INPRESEC

intelligent predictive security

Dragan Pleskonjic

Predict – Prepare – Prevent – Detect

PARADIGM SHIFT IN INFORMATION SECURITY AND PRIVACY WITH ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Dragan Pleskonjic

INPRESEC Initiator / Founder

- Rich experience in creating and managing start-ups, new businesses development
- Leading management positions in international corporations
- Expertise in information security, computer software and networks industry
- Prolific academic career: Adjunct Professorship, authorship of books, scientific papers and journals' articles
- Scientific and security leader, researcher, advisor, architect
- Inventor with a set of U.S. patents granted and several patent applications pending (USPTO, CIPO, EPO, WIPO)
- Entrepreneur, leader, motivator, visionary

Information Security

- ***Information security is complicated, and hard to get right. I'm an expert in the field, and it's hard for me. It's hard for the director of the CIA. And it's hard for you.***

  — Bruce Schneier, called a "security guru" by The Economist

- ***Machine learning plays a part in every stage of your life.***

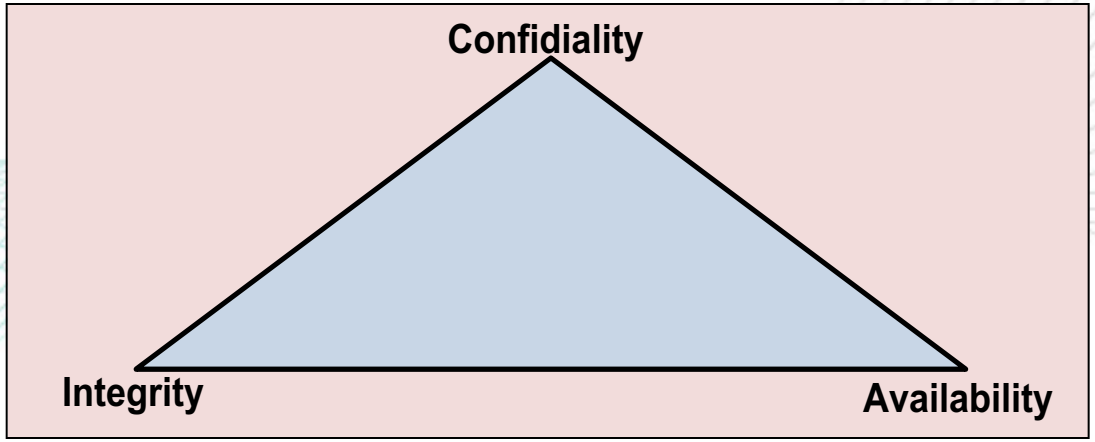
  Pedro Domingos, Professor and author of book:

  — „**The Master Algorithm**: How the Quest for the Ultimate Learning Machine Will Remake Our World“

- **Confidentiality, Integrity, Availability (CIA)**

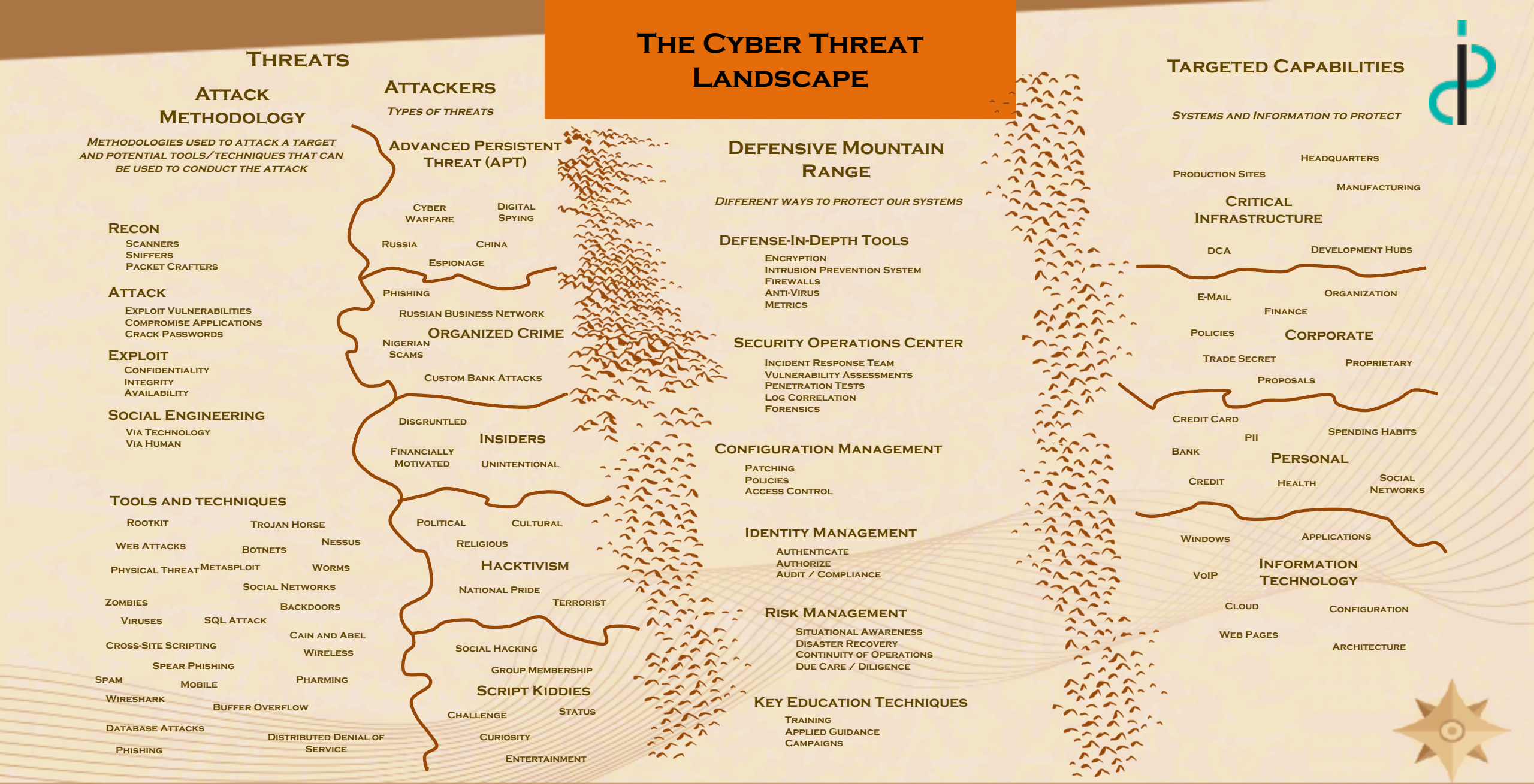
  + authenticity, accountability, non-repudiation, reliability

- **DAD** - ***D****isclosure*, ***A****lteration* ***D****estruction*



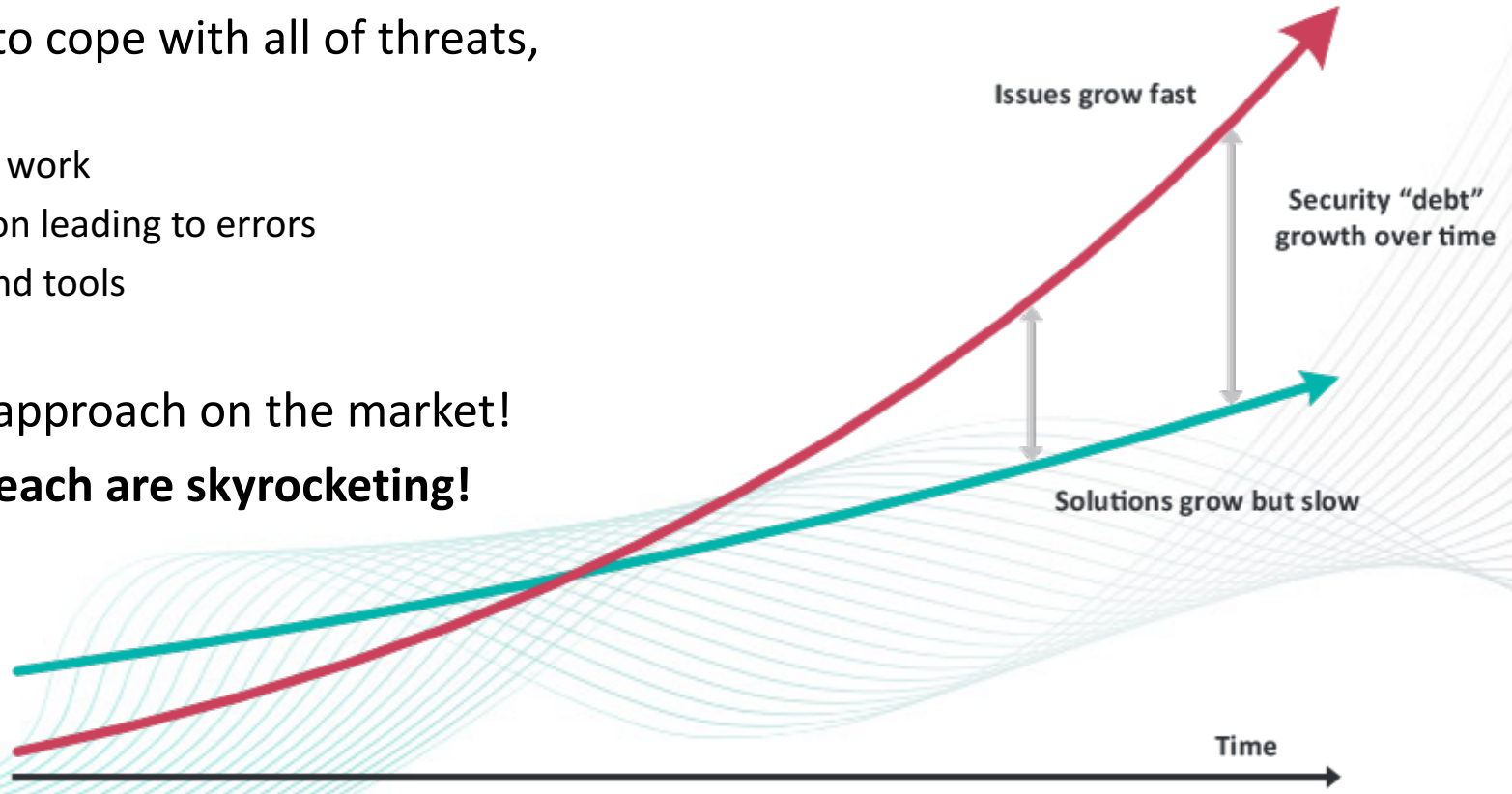
SQL Injection on car license plates

The Cyber Threat Landscape

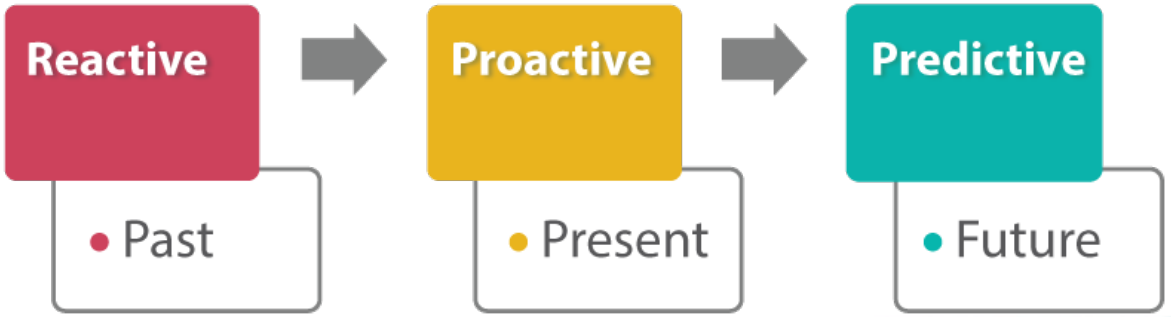


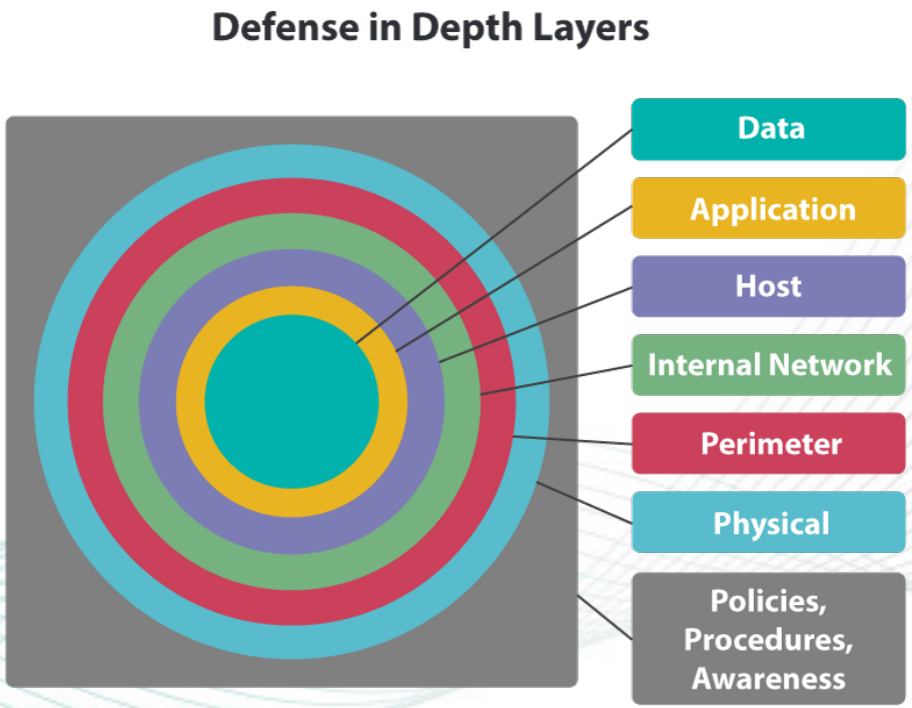
The Problem

- Number of security breaches is **rapidly increasing**
- Organizations are not able to cope with all of threats, attacks and risks any more:
  - significant amount of manual work
  - lack of focus and concentration leading to errors
  - lack of skilled professionals and tools
  - increasing cost
- There is no true predictive approach on the market!
- **Late detection - costs of breach are skyrocketing!**

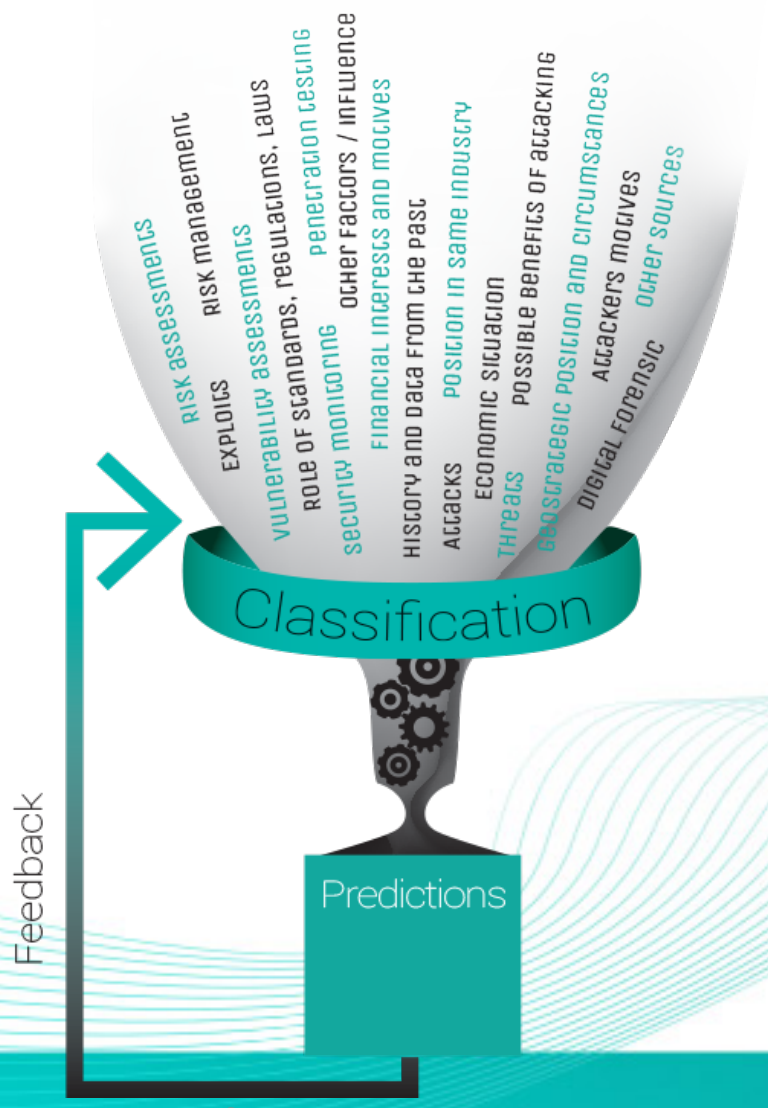


Shift the Paradigm and Defense in Depth





Our Approach



- **INPRESEC's**
  INTELLIGENT PREDICTIVE SECURITY

  Artificial Intelligence
  Machine Learning
  Predictive Analytics
  Big Data
  Σ Threat Intelligence

  BETTER INFORMATION SECURITY

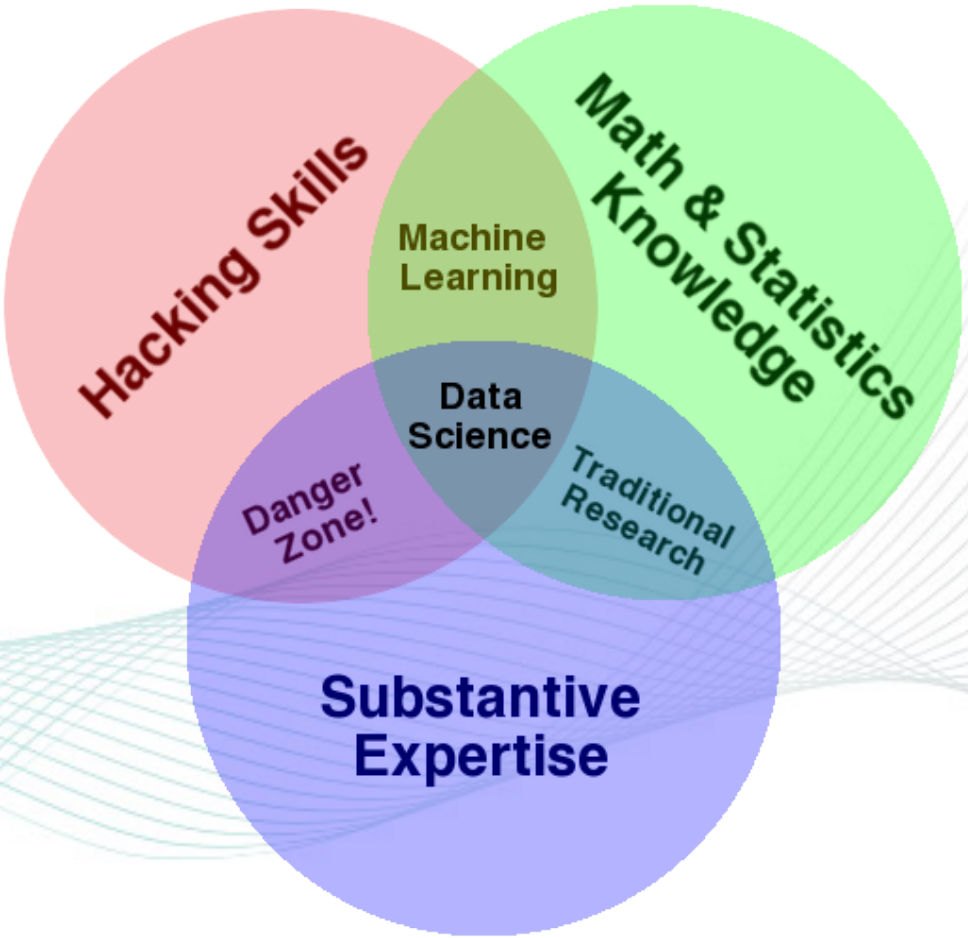
- **CHALLENGES WE ARE ADDRESSING**

  Classification
  Prediction

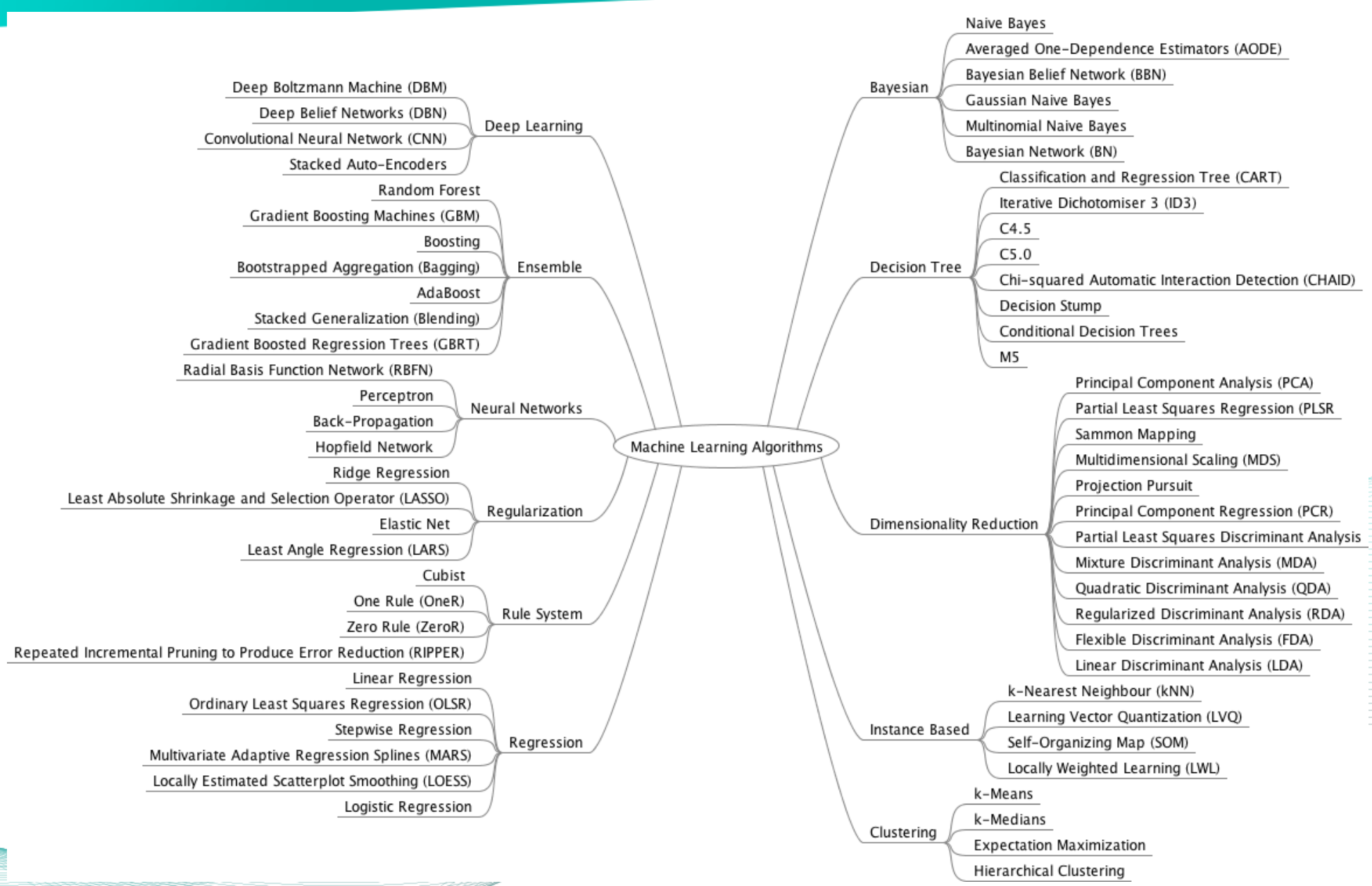
Machine Learning

- Supervised
- Unsupervised
- Reinforcement Learning

- Principal Component Analysis (PCA)



Source: http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram



Machine Learning Tools

- **Keras: The Python Deep Learning library**
  - **TensorFlow** - An open-source software library for Machine Intelligence
  - **Theano** - Python library that allows you to define, optimize, and evaluate mathematical expressions involving multi-dimensional arrays efficiently.
- **scikit-learn** - Machine Learning in Python
- **Matlab** - Statistics and Machine Learning Toolbox
- **Weka** - Collection of machine learning algorithms for data mining tasks.
- **NeuroSolutions** - Neural Network Software
- **Apache Mahout™** - Scalable machine learning and data mining
- **Appache Spark™ Machine Learning Library (MLlib)** - scalable machine learning library consisting of common learning algorithms and utilities, including classification, regression, clustering, collaborative filtering, dimensionality reduction, as well as underlying optimization primitives
- **…**

Our Solution

Integrated and automated workflow – learning in the lab and production

- **Classification of events** ALLOWED/NOT ALLOWED
  Action based on the result
- Common Platform & modules for:
  - Intrusion
  - Data Leak
  - Fraud
  - Malware
  - Malfunction
  - …
- **Prediction**

- **Solution components:**
  SENSOR, AGENT,
  SERVER, ADMIN,
  TRAINER, PREDICTION MODULE
- **Deployment:**
  - Service model: Security as an INPRESEC hosted and managed service
  - Product model: hosted by client, serviced by us

Key INPRESEC Solution Elements

Patent applications in progress

INPRESEC Sensor

- **Software, can be appliance** analyses network traffic & possible security violations, classification based on Machine Learning (ML) - **network-based system**

INPRESEC Server

- **Software - integrates functions of sensors & agents**
- Collects data from Sensors & Agents, analysis, classifying, learning & correlation and actions, based on ML
- Can be linked to SOC / CERT centers or to other security elements (AV, DLP, SIEM,...)

INPRESEC Trainer

- **Software** – training system based on ML

INPRESEC Agent

- **Software** installed on a computer (server, desktop, laptop), mobile device (smart phone, tablet etc.) or network devices (routers, firewalls, etc.), classification based on ML – **host based system**

INPRESEC Admin

- Dashboard, Configuration Console, Management, Monitoring & Reporting Tools.
- Sends alerts or other info through various communication means

INPRESEC Prediction module

- **Software** – data feed with probabilities of security events in future,
- Prediction based on various data sources, Threat Intelligence (TI), predictive analytics and ML

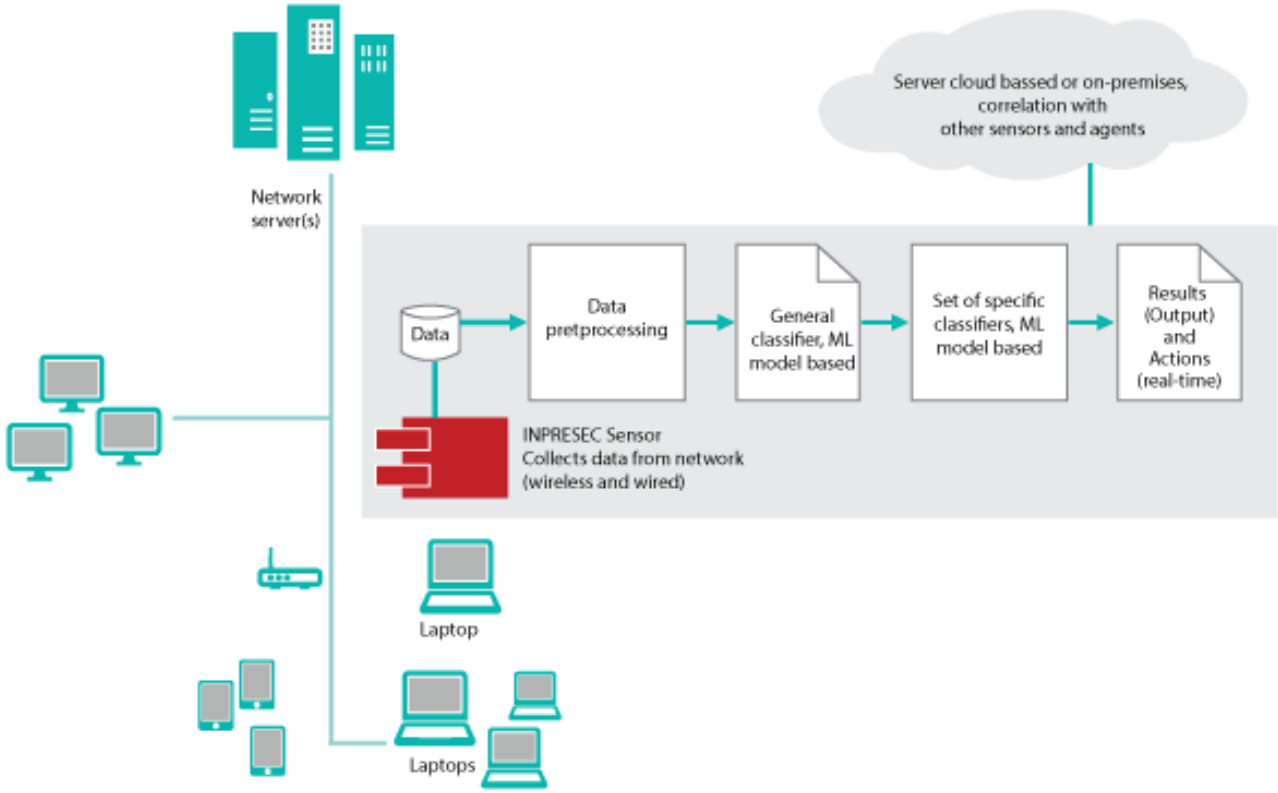
Solution Components - Sensor

SENSOR, AGENT,

SERVER, ADMIN,

TRAINER, PREDICTION MODULE

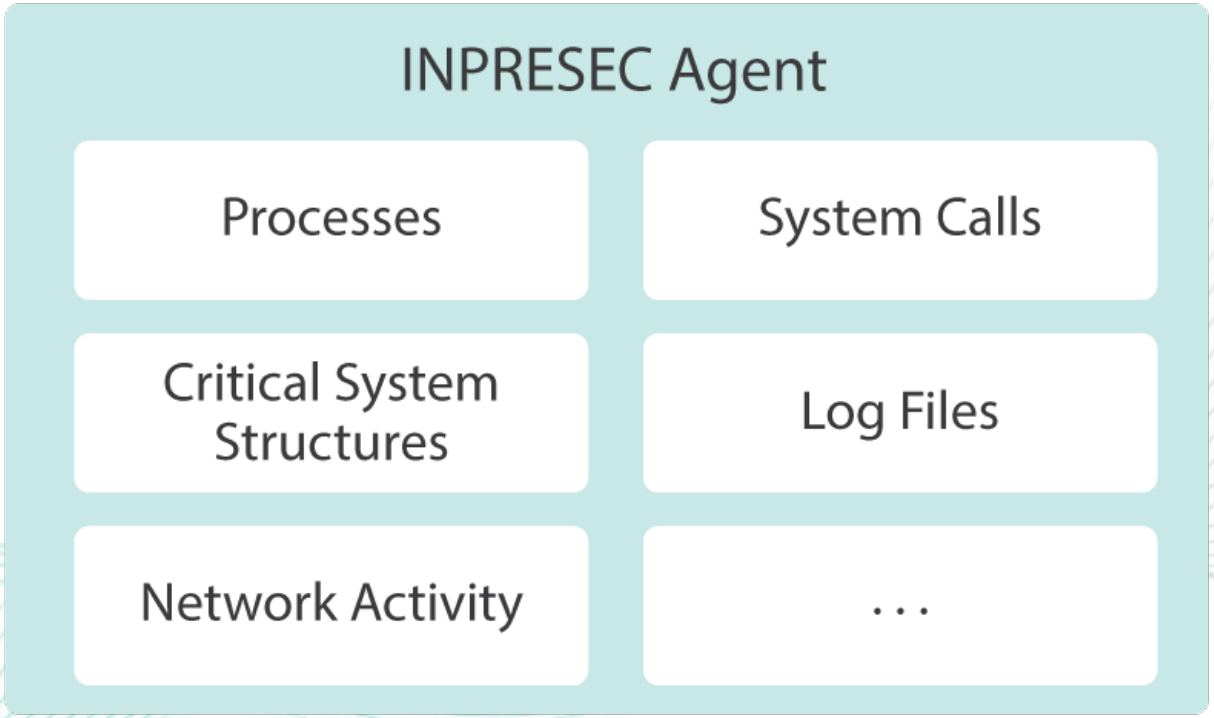
- **Software, can be appliance** analyses network traffic & possible security violations, classification based on Machine Learning (ML) - **network-based system**



Solution Components - Agent

SENSOR, **AGENT**, TRAINER
SERVER, ADMIN,
TRAINER, PREDICTION MODULE

- **Software** installed on a computer (server, desktop, laptop), mobile device (smart phone, tablet etc.) or network devices (routers, firewalls, etc.), classification based on ML – **host based system**



Solution Components – Server and Admin

SENSOR, AGENT, TRAINER

SERVER, ADMIN,

TRAINER, PREDICTION MODULE

- **Software - integrates functions of sensors & agents**
  - Collects data from Sensors & Agents, analysis, classifying, learning & correlation and actions, based on ML
  - Can be linked to SOC / CERT centers or to other security elements (AV, DLP, SIEM,...)

SENSOR, AGENT,

SERVER, ADMIN,

TRAINER, PREDICTION MODULE

- **Software -** Dashboard, Configuration Console, Management, Monitoring & Reporting Tools.
  - Sends alerts or other info through various communication means

Solution Components – Trainer and Prediction Module

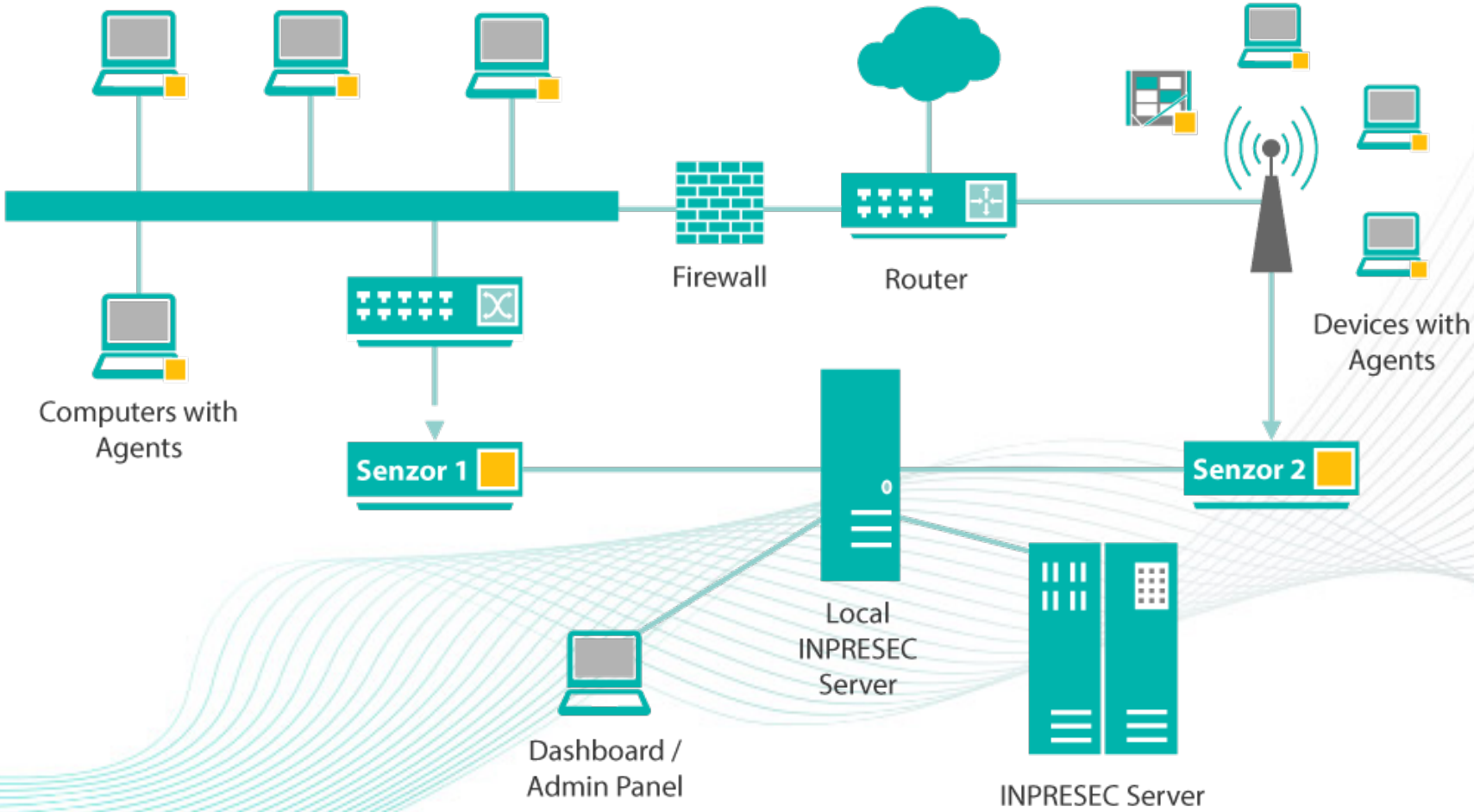
SENSOR, AGENT,
SERVER, ADMIN,
TRAINER, PREDICTION MODULE

- **Software** – training system based on ML
  - Uses "security analyst in the loop" annotations as additional input to datasets
  - Creates new models based on inputs from live system and annotated vectors
  - When new model with better accuracy is created, posts it to server for download by sensors and agents
  - By machine learning, system provides continual improvement adapting to variety of threats, attacks, as well as specific requirements that customers may have.

SENSOR, AGENT,
SERVER, ADMIN,
TRAINER, **PREDICTION MODULE**

- **Software** – data feed with probabilities of security events in future,
  - Prediction based on various data sources, Threat Intelligence (TI), predictive analytics and ML
  - Using various parameters and input data from set of internal and external sources, it analyses them and, through set of our proprietary algorithms, gives probabilities of possible threats and attacks.
  - These data will be later distributed as input to our system and help to set alert levels, thresholds, prevention measures etc.

Possible Network Layout



Machine Learning -> Increased Efficiency

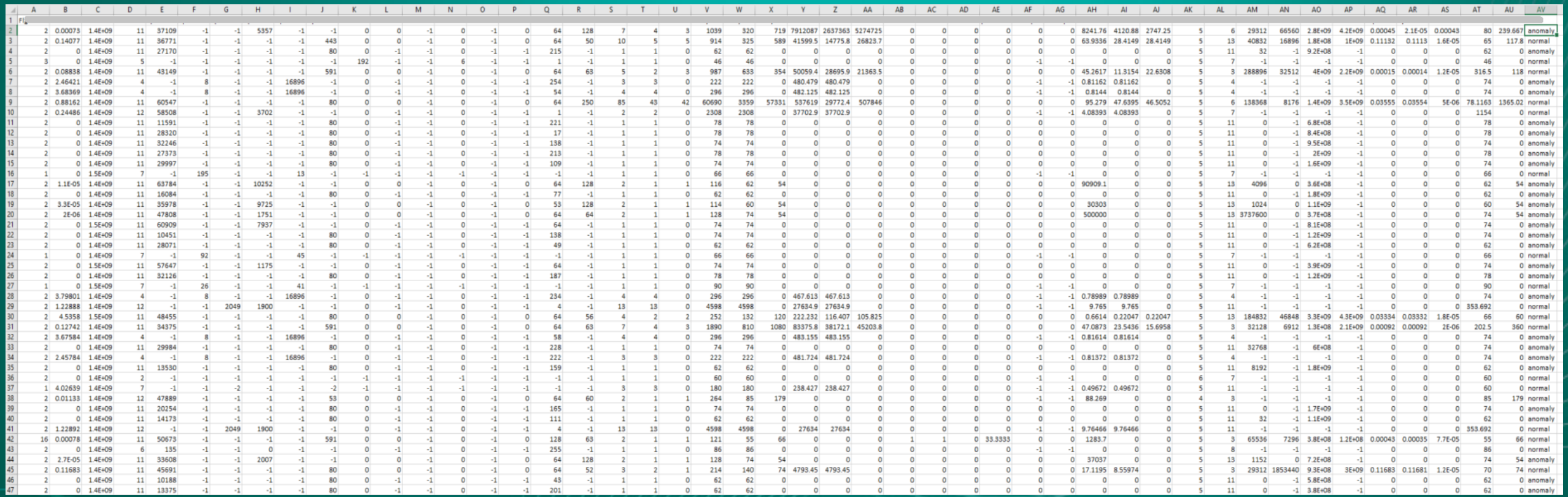
- „**Security analyst in the loop**“ concept
  - Supervised learning – solution becomes more and more clever during time and requires less human intervention
- Decrease grey area during time, eventually to reach A = B
- Team focuses and more innovative and interesting work

Example of decision scale

Datasets – how to obtain/create

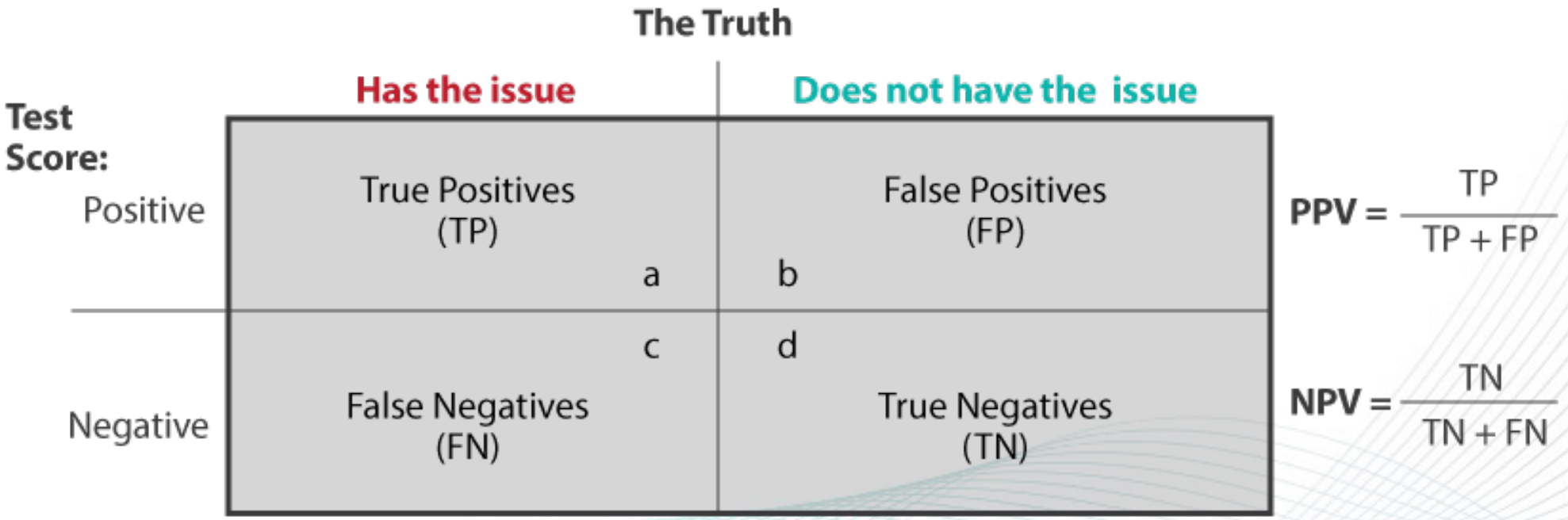
- Various IDS/IPS data sets and test vectors available on Internet
- Created by us from:
  - Testing environments
  - Real environments
- Created by us – dataset generation scripts:
  - “clean” ones i.e. regular, no intrusions or other issues
  - With anomalies, attack, intrusions, data leaks, malware and similar malfunctions

Labeled dataset example (redacted)



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV
Fl																																															
2	0.00073	1.4E+09	11	37109	-1	-1	5357	-1	-1	0	0	-1	0	-1	0	64	128	7	4	3	1039	320	719	7912087	2637363	5274725	0	0	0	0	0	0	8241.76	4120.88	2747.25	5	6	29312	66560	2.8E+09	4.2E+09	0.00045	2.1E-05	0.00043	80	239.667	anomaly
2	0.14077	1.4E+09	11	36771	-1	-1	-1	-1	443	0	0	-1	0	-1	0	64	50	10	5	5	914	325	589	41599.5	14775.8	26823.7	0	0	0	0	0	0	63.9336	28.4149	28.4149	5	13	40832	16896	1.8E+08	1E+09	0.11132	0.1113	1.6E-05	65	117.8	normal
2	0	1.4E+09	11	27170	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	215	-1	1	1	0	62	62	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	32	-1	9.2E+08	-1	0	0	0	62	0	anomaly
3	0	1.4E+09	5	-1	-1	-1	-1	-1	-1	192	-1	-1	6	-1	-1	1	-1	1	1	0	46	46	0	0	0	0	0	0	0	0	-1	-1	0	0	0	5	7	-1	-1	-1	-1	0	0	0	46	0	normal
2	0.08838	1.4E+09	11	43149	-1	-1	-1	-1	591	0	0	-1	0	-1	0	64	63	5	2	3	987	633	354	50059.4	28695.9	21363.5	0	0	0	0	0	0	45.2617	11.3154	22.6308	5	3	288896	32512	4E+09	2.2E+09	0.00015	0.00014	1.2E-05	316.5	118	normal
2	2.46421	1.4E+09	4	-1	8	-1	-1	16896	-1	0	-1	-1	0	-1	-1	254	-1	3	3	0	222	222	0	480.479	480.479	0	0	0	0	0	-1	-1	0.81162	0.81162	0	5	4	-1	-1	-1	-1	0	0	0	74	0	anomaly
2	3.68369	1.4E+09	4	-1	8	-1	-1	16896	-1	0	-1	-1	0	-1	-1	54	-1	4	4	0	296	296	0	482.125	482.125	0	0	0	0	0	-1	-1	0.8144	0.8144	0	5	4	-1	-1	-1	-1	0	0	0	74	0	anomaly
2	0.88162	1.4E+09	11	60547	-1	-1	-1	-1	80	0	0	-1	0	-1	0	64	250	85	43	42	60690	3359	57331	537619	29772.4	507846	0	0	0	0	0	0	95.279	47.6395	46.5052	5	6	138368	8176	1.4E+09	3.5E+09	0.03555	0.03554	5E-06	78.1163	1365.02	normal
2	0.24486	1.4E+09	12	58508	-1	-1	3702	-1	-1	0	-1	-1	0	-1	-1	1	-1	2	2	0	2308	2308	0	37702.9	37702.9	0	0	0	0	0	-1	-1	4.08393	4.08393	0	5	7	-1	-1	-1	-1	0	0	0	1154	0	normal
2	0	1.4E+09	11	11591	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	221	-1	1	1	0	78	78	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	0	-1	6.8E+08	-1	0	0	0	78	0	anomaly
2	0	1.4E+09	11	28320	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	17	-1	1	1	0	78	78	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	0	-1	8.4E+08	-1	0	0	0	78	0	anomaly
2	0	1.4E+09	11	32246	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	138	-1	1	1	0	74	74	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	0	-1	9.5E+08	-1	0	0	0	74	0	anomaly
2	0	1.4E+09	11	27373	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	213	-1	1	1	0	78	78	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	0	-1	2E+09	-1	0	0	0	78	0	anomaly
2	0	1.4E+09	11	29997	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	109	-1	1	1	0	74	74	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	0	-1	1.6E+09	-1	0	0	0	74	0	anomaly
1	0	1.5E+09	7	-1	195	-1	-1	13	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	0	66	66	0	0	0	0	0	0	0	0	-1	-1	0	0	0	5	7	-1	-1	-1	-1	0	0	0	66	0	normal
2	1.1E-05	1.4E+09	11	63784	-1	-1	10252	-1	-1	0	0	-1	0	-1	0	64	128	2	1	1	116	62	54	0	0	0	0	0	0	0	0	0	90909.1	0	0	5	13	4096	0	3.6E+08	-1	0	0	0	62	54	anomaly
2	0	1.4E+09	11	16084	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	77	-1	1	1	0	62	62	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	0	-1	1.8E+09	-1	0	0	0	62	0	anomaly
2	3.3E-05	1.4E+09	11	35978	-1	-1	9725	-1	-1	0	0	-1	0	-1	0	53	128	2	1	1	114	60	54	0	0	0	0	0	0	0	0	0	30303	0	0	5	13	1024	0	1.1E+09	-1	0	0	0	60	54	anomaly
2	2E-06	1.4E+09	11	47808	-1	-1	1751	-1	-1	0	0	-1	0	-1	0	64	64	2	1	1	128	74	54	0	0	0	0	0	0	0	0	0	500000	0	0	5	13	3737600	0	3.7E+08	-1	0	0	0	74	54	anomaly
2	0	1.5E+09	11	60909	-1	-1	7937	-1	-1	0	0	-1	0	-1	-1	64	-1	1	1	0	74	74	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	0	-1	8.1E+08	-1	0	0	0	74	0	anomaly
2	0	1.4E+09	11	10451	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	138	-1	1	1	0	74	74	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	0	-1	1.2E+09	-1	0	0	0	74	0	anomaly
2	0	1.4E+09	11	28071	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	49	-1	1	1	0	62	62	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	0	-1	6.2E+08	-1	0	0	0	62	0	anomaly
1	0	1.4E+09	7	-1	92	-1	-1	45	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	0	66	66	0	0	0	0	0	0	0	0	-1	-1	0	0	0	5	7	-1	-1	-1	-1	0	0	0	66	0	normal
2	0	1.5E+09	11	57647	-1	-1	1175	-1	-1	0	-1	-1	0	-1	-1	64	-1	1	1	0	74	74	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	0	-1	3.9E+09	-1	0	0	0	74	0	anomaly
2	0	1.4E+09	11	32126	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	187	-1	1	1	0	78	78	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	0	-1	1.2E+09	-1	0	0	0	78	0	anomaly
1	0	1.5E+09	7	-1	26	-1	-1	41	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	0	90	90	0	0	0	0	0	0	0	0	-1	-1	0	0	0	5	7	-1	-1	-1	-1	0	0	0	90	0	normal
2	3.79801	1.4E+09	4	-1	8	-1	-1	16896	-1	0	-1	-1	0	-1	-1	234	-1	4	4	0	296	296	0	467.613	467.613	0	0	0	0	0	-1	-1	0.78989	0.78989	0	5	4	-1	-1	-1	-1	0	0	0	74	0	anomaly
2	1.22888	1.4E+09	12	-1	-1	2049	1900	-1	-1	0	-1	-1	0	-1	-1	4	-1	13	13	0	4598	4598	0	27634.9	27634.9	0	0	0	0	0	-1	-1	9.765	9.765	0	5	11	-1	-1	-1	-1	0	0	0	353.692	0	normal
2	4.5358	1.5E+09	11	48455	-1	-1	-1	-1	80	0	0	-1	0	-1	0	64	56	4	2	2	252	132	120	222.232	116.407	105.825	0	0	0	0	0	0	0.6614	0.22047	0.22047	5	13	184832	46848	3.3E+09	4.3E+09	0.03334	0.03332	1.8E-05	66	60	normal
2	0.12742	1.4E+09	11	34375	-1	-1	-1	-1	591	0	0	-1	0	-1	0	64	63	7	4	3	1890	810	1080	83375.8	38172.1	45203.8	0	0	0	0	0	0	47.0873	23.5436	15.6958	5	3	32128	6912	1.3E+08	2.1E+09	0.00092	0.00092	2E-06	202.5	360	normal
2	3.67584	1.4E+09	4	-1	8	-1	-1	16896	-1	0	-1	-1	0	-1	-1	58	-1	4	4	0	296	296	0	483.155	483.155	0	0	0	0	0	-1	-1	0.81614	0.81614	0	5	4	-1	-1	-1	-1	0	0	0	74	0	anomaly
2	0	1.4E+09	11	29984	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	228	-1	1	1	0	74	74	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	32768	-1	6E+08	-1	0	0	0	74	0	anomaly
2	2.45784	1.4E+09	4	-1	8	-1	-1	16896	-1	0	-1	-1	0	-1	-1	222	-1	3	3	0	222	222	0	481.724	481.724	0	0	0	0	0	-1	-1	0.81372	0.81372	0	5	4	-1	-1	-1	-1	0	0	0	74	0	anomaly
2	0	1.4E+09	11	13530	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	159	-1	1	1	0	62	62	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	8192	-1	1.8E+09	-1	0	0	0	62	0	anomaly
2	0	1.4E+09	2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	0	60	60	0	0	0	0	0	0	0	0	-1	-1	0	0	0	6	7	-1	-1	-1	-1	0	0	0	60	0	normal
1	4.02639	1.4E+09	7	-1	-1	-2	-1	-1	-2	-1	-1	-1	-1	-1	-1	-1	-1	3	3	0	180	180	0	238.427	238.427	0	0	0	0	0	-1	-1	0.49672	0.49672	0	5	11	-1	-1	-1	-1	0	0	0	60	0	normal
2	0.01133	1.4E+09	12	47889	-1	-1	-1	-1	53	0	0	-1	0	-1	0	64	60	2	1	1	264	85	179	0	0	0	0	0	0	0	0	-1	88.269	0	0	4	3	-1	-1	-1	-1	0	0	0	85	179	normal
2	0	1.4E+09	11	20254	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	165	-1	1	1	0	74	74	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	0	-1	1.7E+09	-1	0	0	0	74	0	anomaly
2	0	1.4E+09	11	14173	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	111	-1	1	1	0	62	62	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	32	-1	1.1E+09	-1	0	0	0	62	0	anomaly
2	1.22892	1.4E+09	12	-1	-1	2049	1900	-1	-1	0	-1	-1	0	-1	-1	4	-1	13	13	0	4598	4598	0	27634	27634	0	0	0	0	0	-1	-1	9.76466	9.76466	0	5	11	-1	-1	-1	-1	0	0	0	353.692	0	normal
16	0.00078	1.4E+09	11	50673	-1	-1	-1	-1	591	0	0	-1	0	-1	0	128	63	2	1	1	121	55	66	0	0	0	1	1	0	33.3333	0	0	1283.7	0	0	5	3	65536	7296	3.8E+08	1.2E+08	0.00043	0.00035	7.7E-05	55	66	normal
2	0	1.4E+09	6	135	-1	-1	0	-1	-1	0	0	-1	0	-1	-1	255	-1	1	1	0	86	86	0	0	0	0	0	0	0	0	0	-1	0	0	0	5	8	-1	-1	-1	-1	0	0	0	86	0	normal
2	2.7E-05	1.4E+09	11	33608	-1	-1	2007	-1	-1	0	0	-1	0	-1	0	64	128	2	1	1	128	74	54	0	0	0	0	0	0	0	0	0	37037	0	0	5	13	1152	0	7.2E+08	-1	0	0	0	74	54	anomaly
2	0.11683	1.4E+09	11	45691	-1	-1	-1	-1	80	0	0	-1	0	-1	0	64	52	3	2	1	214	140	74	4793.45	4793.45	0	0	0	0	0	0	0	17.1195	8.55974	0	5	3	29312	1853440	9.3E+08	3E+09	0.11683	0.11681	1.2E-05	70	74	normal
2	0	1.4E+09	11	10188	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	43	-1	1	1	0	62	62	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	0	-1	5.8E+08	-1	0	0	0	62	0	anomaly
2	0	1.4E+09	11	13375	-1	-1	-1	-1	80	0	-1	-1	0	-1	-1	201	-1	1	1	0	62	62	0	0	0	0	0	0	0	0	0	0	0	0	0	5	11	0	-1	3.8E+08	-1	0	0	0	62	0	anomaly

Sensitivity and Specificity



Sensitivity

$$\frac{TP}{TP + FN}$$

Or,

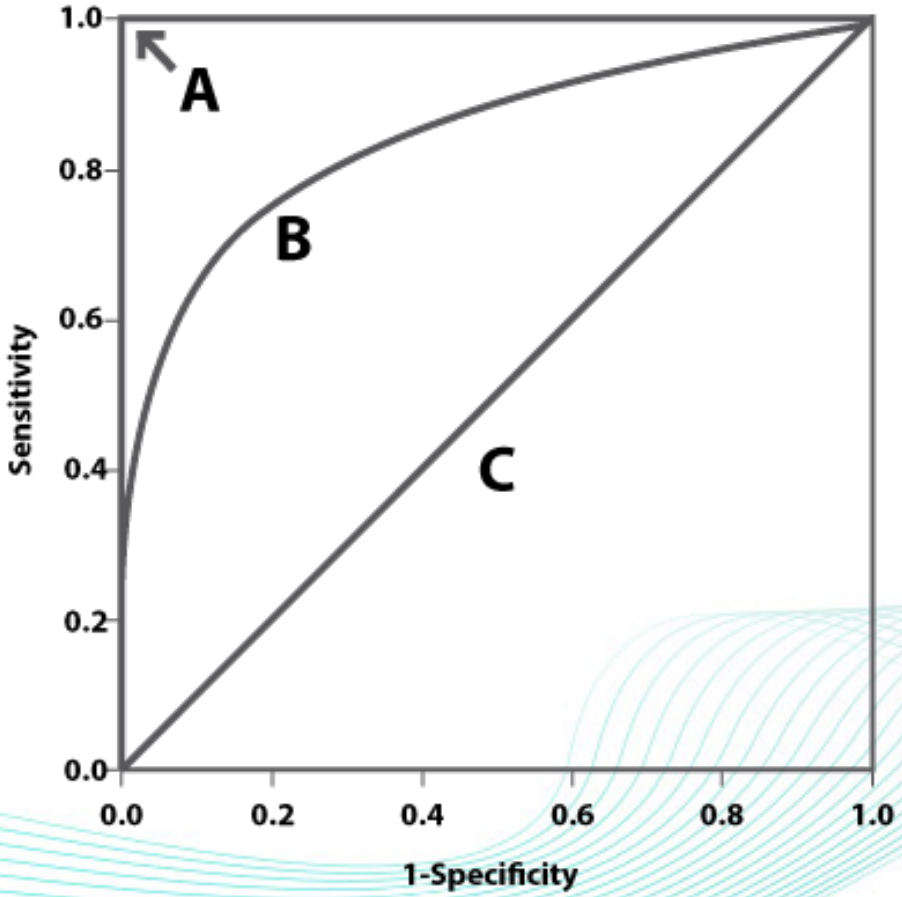
$$\frac{a}{a + c}$$

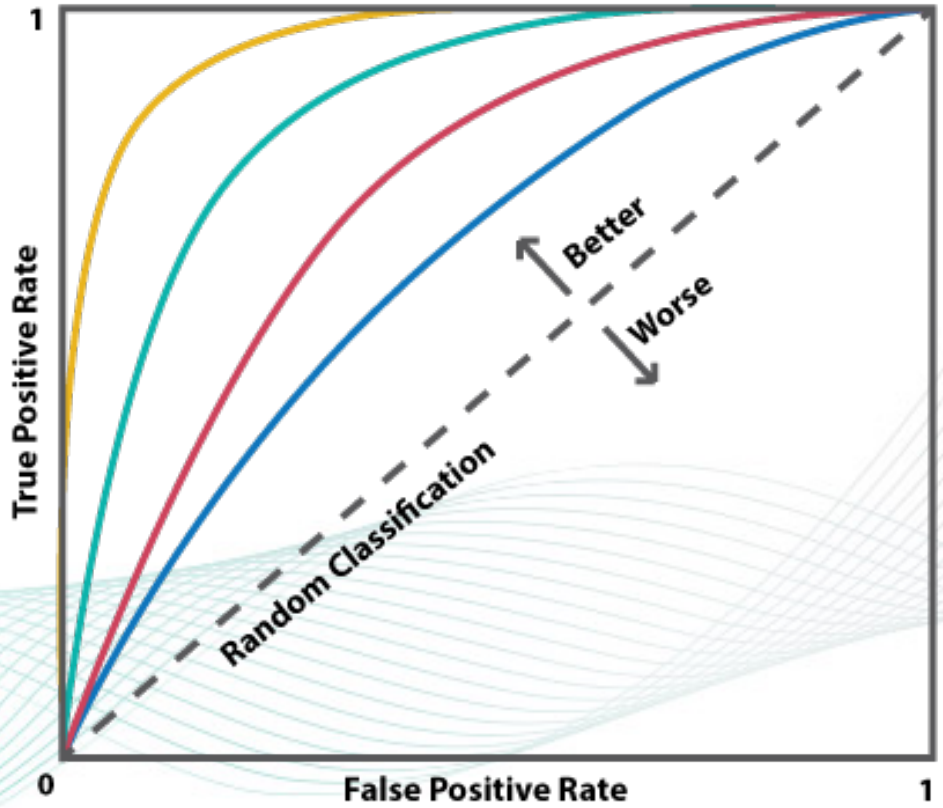
Specificity

$$\frac{TN}{TN + FP}$$

$$\frac{d}{d + b}$$

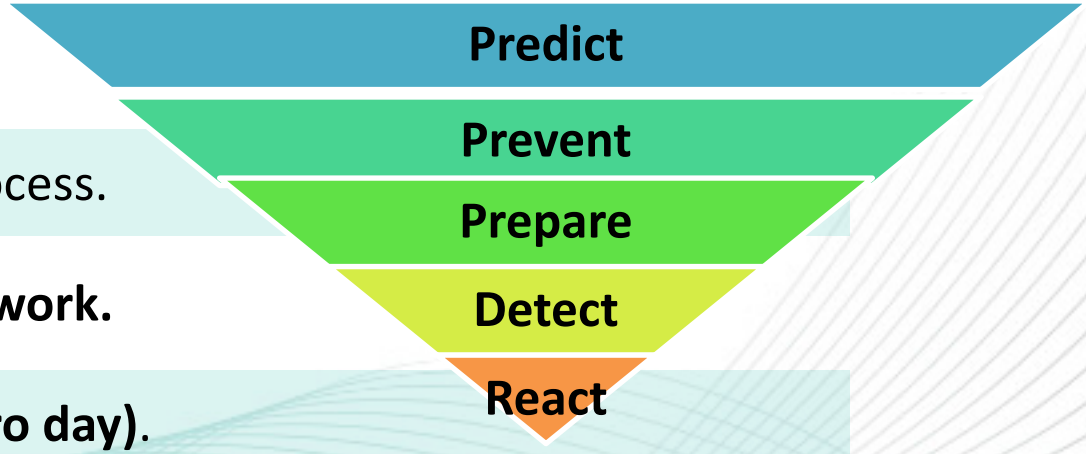
Receiver Operating Characteristic (ROC) curve





Comparison with traditional systems

- **Paradigm shift**: predicts, prevent, prepare - goes ahead of hackers.
- **Better accuracy**, better performances.
- **Automatic learning**, continual improvement process.
- **Lightweight maintenance**. **Removes repetitive work.**
- **Significantly better** in handling **new threats (zero day)**.
- **Multilayer / multilevel**, assures holistic approach.
  - Detects: wide spectra of threats and attacks intrusions, data leak, malware, fraud and other malfunctions.

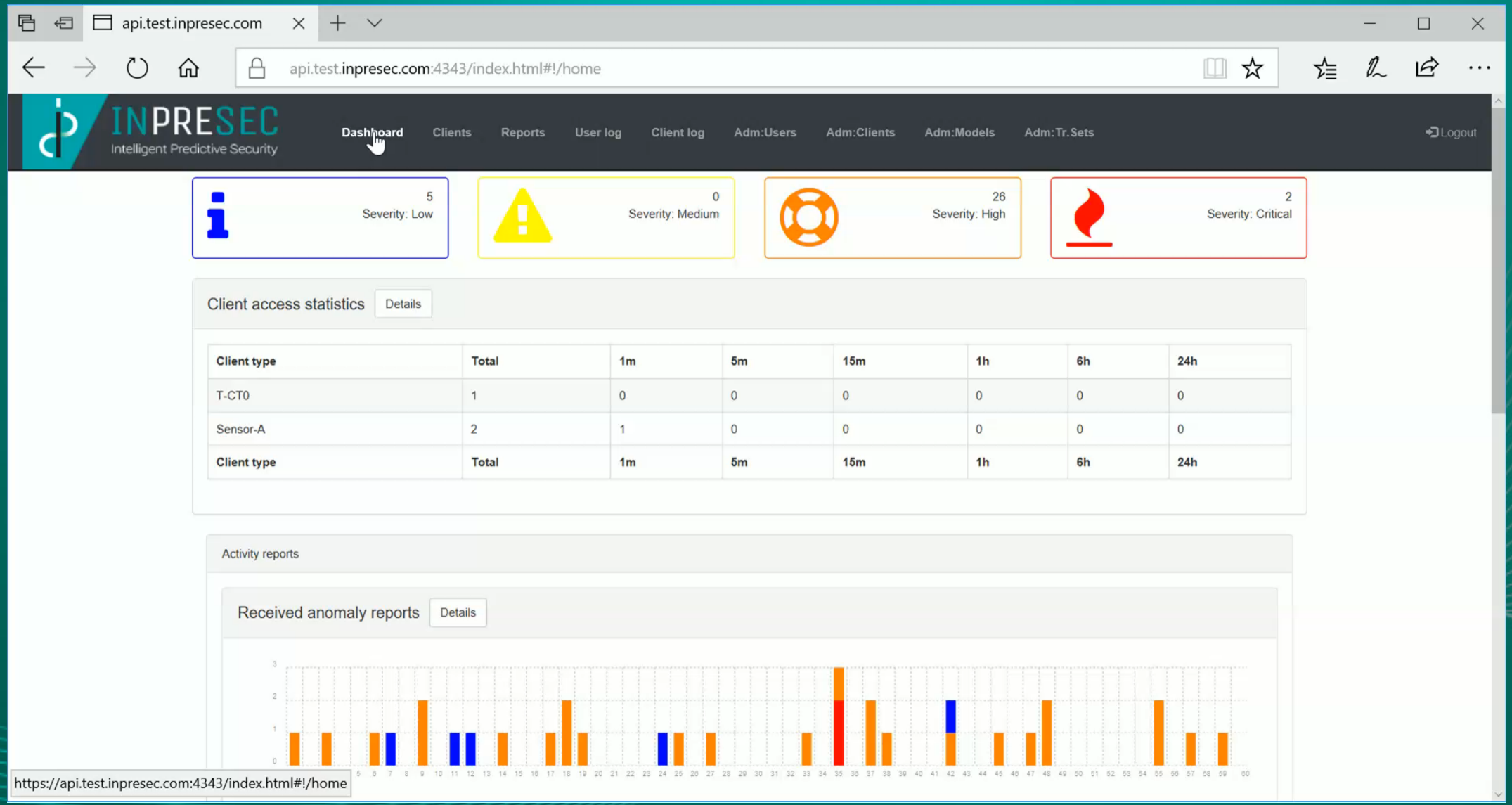


One more thing:

- While the role of ML and AI in cybersecurity is certainly in the early stages and still needs to evolve, hackers will quickly learn to turn machine learning into a distinct advantage

=> **AI & ML** can be misused as **new threat attack vector**

DEMO



THANK YOU, QUESTIONS

Contact:

Dragan Pleskonjic, INPRESEC Initiator and Founder

- E-mail: dragan@conwex.org or dragan@inpresec.com
- Twitter: @DPleskonjic
- Personal web site: www.dragan-pleskonjic.com
- LinkedIn profile: https://www.linkedin.com/in/draganpleskonjic/