Asteroid Diameter Prediction

Vivek Verma
vivekv@arizona.edu
Github

December 9, 2023

Abstract

The project aims to enhance our understanding of asteroid distribution across the solar system, utilizing observational data published JPL's solar system dynamics (SSD) group. Subsequently, it focuses on predicting asteroid diameters through the application of machine learning techniques. The work includes careful selecting of the features from the data, using the correlation matrix plot and recursive feature elimination (RFE). The selected features is used to train a two-layer neural network to model the prediction. The result are compared with a existing analysis and the work shows better results based upon metrics such as mean squared error (MSE) and R-squared(R2). The prediction of the size of the asteroids are based on their brightness, size, orbit configuration around the Sun.

1 INTRODUCTION

The Earth has witnessed numerous amounts of natural disasters, resulting in loss of life and disturbance in habitation. Including the impact from the meteorites. In the past, there have been incidences where impact from the asteroids have caused destructing, like Tunguska and Chelyabinsk. Although the impact frequency of a space rock capable of causing serious damage is low, one or two in decades, but still scientific community spends a considerable effort in detection and classification of near-earth objects (NEO). Even a hit from a asteroid of size of several hundred meter can destroy a big city.

The project delves into the intricate domain of asteroids within our solar system, leveraging the comprehensive observational data meticulously cataloged by NASA's Jet Propulsion Laboratory's Solar System Dynamics (SSD) group. The primary objective of our endeavor is twofold: first, to gain a nuanced understanding of the distribution patterns of these celestial bodies, and second, to predict their diameters—a pivotal factor in assessing their cosmic significance.

Diameter prediction is extremely important for identifying NEAs (Near-Earth Asteroids) and PHAs (Potentially Hazardous Asteroids), since the size of an asteroid can have a big impact on how it interacts with Earth. Smaller asteroids are more likely to burn up in Earth's atmosphere before reaching the surface, while larger asteroids can cause more damage upon impact. By predicting the diameter of NEAs and PHAs, scientists can better assess the potential threat these asteroids pose to Earth and develop strategies to diminish any potential harm. In addition, the diameter prediction can also provide valuable information about the composition, structure, and evolution of asteroids, which can lead to a deeper understanding of the origins of the Solar System we live in.

In our pursuit of these scientific goals, we employ advanced machine learning techniques, serving as the computational backbone for our predictive modeling. This intricate process involves the discerning selection of pertinent features from the observational dataset, facilitated by tools such as correlation matrix plots and recursive feature elimination (RFE). These techniques enable us to distill essential details regarding the brightness, size, and orbital dynamics of asteroids. Central to our methodology is the implementation of a two-layer neural network—a sophisticated computational model that learns and adapts to patterns within the data, enhancing the precision of our predictions. Subsequently, our results are rigorously benchmarked against established analyses, demonstrating superior performance as evidenced by key metrics like mean squared error (MSE) and R-squared (R2).

In essence, our project bridges the realms of science and technology, offering not only insights into the dynamics of asteroids but also showcasing the efficacy of machine learning in unraveling the complexities of

[1]Kaggle link to the project: Link
[2]Github link to the project: Link

celestial phenomena. This journey, grounded in meticulous research and computational acumen, seeks to contribute meaningfully to the ever-expanding frontier of astronomical inquiry.

2 Background

Asteroids are small, rocky objects that revolve around the Sun in elliptical orbits. They are sometimes called minor planets or planetoids and are remnants left over from the early formation of our solar system about 4.6 billion years ago.

2.1 Category

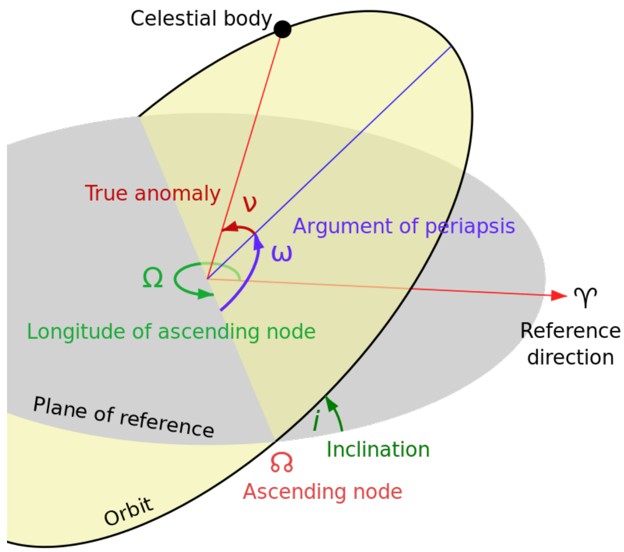
Asteroids range in size from tiny specks to hundreds of kilometers across, and they are located primarily in the asteroid belt between Mars and Jupiter, though, they can also be observed in other regions of the solar system. Most asteroids are irregularly shaped, only some of them are close to spherical, and they are often pitted or cratered. Basically, the diameter of an asteroid of irregular shape is usually represented by its equivalent diameter, which is the diameter of a sphere that has the same volume as the asteroid. This equivalent diameter is a handy way to compare the size of different asteroids, but it should be noted that it does not necessarily represent the actual physical size or shape of the asteroid. In some cases, the equivalent diameter may be an overestimate or an underestimate of the true size of the asteroid, depending on its shape and density.

Table 2.2 describe the category of the asteroid based upon the orbit size.

	Name	Description	Details
MBA	Main-belt Asteroid	Objects orbiting between Mars and Jupiter in the main portion of the asteroid belt.	2.0 au < a < 3.2 au; q > 1.666 au
OMB	Outer Main-belt Asteroid	Objects orbiting between Mars and Jupiter in the outer reaches of the main asteroid belt.	3.2 au < a < 4.6 au
TJN	Jupiter Trojan	Objects trapped in Jupiter's L4/L5 in Lagrange points, share Jupiter's orbit around the sun.	4.6 au < a < 5.5 au, e < 0.3
IMB	Inner Main-belt Asteroid	Objects orbiting between Mars and Jupiter within the inner portion of the asteroid belt.	a < 2.0 au; q > 1.666 au
APO	Apollo	Near-Earth asteroids whose orbit crosses the orbit of Earth.	a > 1.0 au; q < 1.017 au
MCA	Mars-crossing Asteroids	Objects with an orbit that crosses the orbit of Mars.	1.3 au < q < 1.666 au; a < 3.2 au
AMO	Amor	Near-Earth asteroids whose orbit approaches the orbit of Earth but does not cross it	a > 1.0 au; 1.017 au < q < 1.3 au
ATE	Aten	Near-Earth asteroids whose orbit could bring it in close proximity to Earth.	a < 1.0 au; ad > 0.983 au
CEN	Centaur	Objects with an orbit between Jupiter and Neptune.	5.5 au < a < 30.1 au
TNO	Trans-Neptunian Object	Objects with orbits outside Neptune.	a > 30.1 au
AST	Asteroid (other)	Asteroid orbit not matching any defined orbit class.	

2.2 Orbital elements

The orbit of a small object orbiting a massive object can be can be approximated by the restricted two body problem. The massive object is consider to be stationary, which is Sun, and the lighter object is assumed to orbit around it in a elliptic orbit. The orbit of such system may be described using the set values known as orbital elements: a set of parameters that elegantly define the intricate path an asteroid traces through space. Comprising six key elements—semi-major axis (a), eccentricity (e), inclination (i), argument of periapsis (ω), longitude of ascending node (Ω), and mean anomaly (M)—these values intricately shape the elliptical trajectory an asteroid follows around the Sun. The semi-major axis, denoted by a, serves as a cosmic ruler, representing half the length of the longest diameter of the elliptical orbit. Eccentricity (e) captures the elongation of the orbit, ranging from a perfect circle ($e = 0$) to a stretched ellipse ($e > 0$). Inclination (i) defines the tilt of the orbital plane concerning the reference plane, offering insights into the asteroid's orientation in space. The argument of periapsis (ω), longitude of ascending node (Ω), and mean anomaly (M) further refine the orbit, specifying the asteroid's position at a particular point in time. Together, these orbital elements unfold a celestial choreography, allowing astronomers to decipher the unique shape and size of each asteroid's cosmic path around the Sun.



2.3 Brightness

The brightness of asteroids stands as a key observable, providing astronomers with crucial insights into their composition, reflectivity, and intrinsic characteristics. This luminosity is a dynamic interplay of several parameters, and it is reported using following parameters

1 Absolute Magnitude (H): This fundamental parameter gauges the intrinsic brightness of an asteroid if observed from a standard distance of one astronomical unit (AU) away. A lower absolute magnitude signifies a more luminous object, revealing information about its size and reflective properties.

2 Albedo :Albedo, or reflectivity, quantifies the amount of sunlight an asteroid reflects. A high albedo indicates a more reflective surface, contributing to increased brightness. Albedo values range from 0 (perfect absorber) to 1 (perfect reflector).

3 Methodology

The data set is obtained from the source, which is taken from NASA's Jet Propulsion Laboratory's Solar System Dynamics (SSD) group.

A pre-written code is taken from the Kaggle, project link, which predict the asteroid diameter based upon the data available. The work trains three model namely, multilayer perceptron (MLP), CatBoost, and LGBM, and compares the prediction error for test data set. The result suggest that 1st model, MLP gives better results than other.

My work include the modifying the features selection used for the prediction based upon understanding of the feature capability to influence the output column i.e. diameter of the asteroids. I used techniques such as correlation matrix map and recursive feature elimination (RFE), while the existing work only used correlation map. The knowledge about orbital mechanics also helped on eliminating the redundant feature, for example time period is a function semi-major axis, thus having semi-major axis as feature is enough. Plots of features with each other and also with the output variable gave an insight on importance of a particular feature. The feature which have deterministic relation with other features can also be eliminated since it not providing any new information.

After selection of feature, a 2-layered neural network is build with first hidden layer having 64 neurons and second hidden layer having 32 neurons and the output was a scalar value. The mean squared error and R-squared metric are calculate to measure the accuracy of the model. Along with these the same model were also used for prediction with the new feature that were selected for this work. The performance of the 2-layered neural network and the other three model developed in the existing work with new feature set are reported in the Table.

3.1 Data Description

The data sets contain records of 839714 asteroids. The information about each column contained in the data is described in the Table 1. The data is disturbed in a ratio of 80:20 as a training data and testing data. The Training data is further distributed in a same ration 80:20 to get a validation data set.

Name	**Description**
name	object full name
a	semi-major axis (au)
e	eccentricity
i	inclination; angle with respect to x-y elliptic plane (deg)
om	longitude of the ascending node (deg)
w	argument of perihelion (deg)
q	perihelion distance (au)
ad	aphelion distance (au)
per_y	orbital period (years)
data_arc	number of days spanned by the data arc (d)
condition_code	orbit condition code
n_obs_used	number of observations used
H	absolute magnitude parameter
neo	Near-Earth Object flag (Y/N)
pha	Potentially Hazardous Asteroid flag (Y/N)
diameter	object diameter (from equivalent sphere) (km)
extent	object bi/tri-axial ellipsoid dimensions (km)
albedo	geometric albedo
rot_per	rotation period (h)
GM	standard gravitational parameter, product of the mass (M) and gravitational constant (G)
BV	color index B-V magnitude difference
UB	color index U-B magnitude difference
IR	color index I-R magnitude difference
spec_B	spectral taxonomic type (SMASSII)
spec_T	spectral taxonomic type (Tholen)
G	magnitude slope parameter (default is 0.15)
moid	Earth minimum orbit intersection distance (au)
class	asteroid orbit class (e.g., MBA, OMB)
n	mean motion (deg/d)
per	orbital period (d)
ma	mean anomaly (deg)

Table 1: Description of Asteroid Orbital Elements

Two-Layer Neural Network: A Two-Layer neural network which is subset of multilayer perceptron (MLP) represents a simplified yet powerful neural network architecture, comprising an input layer, two hidden layer, and an output layer. An MLP with 2 hidden layers consists of three main layers: the input layer, two hidden layers, and the output layer. Neurons in each layer are interconnected, forming a network that processes input data through a series of transformations. Activation functions, ReLU (Rectified Linear Unit) for hidden layers is used. This non-linearity introduced due to ReLU enables the network to learn and represent complex, non-linear mappings between input and output.

The model have two hidden layers (one with 64 neurons and the other one with 32 neurons) both with ReLU activation function and an output layer representing scalar value - our predicted diameter. As for measuring the model's performance, MSE loss and Adam optimization is utilised. The model is trained for 50 epochs and use early stopping to stop after 10 epochs if the performance doesn't get better, in order to prevent over-fitting and optimize model performance.

3.2 Data Pre-Processing

The dataset comprising 839714 asteroids records with 31 variables, underwent thorough data preprocessing to ensure the reliability and accuracy of the machine learning models. Key preprocessing steps included:

Missing Value Analysis: First the rows with missing diameter value are removed, after removing the value the data set size reduced to 137636. After removing the rows, other columns are checked for missing entries. The columns with missing values are:

Feature name	# of missing rows
H	747
Albedo	1230

Table 2: Showing the missing statistics

Missing Values Imputation: Median imputation is used to fill for the missing numbers and also the imputation method is successfully capturing the patterns in the data.

3.2.1 Feature Selection

Using the correlation map the feature with highest correlation with diameter of the asteroid is H, which is expected. With the assumption of same material the H value will have increase with size. the lager the

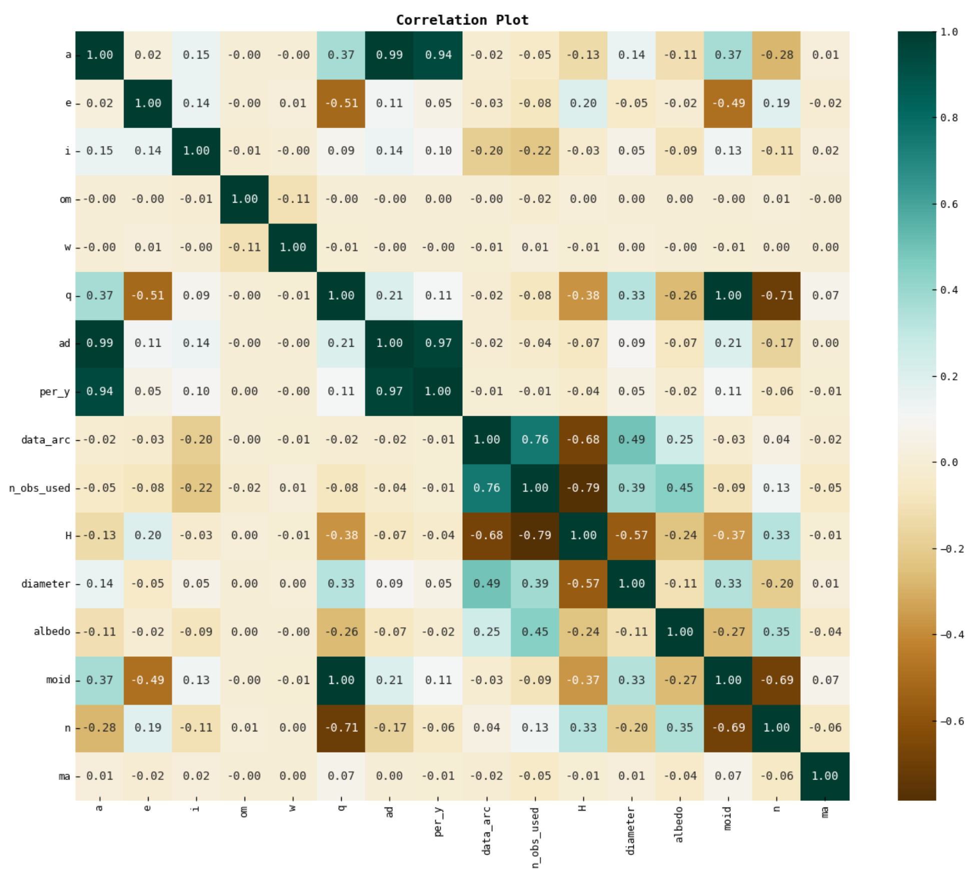


Figure 1: Showing the correlation plot.

asteroid . The final features that are selected based upon correlation map, RFE, and mutual relationship are: a, e, i, q, $data\_arc$, $condition\_code$, n_obs_used, $albedo$, H.

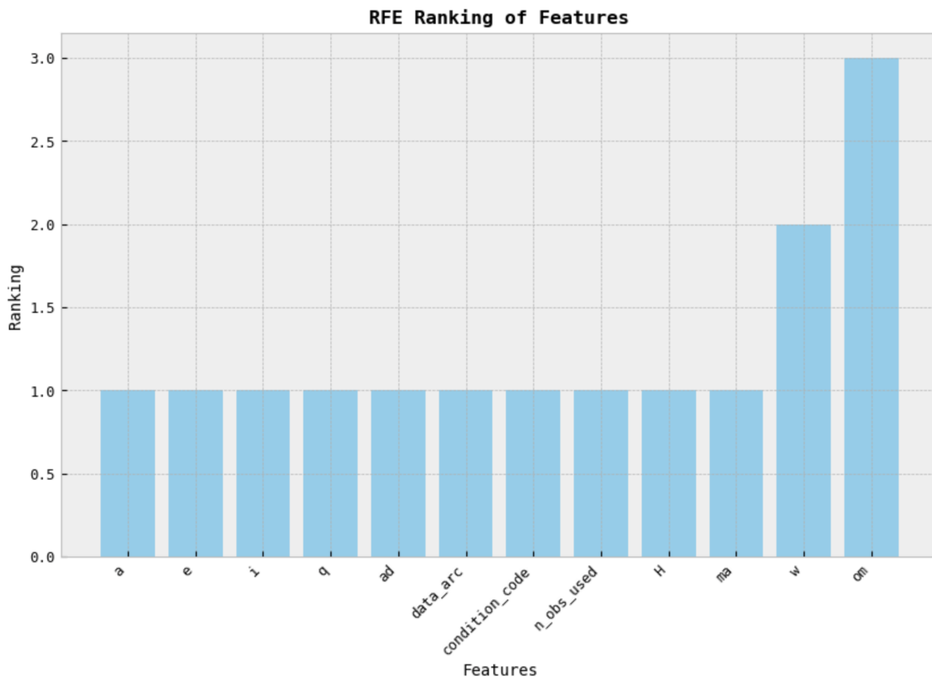


Figure 2: Showing RFE plot.

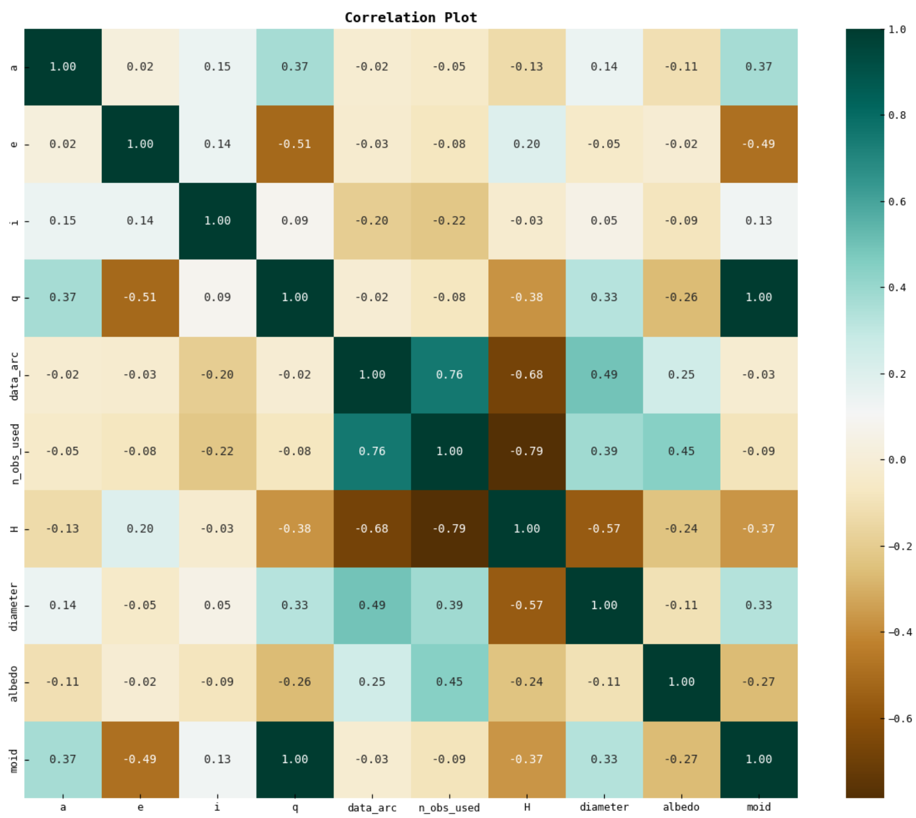


Figure 3: Showing the missing statistics.

3.3 Performance Evaluation

We have Evaluated the models based on the R^2 and MSE(Mean Square Error) Values.

4 Results

The following prediction error values are evaluated for the feature selected in this work. The second column shows the result for the model implemented in the work and next three are the model used in the existing work. Further, Table 4 shows the prediction error values for the feature and model used in the existing work.

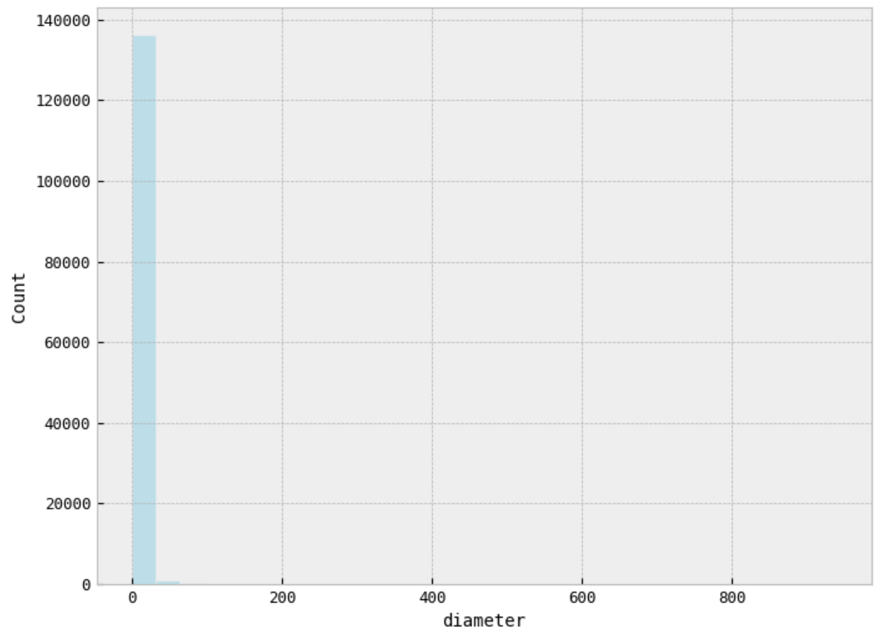


Figure 4: Showing the distribution for the diameter of asteroids.

Models	2-Layered NN	MLP [1]	CatBoost[1]	LGBM??1]
MSE	1.369	1.369	2.212	2.145
R^2	0.981	0.981	0.969	0.970

Table 3: Performance Evaluation of All models

Models	MLP [1]	CatBoost[1]	LGBM??1]
MSE	1.893	2.128	5.632
R^2	0.978	0.975	0.933

Table 4: Performance Evaluation of All models (existing work)

The 2-layered neural network and MLP from the reference work yield same result as both are 2 layer neural network with different amount neurons in each hidden layer. The neural network used in this work has more neurons, the latter has 50 neurons for 1st hidden layer and 10 neurons for 2nd hidden layer. It seems the the number of the layers and neuron were enough for the latter model to generate same level of accuracy.

5 Conclusion

In this project, we aimed to develop a model for predicting diameter for asteroids, a critical parameter in categorising the threat for the space rocks. The implemented 2-layered neural netword model along the modified features from the reference work yield better prediction. The comparison criteria were based on two key performance metrics: MSE and R^2.

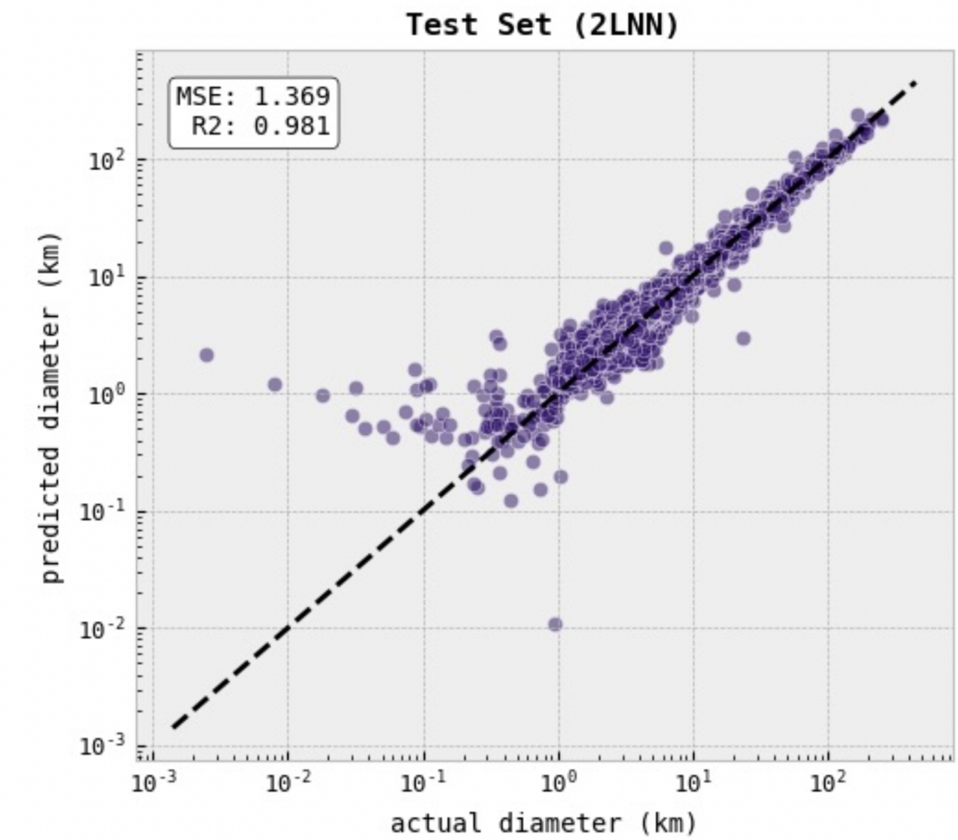


Figure 5: Showing the test data prediction.

References

[1] `https://www.kaggle.com/code/blulypsee/asteroid-diameter-prediction-mlp-catboost-lgbm`.

Rabeendran, Amandin & Denneau, Larry. (2021). A Two-Stage Deep Learning Detection Classifier for the ATLAS Asteroid Survey.

V. Bahel, P. Bhongade, J. Sharma, S. Shukla and M. Gaikwad, "Supervised Classification for Analysis and Detection of Potentially Hazardous Asteroid," 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), Nagpur, India, 2021, pp. 1-4, doi: 10.1109/ICCICA52458.2021.9697222.

C. Rosu and V. Bacu, "Asteroid Image Classification Using Convolutional Neural Networks," 2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj- Napoca, Romania, 2021, pp. 3-10, doi: 10.1109/ICCP53602.2021.9733484.

Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 12, pp. 6999-7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.