CONFERENCIA DE POSGRADO

Clasificación de ciberataques mediante un modelo de ensemble learning

Cátedra VIEWNEXT
AN IBM SUBSIDIARY

Cátedra UEx

Andrés Caro Lindo
Director
Cátedra VIEWNEXT-UEx

Presentación

Cátedra VIEWNEXT AN IBM SUBSIDIARY UEX — Seguridad y Auditoría de Sistemas Software

UNIVERSIDAD DE EXTREMADURA

gim — Grupo de Ingeniería de Medios

Investigación en Informática Forense y Ciberderecho — MOOCUEx

Derecho Tecnológico & Informática Forense

JNIC 2019 — Jornadas Nacionales de Investigación en Ciberseguridad

Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Introducción

MACHINE LEARNING

¿Extracción de conocimiento a partir de

DATASETS

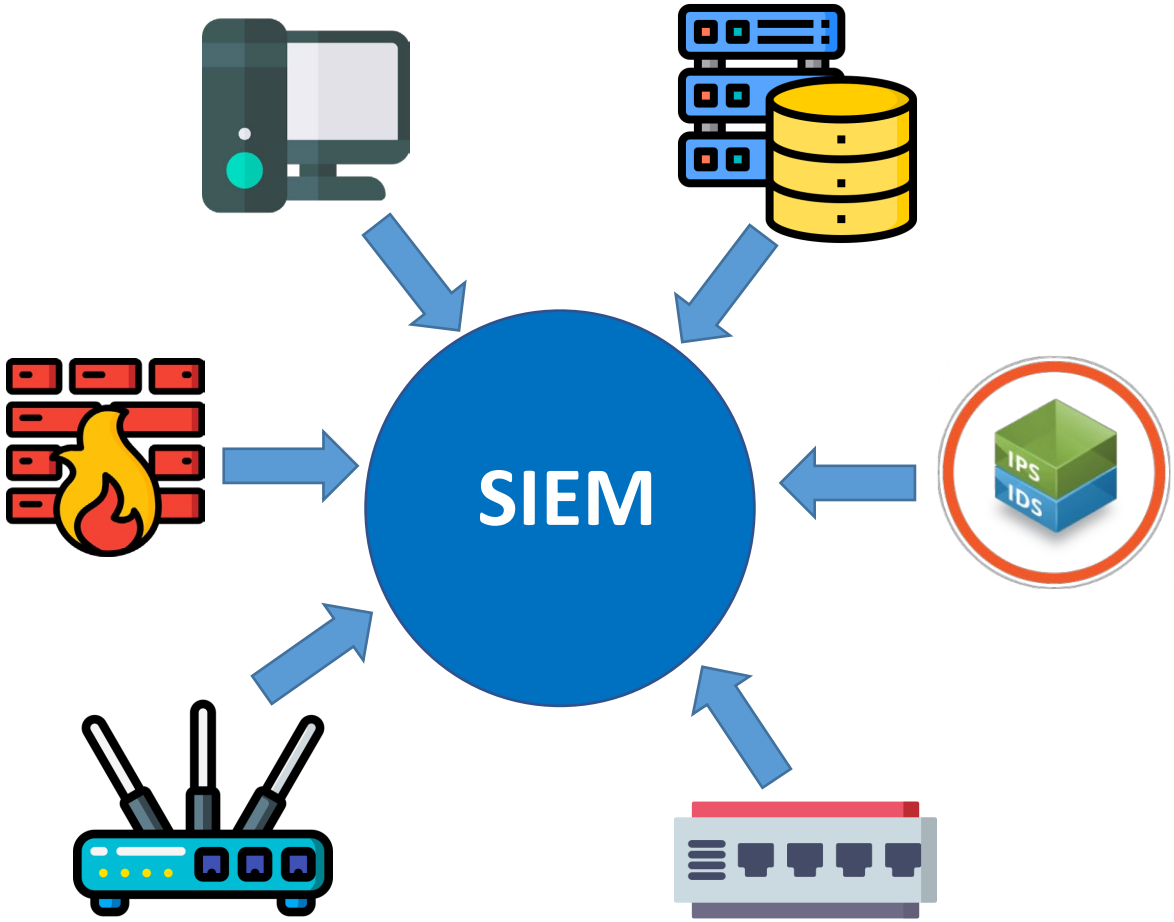
fuentes de datos procedentes de la

SIEM

monitorización de eventos de seguridad?

Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Introducción



Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Dataset

UNSW-NB15

Cyber Range Lab at UNSW Canberra

	Train	Train (%)	Test	Test (%)
Analysis	2000	1.14	677	0.82
Backdoor	1746	1.00	583	0.71
DoS	12264	6.99	4089	4.97
Exploits	33393	19.04	11132	13.52
Fuzzers	18184	10.37	6062	7.36
Generic	40000	22.81	18871	22.92
Normal	56000	31.94	37000	44.94
Reconnaissance	10491	5.98	3496	4.25
Shellcode	1133	0.65	378	0.46
Worms	130	0.07	44	0.05
TOTAL	**175341**	**100%**	**82332**	**100%**

Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Dataset

Técnicas de balanceo

- **OVERSAMPLING** - GENERAR MÁS MUESTRAS, EN LAS CLASES MINORITARIAS
  - **SMOTE** (SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE)
  - SMOTE-BORDERLINE1 (**BLINE-1**)
  - ADAPTATIVE SYNTHETHIC (**ADASYN**)

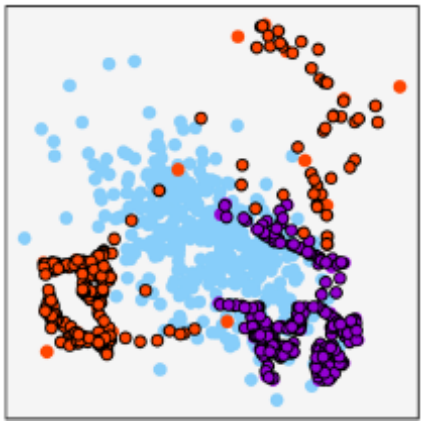
- **UNDERSAMPLING** - REDUCIR EL NÚMERO DE MUESTRAS
  - RANDOM UNDERSAMPLING (**RU**)
  - CLUSTER CENTROIDS (**CC**)
  - NEAR MISS (NM)
  - REPEATED EDITED NEAREST NEIGHBOUR (**RENN**)
  - TOMEK LINKS (**TKL**)

Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Dataset

Técnicas de balanceo

- **OVERSAMPLING** - GENERAR MÁS MUESTRAS, EN LAS CLASES MINORITARIAS
  - **SMOTE** (SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE)

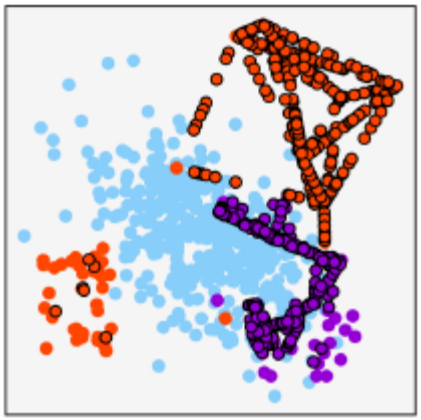


Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Dataset

Técnicas de balanceo

- **OVERSAMPLING** - GENERAR MÁS MUESTRAS, EN LAS CLASES MINORITARIAS
  - SMOTE-BORDERLINE1 (**BLINE-1**)

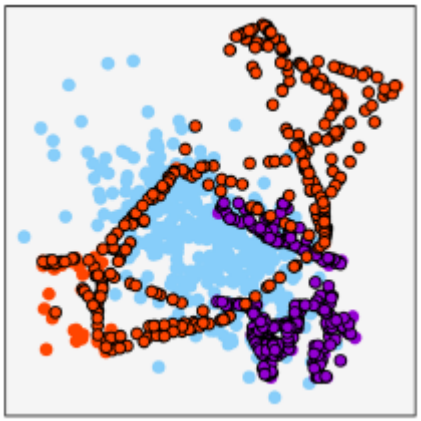


Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Dataset

Técnicas de balanceo

- **OVERSAMPLING** - GENERAR MÁS MUESTRAS, EN LAS CLASES MINORITARIAS
  - ADAPTATIVE SYNTHETHIC (**ADASYN**)



Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Dataset

Técnicas de balanceo

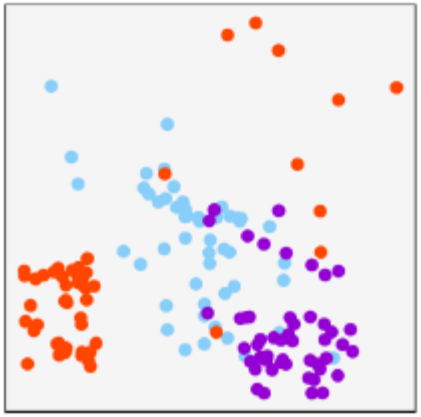
- **UNDERSAMPLING** - REDUCIR EL NÚMERO DE MUESTRAS
  - RANDOM UNDERSAMPLING (**RU**)
  - CLUSTER CENTROIDS (**CC**)
  - NEAR MISS (**NM**)
  - REPEATED EDITED NEAREST NEIGHBOUR (**RENN**)
  - TOMEK LINKS (**TKL**)

Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Dataset

Técnicas de balanceo

- **UNDERSAMPLING** - REDUCIR EL NÚMERO DE MUESTRAS
  - RANDOM UNDERSAMPLING (**RU**)

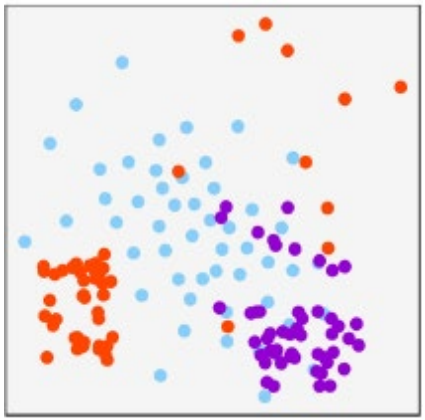


Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Dataset

Técnicas de balanceo

- **UNDERSAMPLING** - REDUCIR EL NÚMERO DE MUESTRAS
  - CLUSTER CENTROIDS (**CC**)



Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Dataset

Técnicas de balanceo

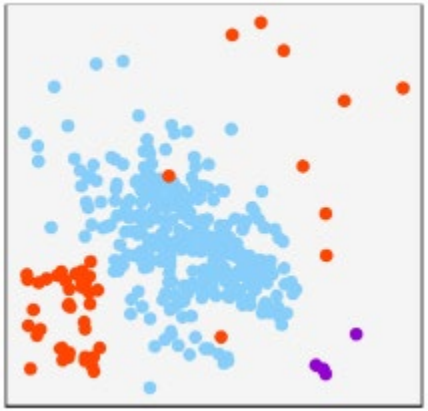
- **UNDERSAMPLING** - REDUCIR EL NÚMERO DE MUESTRAS
  - NEAR MISS (**NM**)

Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Dataset

Técnicas de balanceo

- **UNDERSAMPLING** - REDUCIR EL NÚMERO DE MUESTRAS
  - REPEATED EDITED NEAREST NEIGHBOUR (**RENN**)

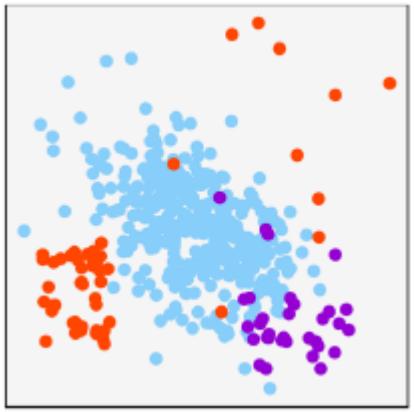


Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Dataset

Técnicas de balanceo

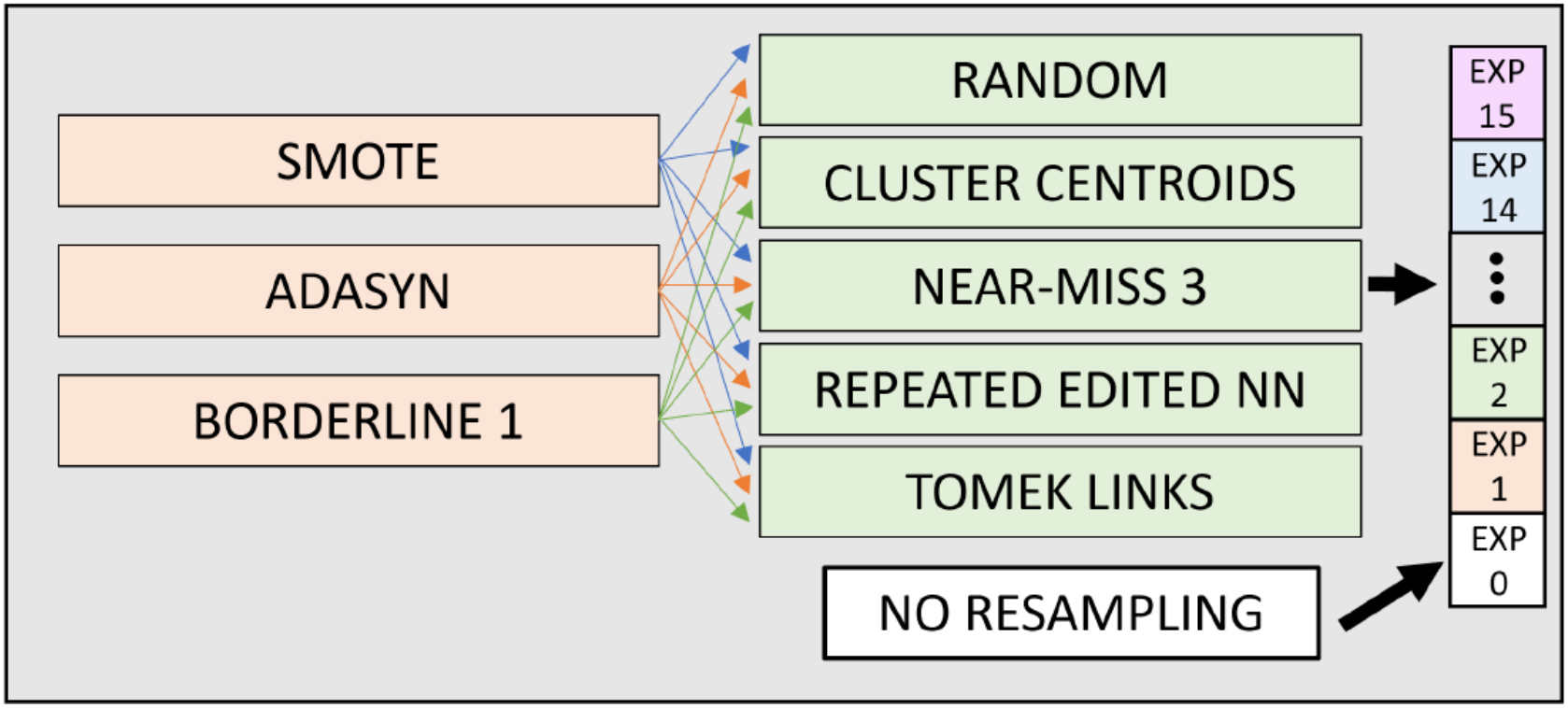
- **UNDERSAMPLING** - REDUCIR EL NÚMERO DE MUESTRAS
  - TOMEK LINKS (**TKL**)



Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Experimentos

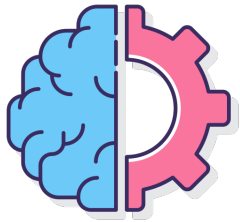
Técnicas de balanceo



Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Experimentos

Machine Learning



- SUPPORT VECTOR MACHINE (SVM)
- DECISION TREES (DT)
- K NEAREST NEIGHBORS (KNN)
- NAÏVE BAYES (NB)

UNSW-NB15

SUPPORT VECTOR MACHINE

DECISION TREE

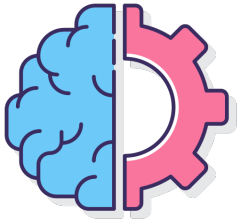
K-NEARES NEIGHBORS

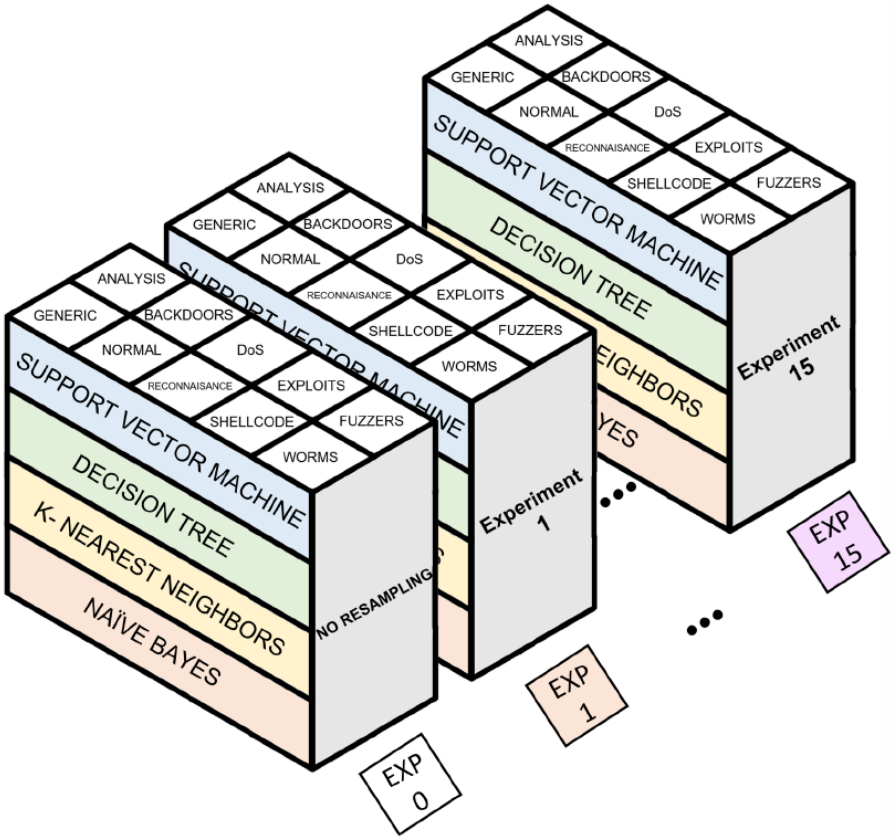
NAÏVE BAYES

Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Experimentos

ML + BALANCEO

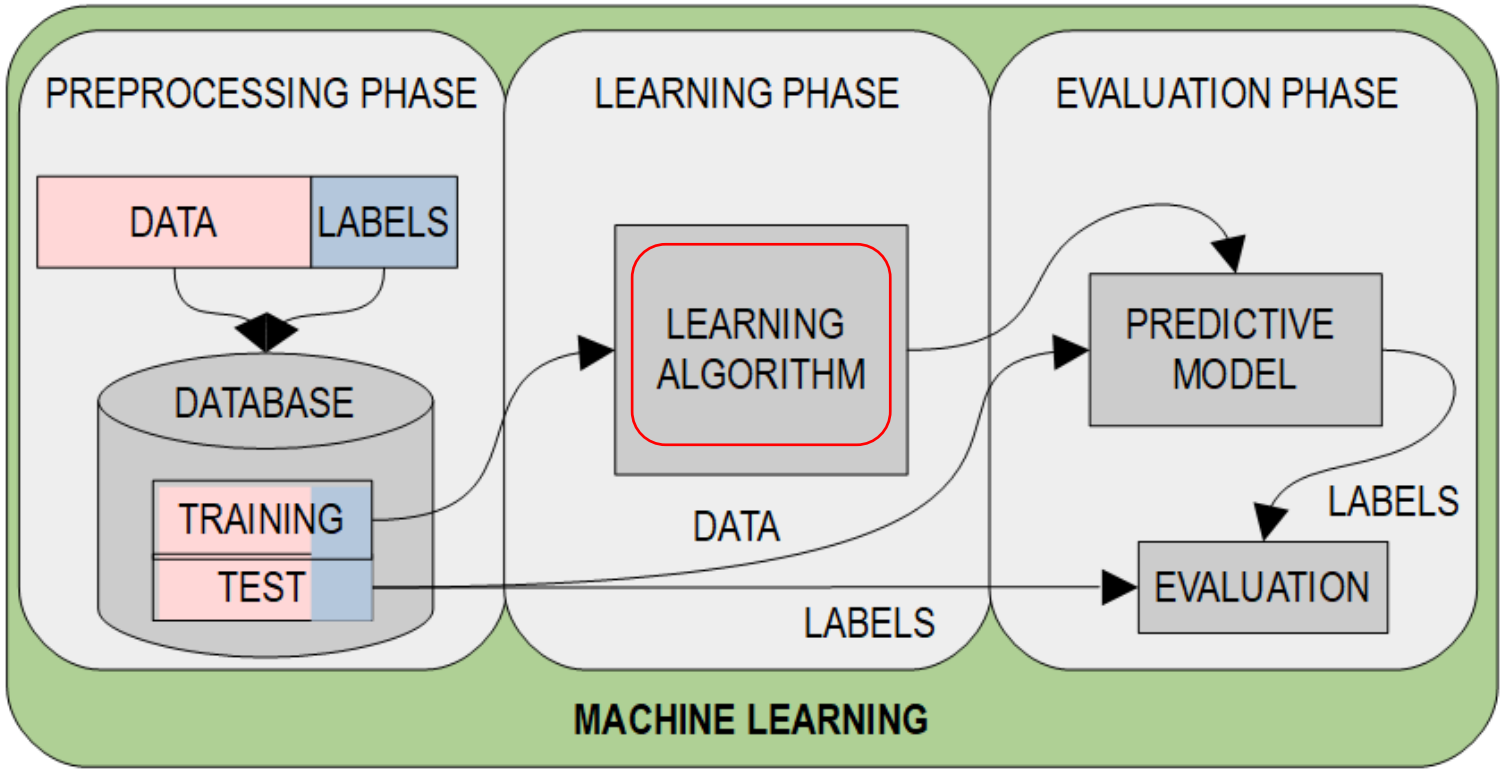




Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Experimentos

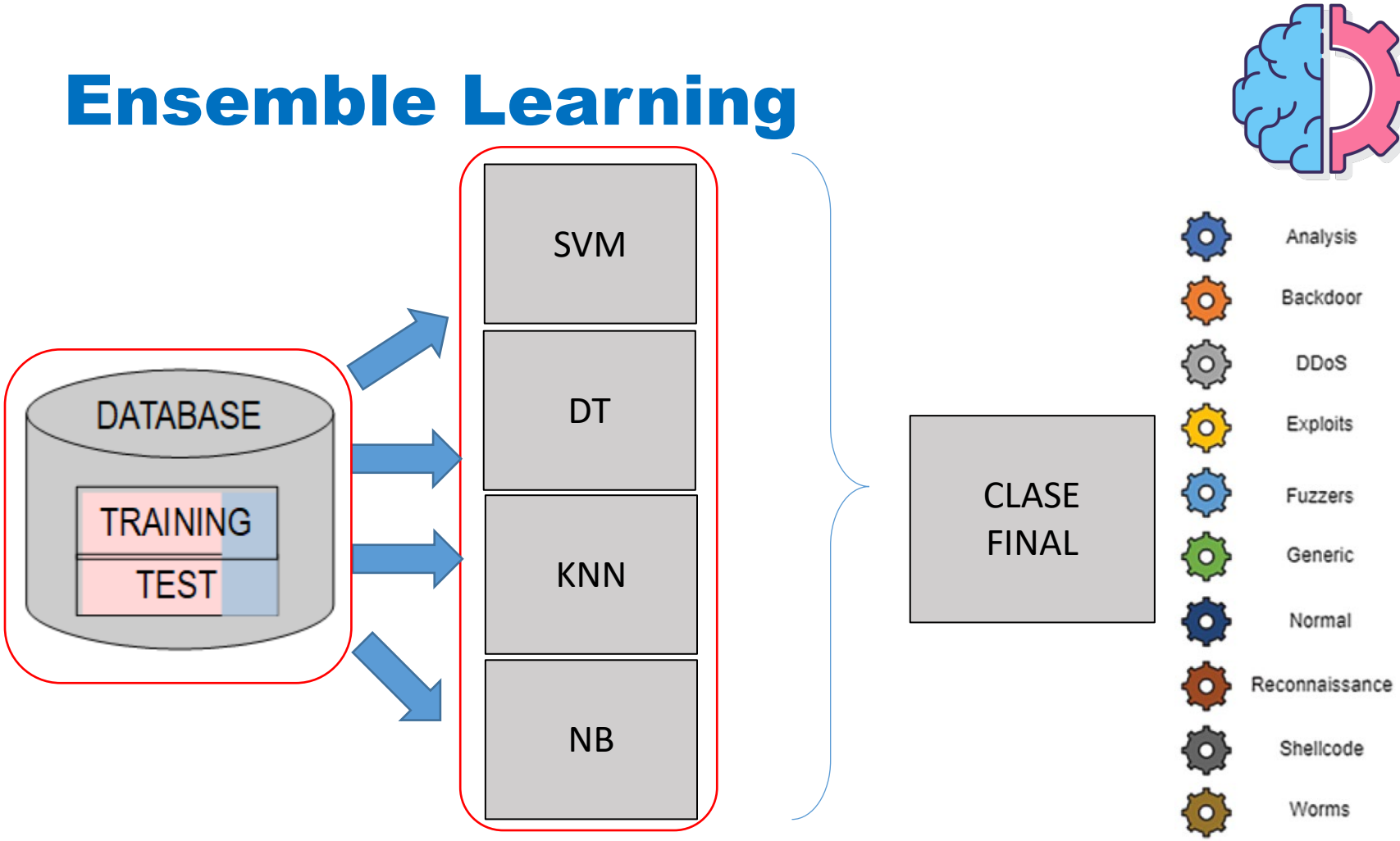
Machine Learning



Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Experimentos

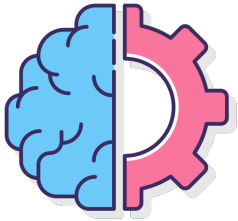
Ensemble Learning

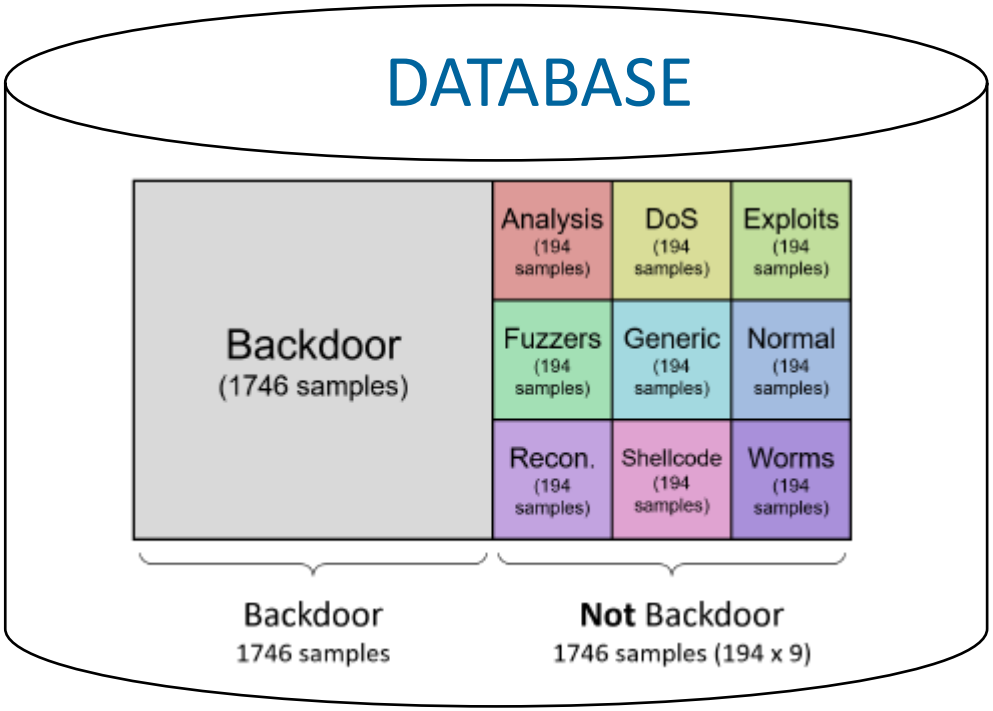


Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Experimentos

Ensemble Learning

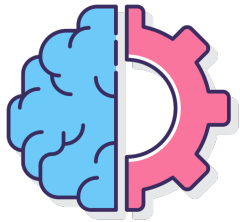




Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Experimentos

Ensemble Learning



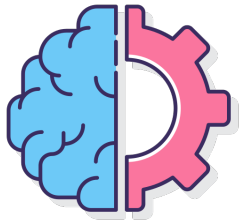
Medicina general

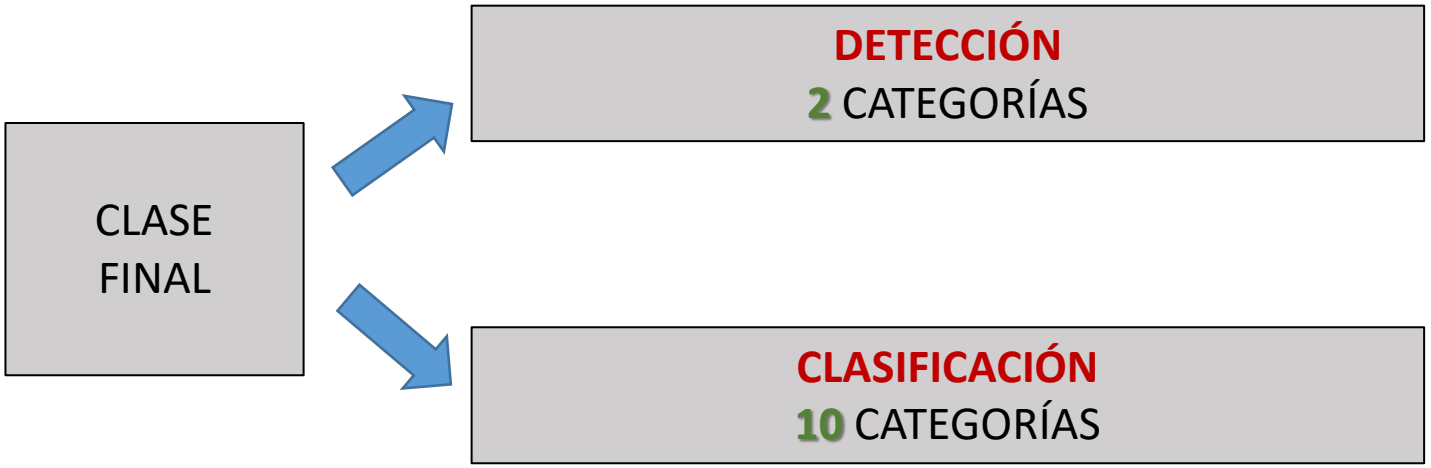
- Oftalmología
- Otorrinolaringología
- Alergología
- Digestivo
- Cardiología
- Dermatología

Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Experimentos

Ensemble Learning

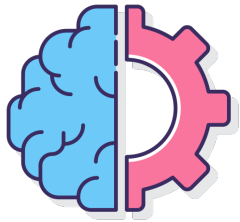


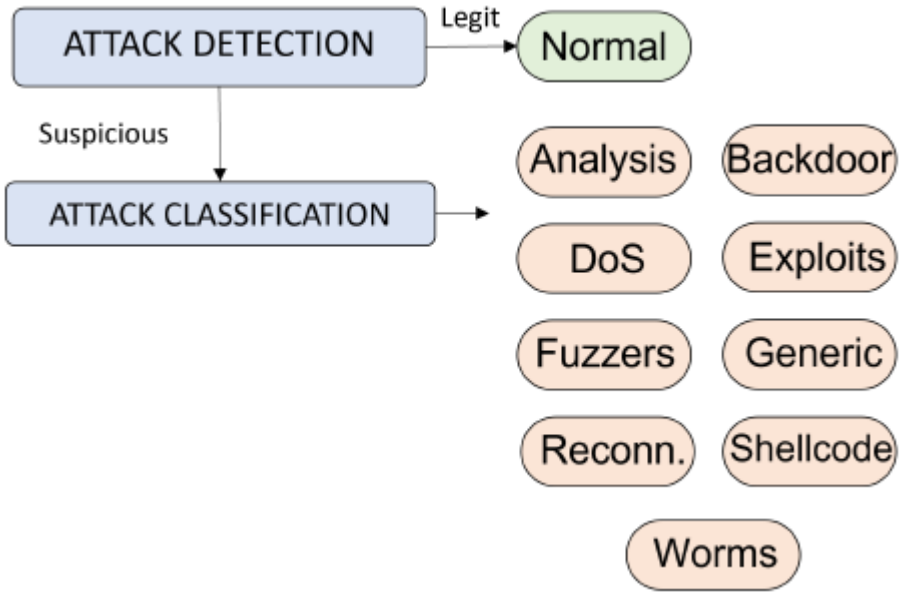


Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Experimentos

Ensemble Learning

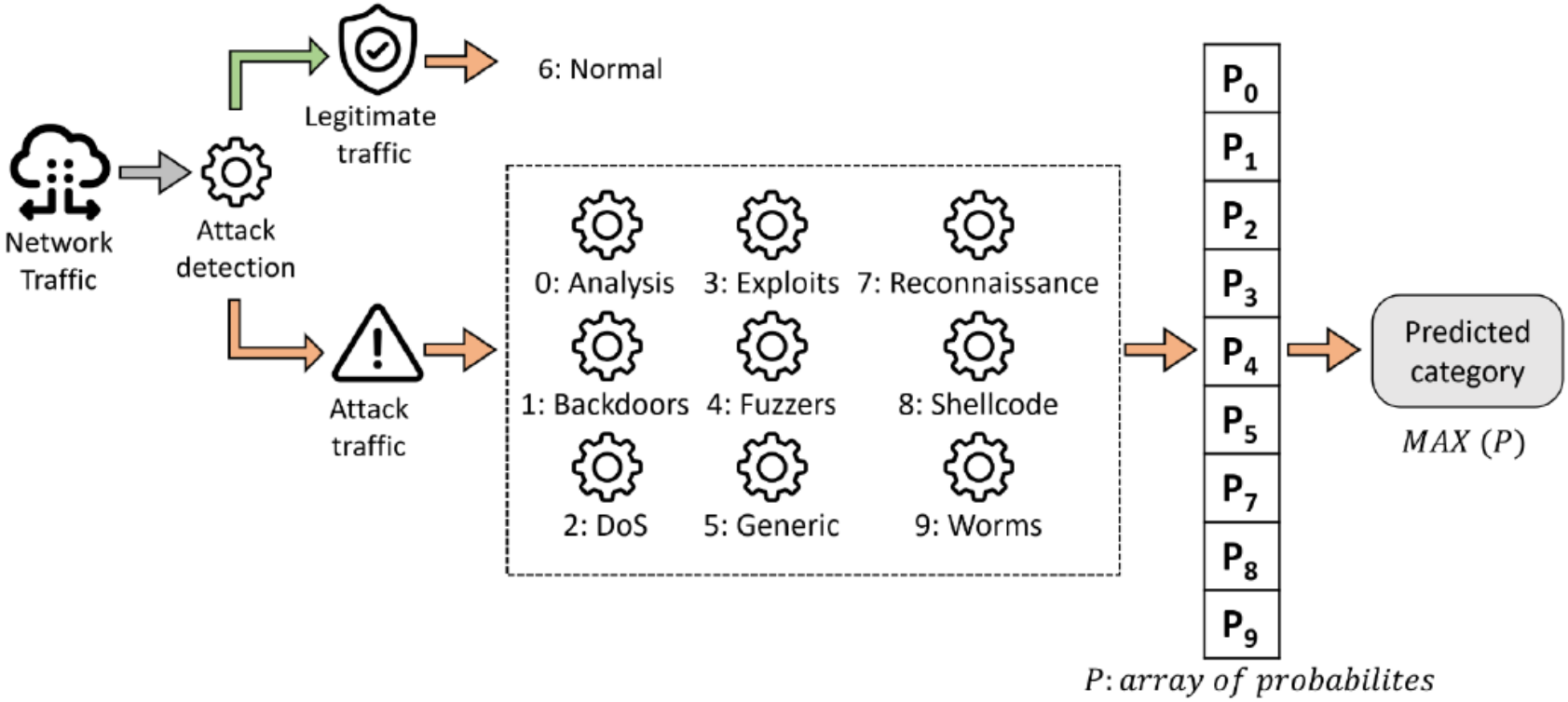




Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Experimentos

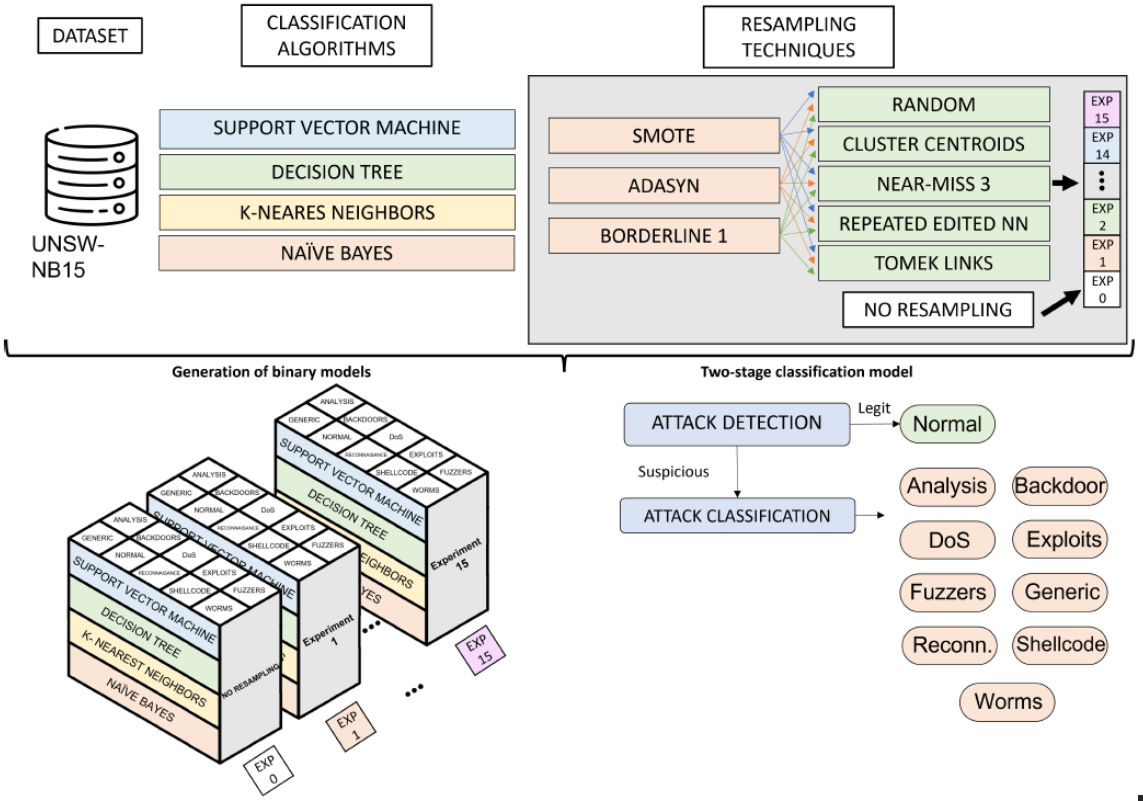
Ensemble Learning



Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Experimentos

Ensemble Learning



Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Resultados

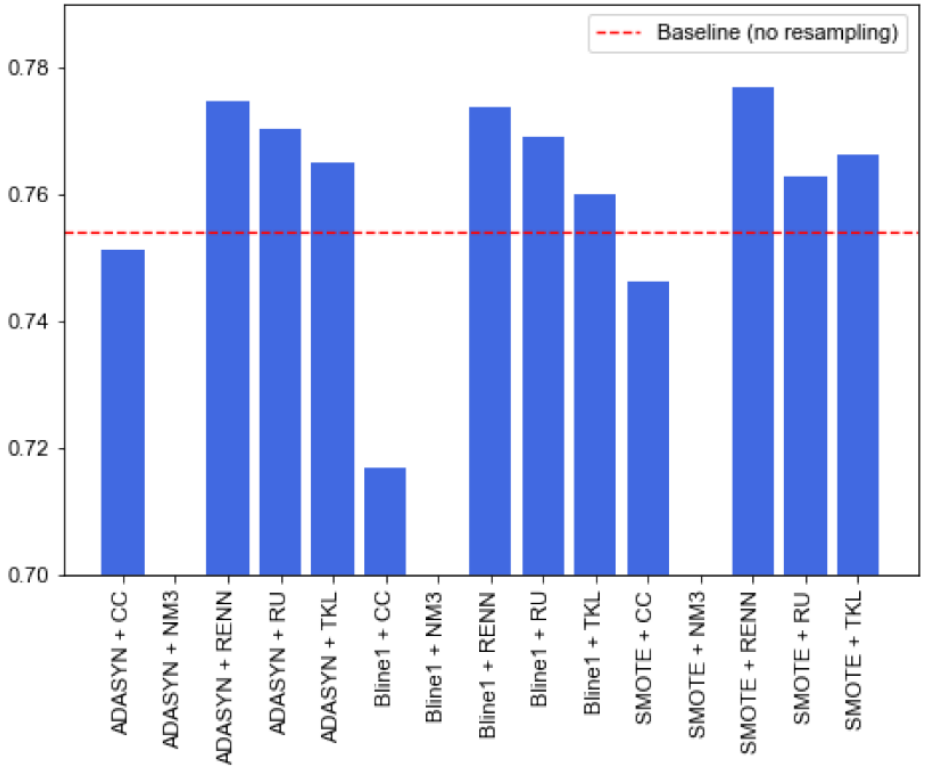
Undersampling - Oversampling

	Composition									
Models	An.	Ba.	DoS	Ex.	Fu.	Ge.	No.	Re.	Sh.	Wo.
Analysis	-	US	US	US	US	US	US	US	US	OS
Backdoor	US	-	US	US	US	US	US	US	US	OS
DoS	US	US	-	US	US	US	US	US	OS	OS
Exploits	OS	OS	US	-	US	US	US	US	OS	OS
Fuzzers	OS	OS	US	US	-	US	US	US	OS	OS
Generic	OS	OS	US	US	US	-	US	US	OS	OS
Normal	OS	OS	US	US	US	US	-	US	OS	OS
Reconnaissance	US	US	US	US	US	US	US	-	OS	OS
Shellcode	US	US	US	US	US	US	US	US	-	US
Worms	US	US	US	US	US	US	US	US	US	-

Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Resultados

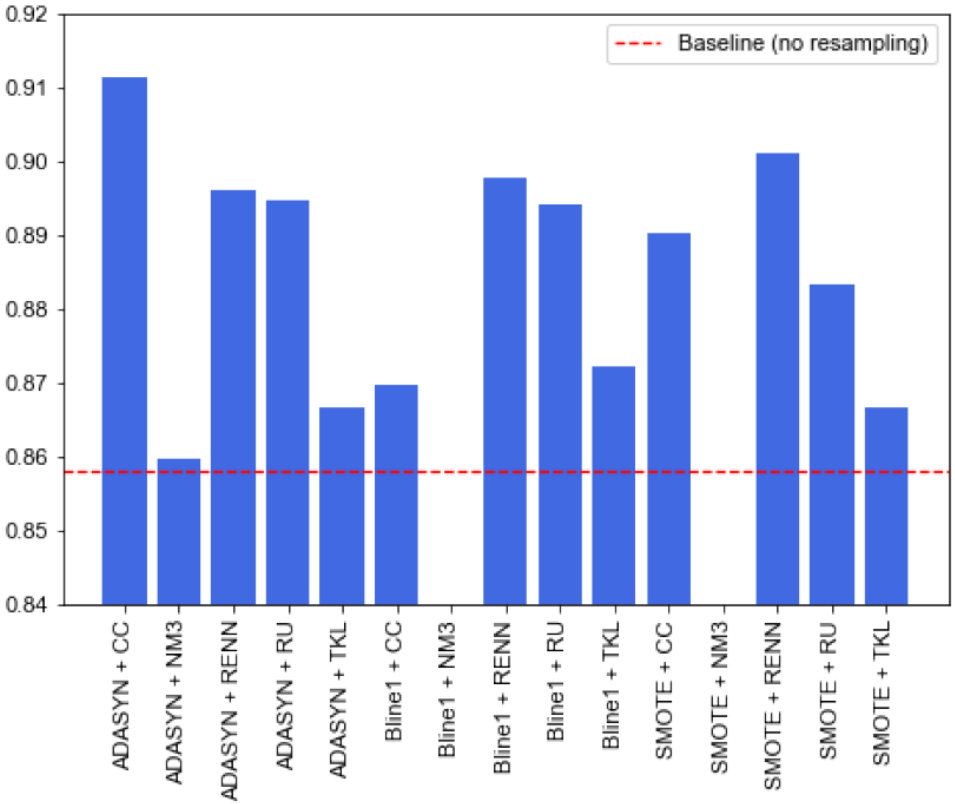
Multiclass classification



Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Resultados

Clasificación en dos pasos



Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

Resultados

Clasificación en dos pasos

Category	Acc.	Prec.	Recall	F1	FPR	Train Time
Analysis	0.8312	0.8388	0.8312	0.8298	0.1723	105 ms
Backdoor	0.8553	0.8747	0.8553	0.8530	0.1502	98 ms
DoS	0.7907	0.7936	0.7907	0.7909	0.2076	10.6 s
Exploits	0.7983	0.7975	0.7983	0.7963	0.2157	692 ms
Fuzzers	0.8127	0.8247	0.8127	0.8128	0.1823	47.4 s
Generic	0.9835	0.9839	0.9835	0.9835	0.0144	2 s
Normal	0.9112	0.9112	0.9112	0.9112	0.0952	2.4 s
Reconn.	0.9183	0.9255	0.9183	0.9183	0.0785	170 ms
Shellcode	0.9643	0.9644	0.9643	0.9643	0.0357	88 ms
Worms	0.8876	0.8884	0.8876	0.8876	0.1126	58 m

Andrés Caro
Universidad de Extremadura
@_AndresCaro_
andresc@unex.es

MUCHAS GRACIAS

Andrés Caro
Universidad de Extremadura

andresc@unex.es
@_AndresCaro_
www.linkedin.com/in/andrescaro/