

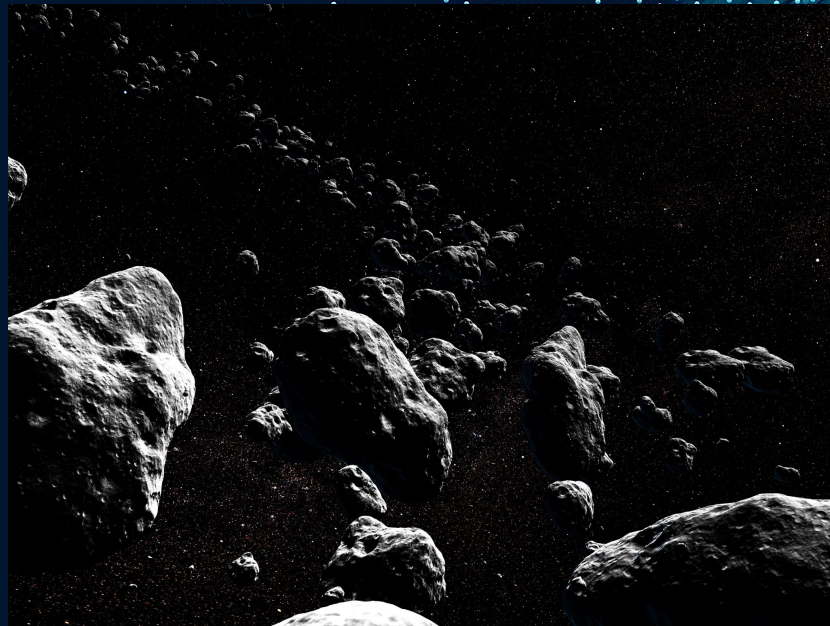


# **PREDICȚIA DIAMETRELOR ASTEROIZILOR**

Radu Mihalache  
Luca Costin  
Delia Grigoriță  
Iulian Muntean  
Claudiu Stan

# ASTEROIZI

- *planetoizi*, orbitează în jurul soarelui
- forme variate
- diametru de 1 m - 1000km
- pietroase, metalice sau de gheață
- 1M+ descoperiti



# MOTIVATIE

- Evaluarea potențialelor riscuri de impact
- Dezvolta astrometria & astrogeologia
- Potențiale misiuni de exploatare
- Vrem sa stim tot





# PROCESAREA DATELOR

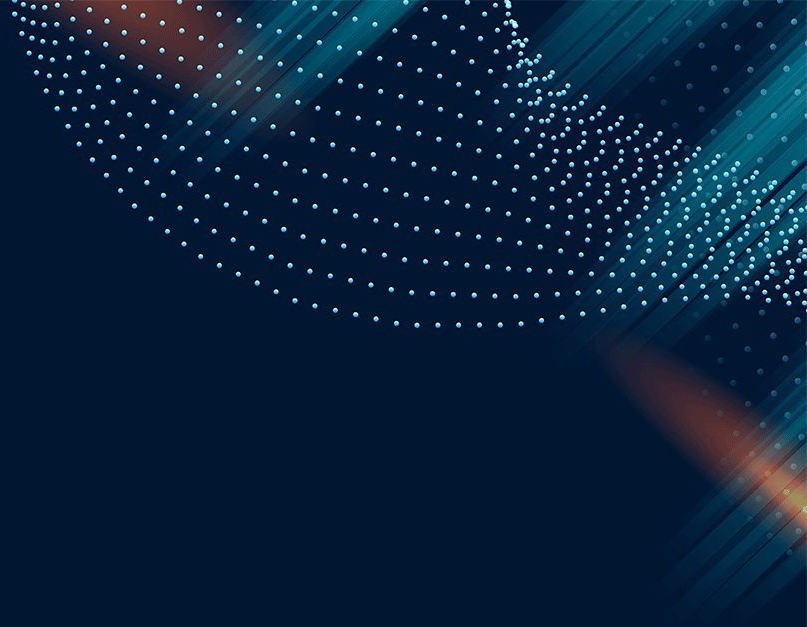
- Pentru unele instanțe, unele campuri sunt goale -> completat cu Mean și Most Frequent
- Datele categoricale au fost convertite in format one-hot
- Au fost selectate 88 de feature-uri din 188 pentru o convergenta mai rapidă
- Datele au fost scalate, în special pentru rețelele neuronale

n_obs	usr	H	albedo	rot_per	GM	BV	UB	G	moid	n	per	ma	neo#N	neo#Y	pha#N	pha#Y	spec_B#B	spec_B#C	spec_B#Cl	spec_B#C	spec
1002	3.34	0.09	9.07417	62.6284	0.713	0.426	0.12	1.59478	0.213885	1683.146	77.3721	1	0	1	0	0	0	1	0	0	
8490	4.13	0.101	7.8132	14.3	0.635	0.284	0.11	1.23324	0.213503	1686.156	59.69913	1	0	1	0	1	0	0	0	0	
7104	5.33	0.214	7.21	7.821928	0.824	0.433	0.32	1.03454	0.226019	1592.787	34.92502	1	0	1	0	0	0	0	0	0	
9325	3.2	0.4228	5.342128	17.8	0.782	0.492	0.32	1.13948	0.271609	1325.433	95.86194	1	0	1	0	0	0	0	0	0	
2916	6.85	0.274	16.806	7.821928	0.826	0.411	0.178739	1.09589	0.238632	1508.6	282.3663	1	0	1	0	0	0	0	0	0	
6034	5.71	0.2679	7.2745	7.821928	0.822	0.399	0.24	0.973965	0.260972	1379.46	86.19792	1	0	1	0	0	0	0	0	0	
5206	5.51	0.2766	7.139	7.821928	0.855	0.484	0.178739	0.8461	0.267535	1345.619	140.4197	1	0	1	0	0	0	0	0	0	
2744	6.49	0.226	12.865	7.821928	0.885	0.489	0.28	0.874176	0.301681	1193.314	194.8829	1	0	1	0	0	0	0	0	0	
2649	6.28	0.118	5.079	7.821928	0.858	0.496	0.17	1.10691	0.267484	1345.875	276.8616	1	0	1	0	0	0	0	0	0	
3409	5.43	0.0717	27.63	7	0.696	0.351	0.178739	1.77839	0.177007	2033.818	152.1849	1	0	1	0	0	1	0	0	0	
5492	6.55	0.191	13.7204	7.821928	0.837	0.417	0.178739	1.19322	0.256524	1403.375	278.9307	1	0	1	0	0	0	0	0	0	
3090	7.24	0.163	8.6599	7.821928	0.874	0.515	0.22	0.824953	0.276353	1302.68	133.3359	1	0	1	0	0	0	0	0	0	
2385	6.74	0.07	7.045	7.821928	0.745	0.452	0.178739	1.43633	0.238391	1510.123	187.4885	1	0	1	0	0	0	0	0	0	
2755	6.3	0.159	15.028	7.821928	0.833	0.388	0.178739	1.17966	0.237067	1518.561	164.9359	1	0	1	0	0	0	0	0	0	
2501	5.28	0.248	6.083	7.821928	0.839	0.451	0.23	1.19485	0.229238	1570.418	283.3877	1	0	1	0	0	0	0	0	0	
2364	5.9	0.1203	4.196	1.53	0.729	0.299	0.2	1.5358	0.197142	1826.093	288.3359	1	0	1	0	0	0	0	0	0	
3666	7.76	0.193	12.27048	7.821928	0.829	0.438	0.178739	1.12981	0.253843	1418.199	303.3644	1	0	1	0	0	0	0	0	0	
5082	6.51	0.181	11.57	7.821928	0.854	0.425	0.25	0.813258	0.283179	1271.281	267.2544	1	0	1	0	0	0	0	0	0	
3316	7.13	0.037	7.4432	7.821928	0.719	0.324	0.1	1.06213	0.258164	1394.461	197.3386	1	0	1	0	0	0	0	0	0	
2481	6.5	0.241	8.098	7.821928	0.854	0.463	0.25	1.08461	0.263474	1366.36	117.6951	1	0	1	0	0	0	0	0	0	

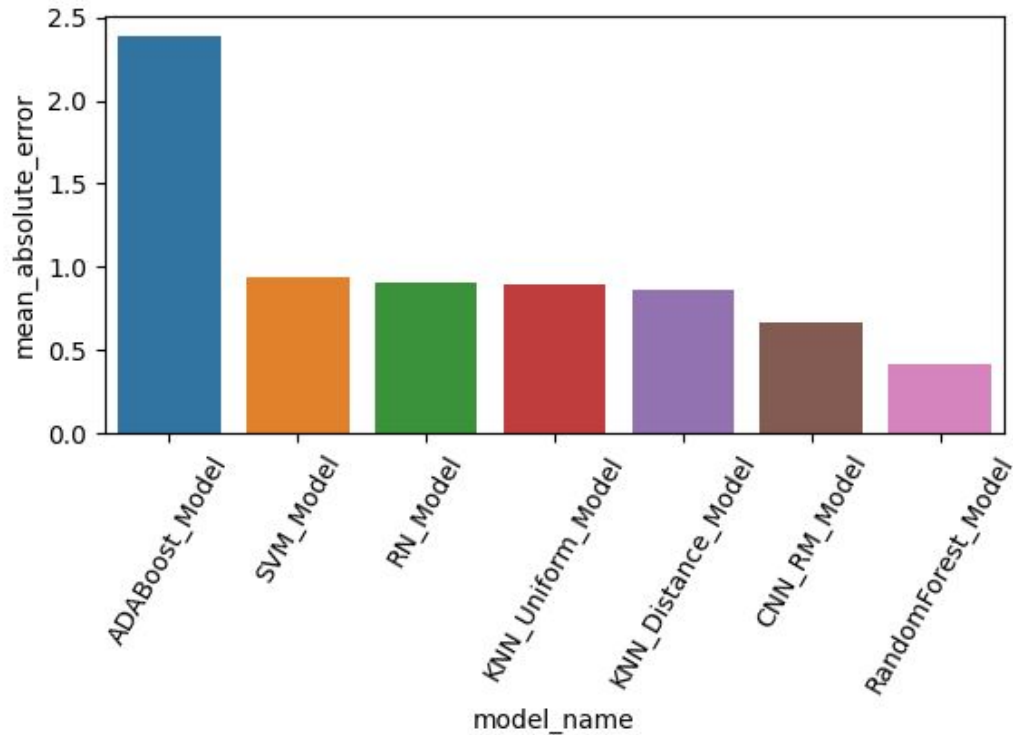
---

# MODELE

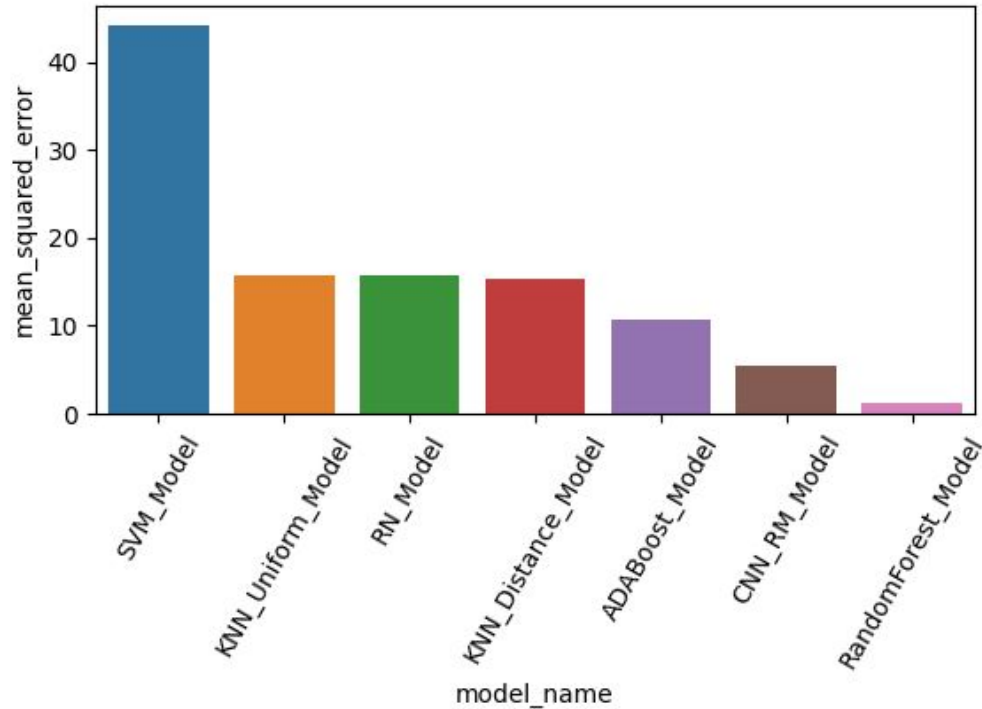
1. RN clasic
2. RN convolutional
3. SVM
4. AdaBoost
5. K-NN uniform
6. K-NN distance
7. Random Forest



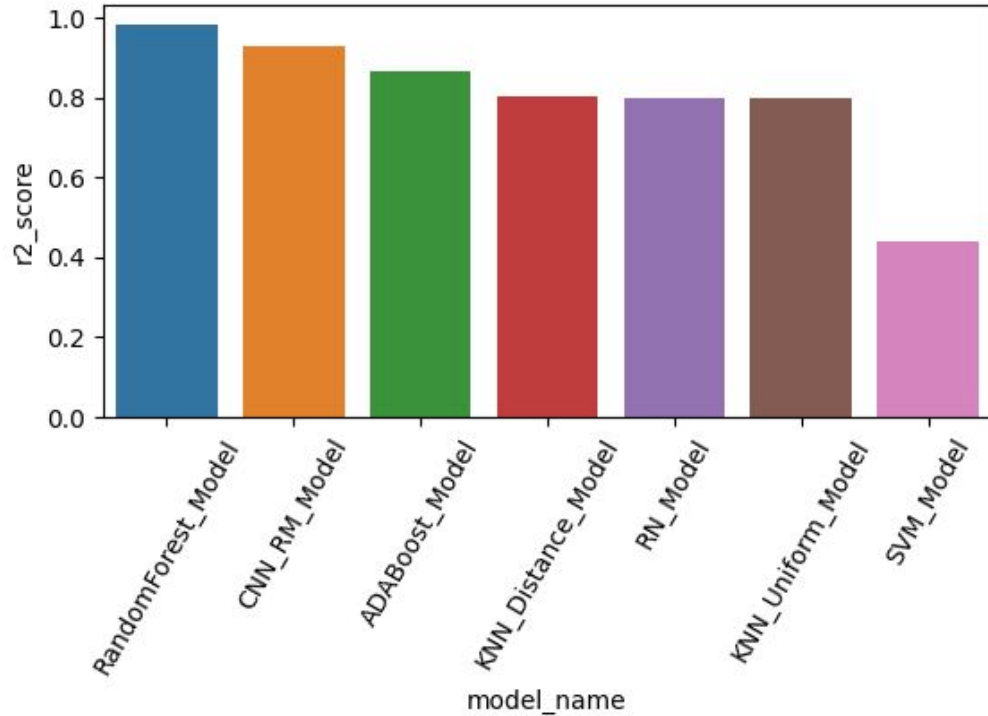
## Models comparison on mean absolute error



## Models comparison on mean squared error

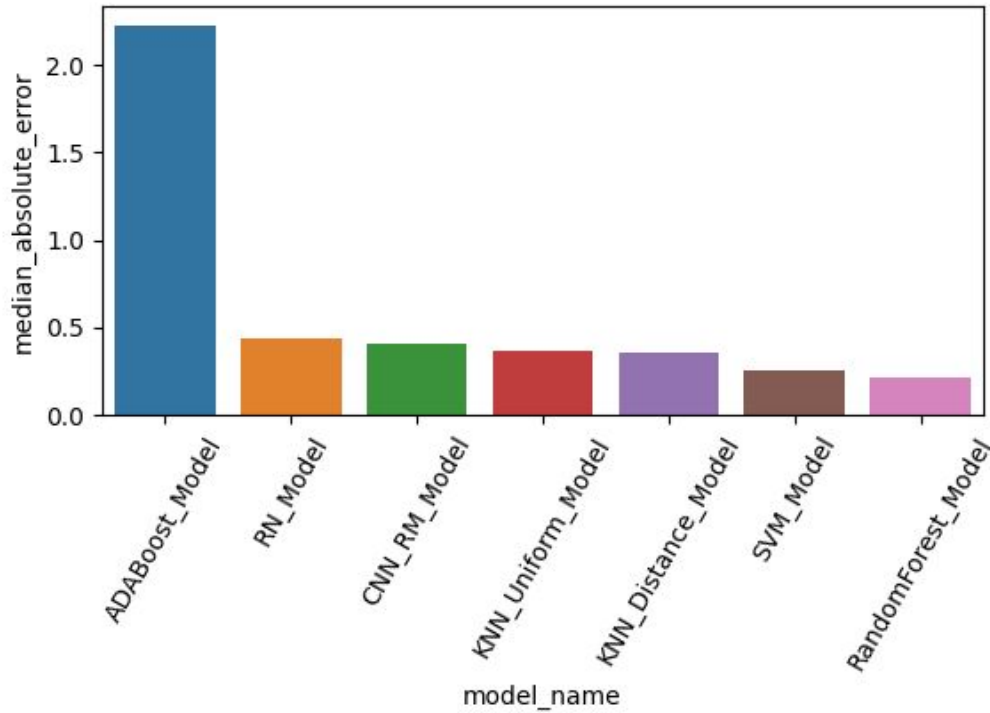


## Models comparison on r2\_score



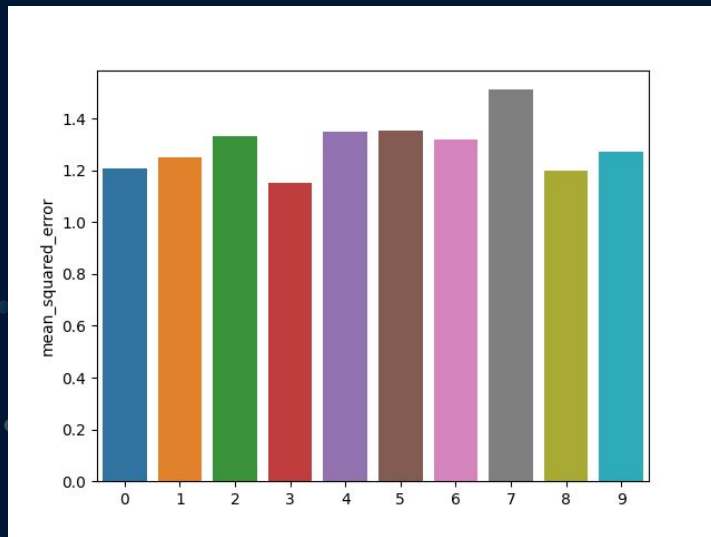


## Models comparison on median absolute error



# Random Forest

Random Forest construiește mai mulți arbori de decizie la antrenare și oferă media predicțiilor ancestor arbori. Acest algoritm produce rezultate bune pentru seturi de date care combină caracteristici categoricale și caracteristici continue.

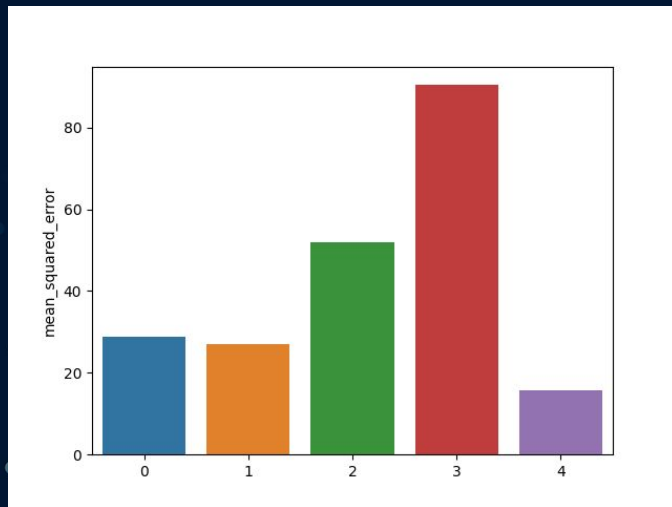


## RN clasic

Rețeaua neuronală are în componența sa 3 straturi ascunse de 1024, 512 respectiv 64 de neuroni.

Funcția folosită pe acestea este selu, rata de regularizare L2 fiind 0.01 pentru activare respectiv bias.

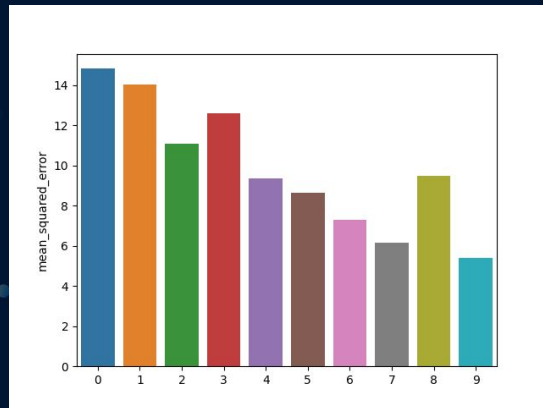
Funcția de activare pentru neuronul de ieșire este relu



# RN convolutional

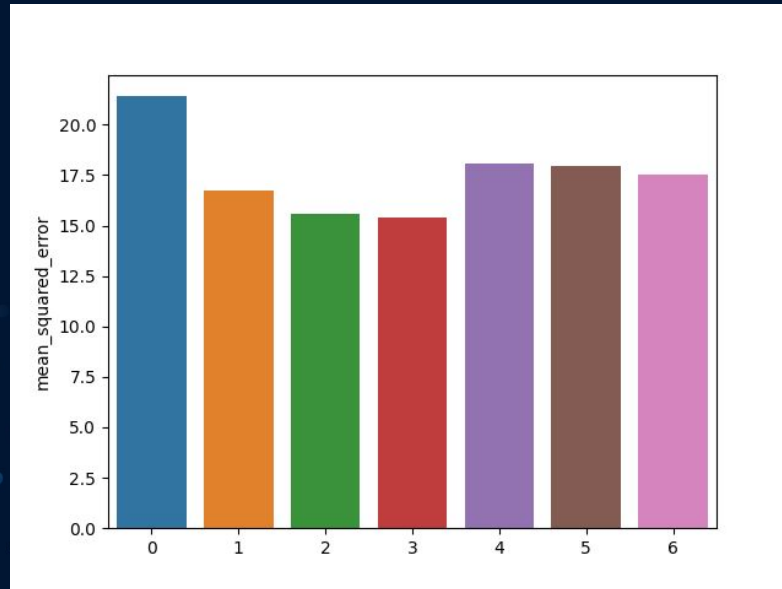
Reteua convolutionala folosita are o sectiune in care aplica un kernel de dimensiune 2 de 65 de ori pe caracteristicile de intrare. Facand acest lucru, instatele in care una din caracteristici este outlier devin mai balansate.

Output-ul sectiunii de convolutie este luat drept input de o retea neuronală denisa cu 3 straturi de 512, 256 si 1 neuron. Funcția de activare pentru primele doua straturi este selu iar pentru stratul de output este relu.



# K-NN

Doua modele diferite, în funcție de cum se considera votul majoritar: "uniform" - valoarea va fi data de valorile celor mai apropiate k puncte în mod majoritar, "distance" - valoarea va fi data de valorile ponderate celor mai apropiate k puncte, invers proportional cu distanta pana la ele.





# CONCLUZII

1.În urma folosirii mai multor modele de învățare automată, remarcăm ca acestea au utilitate diferită în raport cu ce date ne sunt oferite (sensibilitatea la zgomote, putere de generalizare). Particularizând pe exemplele prezentate, *Random Forest* este cea mai buna alegere.

2.Importanta formatarii datelor și efectul asupra performanței (acuratetii)

3.Diferenta de corectitudine între o *Retea Neuronala Clasica* și o *Retea Neuronala Convolutionala*, de asemenea dificultatea de calibrare a acestora

4.O comparatie sumara între K-NN uniform și K-NN Shepard

