

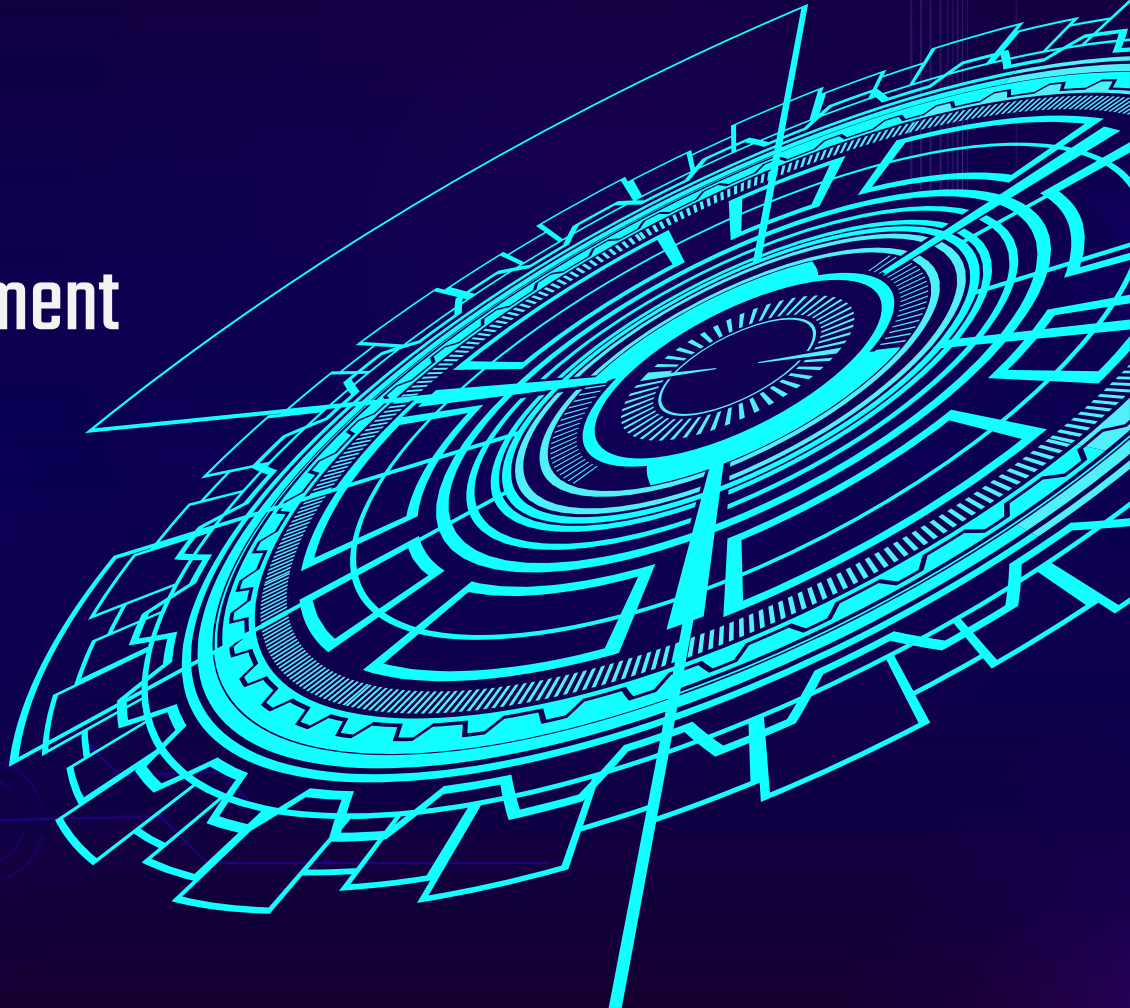
PUBG - Finish Placement Prediction

Advanced Machine Learning
A.A. 2019 - 2020

Palazzi Bruno 806908



Villa Giacomo 807462



CONTENUTI

01

Introduzione al problema



02

Descrizione Dataset



03

Ottimizzazione



04

Valutazione Performance



05

Conclusioni e possibili sviluppi futuri





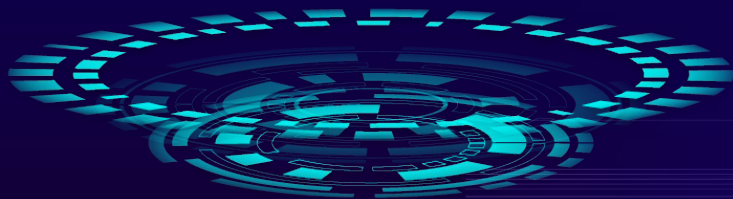
INTRODUZIONE



Videogiocare è diventato
un **business** con l'avvento
dello streaming



Campionati mondiali con
montepremi in denaro
sono sempre più comune



INTRODUZIONE

PUBG

(**P**layer**U**nknown's
Battle**G**rounds) è
un massively
multiplayer online
game



Differenti modalità di
gioco disponibili



I **player** atterrano in un
punto della **mappa** senza
armi né protezioni

PRESS
START



DATASET

Il **dataset** contiene
diverse informazioni:

- damageDealt
- kills
- walkDistance
- weaponsAcquired
- matchType
- ...

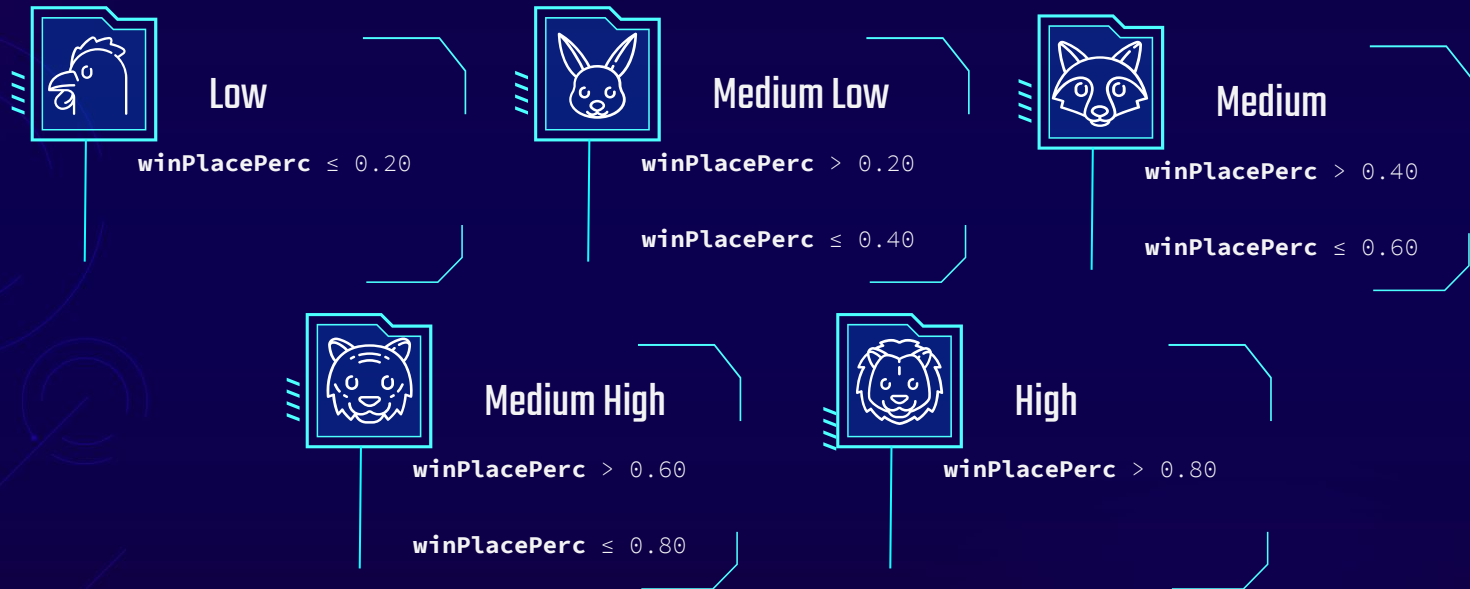


Ci siamo concentrati
solo sulle **partite
single player.**

Il **task iniziale** era
una **regressione
lineare.**

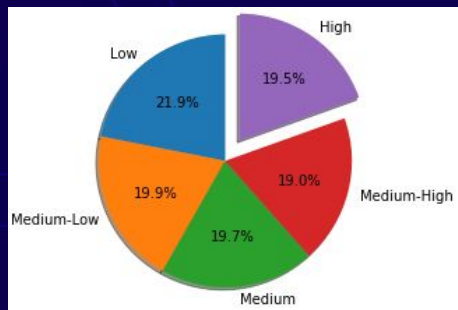
Abbiamo **definito delle
classi** in funzione del
valore da predire.

DATASET



DATASET

Distribuzione delle classi



Classi bilanciate

Correlazione attributi con classe

label	0.12	0.63	0.49	0.34	0.4	-0.77	0.01	0.48	0.43	0.430	0.006	0.0062	0.3	0.052	0.17	0.053	0.81	0.67	0.009	1
assists																				
boosts																				
damageDealt																				
headshotKills																				
heals																				
killPlace																				
killPoints																				
kills																				
killStreaks																				
longestKill																				
matchDuration																				
numGroups																				
rideDistance																				
roadKills																				
swimDistance																				
vehicleDestroys																				
walkDistance																				
weaponsAcquired																				
winPoints																				
label																				

Alta correlazione tra i dati





OTTIMIZZAZIONE

Basata su un approccio SMBO
Ha riguardato gli optimizer:

- **SGD** (Learning rate, momentum)
- **RMSprop** (Learning rate, rho)
- **Adagrad** (Learning rate)
- **Adadelta** (Learning rate, rho)

Ma anche layers nascosti e neuroni:

- **Primo** (da 150 a 300 neuroni)
- **Secondo** (da 100 a 250 neuroni)
- **Terzo** (da 50 a 100 neuroni)

Accuracy su 3-Cv come funzione nascosta.

5 punti iniziali e **30 selezionati** con **Expectation of Improvement**.

Random Forest come modello surrogato. 





OTTIMIZZAZIONE

Le **configurazioni** condividono:

- **Layers Dropout** (40% dei neuroni)
- **Early Stopping** (patience pari a 5)
- **Loss Function** (categorical cross-entropy)
- **Funzione output** (softmax)
- **Attivazione hidden** (ReLU)



Ma anche:

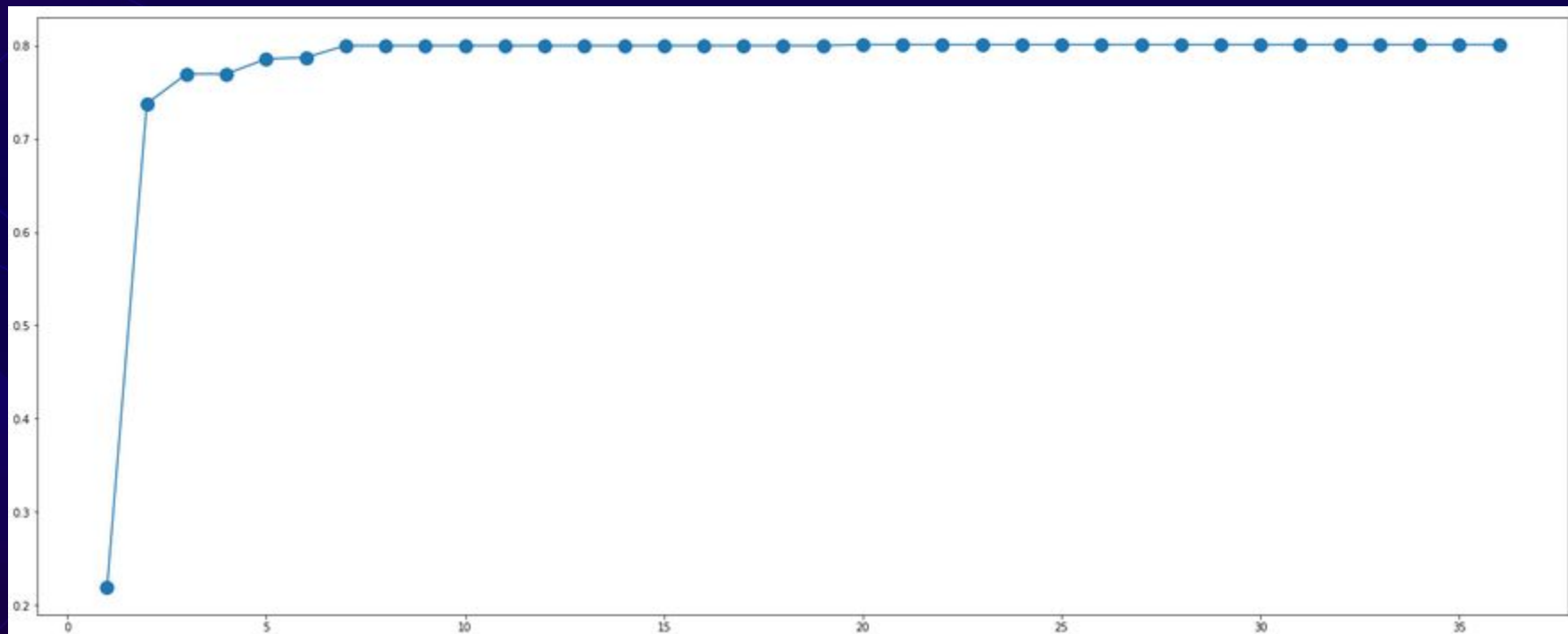
- **Neuroni input** (19 neuroni)
- **Neuroni output** (5 classi)

Si ipotizzano **buone performance...**



OTTIMIZZAZIONE

Best Seen ottenuto date le 35 iterazioni





VALUTAZIONI PERFORMANCE

Classe	Precision	Recall
Low	0.943	0.920
Medium Low	0.804	0.817
Medium	0.741	0.761
Medium High	0.677	0.676
High	0.822	0.807



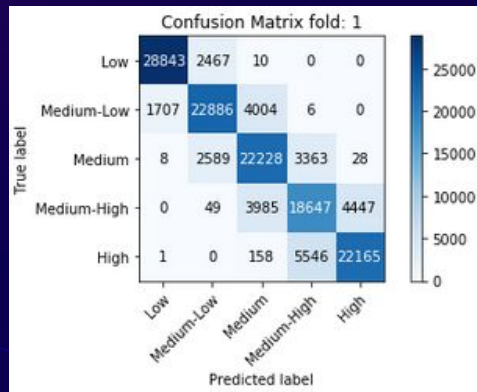
Configurazione migliore:

- **Ottimizzatore:** *Adadelta*
- **Learning rate:** *0.09644*
- **Rho:** *0.88094*
- **Layers nascosti:** *1*
- **Neuroni:** *228*

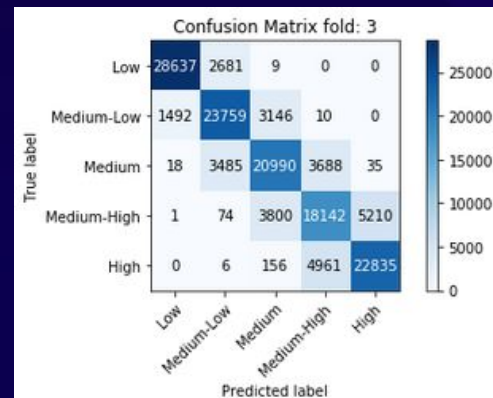
L'**accuracy media**
ottenuta è pari a
0.80.



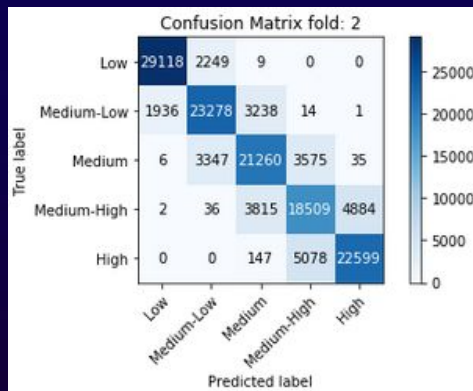
VALUTAZIONI PERFORMANCE



Tendenza
ottimistica
nell'inferenza.



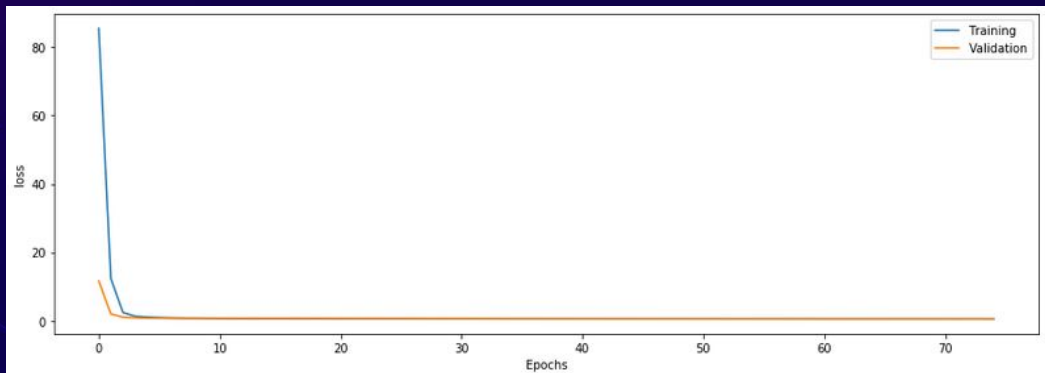
Si **sbaglia** meno
su **Low** che su
High.



Le **classi opposte**
non vengono **mai**
confuse.

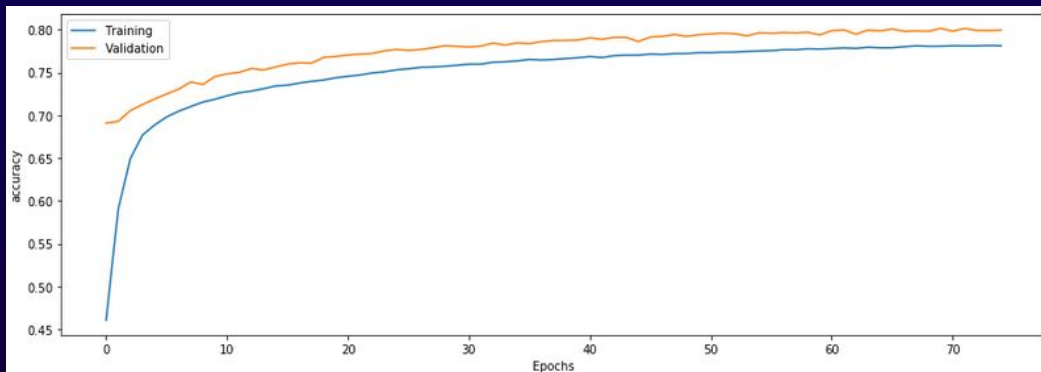


VALUTAZIONI PERFORMANCE



Early stopping
definito sul
validation set.

Buona capacità di
generalizzazione
su dati mai visti.





CONCLUSIONI E POSSIBILI SVILUPPI FUTURI



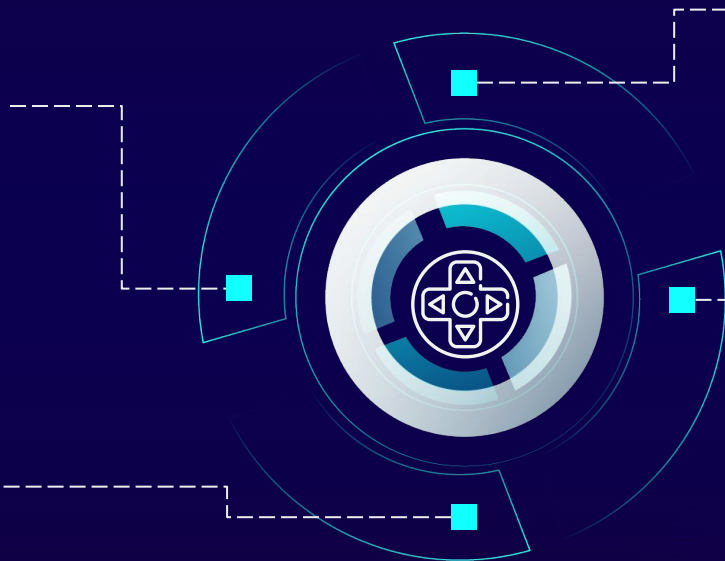
Adadelta

Finestra sugli update del gradiente premiata.



Troppo Ottimismo

Sarebbe **meglio sottomistare** le capacità che sovrastimare.



SGD

Non converge
(troppo lento),
early stopping
taglia.

Funzione nascosta

Potrebbero esserci
**aree con ottimi
migliori.**



CONCLUSIONI E POSSIBILI SVILUPPI FUTURI


Early Stopping

 Cosa **succede** se lo **togliamo**?


Più Budget

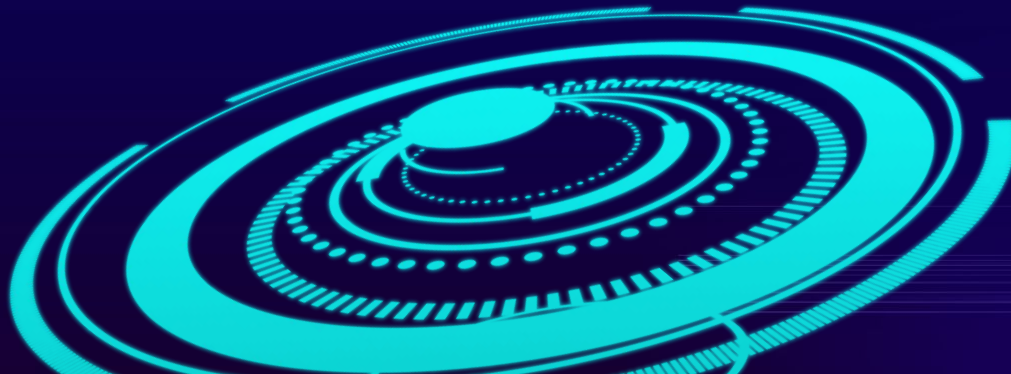
 **30 punti** per **6** **variabili...**

Confronti

Adadelta è veramente il migliore? 

Analisi

Molte **analisi** **possibili.** 

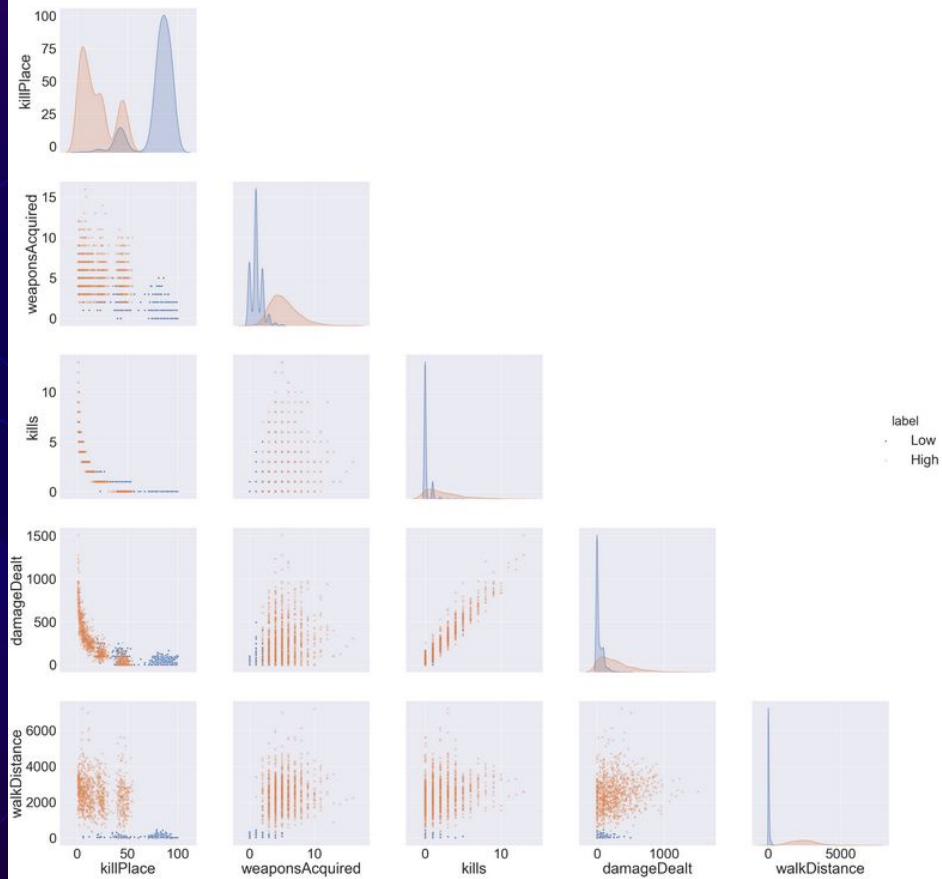




Grazie per l'attenzione.



DATASET



Classe High vs Classe Low

Le classi opposte **abbastanza distinguibili...**



... non dovrebbero dare problemi. **Ma le classi intermedie?**

