PaAC Asteroid Project Report
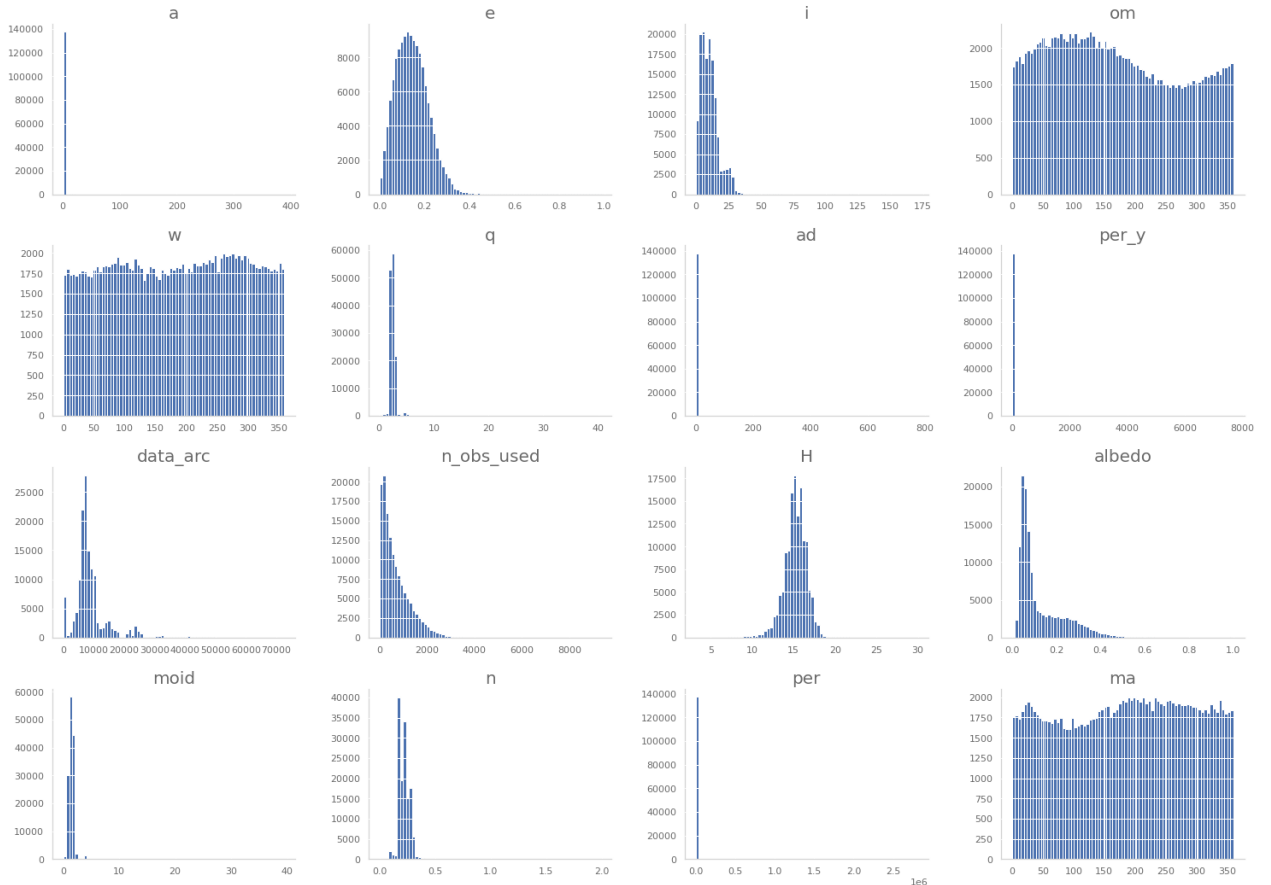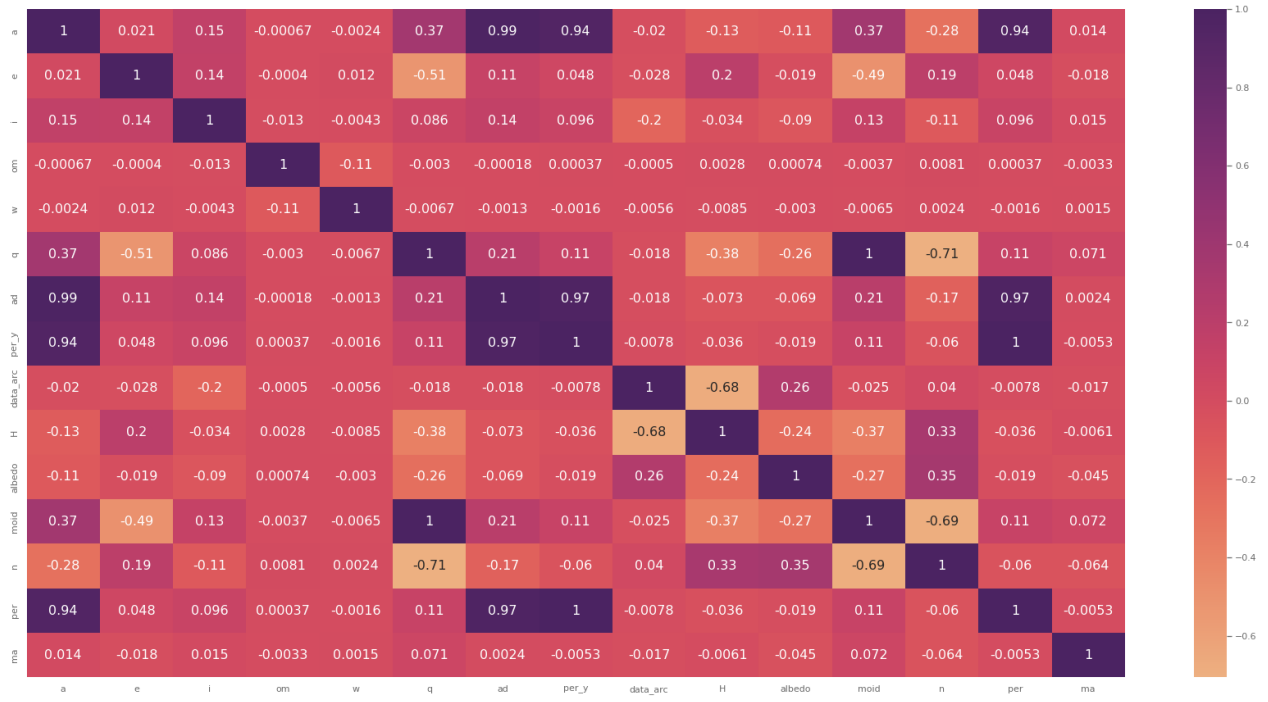
Adityan S

The used data set officially maintained by Jet Propulsion Laboratory of California Institute of Technology which is an organization under NASA.

JPL Small-Body Database Search Engine

1   Dataset Description

The feature refined dataset contains 32 features and 136005 records. The Crossplot and Correlation Plot are as follows

	a	e	i	om	w	q	ad	per_y	data_arc	H	albedo	moid	n	per	ma
a	1	0.021	0.15	-0.00067	-0.0024	0.37	0.99	0.94	-0.02	-0.13	-0.11	0.37	-0.28	0.94	0.014
e	0.021	1	0.14	-0.0004	0.012	-0.51	0.11	0.048	-0.028	0.2	-0.019	-0.49	0.19	0.048	-0.018
i	0.15	0.14	1	-0.013	-0.0043	0.086	0.14	0.096	-0.2	-0.034	-0.09	0.13	-0.11	0.096	0.015
om	-0.00067	-0.0004	-0.013	1	-0.11	-0.003	-0.00018	0.00037	-0.0005	0.0028	0.00074	-0.0037	0.0081	0.00037	-0.0033
w	-0.0024	0.012	-0.0043	-0.11	1	-0.0067	-0.0013	-0.0016	-0.0056	-0.0085	-0.003	-0.0065	0.0024	-0.0016	0.0015
q	0.37	-0.51	0.086	-0.003	-0.0067	1	0.21	0.11	-0.018	-0.38	-0.26	1	-0.71	0.11	0.071
ad	0.99	0.11	0.14	-0.00018	-0.0013	0.21	1	0.97	-0.018	-0.073	-0.069	0.21	-0.17	0.97	0.0024
per_y	0.94	0.048	0.096	0.00037	-0.0016	0.11	0.97	1	-0.0078	-0.036	-0.019	0.11	-0.06	1	-0.0053
data_arc	-0.02	-0.028	-0.2	-0.0005	-0.0056	-0.018	-0.018	-0.0078	1	-0.68	0.26	-0.025	0.04	-0.0078	-0.017
H	-0.13	0.2	-0.034	0.0028	-0.0085	-0.38	-0.073	-0.036	-0.68	1	-0.24	-0.37	0.33	-0.036	-0.0061
albedo	-0.11	-0.019	-0.09	0.00074	-0.003	-0.26	-0.069	-0.019	0.26	-0.24	1	-0.27	0.35	-0.019	-0.045
moid	0.37	-0.49	0.13	-0.0037	-0.0065	1	0.21	0.11	-0.025	-0.37	-0.27	1	-0.69	0.11	0.072
n	-0.28	0.19	-0.11	0.0081	0.0024	-0.71	-0.17	-0.06	0.04	0.33	0.35	-0.69	1	-0.06	-0.064
per	0.94	0.048	0.096	0.00037	-0.0016	0.11	0.97	1	-0.0078	-0.036	-0.019	0.11	-0.06	1	-0.0053
ma	0.014	-0.018	0.015	-0.0033	0.0015	0.071	0.0024	-0.0053	-0.017	-0.0061	-0.045	0.072	-0.064	-0.0053	1

The Diameter can be theoretically calculated by Absolute Magnitude H and Albedo a. Hence we add this Estimated diameter d_{est} as a feature to refine our dataset.

$$d_{est} = 10^{3.123 - 0.5 \log_{10} a - 0.2H}$$
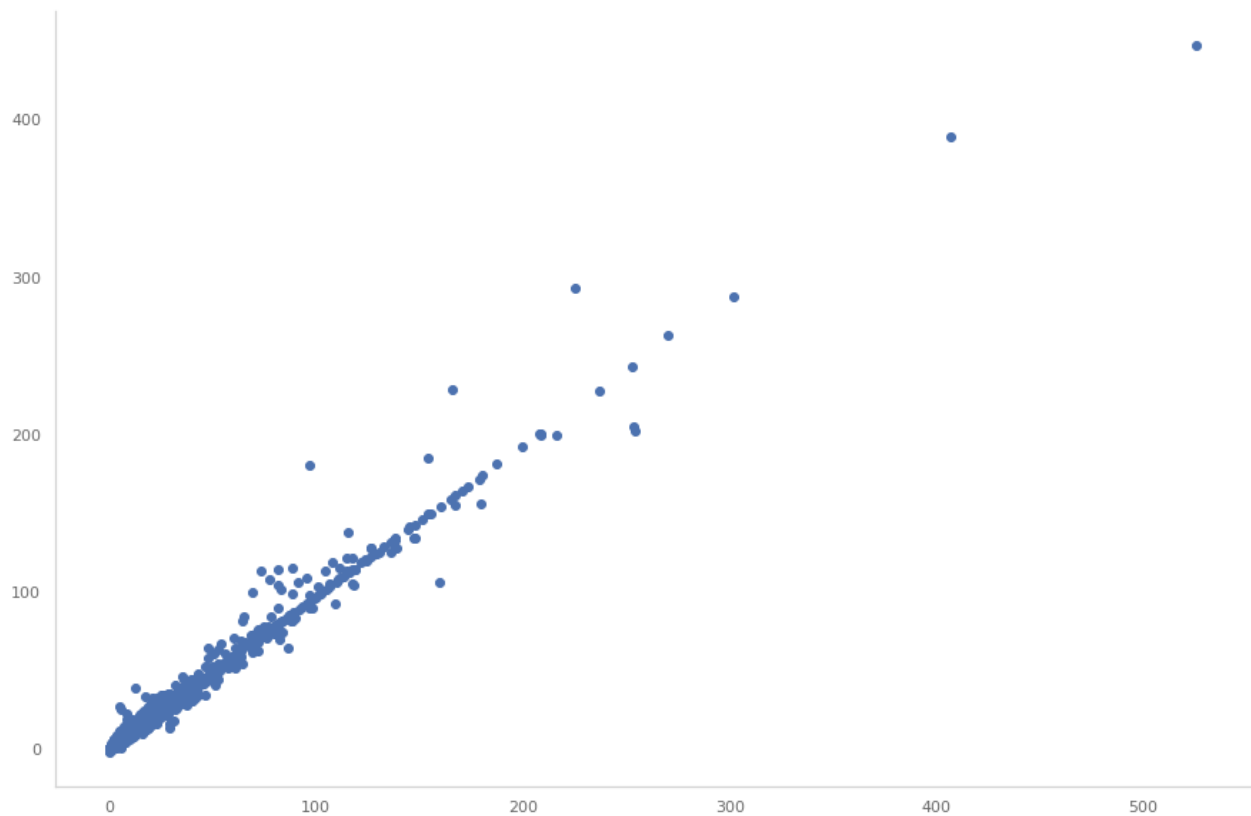
Reference : NASA

2 Model Implementation

A total of 3 models have been trained. They are as follows,

2.1 Linear Regression

For Linear Regression with no Regularization, the d_{pred} vs d_{real} is as follows
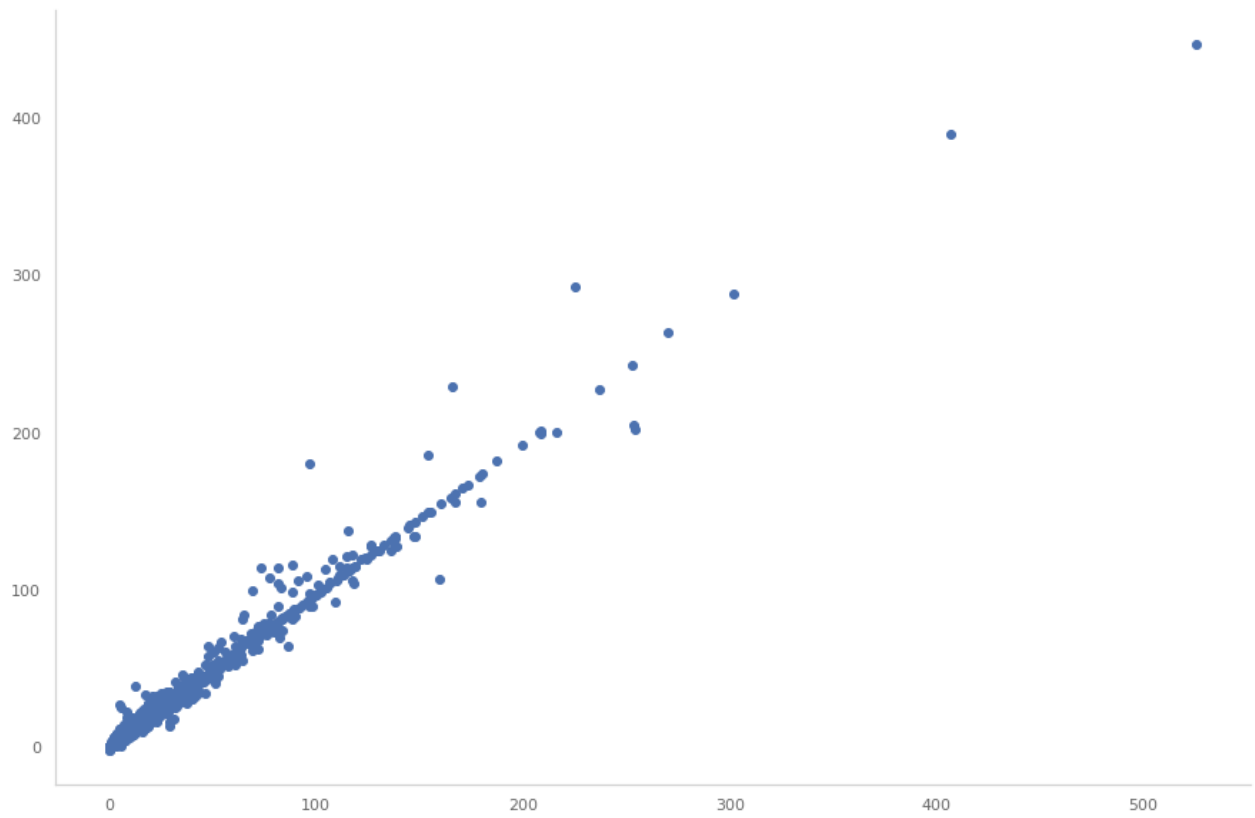
- Mean Squared Error : 1.668289980658958

- r2 Score : 0.9804787292677338

2.2 Linear Regression with L2 Regularization (Ridge Regression)

For Linear Regression with L2 Regularization and Best Parameter $\alpha = 100$, the d_{pred} vs d_{real} is as follows

- Mean Squared Error : 1.6745546130261924

- r2 Score : 0.9794149846560747

2.3   Multilayer Perceptron Neural Network

For Multilayer Perceptron Neural Network with ReLU (rectified linear unit) activation function, the d_{pred} vs d_{real} is as follows

- Mean Squared Error : 1.6141165334766034

- r2 Score : 0.9815957093790618