

CLASSIFICATION OF NEAR-EARTH OBJECTS

A cura di Biancini Mattia, Gherardi Marco, Monti Lorenzo



Fonte

4687

Istanze

40

Attributi

HAZARDOUS

TRUE

FALSE

Dataset originale

Dataset originale → Analisi e Pulizia

Dataset originale → Analisi e Pulizia → PCA

Dataset originale → **Analisi e Pulizia** → **PCA** →



Dataset sporcati

Dataset originale → **Analisi e Pulizia** → **PCA** →

Addestramento



Dataset sporcati

Dataset originale → **Analisi e Pulizia** → **PCA** →

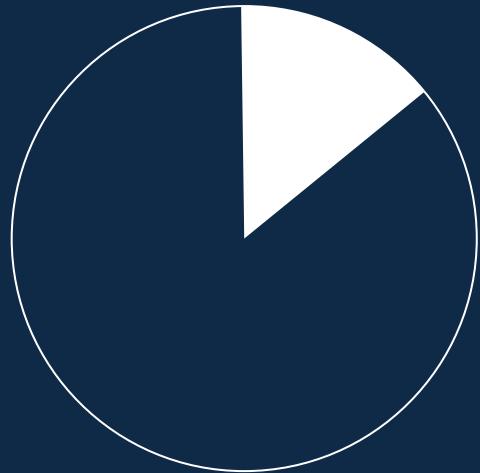
Dataset sporcati → **Pulizia**

Addestramento



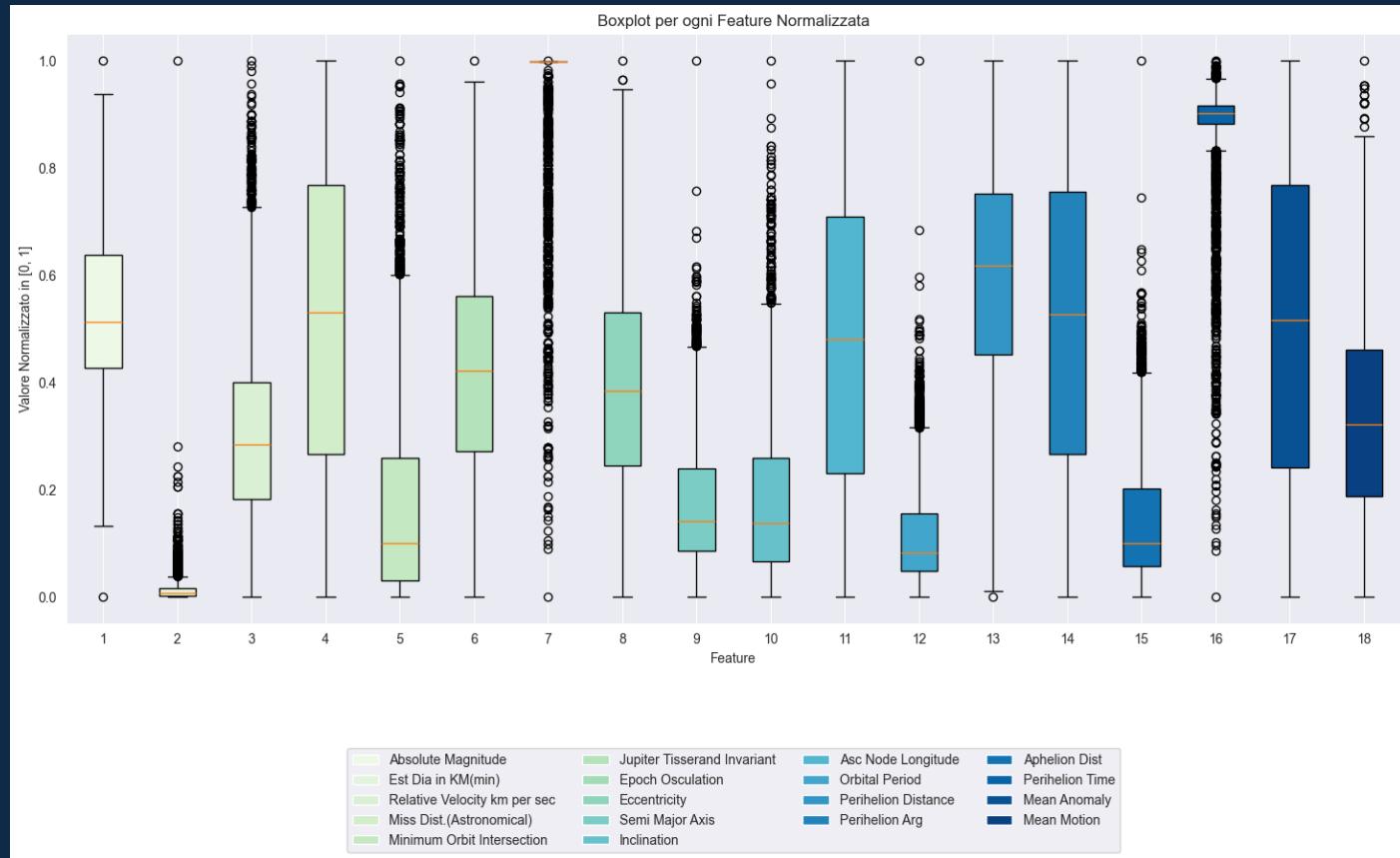
1

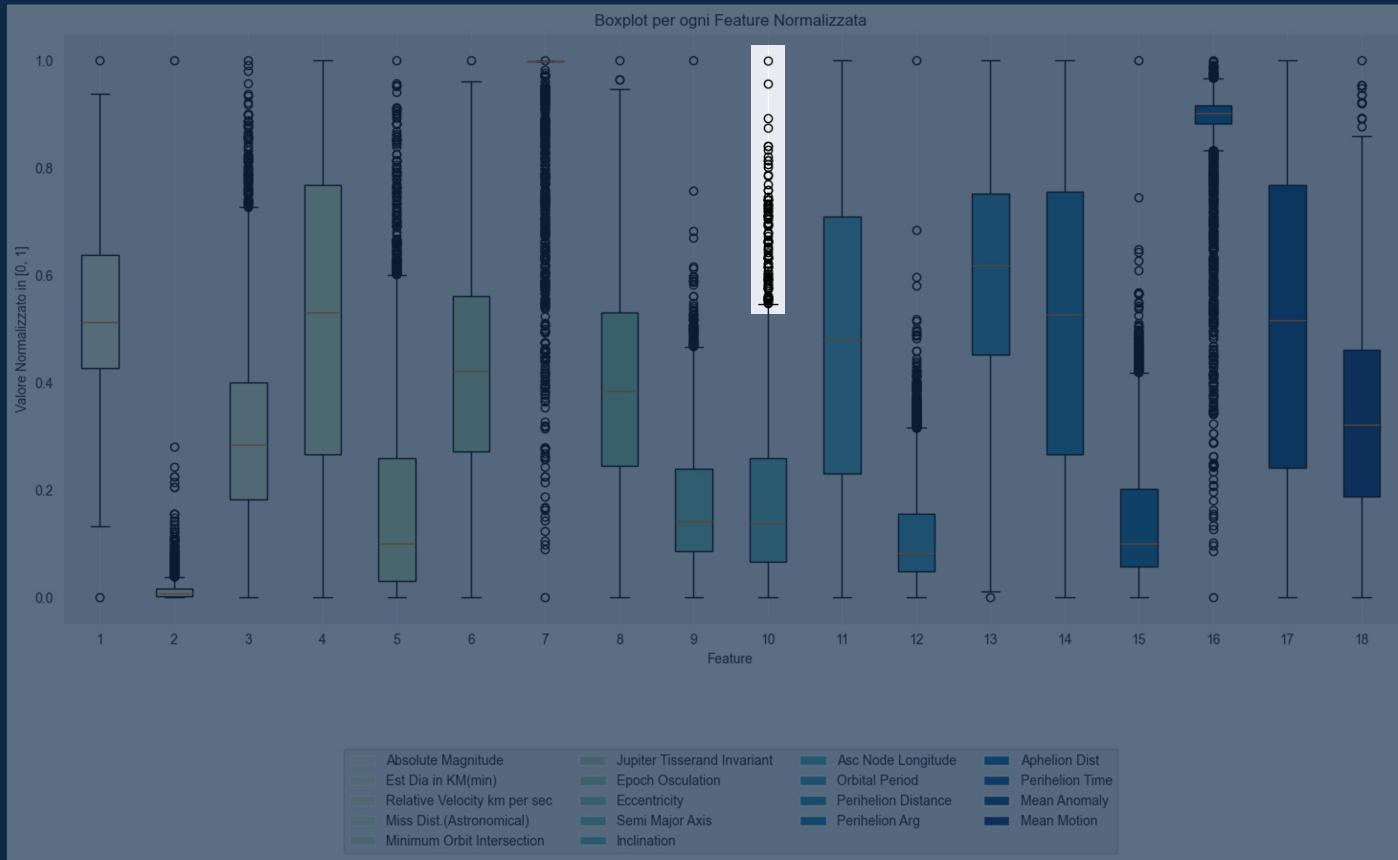
Analisi e Pulizia

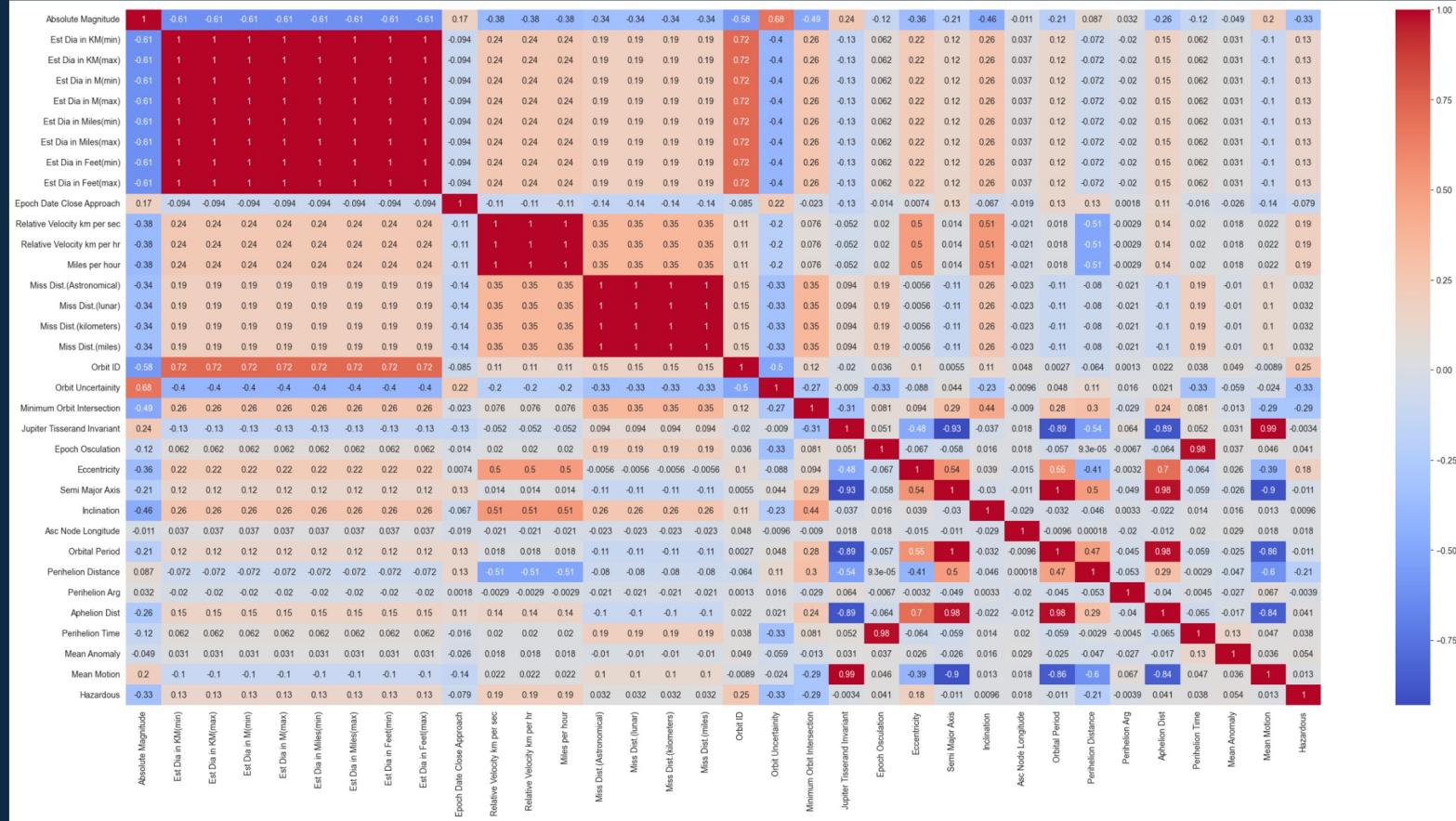


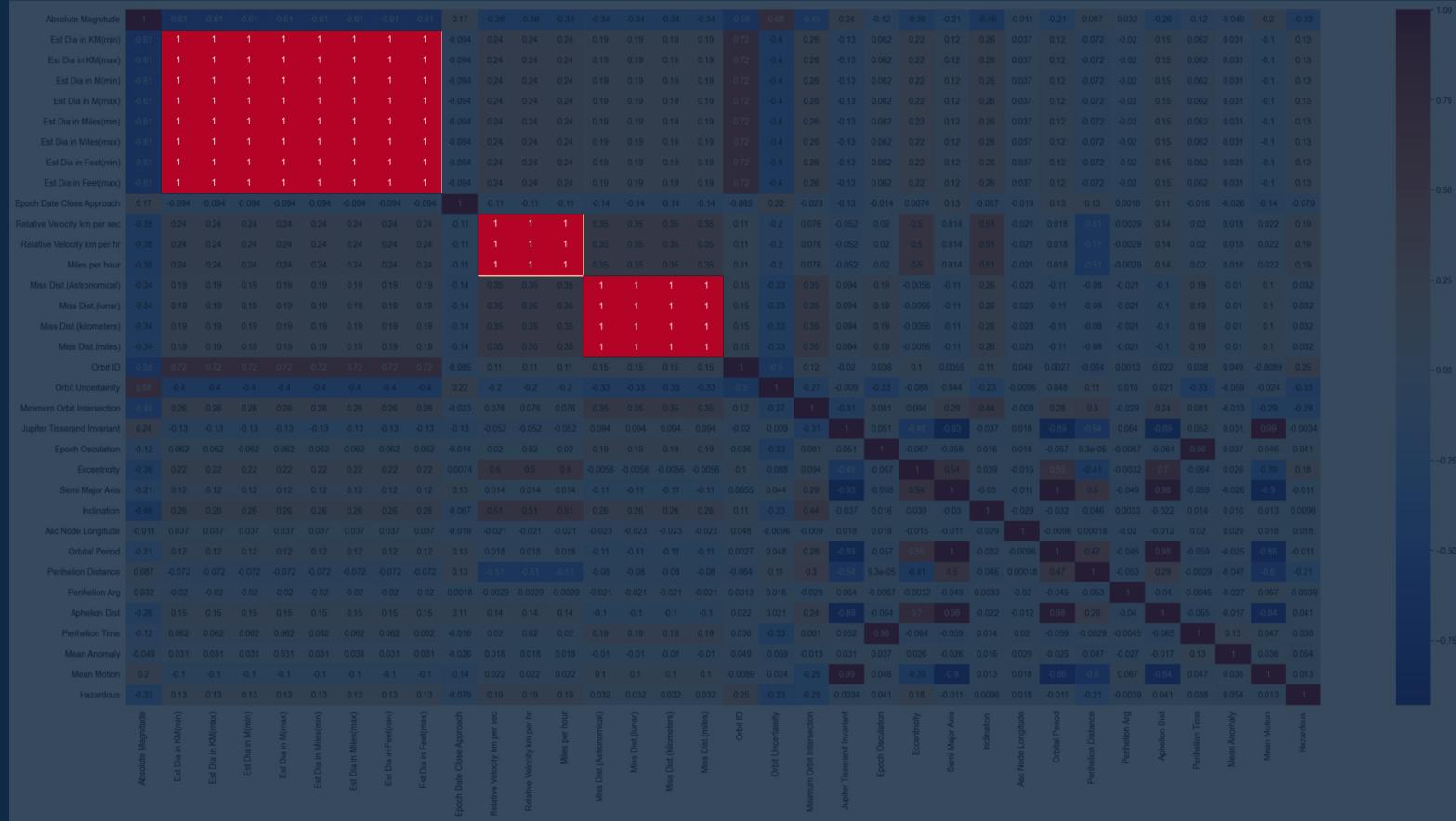
~ 16%
HAZARDOUS TRUE

UNBALANCED DATASET

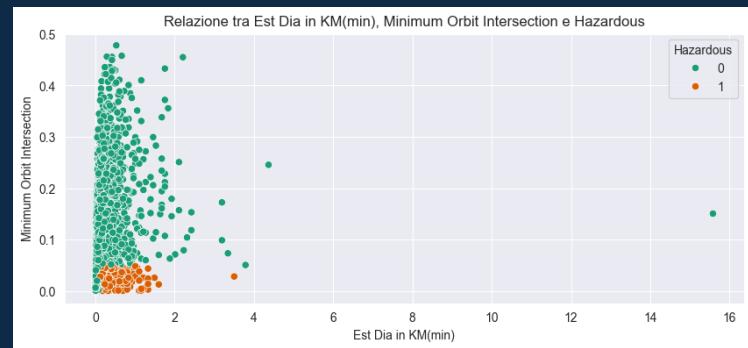
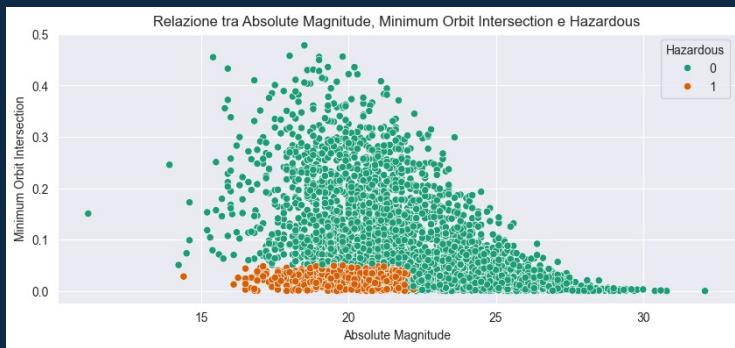




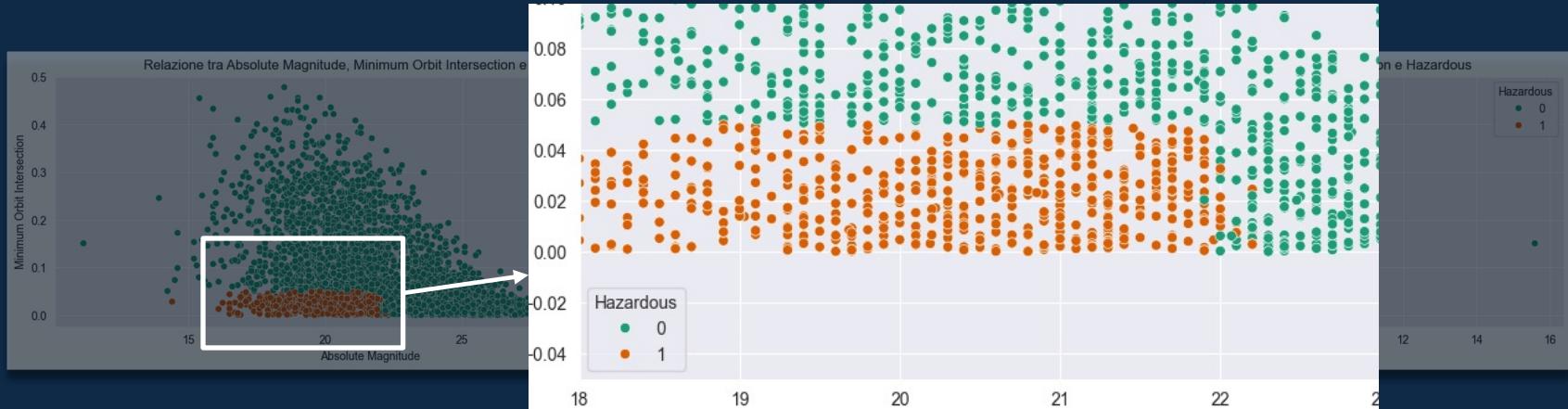




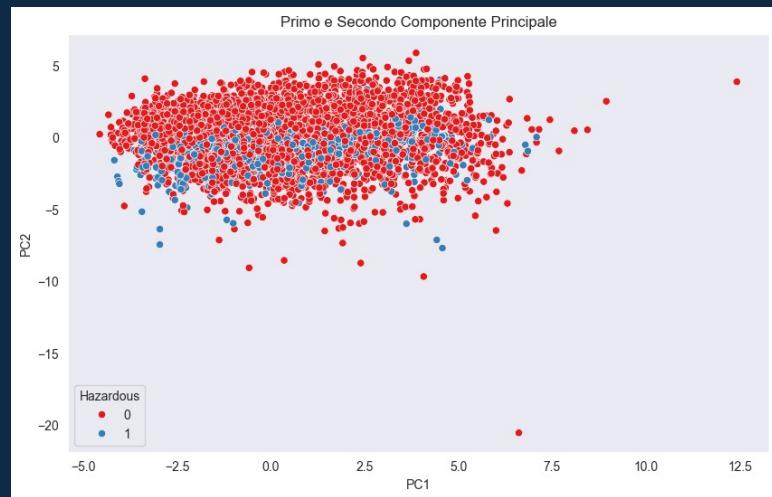
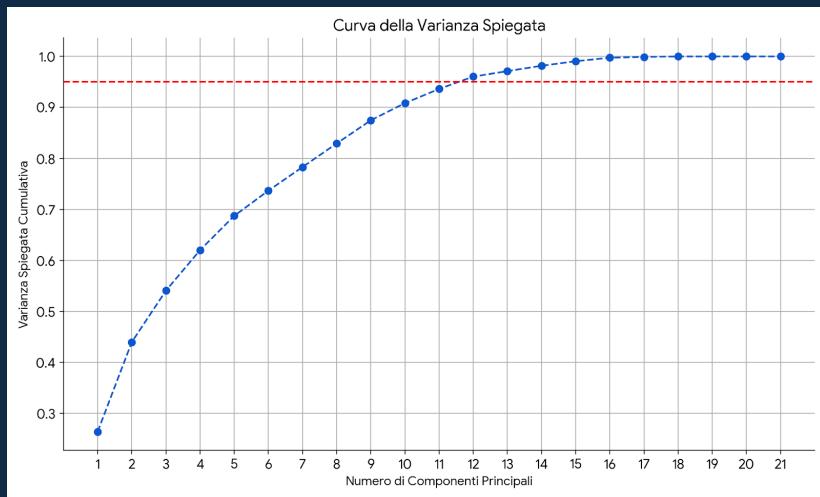
Minimum Orbit Intersection

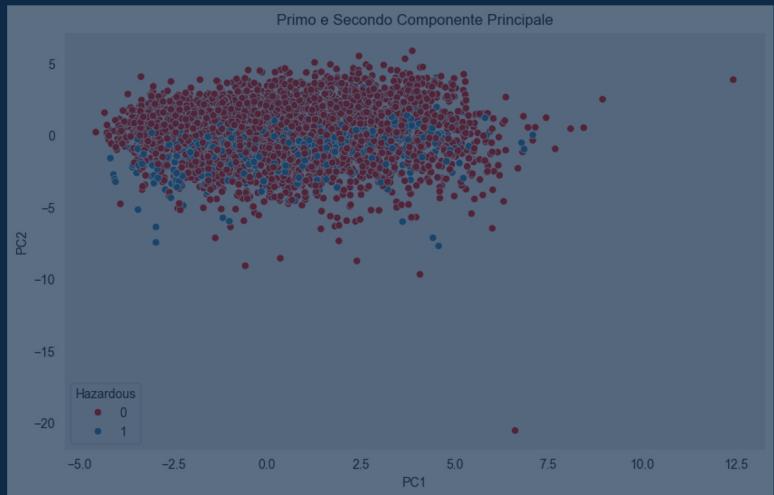
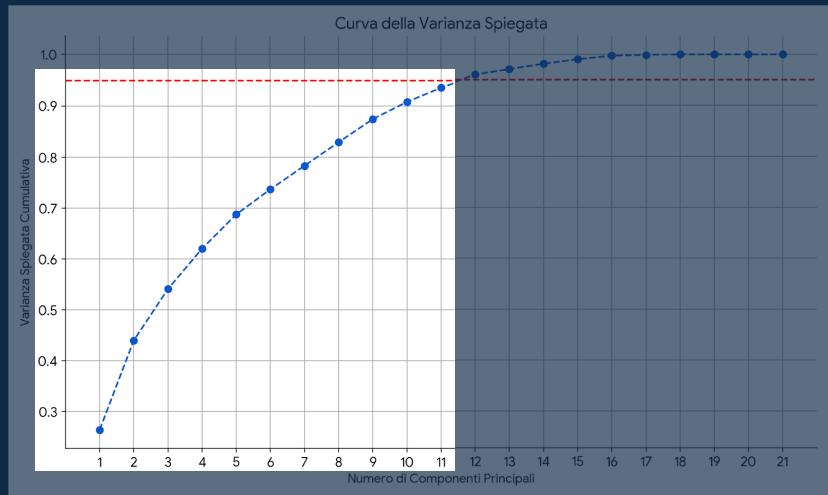


Minimum Orbit Intersection



2 PCA





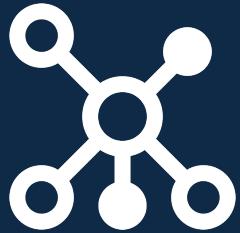
PC	Absolute Magnitude	Est Dia in KM(min)	Epoch Date Close Approach	Relative Velocity km per sec	Miss Dist. (Astronomical)	Orbit ID	Orbit Uncertainty	Minimum Orbit Intersection	Jupiter Tisserand Invariant	Epoch Osculation	Eccentricity	Semi Major Axis	Inclination	Asc Node Longitude	Orbital Period	Perihelion Distance	Perihelion Arg	Aphelion Dist	Perihelion Time	Mean Anomaly	Mean Motion
1	-0.167	0.113	0.050	0.053	-0.012	0.057	-0.037	0.174	-0.406	-0.021	0.251	0.412	0.043	-0.006	0.405	0.191	-0.029	0.405	-0.022	-0.010	-0.392
2	0.415	-0.338	0.147	-0.288	-0.287	-0.319	0.388	-0.201	-0.079	-0.182	-0.109	0.096	-0.278	-0.011	0.094	0.195	0.003	0.057	-0.183	-0.049	-0.103
3	0.045	-0.065	0.051	-0.327	0.071	-0.032	-0.149	0.136	-0.045	0.531	-0.327	0.025	-0.102	0.025	0.015	0.372	-0.035	-0.067	0.531	0.044	-0.087
4	-0.133	0.209	-0.040	-0.310	0.031	0.267	-0.061	0.275	-0.001	-0.377	-0.410	-0.047	0.133	0.022	-0.062	0.414	-0.033	-0.157	-0.383	-0.094	-0.039
5	0.051	-0.323	0.019	0.270	0.353	-0.473	0.157	0.382	-0.025	-0.012	-0.097	-0.018	0.478	-0.173	-0.021	0.101	-0.011	-0.045	-0.024	-0.139	-0.013
6	-0.007	0.080	0.176	-0.043	-0.024	0.108	-0.037	-0.039	0.009	0.072	0.014	0.001	-0.061	-0.481	0.003	-0.019	0.655	0.006	0.026	-0.524	0.007
7	0.008	0.028	0.349	0.055	-0.076	-0.008	0.112	0.104	0.008	0.001	-0.020	-0.007	0.204	0.644	-0.005	0.016	0.583	-0.012	0.024	0.217	0.019
8	-0.016	0.079	0.467	0.004	-0.132	0.077	0.057	0.021	0.007	-0.043	-0.030	-0.025	0.150	-0.546	-0.028	0.015	-0.009	-0.032	0.017	0.650	0.012
9	0.013	0.147	0.698	0.085	-0.027	0.114	0.092	-0.025	0.020	0.076	0.009	-0.038	0.071	0.146	-0.040	-0.057	-0.470	-0.028	0.033	-0.446	0.027
10	0.047	-0.069	0.241	-0.078	0.762	0.046	-0.029	0.094	0.046	-0.123	0.086	0.021	-0.521	0.032	0.029	-0.042	0.060	0.034	-0.109	0.135	0.029
11	0.278	0.384	-0.170	0.319	0.301	0.200	0.465	-0.400	-0.096	0.077	-0.217	0.026	0.116	-0.021	0.013	0.196	0.030	-0.021	0.084	0.038	-0.112
12	0.086	0.365	-0.149	-0.249	-0.084	-0.012	0.540	0.557	0.150	0.094	0.208	0.004	-0.102	-0.042	0.036	-0.176	-0.034	0.049	0.103	-0.010	0.179
13	-0.141	-0.085	0.005	-0.386	0.197	0.033	0.060	-0.326	0.325	0.015	0.003	0.229	0.376	0.000	0.367	0.006	-0.022	0.250	0.001	-0.013	0.420
14	0.146	-0.597	-0.061	0.118	-0.043	0.720	0.188	0.189	0.015	0.045	0.060	0.013	0.058	-0.008	0.021	0.006	-0.020	0.012	0.050	-0.013	0.021
15	0.636	0.173	0.005	0.277	-0.107	0.013	-0.427	0.215	0.207	-0.038	-0.130	0.155	-0.040	-0.004	0.245	0.062	-0.011	0.155	-0.050	0.010	0.267
16	0.493	0.072	-0.020	-0.461	0.180	0.079	-0.202	-0.054	-0.235	-0.009	0.300	-0.109	0.386	0.014	-0.183	-0.222	-0.007	-0.064	-0.015	-0.000	-0.252
17	0.027	0.028	0.001	0.067	-0.018	-0.030	-0.028	-0.068	0.155	-0.000	0.654	-0.042	0.001	0.001	-0.124	0.667	-0.002	-0.216	0.007	-0.007	0.163
18	0.003	0.001	0.001	-0.003	0.002	-0.002	-0.011	-0.005	-0.002	-0.704	-0.005	0.003	0.005	0.000	0.003	-0.004	-0.003	0.005	0.707	-0.064	0.007
19	0.002	0.005	0.001	0.011	-0.005	-0.004	-0.002	-0.001	0.459	-0.004	0.058	-0.301	0.018	0.000	0.577	-0.056	-0.000	-0.317	0.003	0.000	-0.508
20	0.005	0.001	0.000	0.005	-0.001	0.001	-0.002	0.002	0.595	-0.003	-0.021	0.319	0.029	0.000	-0.491	0.027	-0.001	0.344	0.002	-0.000	-0.428
21	-0.000	0.000	-0.000	0.000	-0.000	-0.000	-0.000	0.000	0.000	-0.000	0.000	0.730	0.000	-0.000	0.000	-0.169	-0.000	-0.662	0.000	-0.000	-0.000

PC	Absolute Magnitude	Est Dia in KM(min)	Epoch Date Close Approach	Relative Velocity km per sec	Miss Dist. (Astronomical)	Orbit ID	Orbit Uncertainty	Minimum Orbit Intersection	Jupiter Tisserand Invariant	Epoch Osculation	Eccentricity	Semi Major Axis	Inclination	Asc Node Longitude	Orbital Period	Perihelion Distance	Perihelion Arg	Aphelion Dist	Perihelion Time	Mean Anomaly	Mean Motion
1	-0.167	0.113	0.050	0.053	-0.012	0.057	-0.037	0.174	-0.406	-0.021	0.251	0.412	0.043	-0.006	0.405	0.191	-0.029	0.405	-0.022	-0.010	-0.392
2	0.415	-0.338	0.147	-0.288	-0.287	-0.319	0.388	-0.201	-0.079	-0.182	-0.109	0.096	-0.278	-0.011	0.094	0.195	0.003	0.057	-0.183	-0.049	-0.103
3	0.045	-0.065	0.051	-0.327	0.071	-0.032	-0.149	0.136	-0.045	0.531	-0.327	0.025	-0.102	0.025	0.015	0.372	-0.035	-0.067	0.531	0.044	-0.087
4	-0.133	0.209	-0.040	-0.310	0.031	0.267	-0.061	0.275	-0.001	-0.377	-0.410	-0.047	0.133	0.022	-0.062	0.414	-0.033	-0.157	-0.383	-0.094	-0.039
5	0.051	-0.323	0.019	0.270	0.353	-0.473	0.157	0.382	-0.025	-0.012	-0.097	-0.018	0.478	-0.173	-0.021	0.101	-0.011	-0.045	-0.024	-0.139	-0.013
6	-0.007	0.080	0.176	-0.043	-0.024	0.108	-0.037	-0.039	0.009	0.072	0.014	0.001	-0.061	-0.481	0.003	-0.019	0.655	0.006	0.026	-0.524	0.007
7	0.008	0.028	0.349	0.055	-0.076	-0.008	0.112	0.104	0.008	0.001	-0.020	-0.007	0.204	0.644	-0.005	0.016	0.583	-0.012	0.024	0.217	0.019
8	-0.016	0.079	0.467	0.004	-0.132	0.077	0.057	0.021	0.007	-0.043	-0.030	-0.025	0.150	-0.546	-0.028	0.015	-0.009	-0.032	0.017	0.650	0.012
9	0.013	0.147	0.698	0.085	-0.027	0.114	0.092	-0.025	0.020	0.076	0.009	-0.038	0.071	0.146	-0.040	-0.057	-0.470	-0.028	0.033	-0.446	0.027
10	0.047	-0.069	0.241	-0.078	0.762	0.046	-0.029	0.094	0.046	-0.123	0.086	0.021	-0.521	0.032	0.029	-0.042	0.060	0.034	-0.109	0.135	0.029
11	0.278	0.384	-0.170	0.319	0.301	0.200	0.465	-0.400	-0.096	0.077	-0.217	0.026	0.116	-0.021	0.013	0.196	0.030	-0.021	0.084	0.038	-0.112
12	0.086	0.365	-0.149	-0.249	-0.084	-0.012	0.540	0.557	0.150	0.094	0.208	0.004	-0.102	-0.042	0.036	-0.176	-0.034	0.049	0.103	-0.010	0.179
13	-0.141	-0.085	0.005	-0.386	0.197	0.033	0.060	-0.326	0.325	0.015	0.003	0.229	0.376	0.000	0.367	0.006	-0.022	0.250	0.001	-0.013	0.420
14	0.146	-0.597	-0.061	0.118	-0.043	0.720	0.188	0.189	0.015	0.045	0.060	0.013	0.058	-0.008	0.021	0.006	-0.020	0.012	0.050	-0.013	0.021
15	0.636	0.173	0.005	0.277	-0.107	0.013	-0.427	0.215	0.207	-0.038	-0.130	0.155	-0.040	-0.004	0.245	0.062	-0.011	0.155	-0.050	0.010	0.267
16	0.493	0.072	-0.020	-0.461	0.180	0.079	-0.202	-0.054	-0.235	-0.009	0.300	-0.109	0.386	0.014	-0.183	-0.222	-0.007	-0.064	-0.015	-0.000	-0.252
17	0.027	0.028	0.001	0.067	-0.018	-0.030	-0.028	-0.068	0.155	-0.000	0.654	-0.042	0.001	0.001	-0.124	0.667	-0.002	-0.216	0.007	-0.007	0.163
18	0.003	0.001	0.001	-0.003	0.002	-0.002	-0.011	-0.005	-0.002	-0.704	-0.005	0.003	0.005	0.000	0.003	-0.004	-0.003	0.005	0.707	-0.064	0.007
19	0.002	0.005	0.001	0.011	-0.005	-0.004	-0.002	-0.001	0.459	-0.004	0.058	-0.301	0.018	0.000	0.577	-0.056	-0.000	-0.317	0.003	0.000	-0.508
20	0.005	0.001	0.000	0.005	-0.001	0.001	-0.002	0.002	0.595	-0.003	-0.021	0.319	0.029	0.000	-0.491	0.027	-0.001	0.344	0.002	-0.000	-0.428
21	-0.000	0.000	-0.000	0.000	-0.000	-0.000	-0.000	0.000	0.000	-0.000	0.000	0.730	0.000	-0.000	0.000	-0.169	-0.000	-0.662	0.000	-0.000	-0.000

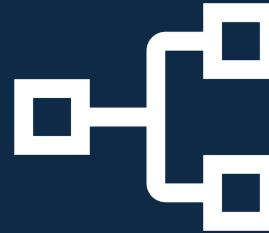
3

Machine Learning

MODELLI UTILIZZATI



Deep Neural Network



Decision Tree

DEEP NEURAL NETWORK

SUPERVISIONATO

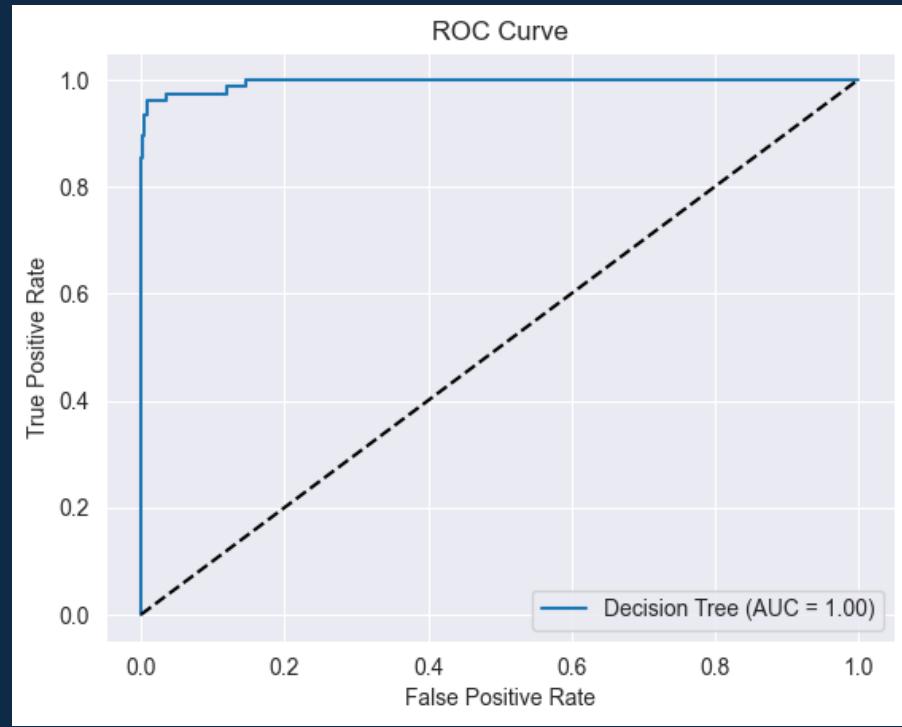
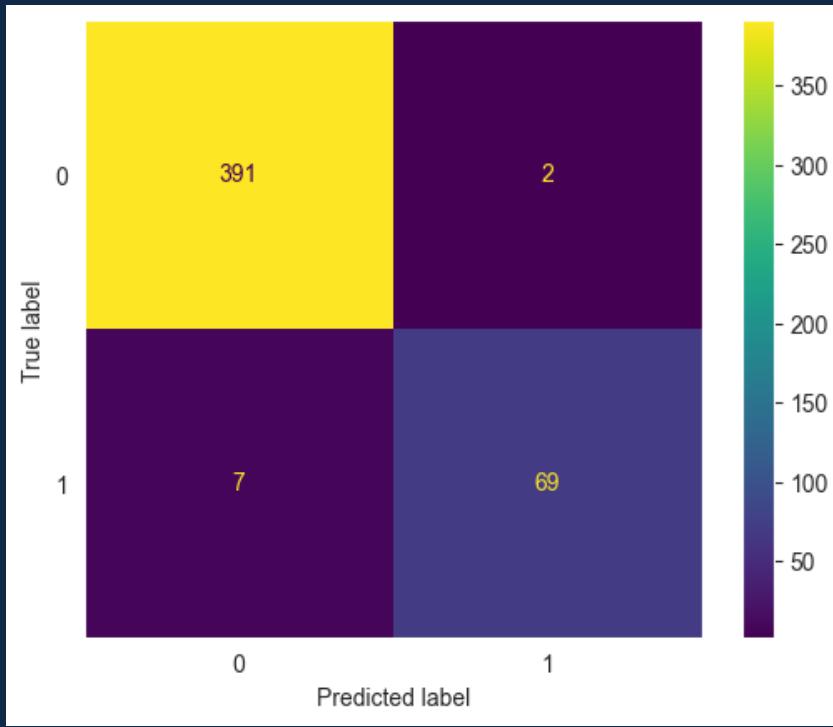
In linea con gli obiettivi

F1 SCORE

Parametro di addestramento

REPRESENTATION LEARNING





DECISION TREE

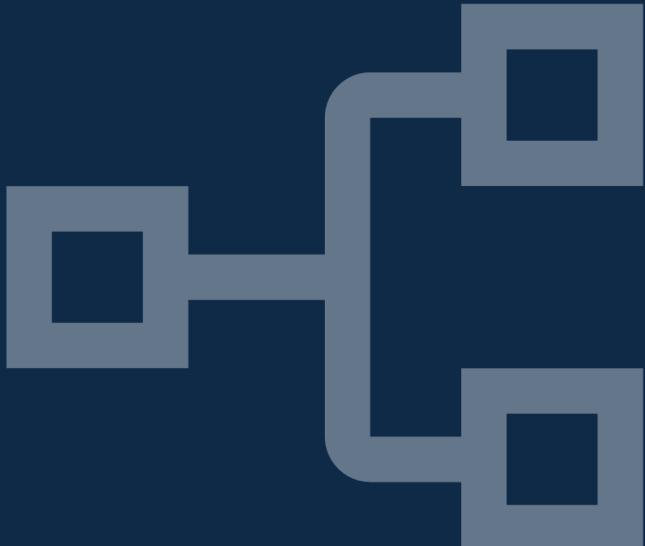
SUPERVISIONATO

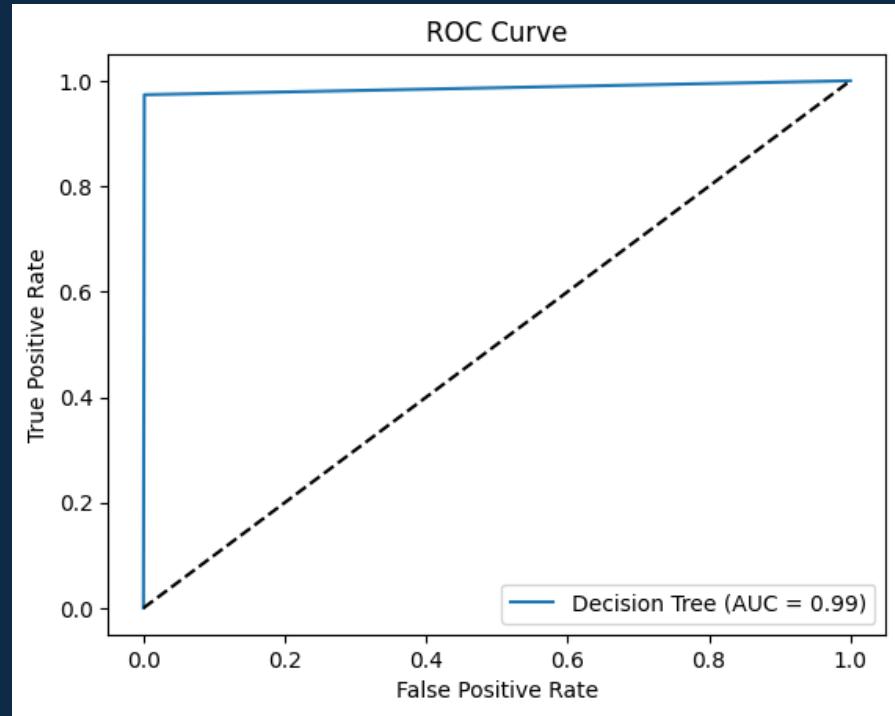
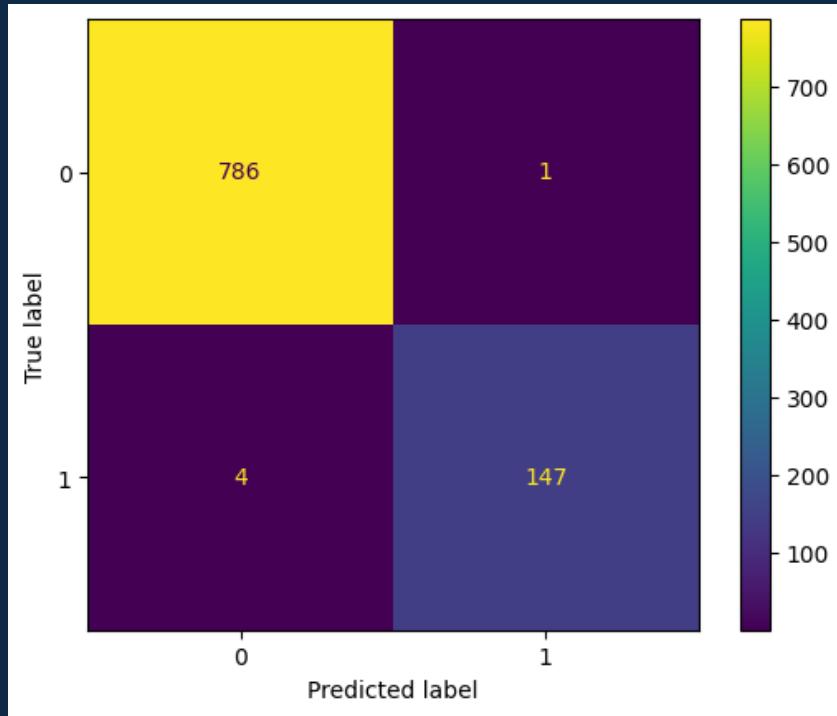
In linea con gli obiettivi

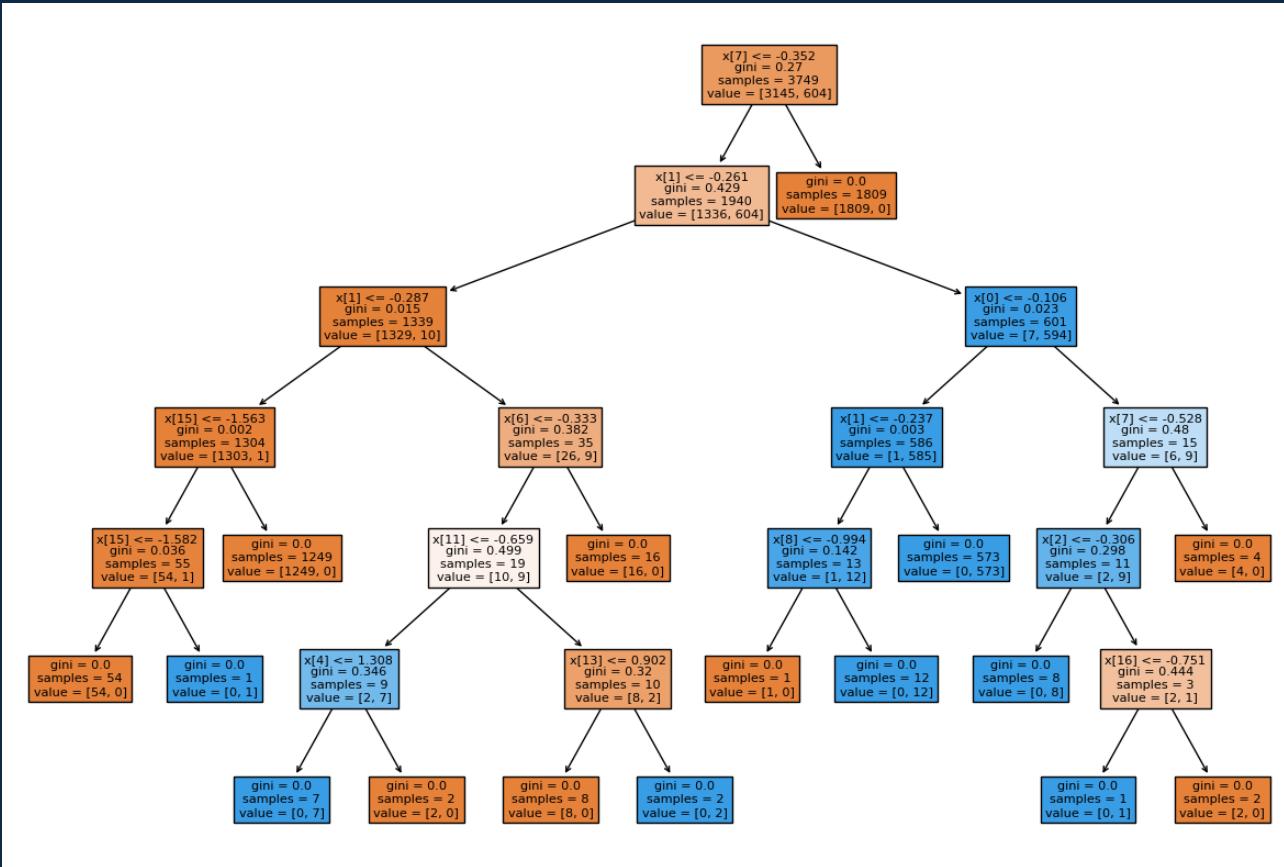
F1 SCORE

Su dataset voluminosi

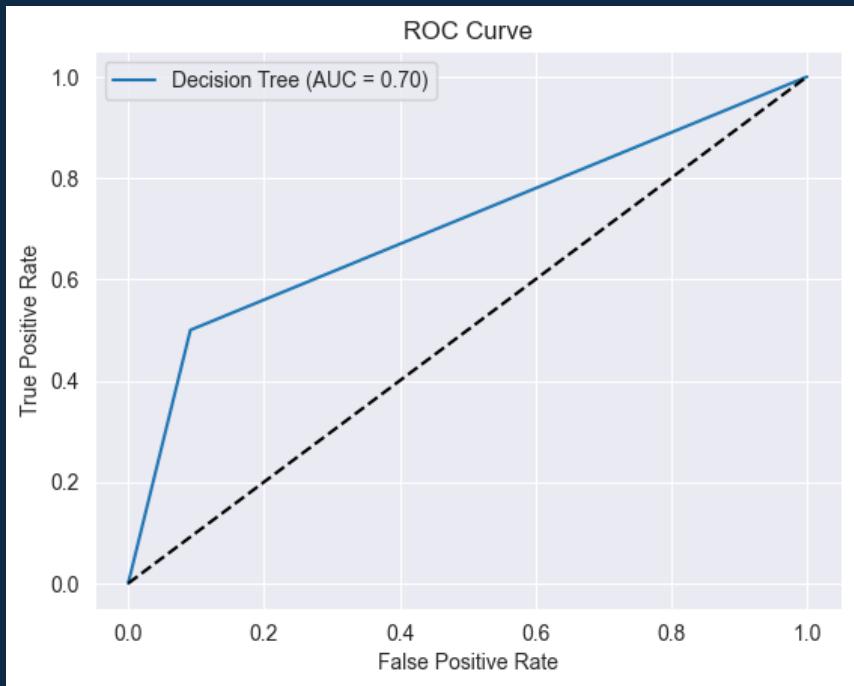
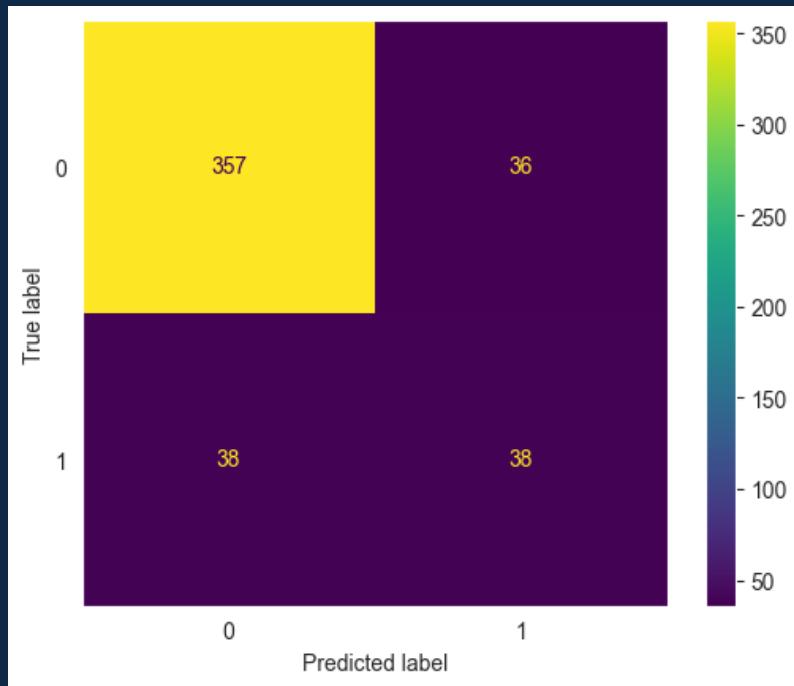
TRADITIONAL ML



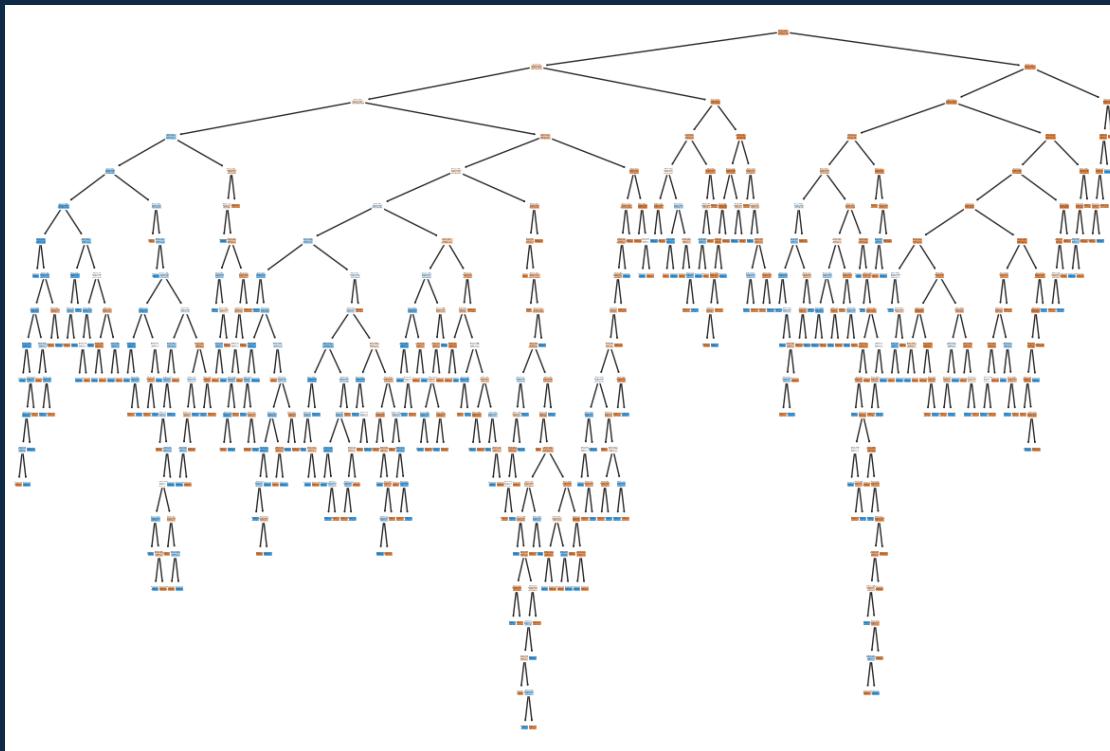




DECISION TREE + PCA



DECISION TREE + PCA



4

Sporcare i Dataset

Null

Inconsistent

Outliers

Drop features

0	4.48588	7.85937	5.14026
1	1.77006	7.58746	9.59453
2	8.30209	8.95255	2.52817
3	1.55749	7.22025	1.24211
4	5.81451	8.52578	7.65999
5	1.12598	1.60386	9.7734
6	7.61767	1.15726	4.77375
7	6.83522	3.59158	3.36712
8	2.58848	7.93221	7.74431
9	8.45246	2.63733	8.16675

Null

Inconsistent

Outliers

Drop features

0	4.48588	7.85937	5.14026
1	1.77006	7.58746	9.59453
2	8.30209	8.95255	2.52817
3	1.55749	7.22025	1.24211
4	5.81451	8.52578	7.65999
5	1.12598	1.60386	9.7734
6	7.61767	1.15726	4.77375
7	6.83522	3.59158	3.36712
8	2.58848	7.93221	7.74431
9	8.45246	2.63733	8.16675

0	4.48588	7.85937	nan
1	nan	nan	nan
2	8.30209	8.95255	nan
3	1.55749	7.22025	1.24211
4	5.81451	nan	7.65999
5	nan	1.60386	9.7734
6	nan	nan	nan
7	nan	nan	nan
8	2.58848	7.93221	7.74431
9	nan	nan	8.16675

Null

Inconsistent

Outliers

Drop features

0	4.48588	7.85937	5.14026
1	1.77006	7.58746	9.59453
2	8.30209	8.95255	2.52817
3	1.55749	7.22025	1.24211
4	5.81451	8.52578	7.65999
5	1.12598	1.60386	9.7734
6	7.61767	1.15726	4.77375
7	6.83522	3.59158	3.36712
8	2.58848	7.93221	7.74431
9	8.45246	2.63733	8.16675

0	4.48588	17.275	-2.74385
1	1.77006	7.58746	40.7302
2	-1.73676	-3.13565	74.0342
3	62.0458	41.2071	1.24211
4	-2.3735	8.52578	7.65999
5	-3.01534	1.60386	55.3104
6	7.61767	1.15726	4.77375
7	6.83522	33.4323	-1.3759
8	-1.44999	-1.67514	7.74431
9	8.45246	2.63733	8.16675

Null

Inconsistent

Outliers

Drop features

0	4.48588	7.85937	5.14026
1	1.77006	7.58746	9.59453
2	8.30209	8.95255	2.52817
3	1.55749	7.22025	1.24211
4	5.81451	8.52578	7.65999
5	1.12598	1.60386	9.7734
6	7.61767	1.15726	4.77375
7	6.83522	3.59158	3.36712
8	2.58848	7.93221	7.74431
9	8.45246	2.63733	8.16675

0	14.4666	7.85937	13.9618
1	0.34673	7.58746	9.59453
2	8.30209	8.95255	-0.891246
3	1.55749	14.1388	1.24211
4	14.66662	13.7924	7.65999
5	1.12598	1.60386	9.7734
6	10.8217	2.48277	10.3785
7	6.83522	12.7161	0.762408
8	2.58848	7.93221	7.74431
9	12.2982	1.05325	10.4094

Null

Inconsistent

Outliers

Drop features

0	4.48588	7.85937	5.14026
1	1.77006	7.58746	9.59453
2	8.30209	8.95255	2.52817
3	1.55749	7.22025	1.24211
4	5.81451	8.52578	7.65999
5	1.12598	1.60386	9.7734
6	7.61767	1.15726	4.77375
7	6.83522	3.59158	3.36712
8	2.58848	7.93221	7.74431
9	8.45246	2.63733	8.16675

0	7.85937	5.14026
1	7.58746	9.59453
2	8.95255	2.52817
3	7.22025	1.24211
4	8.52578	7.65999
5	1.60386	9.7734
6	1.15726	4.77375
7	3.59158	3.36712
8	7.93221	7.74431
9	2.63733	8.16675

~ 100 dataset

Null

Inconsistent
Outliers

Drop features

Dataset sporcati

Dataset sporcati →

Pulizia

Outlier / inconsistent → Null



SHAP

COS'È

**metodo che fornisce spiegazioni coerenti e
interpretabili sui modelli di machine learning.**

OBIETTIVO

**Aiuta a comprendere come e perché un modello
di machine learning prende determinate
decisioni, rendendo i modelli complessi meno
opachi.**

BLACK BOX

Contributo marginale

Per ogni caratteristica, calcola il **contributo marginale**, che è la differenza nella predizione del modello quando la caratteristica è inclusa rispetto a quando non lo è.

Media
combi
I valori di Shapley per una
prendendo la media dei
possibili combinazioni di

BLACK BOX

marginale

colà il contributo marginale,
dizione del modello quando
rispetto a quando non lo è.

Media sulle
combinazioni

I valori di Shapley per una caratteristica sono ottenuti
prendendo la media dei contributi marginali su tutte le
possibili combinazioni delle altre caratteristiche.

I valori di Shapley sono ad
somma dei valori di Shapley
uguale alla differenza tra la
un esempio e la media di
esempi

BLACK BOX

o sulle
nazioni

caratteristica sono ottenuti
tributi marginali su tutte le
elle altre caratteristiche.

Additività

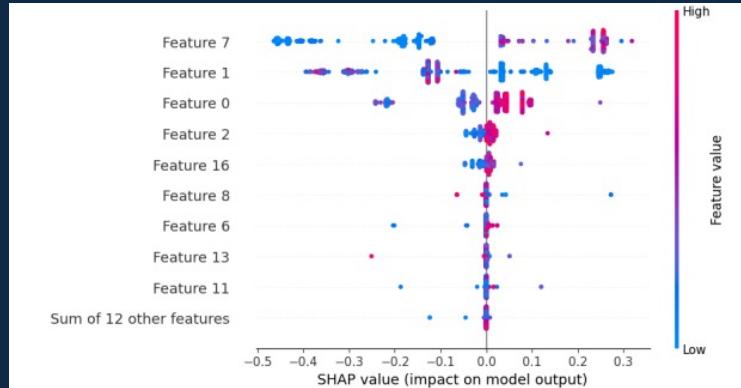
I valori di Shapley sono additivi, il che significa che la somma dei valori di Shapley di tutte le caratteristiche è uguale alla differenza tra la predizione del modello per un esempio e la media delle predizioni su tutti gli esempi.

BEESWARM PLOT

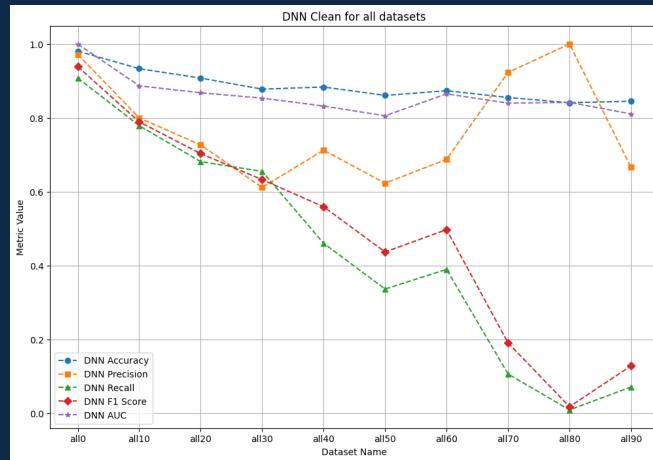
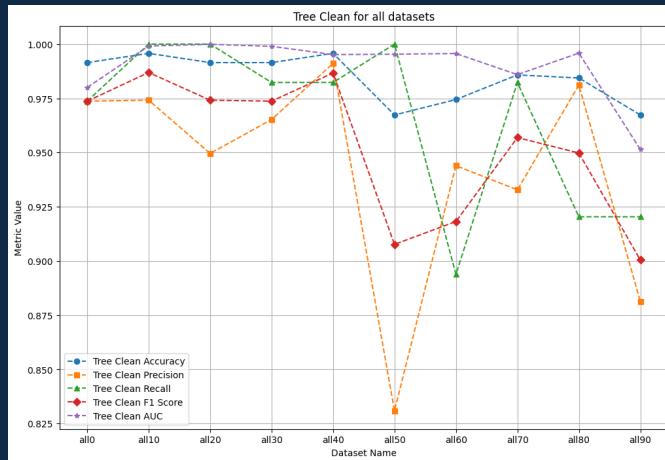
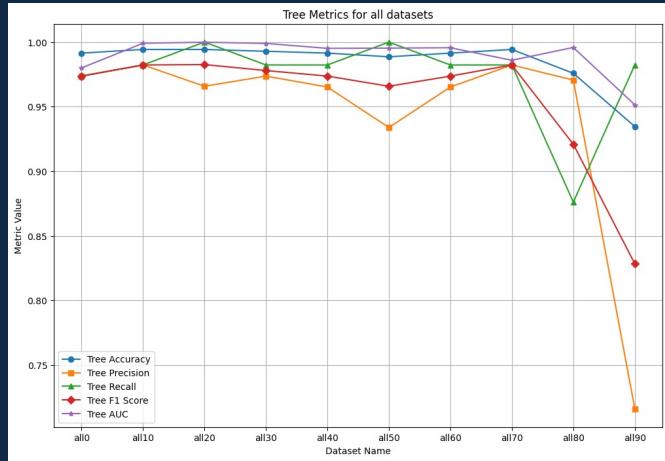
tipo di visualizzazione che aiuta a comprendere come i valori SHAP sono distribuiti per ciascuna caratteristica (feature) nel dataset.

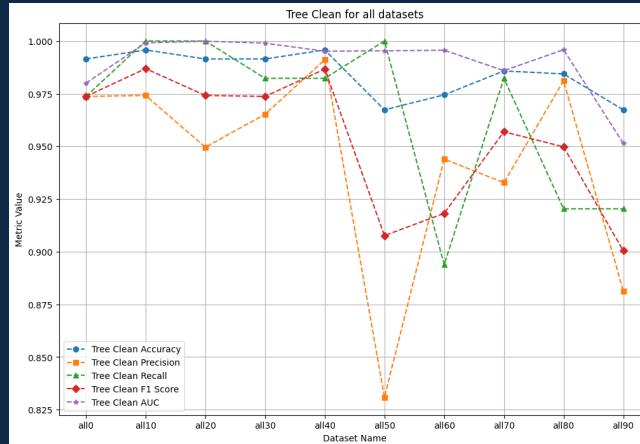
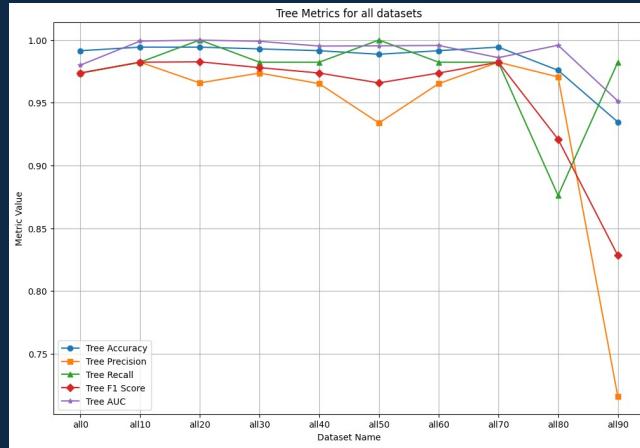


Questo tipo di plot è utile per visualizzare la distribuzione dei contributi delle caratteristiche in modo più dettagliato rispetto ad altri tipi di grafici come i bar plot.



Analisi dei dati raccolti



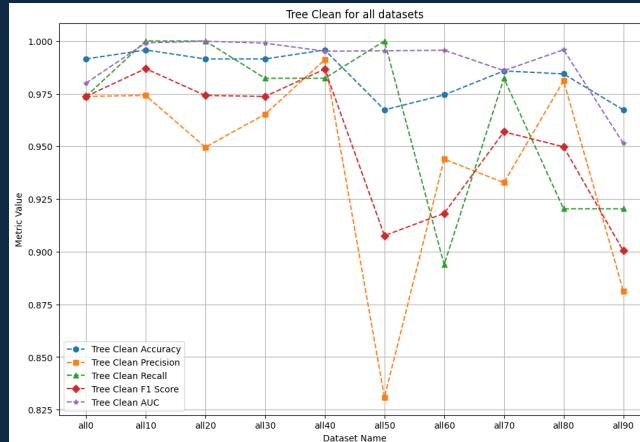
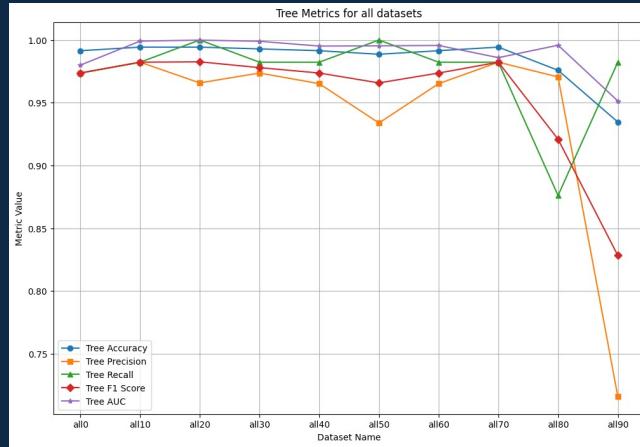


Perdita di informazioni

Imputazione inefficiente

Riduzione della complessità

Impatto su features importanti

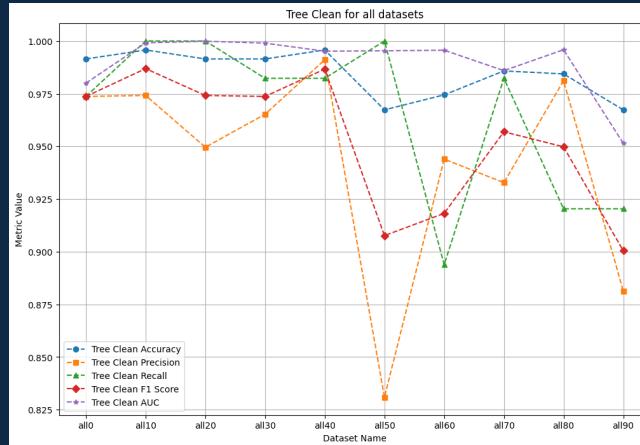
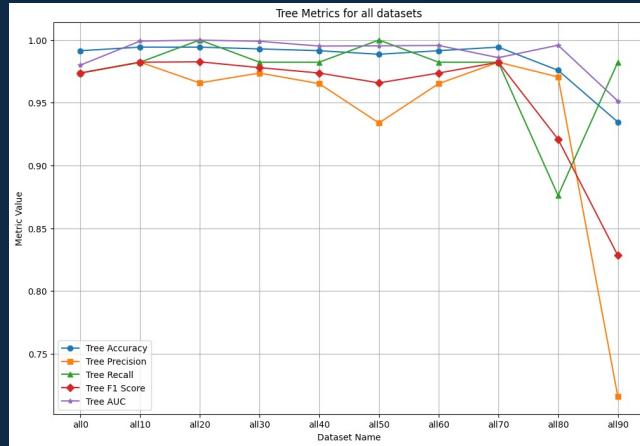


Perdita di informazioni

Imputazione inefficiente

Riduzione della complessità

Impatto su features importanti

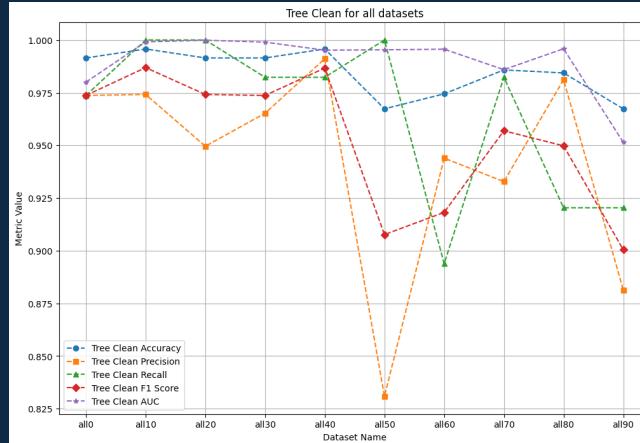
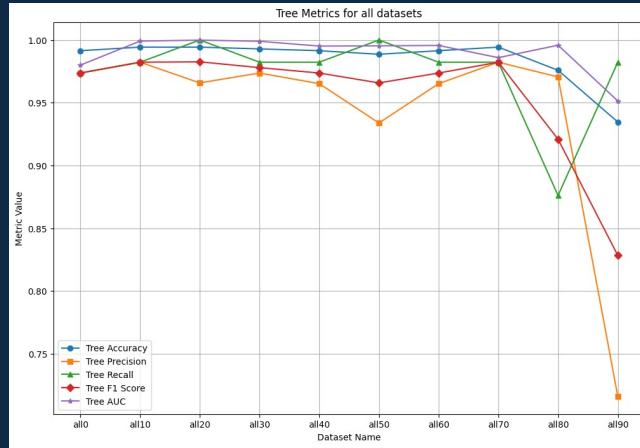


Perdita di informazioni

Imputazione inefficace

Riduzione della complessità

Impatto su features importanti

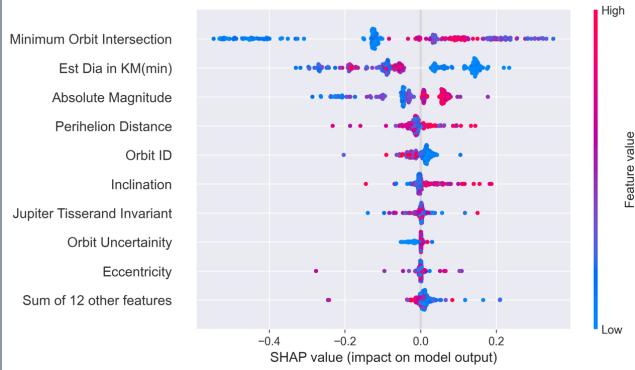
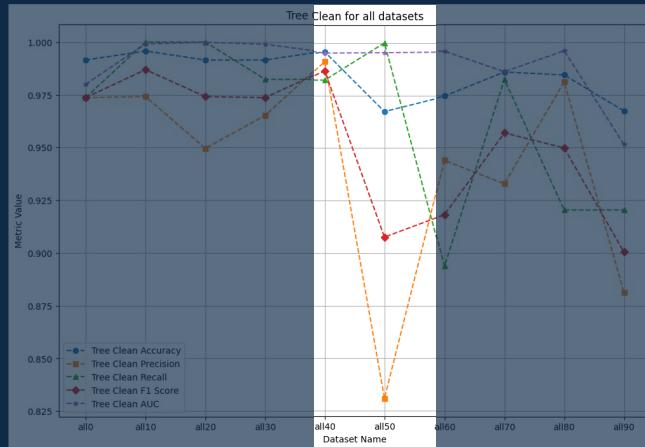
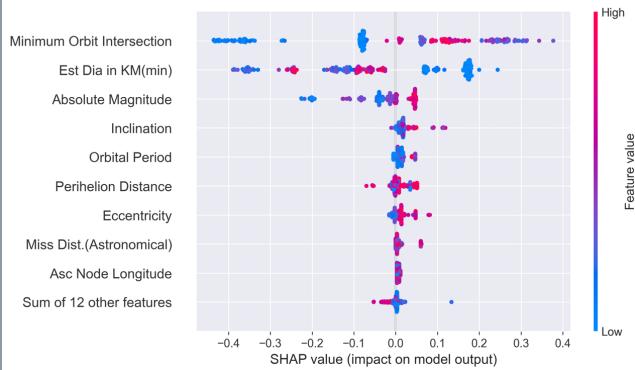
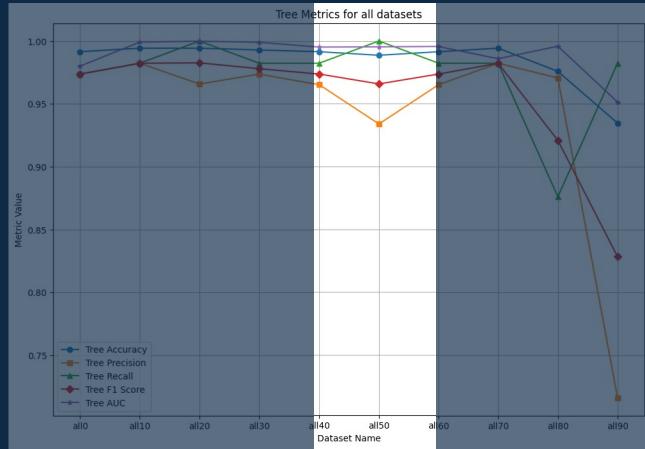


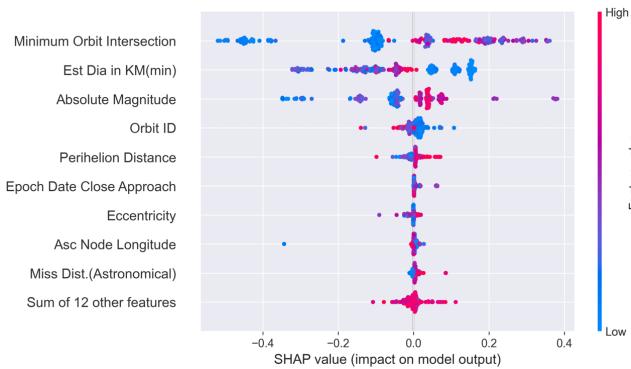
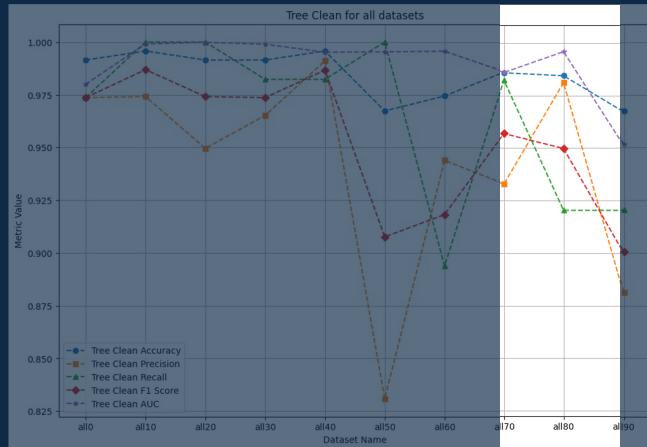
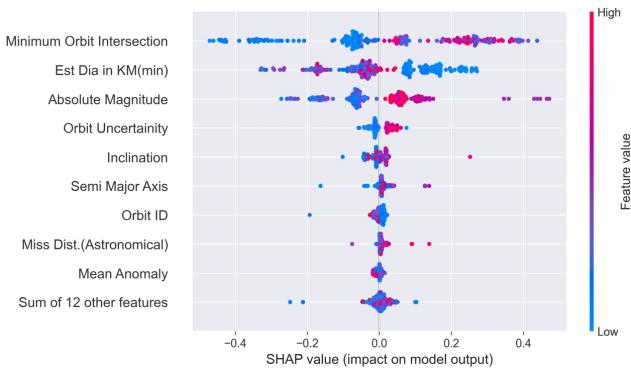
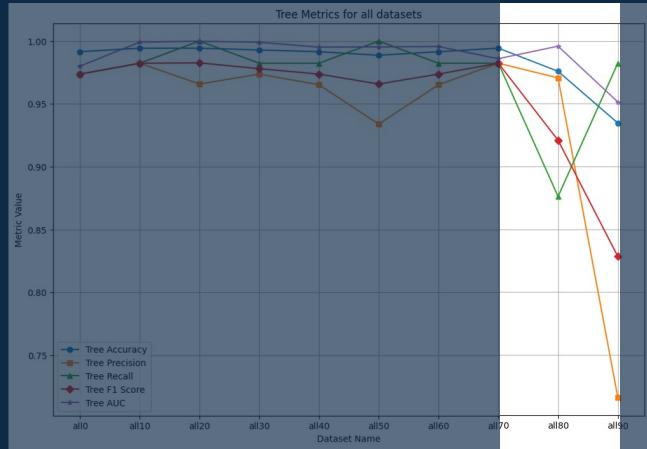
Perdita di informazioni

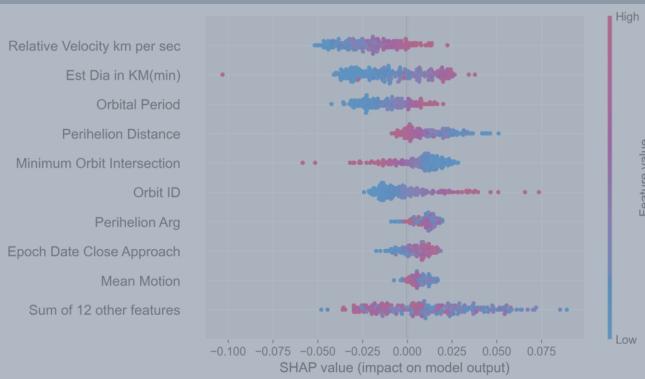
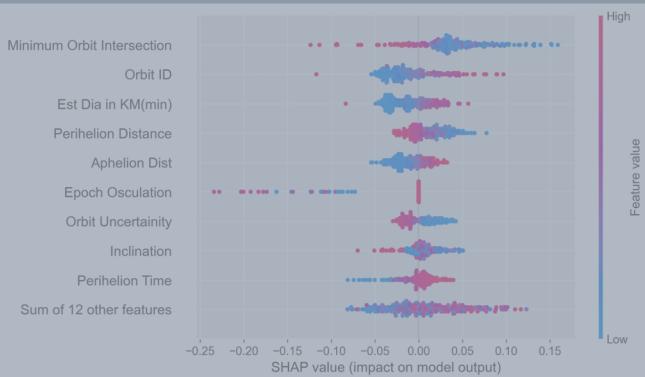
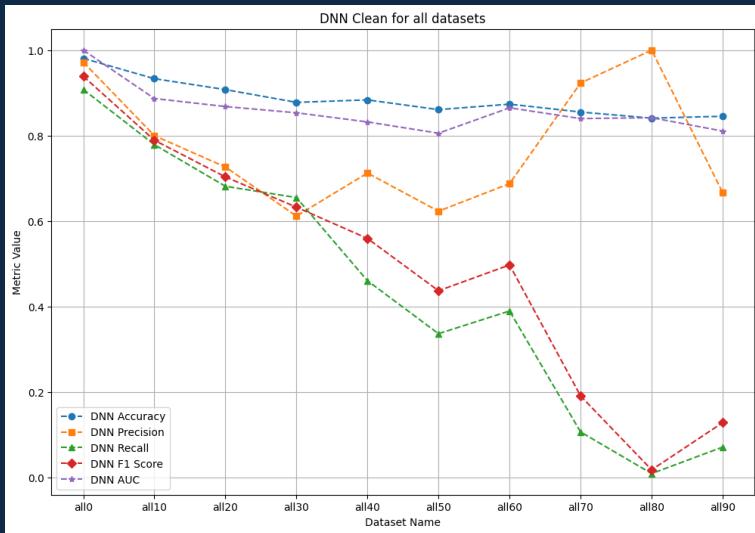
Imputazione inefficiente

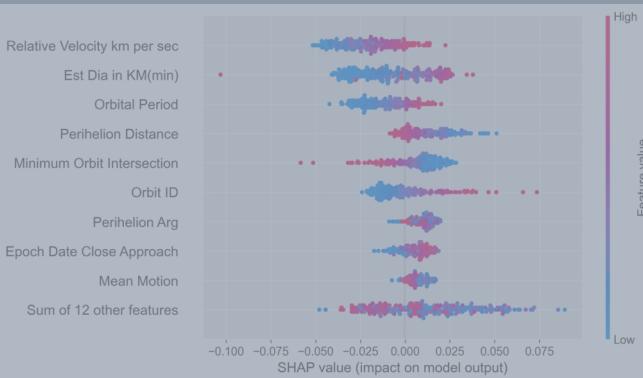
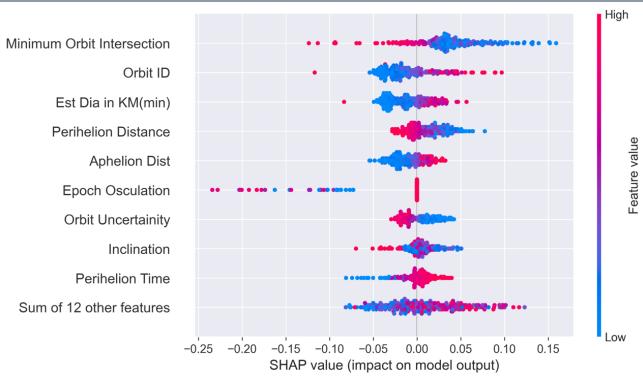
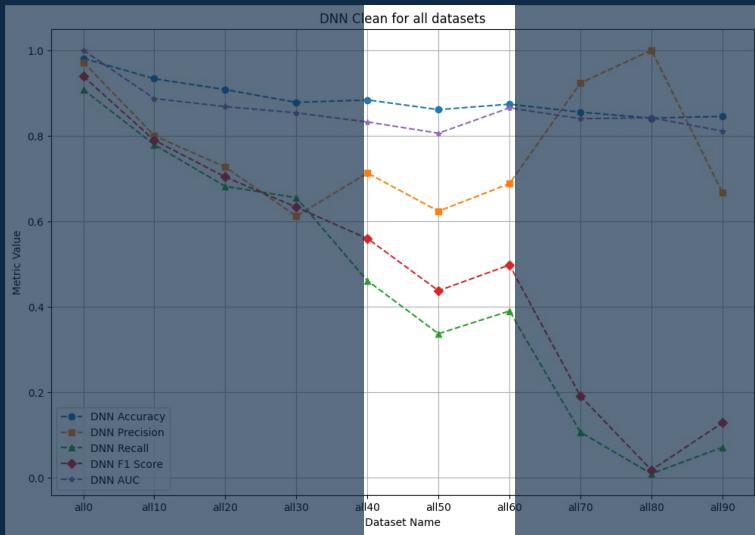
Riduzione della complessità

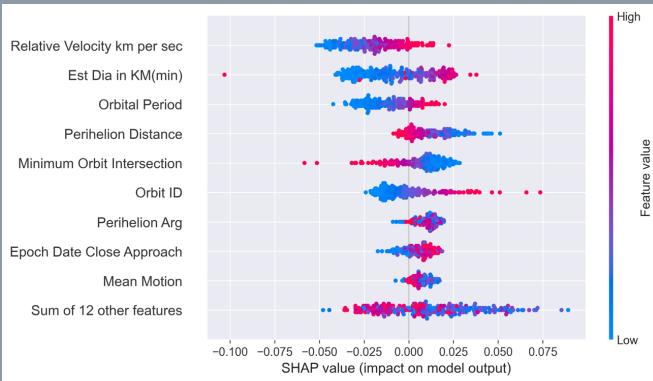
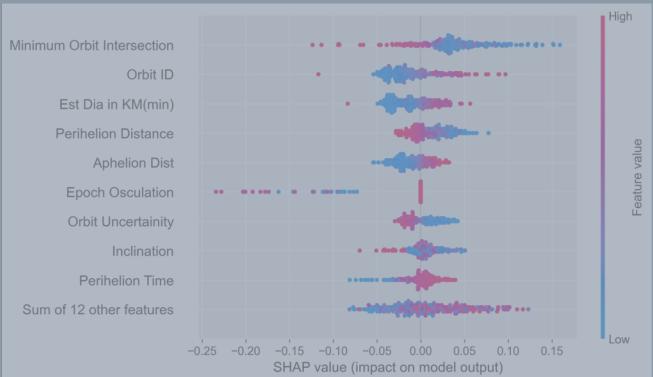
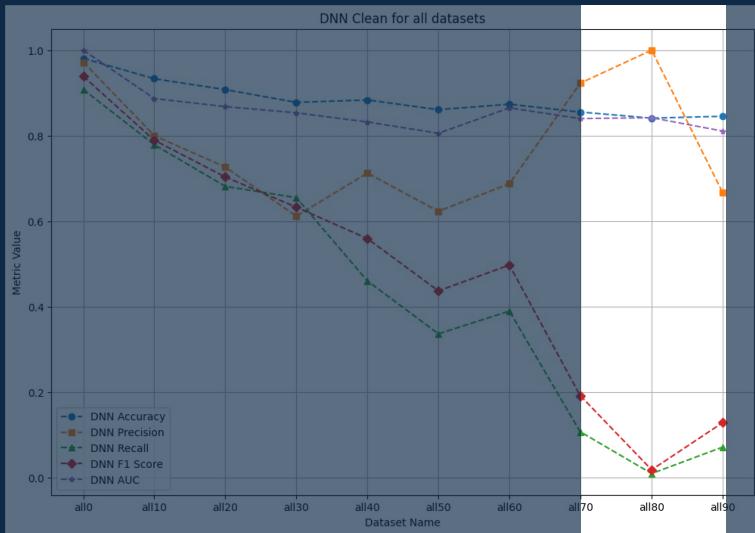
Impatto su features importanti

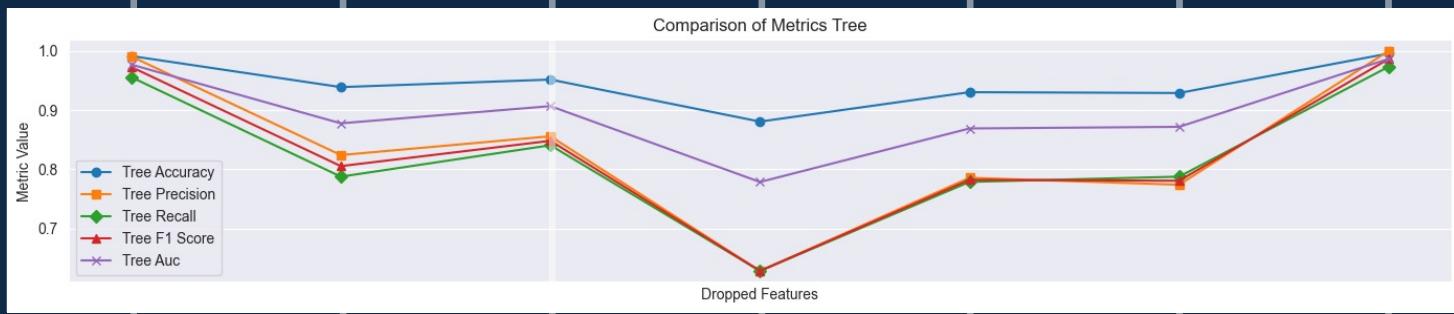
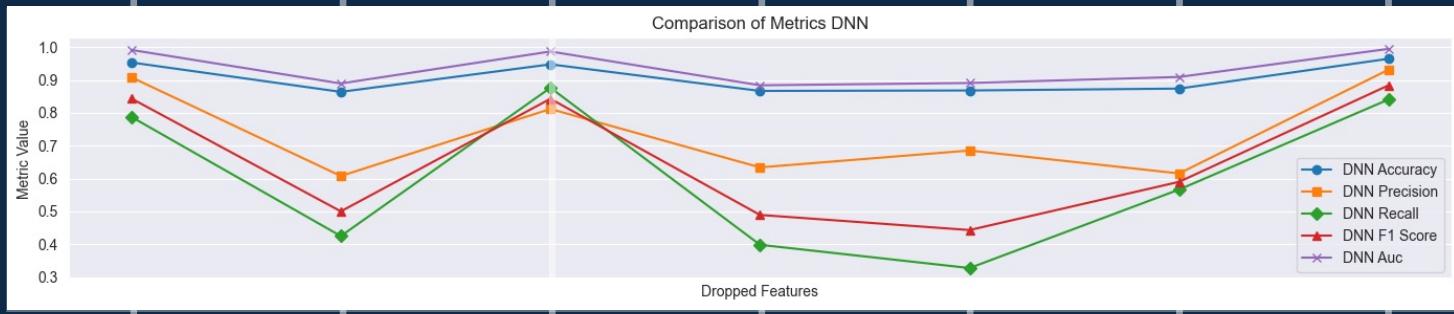












DROP

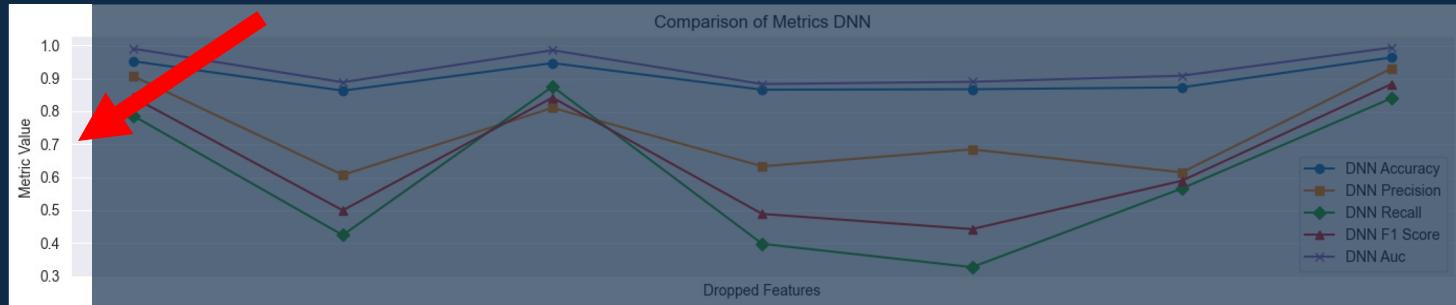
Ab. Mag.
+
min.orb.

Ab. Mag.
+
Est. Dia.

Ab. Mag.
+
min.orb.
+
Est. Dia.

min.orb.

min.orb.
+
Est. Dia.
Est. Dia.



DROP

Ab. Mag.

Ab. Mag. + min.orb.

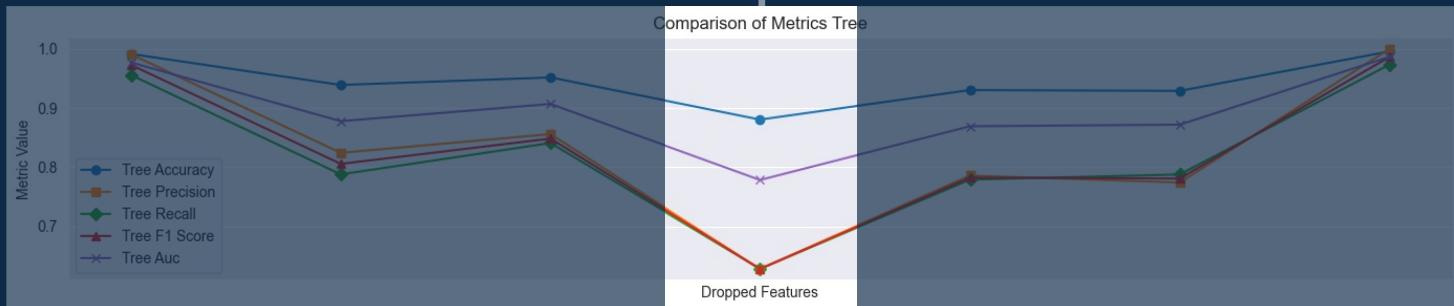
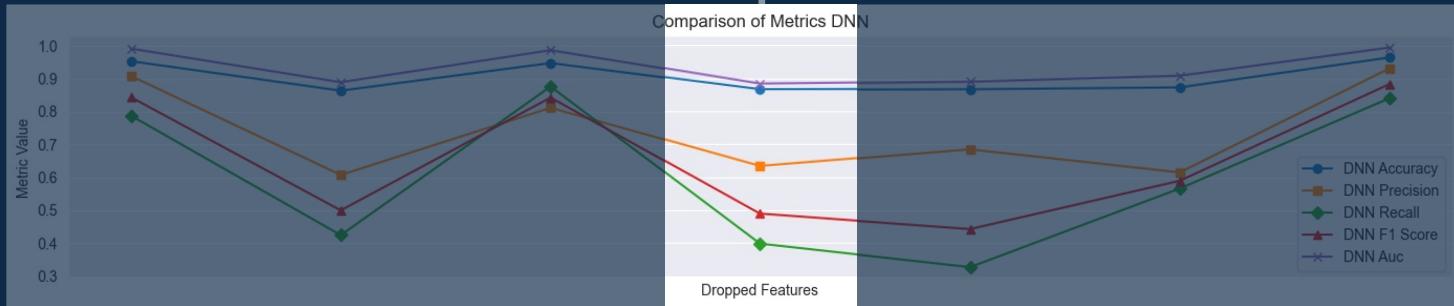
Ab. Mag. + Est. Dia.

Ab. Mag. + min.orb. + Est. Dia.

min.orb.

min.orb. + Est. Dia.

Est. Dia.



DROP

Ab. Mag.

Ab. Mag.
+
min.orb.

Ab. Mag.
+
Est. Dia.

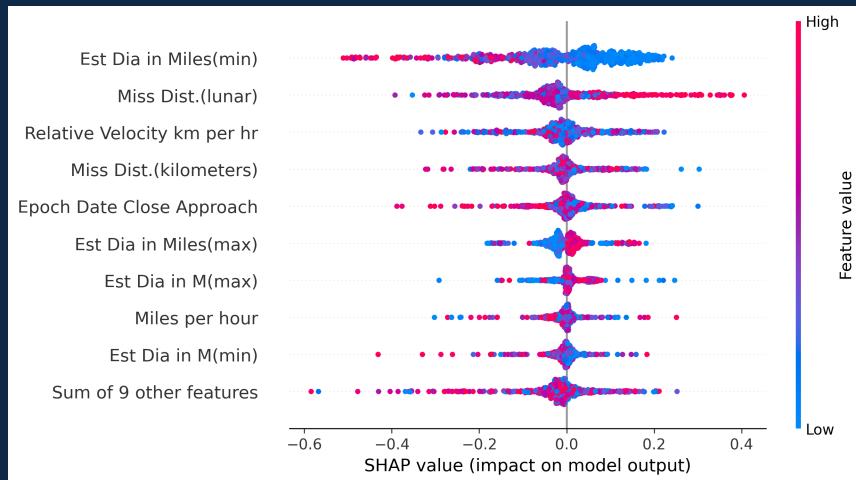
Ab. Mag.
+
min.orb.
+
Est. Dia.

min.orb.

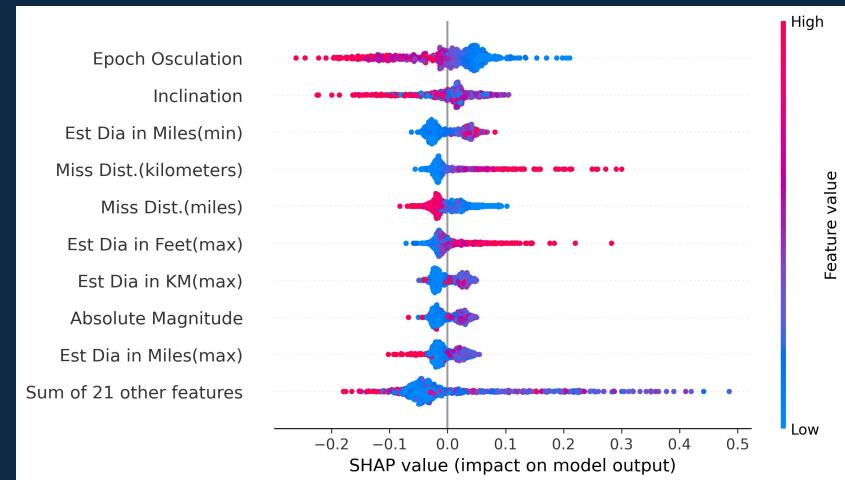
min.orb.
+
Est. Dia.

Est. Dia.

DNN



TREE



**Grazie per
l'attenzione**