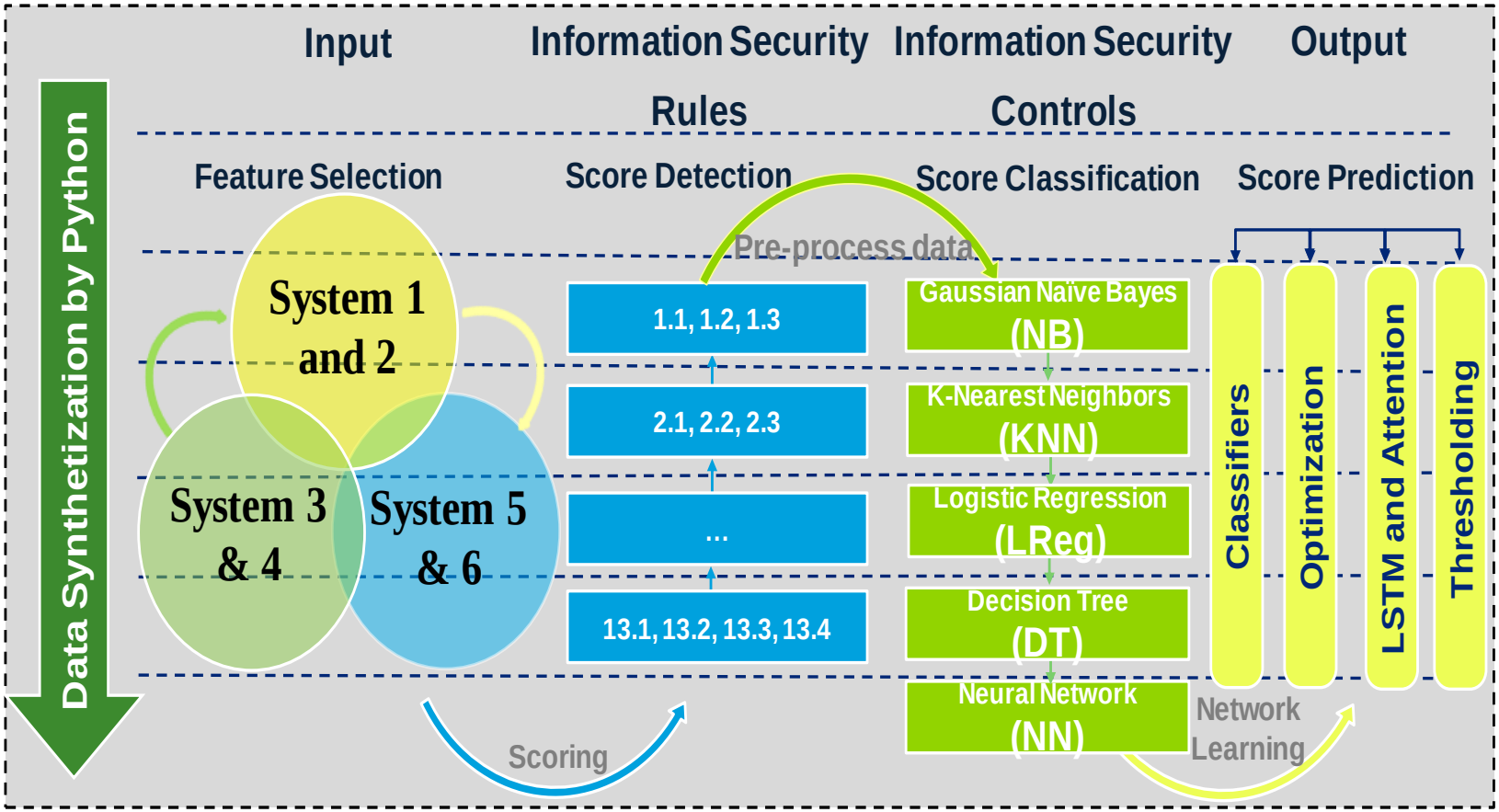
Baseline Information Security Rules (ISRs 2/3)

	ISR_{16} to ISR_{18} Under Process 6 – Network Design
6.1	Are data encrypted to prevent hacker sniffing e-banking networks? (Zachary et al. 2012)
6.2	Is penetration testing used for identifying network vulnerabilities? (Chen and Corriveau 2009)
6.3	Are networks configured to guard against physical attack and unauthorized network intrusion? (Jassal and Sehgal 2013)
	ISR_{19} to ISR_{21} Under Process 7 – Vulnerability Management Controls
7.1	Are malware protection technologies deployed to protect systems (e.g. encryption of code, polymorphism or obfuscation)? (Gharibil and Mirz 2011)
7.2	Is an intrusion prevention system used to analyze traffic control? (Balusamy et al. 2017)
7.3	Are web applications scanned to identify vulnerable instances? (Kaur et al. 2017)
	ISR_{22} to ISR_{24} Under Process 8 – Patch Management Controls
8.1	Are security patches updated regularly for online banking? (Jassal and Sehgal 2013)
8.2	Does patch management include the collection of the latest patches and the management of post-patch conflicts? (Jassal and Sehgal 2013a)
8.3	Are event logs reviewed to confirm the latest patches applied? (Lawati and Ali 2015)
	ISR_{25} to ISR_{27} Under Process 9 – Service Level Management (SLA)
9.1	Are SLAs with each infrastructure provider customized? (Rai and Bunkar 2014)
9.2	Can vendors quickly reallocate computing resources without any downtime based on the SLA? (Popovic and Hocenski 2010)
9.3	Are metrics used to measure the service level of vendors? (Kamongi et al. 2013)
	ISR_{28} to ISR_{30} Under Process 10 – Monitoring Controls
10.1	Are transactions in online banking systems monitored to detect fraud patterns with artificial intelligence or transaction history analysis? (Peotta et al. 2011)
10.2	Are network traffic for mobile banking apps monitored to inspect deep packets and their flow for vulnerability assessment? (Joshi et al. 2018)
10.3	Are removable device and email system vulnerability detected by security monitoring systems? (Kim and Kim 2018)

Baseline Information Security Rules (ISRs 3/3)

	ISR_{31} to ISR_{33} Under Process 11 – Response Controls
11.1	Are cyber security incidents detected, prevented and responded by a computer emergency response team? (Aloul 2012)
11.2	Are cyber-attacks well communicated and documented? (Miller and Rowe 2012)
11.3	Are cyber security incidents reported and forecasted? (Liu et al. 2015)
	ISR_{34} to ISR_{36} Under Process 12 – Capacity and Performance Management Controls
12.1	Is system service capacity optimized? (Xiao and Zhang 2010)
12.2	Is network and web performance tracked to manage system events? (Sun and Chen 2010)
12.3	Is controller performance on server clusters analyzed? (Kusic et al. 2009)
	ISR_{37} to ISR_{40} Under Process 13 – Service Provider Management Controls
13.1	Are outsourced systems managed with a degree of control? (Baldwin et al. 2001)
13.2	Are vendor services designed based on the size of organization? (Tallon 2010)
13.3	Are vendor service failures and service recoveries analyzed? (Michel 2001)
13.4	Are vendor service level guarantees specified in SLA? (Bhoj et al. 2001)

Baseline Learning Model



Baseline
Data Inputs

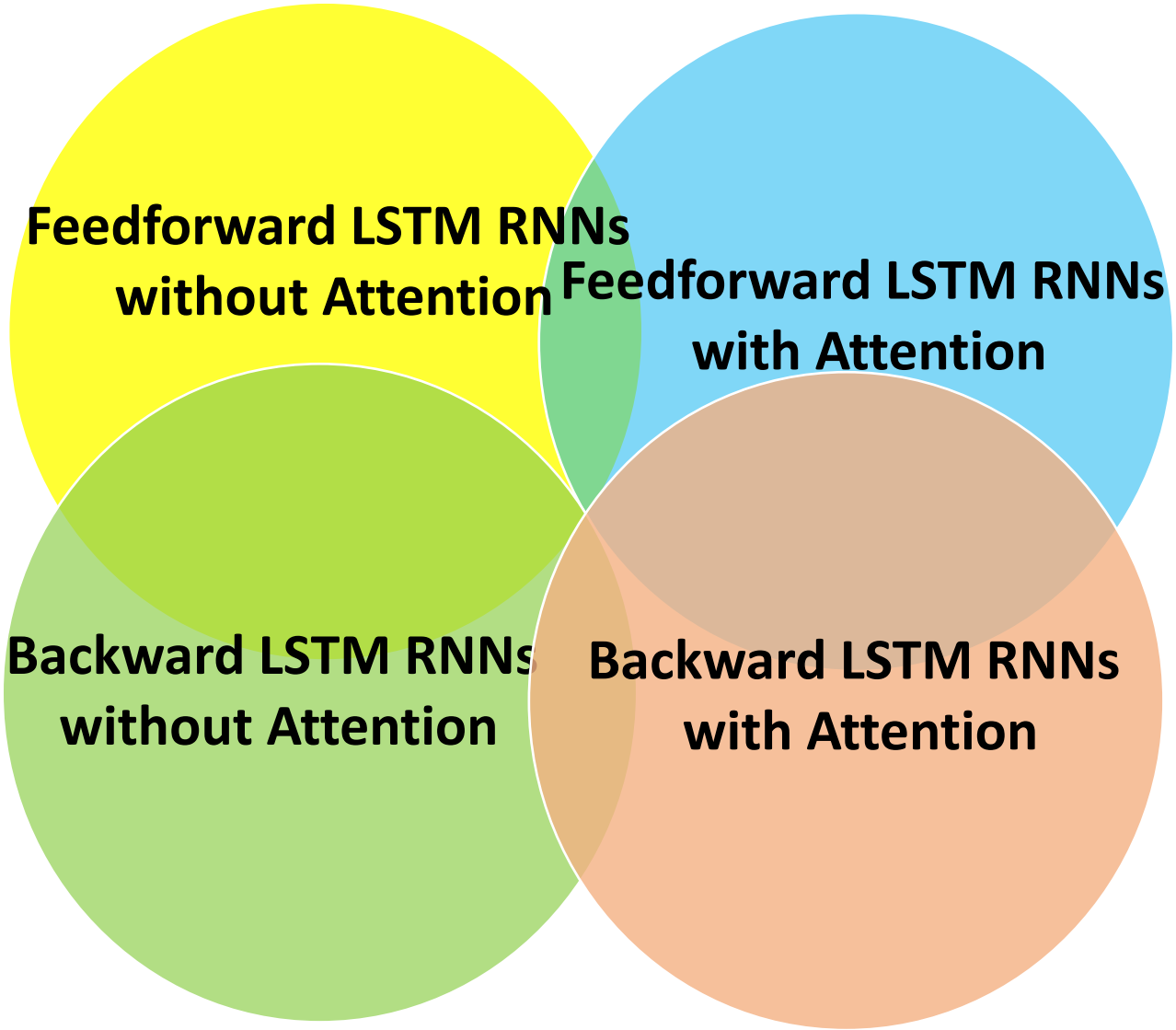
The compliance checklist covering all ISRs is compiled into a program by Python. The program is mapped to scores synthesized in the dataset.

1. Inputs are in a structure of [samples, time steps, features].
2. Samples are compliance levels while the number of time steps is forty (due to 40 ISRs).
3. The number of features is six (assumed Systems).

Baseline

Experiment – Network Methodologies

Input & output shape is 3D – (sample, timesteps & feature)



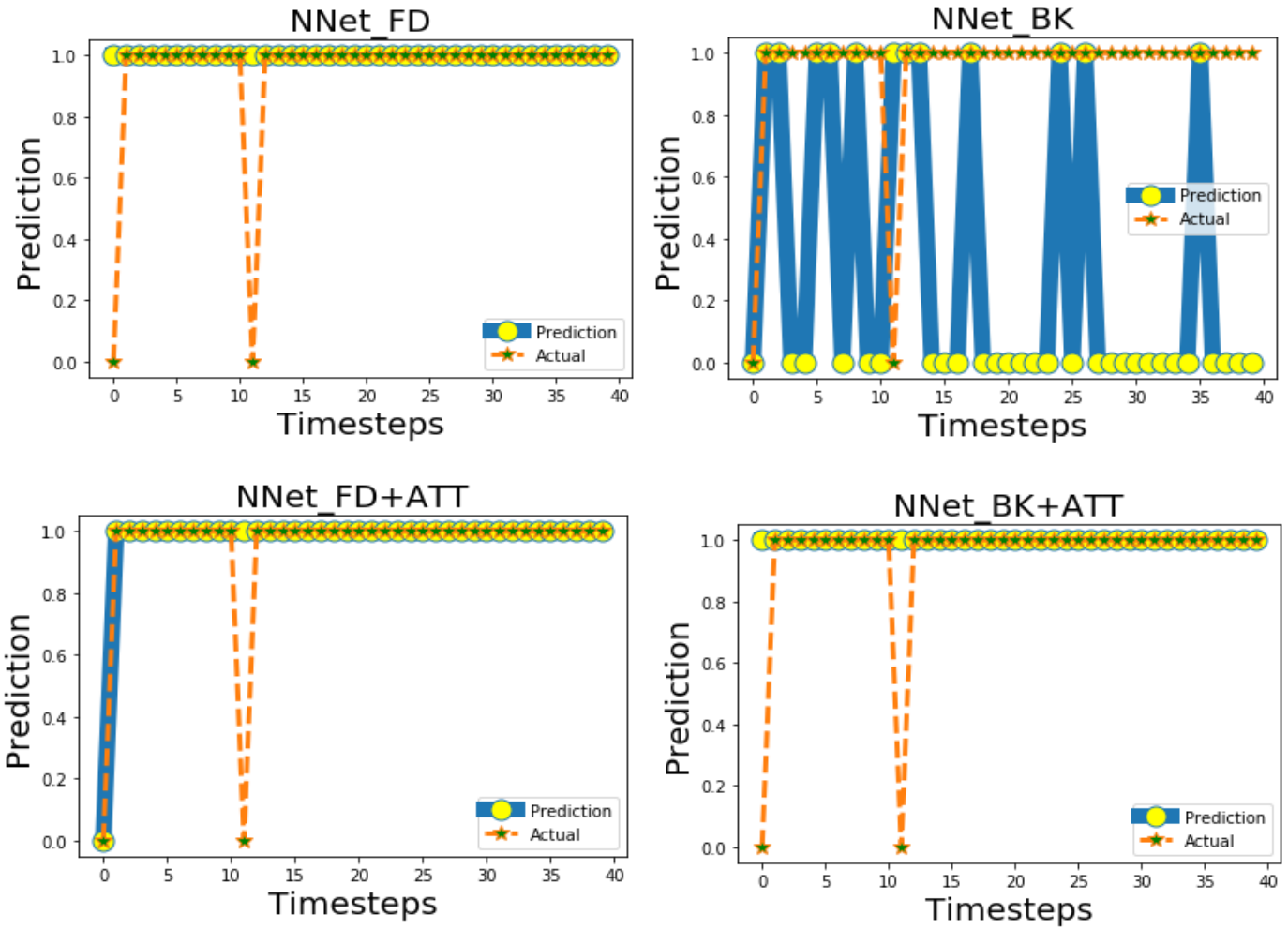
Baseline
Experiment – Result

Comparison Results of Neural Networks

NN	Accuracy	Loss	Precision	Recall	F1-Support
NNet_FD	0.9500	0.4120	0.9025	0.9500	0.9256
NNet_BK	0.3000	0.7167	0.8726	0.3000	0.4213
NNet_FD+ATT	0.9500	0.3603	0.9025	0.9500	0.9256
NNet_BK+ATT	0.9500	0.4543	0.9025	0.9500	0.9256

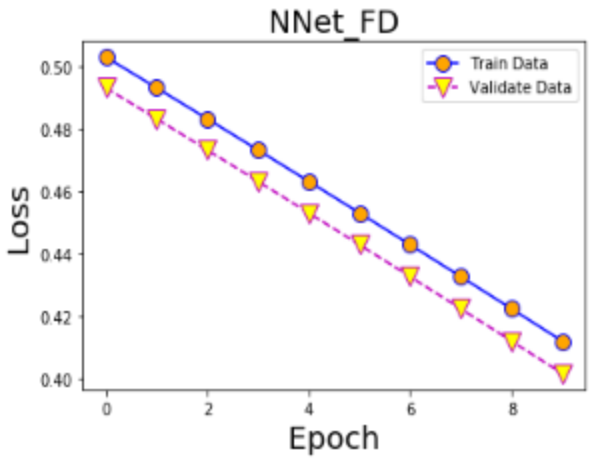
Baseline
Experiment – Analytics 1

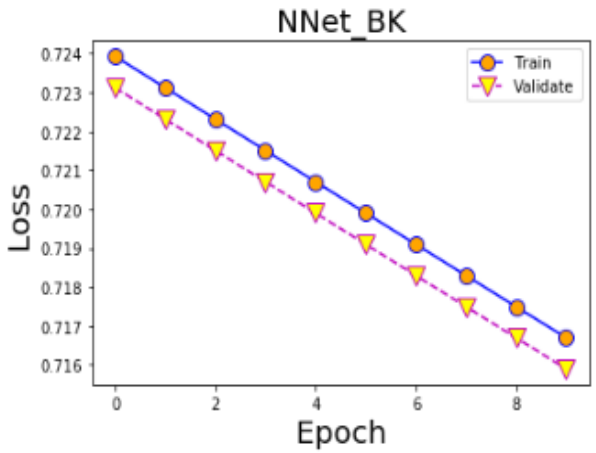
Prediction of Compliance Levels for Neural Networks

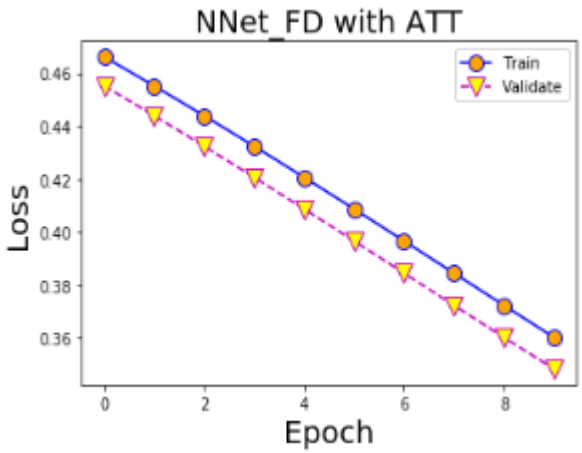


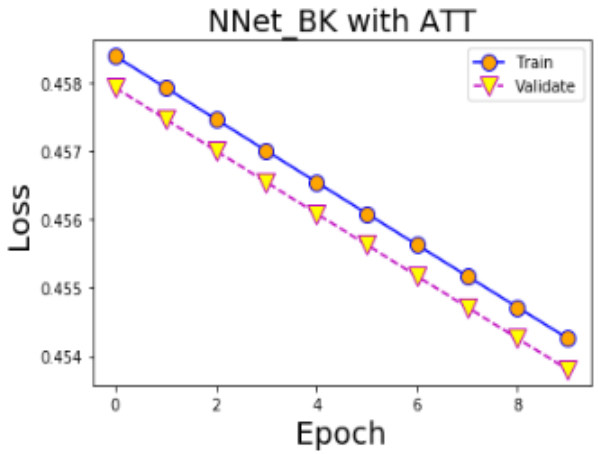
Baseline
Experiment – Analytics 2

Train Loss and Validated Loss of Neural Networks



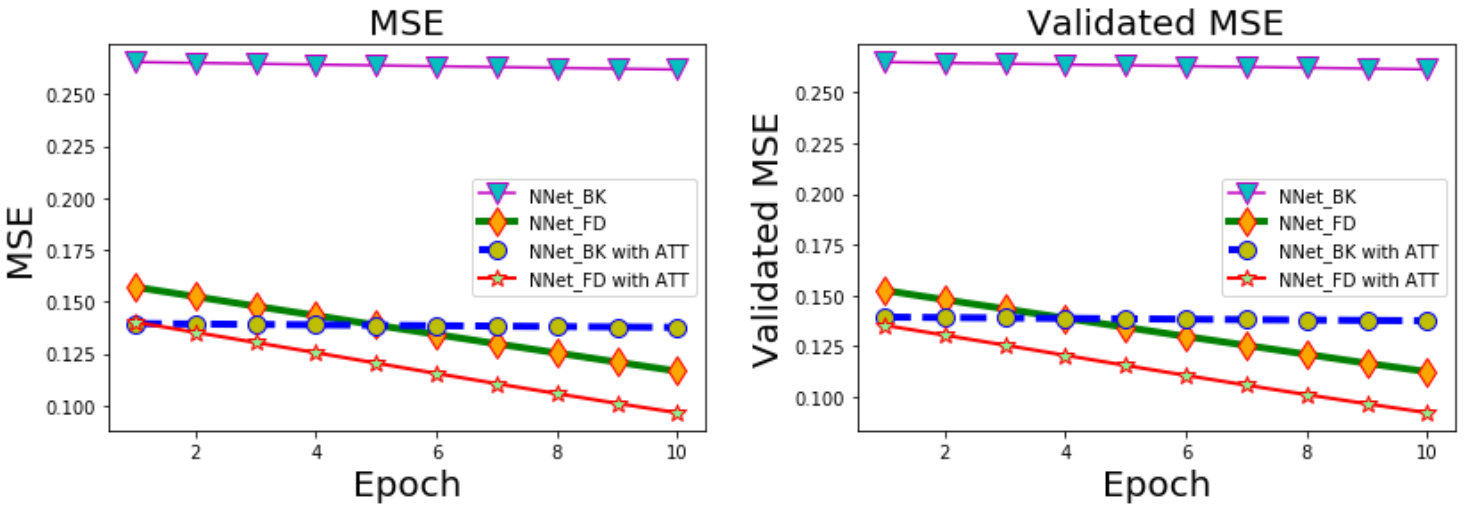






Baseline
Experiment – Analytics 3

MSE and Validated MSE of Neural Networks over Epochs



Baseline
Experiment – Comparative Analysis

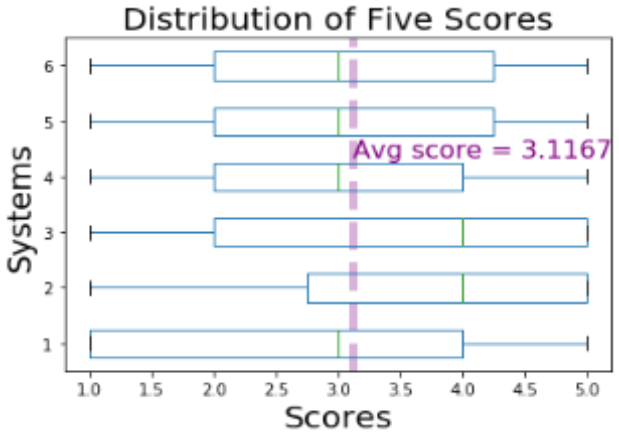
Comparison of Different Networks

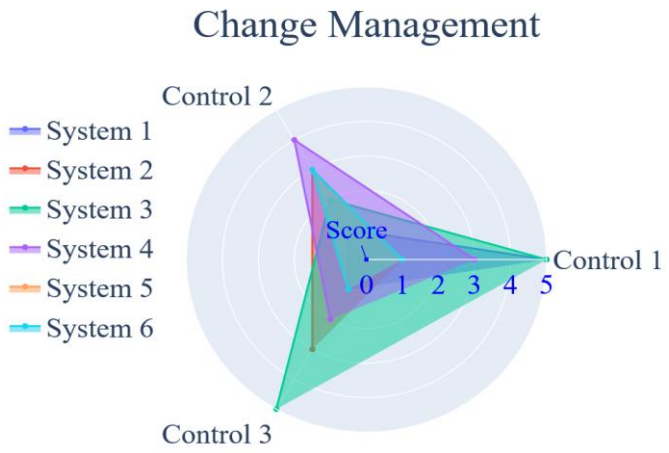
Network Type	Accuracy	Loss	Precision	Recall	F1-Support
NB	0.4286	1.5109	0.3389	0.3333	0.2976
KNN	0.5357	20.6675	0.0208	0.0833	0.0333
LReg	0.5357	2.2851	0.5139	0.4167	0.3556
DT	0.9286	23.1414	0.3833	0.2500	0.2847
NNet_FD+ATT	0.9500	0.3603	0.9025	0.9500	0.9256

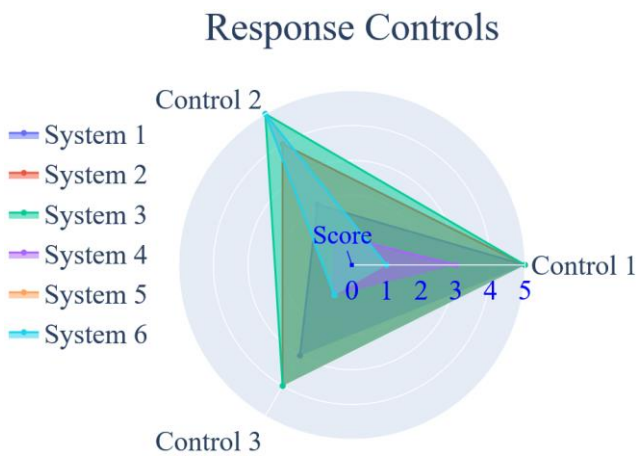
NB – Gaussian Naïve Bayes
KNN – K-Nearest Neighbors
LReg – Logistic Regression
DT – Decision Tree

Baseline
Experiment – Compliance Reports

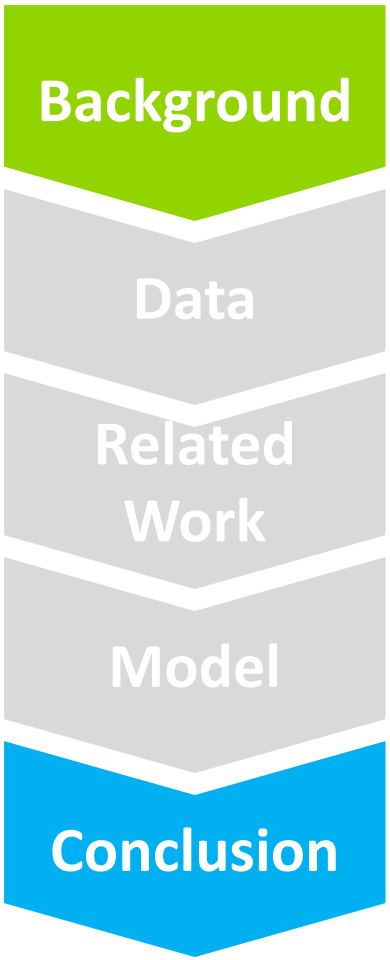
Samples





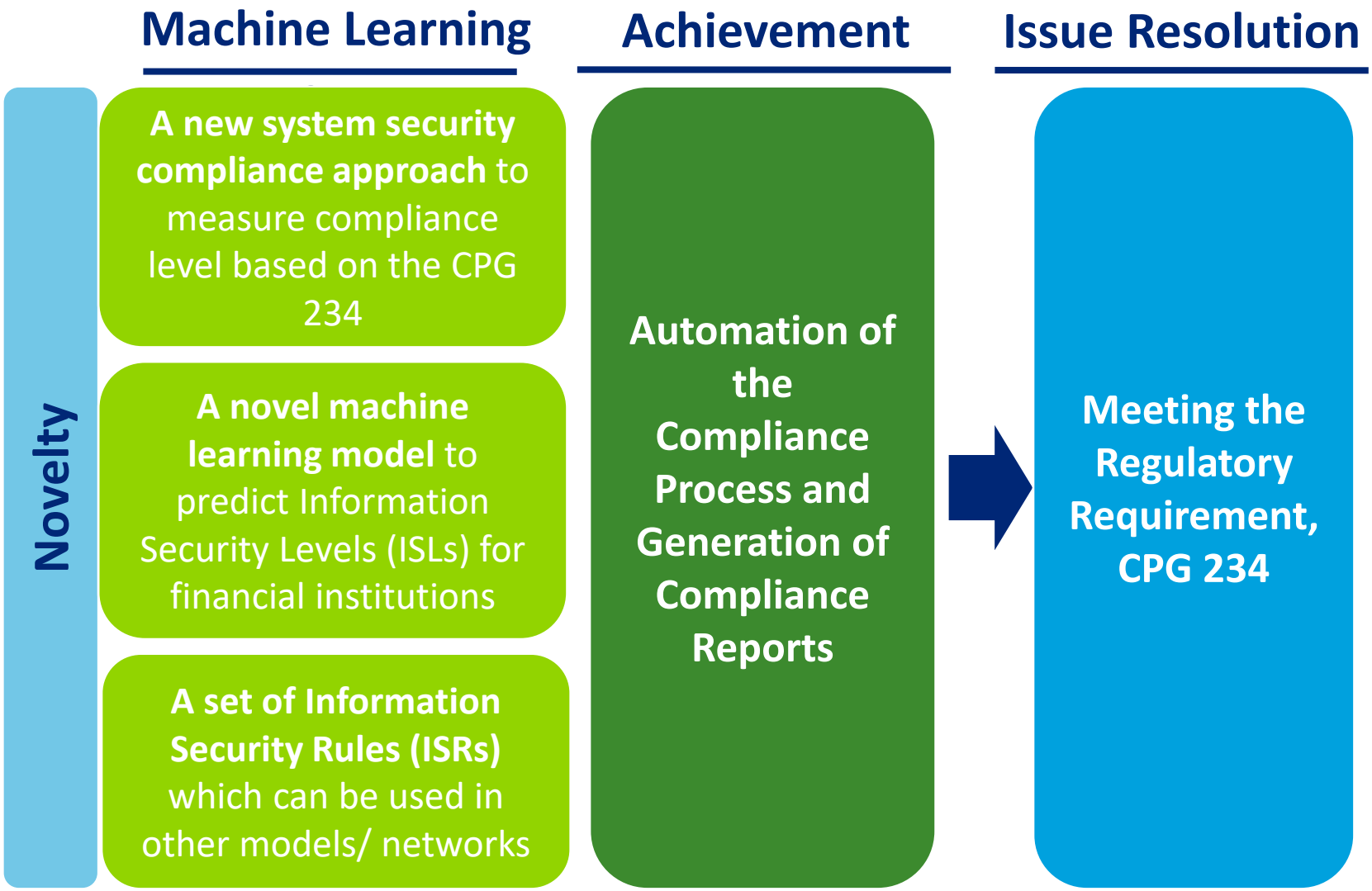


SECTIONS



- Summary of the contribution.

Conclusion
Contributions



End of the Presentation

Thank You