

Inteligencia Computacional

**Actividad de Evaluación 1**

Profesor Jorge Rodríguez Ruíz

José María Aguíñiga Díaz

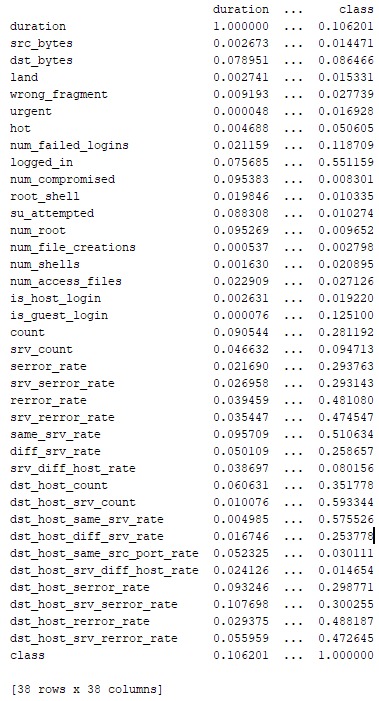
A01376669

Jorge Alexis Rubio Sumano

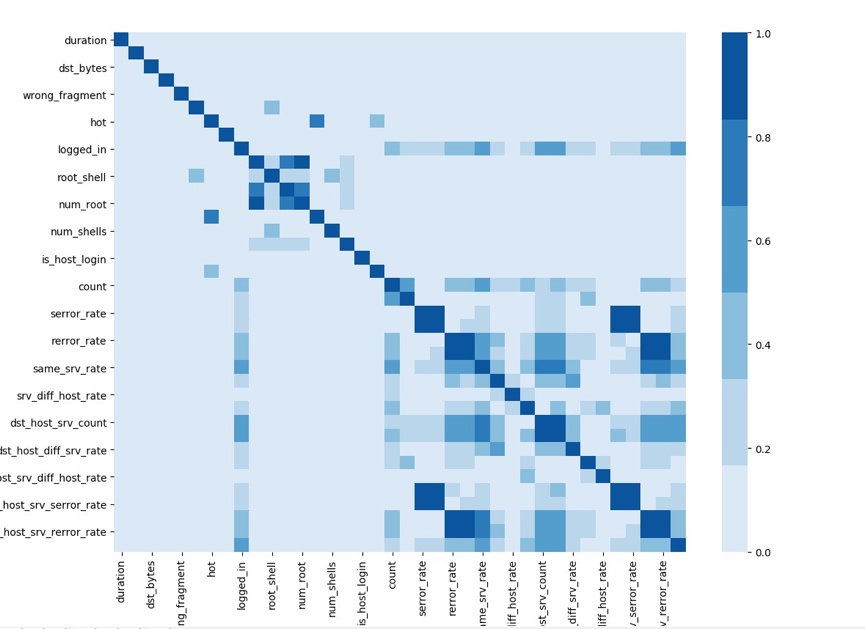
El objetivo de la práctica es aplicar sobre un dataset los modelos de KNN y en nuestro caso redes neuronales para determinar que modelo fue el más preciso, para esto generamos 10-cross fold sobre el mismo data set.

**Correlación**

La correlación fue utilizada para la discriminación de datos irrelevantes a nuestro dataset para tener un dataset limpio. Tomamos como estándar los elementos que tuvieran una correlación mayor a 0.1, los que no cumplieran con dicha condición fueron eliminados.



**Heat Map**



Para finalizar la preparación de los datos cambiamos la última columna correspondiente a la clase a un tipo de dato binario, es decir, anomaly sería ahora 0 y normaly 1.

**Cross-fold**

Para la realización de un 10-cross-fold se revolvió el data set y se dividió por fold en train y test escribiendo dichos data sets en nuevos archivos, todo este proceso iterando 10 veces.

**KNN**

Se realizó el entrenamiento con el modelo KNN en el cual después de iterar en diversas ocasiones con K’s entre 1 y 40 se encontró que los mejores resultados se ubicaban con K’s entre 1 y 10.

Accuracy

K=1

Fold=5  
Accuracy: 0.9079052435453819  
F1: 0.8920773549594511

K=2

Fold=5

Accuracy: 0.9079052435453819  
F1: 0.8920773549594511

K=3

Fold=3

Accuracy: 0.9094135391713246  
F1: 0.8917859035506095

K=4

Fold=9

Accuracy: 0.910478218436696  
F1: 0.8931257282067578

K=5

Fold=3

Accuracy: 0.9054209919261822  
F1: 0.8881896370883157

K=6

Fold=9

Accuracy: 0.9096797089876675  
F1: 0.8911462788708299

K=7

Fold= 7

Accuracy: 0.9056871617425251  
F1: 0.8880463401790416

K=8

Fold=9

Accuracy: 0.9090586460828676  
F1: 0.8897493815209208

K=9

Fold=5

Accuracy: 0.9029367403069826  
F1: 0.8831196581196581

Aquí resaltamos las K’s con mejores accuracy junto con el fold con el cual lograron dicho acontecimiento.

Podemos observar que la K con mejor accuracy presentada es la K=4 con el fold número 5, el resultado obtenido fue: 0. 910478218436696

F1

K=1

Fold=8

Accuracy: 0.9051548221098394  
F1: 0.8887964215125351

K=2

Fold=4

Accuracy: 0.9075503504569249  
F1: 0.8929085303186022

K=3

Fold=3

Accuracy: 0.9094135391713246  
F1: 0.8917859035506095

K=4

Fold=3

F1: 0.8931321869395507

K=5

Fold=3

F1: 0.8881896370883157

K=6

Fold=7

F1: 0.8926280365969083

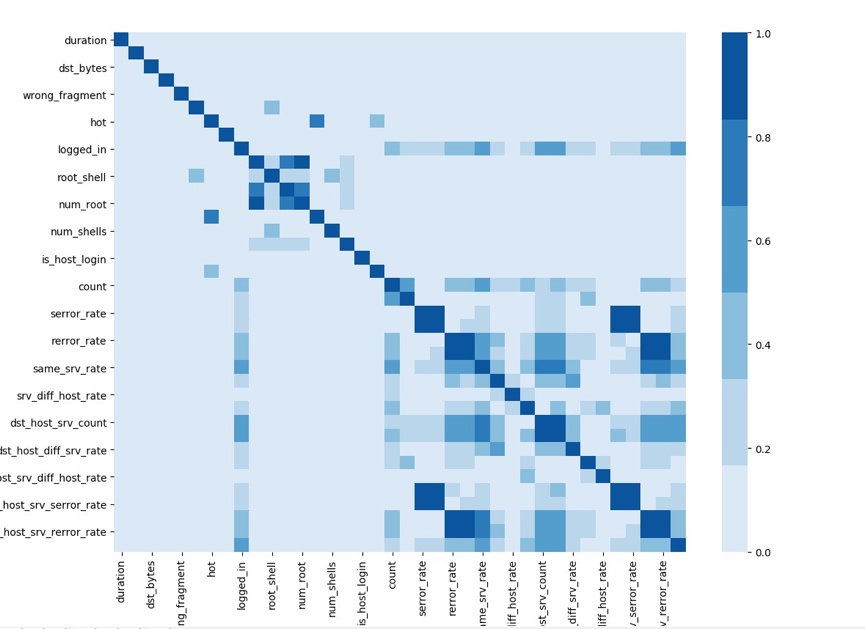
K=7

Fold=7

3. Haz un análisis de correlación para identificar variables que puedan servir o que sean repetidas. Puedes utilizar un subconjunto de los datos para ello.

Eliminamos todas los datos con correlación menor a 0.1 para la limpieza del data set.

Heat Map:



Correlation:

