

Capitolo 1

Risultati

In questo capitolo verranno presentati e discussi i risultati ottenuti nelle diverse configurazioni descritte nei capitoli precedenti.

1.1 Performance Evaluation

Per valutare la qualità dei modelli di ML Supervisionati Deep e non Deep, sono stati utilizzati i seguenti metri di performance:

- **MSE: Mean Squared Error**
Mean Squared Error (MSE) è una media delle differenze al quadrato tra i valori predetti e quelli reali.
- **RMSE: Root Mean Squared Error**
Root Mean Squared Error (RMSE) è la radice quadrata della media delle differenze al quadrato tra i valori predetti e quelli reali.
- **R2: Coefficient of Determination**
Coefficient of Determination (R2) è una misura di quanto i valori predetti siano vicini ai valori reali.
- **MAE: Mean Absolute Error**
MAE è una media delle differenze assolute tra i valori predetti e quelli reali.

1.2 Tecniche di ML Supervisionate non Deep

| Model | MSE | RMSE | MAE |
|-------------------|----------------------|---------------------|----------------------|
| Linear Regression | 0.005277524569837709 | 0.0726465730082136 | 0.05598534700766051 |
| Lasso | 0.005486814781830245 | 0.07407303680712872 | 0.0568421300459931 |
| Ridge | 0.005173317239238808 | 0.0719257759029321 | 0.05530627637007967 |
| K Neighbors | 0.04014675063926314 | 0.20036654071791313 | 0.15758359542154865 |
| SVR | 0.004249807698698538 | 0.06519054915168715 | 0.049401864388909957 |
| Decision Tree R. | 0.02518975360339929 | 0.15871280226685966 | 0.12227319185366432 |
| Random Forest R. | 0.012127163985177493 | 0.11012340343985694 | 0.0845742608309641 |

| Model | R2 |
|-------------------|--------------------|
| Linear Regression | 0.9779884023114186 |
| Lasso | 0.9771154908004298 |
| Ridge | 0.9784230321851388 |
| K Neighbors | 0.8325551853181739 |
| SVR | 0.9822748229629815 |
| Decision Tree R. | 0.8949381068993166 |
| Random Forest R. | 0.9494197987687645 |

Di seguito sono riportati i risultati dei modelli precedenti, ma con l'utilizzo della PCA

| Model | MSE | RMSE | MAE |
|-------------------|-----------------------|---------------------|----------------------|
| Linear Regression | 0.006377020549959092 | 0.07985624928556996 | 0.06110574164990556 |
| Lasso | 0.009433868123483865 | 0.09712810161577269 | 0.07491622588172611 |
| Ridge | 0.0063303117843782316 | 0.07956325649681661 | 0.060759651674386066 |
| K Neighbors | 0.03922817004507678 | 0.1980610260628698 | 0.15566020188713492 |
| SVR | 0.004886523894752017 | 0.06990367583147554 | 0.05318109432495814 |
| Decision Tree R. | 0.060707019543795586 | 0.24638794520794963 | 0.19116443494466778 |
| Random Forest R. | 0.03821002323920518 | 0.19547384285168484 | 0.15058396042155897 |

| Model | R2 |
|-------------------|--------------------|
| Linear Regression | 0.973402604016331 |
| Lasso | 0.9606530472699164 |
| Ridge | 0.9735974178050478 |
| K Neighbors | 0.836386418354838 |
| SVR | 0.9796191952034381 |
| Decision Tree R. | 0.7468020331524539 |
| Random Forest R. | 0.8406329240000876 |

Miglior Classic ML

| Model | C | Epsilon | Gamma | Kernel |
|-------|---|---------|-------|--------|
| SVR | 1 | 0.01 | 0.01 | rbf |

Miglior Classic ML con PCA

| Model | C | Epsilon | Gamma | Kernel |
|-------|-----|---------|-------|--------|
| SVR | 100 | 0.01 | 0.001 | rbf |

1.3 Neural Network

Neural Network Results

| Model | MSE | RMSE | MAE |
|-------------|----------------------|---------------------|---------------------|
| NN | 0.00664011668413877 | 0.0814869105815887 | 0.0629449188709259 |
| NN With PCA | 0.005938166752457619 | 0.07705949991941452 | 0.05929824709892273 |

| Model | R2 |
|-------------|--------------------|
| NN | 0.9796283841133118 |
| NN With PCA | 0.9753079414367676 |

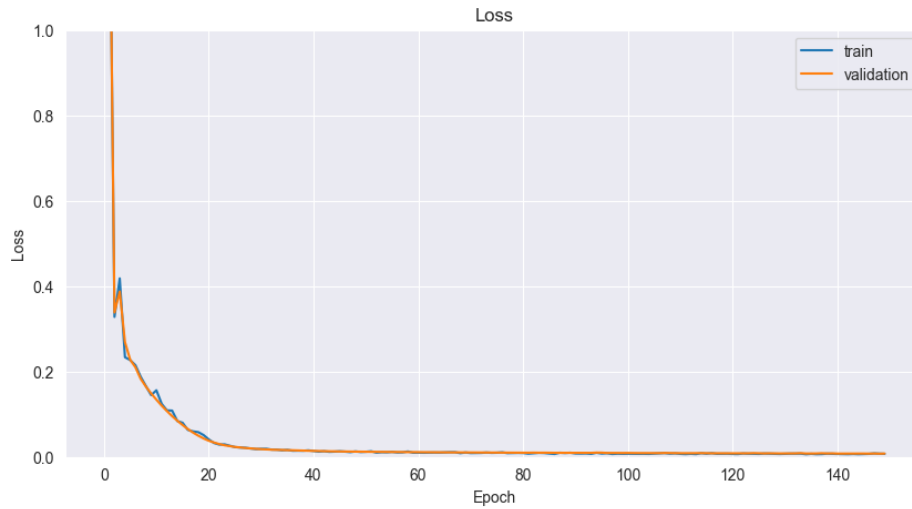
Miglior NN

| Batch Size | Input Layer | Output Layer | lr | Epochs |
|------------|-------------|--------------|------|--------|
| 1024 | 64 | 256 | 0.01 | 400 |

Miglior NN con PCA

| Batch Size | Input Layer | Output Layer | lr | # Epochs |
|------------|-------------|--------------|-------|----------|
| 256 | 128 | 256 | 0.001 | 600 |

Di seguito è riportato un grafico contenente l'andamento della funzione loss del training e validation del miglior modello di Neural Network



1.4 TabNet

TabNet Results

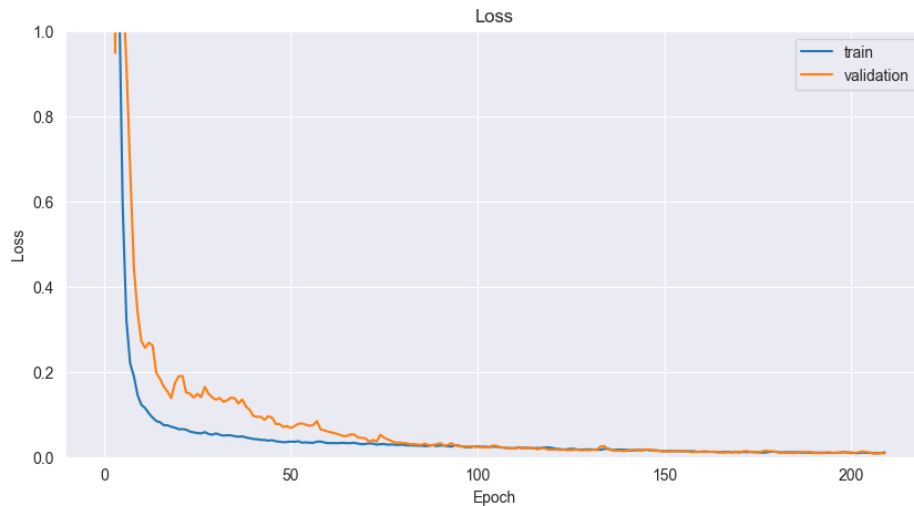
| Model | MSE | RMSE | MAE |
|-----------------|---------------------|--------------------|--------------------|
| TabNet | 0.00795749657941388 | 0.0892048013248944 | 0.0689039959039718 |
| TabNet With PCA | 0.00813222261181444 | 0.0901788368289059 | 0.964001783993923 |

| Model | R2 |
|-----------------|-------------------|
| TabNet | 0.964775228814175 |
| TabNet With PCA | 0.964001783993923 |

Miglior TabNet

| Batch Size | width | step | lr | Max Epochs |
|--------------|-------|------|------|------------|
| 2048 | 8 | 5 | 0.02 | 210 |
| 512 with PCA | 32 | 5 | 0.02 | 150 |

Di seguito è riportato un grafico contenente l'andamento della funzione loss del training e validation del miglior modello di TabNet



1.5 Conclusioni

Nel presente lavoro si è eseguita la pipeline descritta nell'Introduzione, in seguito ad aver acquisito il dataset abbiamo addestrato i modelli sia con dati grezzi sia applicando la PCA confrontando i risultati. Abbiamo notato che in tutti i modelli i risultati con PCA ($n_components = 0.95$) hanno portato a una perdita di informazioni e non un risparmio in termini di tempo sostanziale. Per tutti i modelli di Machine Learning è stato fissato il Random State per rendere possibile la riproducibilità dei risultati. Il tuning degli iperparametri è stato effettuato con il metodo GridSearchCV con 3-fold cross validation per ottimizzare proprio il tempo ed avere un risultato congruo in termini di Mean Test Score tra tutti i modelli di Machine Learning. Il modello dei Tabular Data è risultato molto più efficiente in termini di tempo rispetto alle Neural Network nonostante teniamo sempre in considerazione che i parametri (e di conseguenza i risultati) sono differenti. Possiamo concludere che tutti i modelli (eccezione fatta per KNN e DT) riescono a generalizzare bene sui dati in modo da fare previsioni accurate su dati mai visti in precedenza.