Applications of Machine Learning in Estimating the Minimum Distance of Approach of an NEO

Jayant Mehra
mehrajayant@gmail.com

Dated : 5 November, 2016

1 Abstract

Although the current detection techniques have been able to calculate the minimum distance to which a Near Earth Object (NEO) can approach Earth for thousands of NEOs, there are millions of yet undiscovered NEOs which could pose a threat to Planet Earth. An NEO is considered highly dangerous if the minimum distance between it and the centre of the Earth is less than 0.03 AU. However, only a handful NEOs have been detected prior to entering this danger zone. The immense task of asteroid hunting by conventional techniques is further complicated by a high number of false positives and false negatives. In this report, machine learning algorithms are written to predict the minimum distance upto which an NEO can approach the planet and classify NEOs as whether they are in the danger zone or no based on their physical characteristics. In section 4 of the study, an Artificial Neural Network based on the backpropagation algorithm and a Logistic Classification based on Unconstrained Minimisation using the fminunc function are employed to classify NEOs with an accuracy of 92% and 90% respectively. In section 5 of the report, the Levenberg - Marquardt Algorithm based on an Artificial Neural Network is employed to calculate the minimum distance with a regression R value of 0.79 (Value of 1 being the maximum). All the algorithmic systems developed have low false positive and false negative rates

2 Introduction

Near Earth Objects can cause a cataclysmic damage to Planet Earth and since there are millions of yet undiscovered NEOs, it is indispensable to take measures to predict the minimum distances up to which they can approach Earth.

The NEOs are considered potentially dangerous if they are within 0.03 AU (Astronomical Units) from the centre of the Earth. However, the current techniques, which are based solely on the use of hardware, are unable to detect these NEOs. According to NASA Space Apps, 2016, in many cases, the conventional techniques have failed to identify an NEO even before it entered the planet's atmosphere. To predict the minimum distance and classify NEOs as whether they would enter the 0.03 AU danger zone or no, effective machine learning algorithms are developed.

Machine Learning has numerous benefits in NEO hunting. It eliminates mistakes committed due to human carelessness. There are several databases which contain large amounts of NEO related data. Thus, the algorithms can train themselves further to achieve high accuracies. Moreover, the algorithmic system can easily be shared.

In this report, I employ machine learning algorithms such as the backpropagation in an Artificial Neural Network and Unconstrained Optimisation. The fminunc (UO) is used for a logistic classification model and is coded in MATLAB and python. The ANNs have been trained by the Artificial Neural Network Pattern Recognition and Fitting Tool in MATLAB.

The features used for training the algorithms were carefully chosen. The data for the orbital arc and minimum distance of approach was from the Minor Planet Center Database provided by The International Astronomical Union (IAU). The data for Absolute Magnitude and eccentricity was from the California Institute of Technology's Jet Propulsion Laboratory Small-Body Database Browser.

Report Layout The report is structured as follows. Section 3 contains the research, algorithms used and findings for classifying NEOs based on their minimum distance of approach. Section 4 contains the research, algorithm used and findings for predicting the minimum distance up to which an NEO can approach planet Earth. Section 5, 6 and 7 contain the conclusions, acknowledgements and references.

3 Classification

3.1 Artificial Neural Network

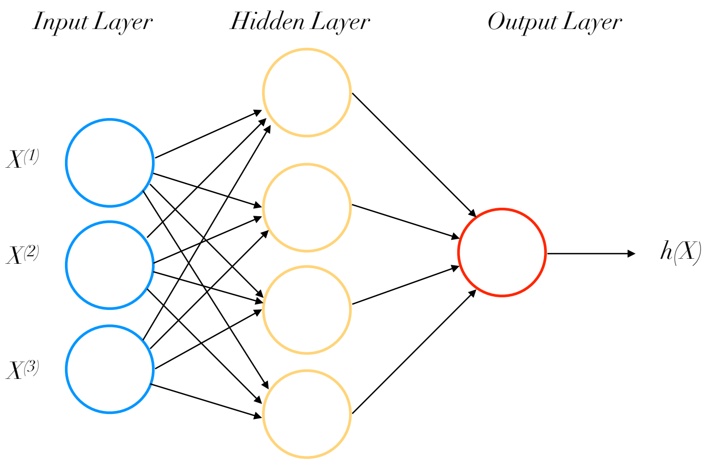

Figure 1: Sample 3-Layer Artificial Neural Network Architecture

3.1.1 Feature Selection

From the Minor Planet Center Database provided by The International Astronomical Union (IAU) and the California Institute of Technology's Jet Propulsion Laboratory Small-Body Database Browser, the following features were used (two at a time) to train the ANN :

S. No.	Feature	Physical Significance
1	Orbital Arc (OA)	Time required to cross an imaginary, arbitrary arc.
2	Absolute Magnitude (H)	Visual magnitude of an NEO 1 AU away, 1 AU from the sun & 0 phase angle
3	Semi-Major Axis (a)	0.5 * major axis of the elliptical orbit
4	Eccentricity (e)	Deviation of the orbit from a perfect circle
5	Node Angle (N)	Longitude of the ascending node
6	Period (t_p)	Rotation Period
7	Inclination (i)	Inclination of the orbit with respect to the Earth.

Table 1 : Features

On the basis of ease of calculability of the physical parameters of new NEOs and accuracy of the ANN, absolute magnitude (H) and eccentricity (e) were selected after each of the features listed above were used by the ANN.

3.1.2 ANN Development

The neural network was developed using an in-built tool in MATLAB. The two features listed above were used as inputs. The ANN was three layered wherein the hidden layer contained seven hundred neurons. Figure 1 illustrates a three layered, 4 hidden layer neuron artificial neural network architecture.
The Backpropagation Propagation Algorithm was used to train the Neural Network.

$$\delta^{(i)} = (\Theta^{(i)})^T \cdot \delta^{(i+1)} .* g'(z^{(i)})$$

The backpropagation algorithm calculates the error (δ) of the last node and moves back.
g' represents the differential of the sigmoid function and
$z^{(i)} = \Theta^2 a^{(i-1)}$
. represents an element by element multiplication*
Note : The above algorithm is not valid for the first node.

3.1.3 Results

The Artificial Neural Network used over 160 sets of data to train itself on classifying the NEOs as dangerous (< 0.03 AU) or no. In a single test run, 2 features were used in the input layer (see Figure 1). The most accurate results, as mentioned above, were produced when Absolute Magnitude (H) and eccentricity (e) were used in the input vector.

A breakup of 70 : 15 : 15 was used in training, validating and testing of the algorithm on the dataset. The results are from the testing period of the ANN.

The ANN had an accuracy of 92% when it was trained over 160 sets of data from the above mentioned sources. The accuracy of the algorithm was 80% (rounded off to the nearest ones) for around 100 sets of data. It gradually increased to, as mentioned above, 92% accuracy. The upward tend in figure 2 makes it certain that the accuracy can be further increased with more data (available in the IAU database). There were just 5 false positives and 8 false negatives.
The Matthews Correlation Coefficient was 0.821 indicating a strong relation between the input data.

Based on the classification, the NEO was either predicated as within 0.03 AU or no.

Output	**Prediction**
0	$>$ 0.03 AU
1	$<$ 0.03 AU

Table 2 : Output Significance

The results have been summarised in Table 3. The result is from the confusion matrix generated by the MATLAB ANN toolbox. Matthews Correlation Coefficient, sensitivity and specificity were calculated from the confusion matrix.

<table>
<tr><td colspan="5">Overall Performance of the ANN</td></tr>
<tr><td colspan="2">Positive</td><td colspan="2">Negative</td><td>Sensitivity : 86.4%</td></tr>
<tr><td>TP : 51</td><td>FP : 5</td><td>TN : 92</td><td>FN : 8</td><td>Specificity : 94.8%</td></tr>
<tr><td colspan="2">Positive Prediction Rate : 91%</td><td colspan="2">Negative Prediction Rate : 92.0%</td><td rowspan="2">Matthews Correlation Coefficient : 0.821</td></tr>
<tr><td colspan="4">Inconclusive : 0</td></tr>
<tr><td colspan="5">Overall Accuracy : 92%</td></tr>
</table>

Table 3 : Summary of the results from the ANN

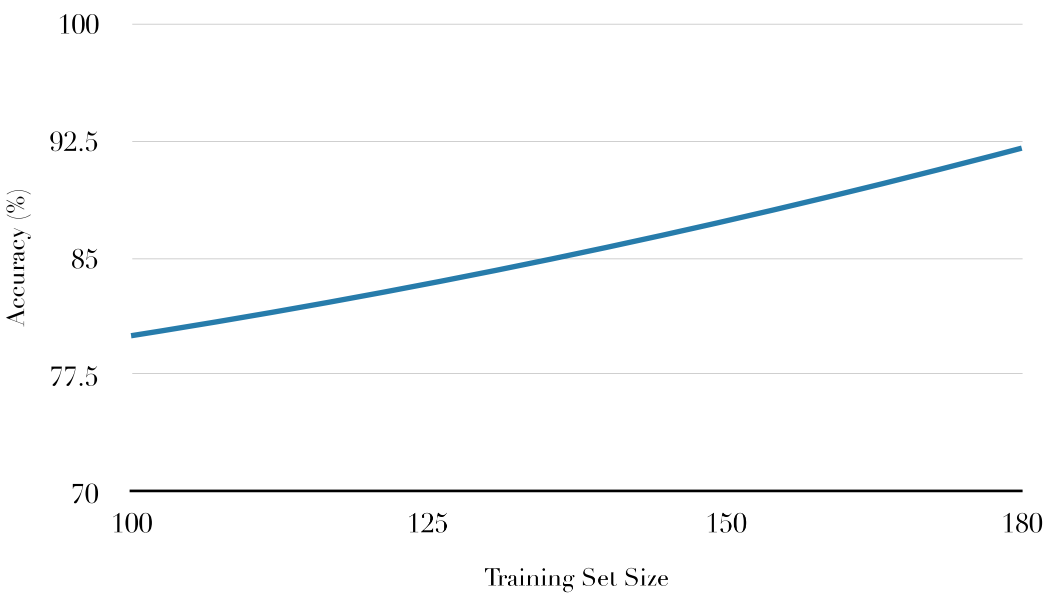

Figure 2 : Comparison of the accuracy of the ANN and the training set size

3.2 Unconstrained Minimisation using fminunc

The fminunc function is based on unconstrained minimisation. It takes in the cost, the gradient and an initial set of parameters as the input and outputs a set of parameters which classify the data.

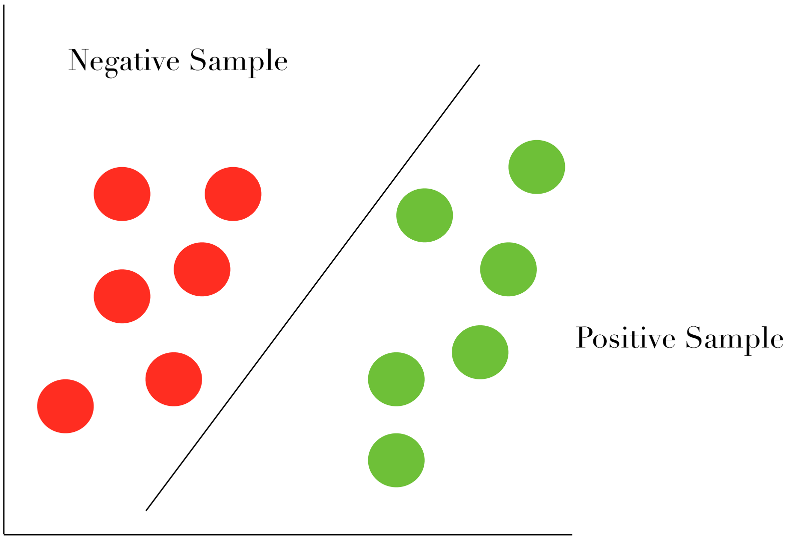

Figure 3 : An example of a logistic classification

3.2.1 Feature Selection

The data for the features was collected from the IAU database and Caltech's JPL browser. The features listed in table 1 were used to classify the data. On the basis of ease of calculability of the physical parameters of new NEOs and accuracy of the ANN, absolute magnitude (H) and eccentricity (e) were selected after each of the features listed above were used by the ANN.

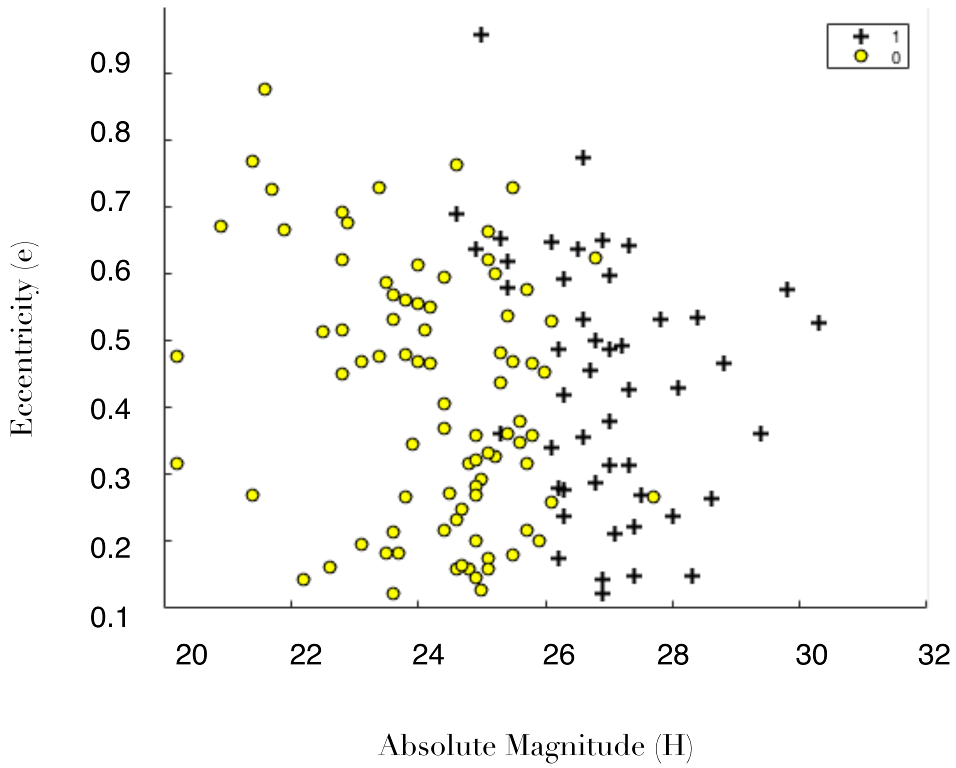

Figure 4 : Eccentricity V/S Absolute Magnitude for the data to be classified (scaled down version)

3.2.2 Algorithm Development

The algorithm to calculate the cost and the gradient was custom coded in MATLAB. MATLAB's inbuilt libraries were used for the fminunc function. The classification model is not regularised.
The system is also under development in Python 3.

3.2.3 Results

The algorithm was used over 180 sets of data to train itself on classifying the NEOs as dangerous (< 0.03 AU) or no. In a single test run, 2 features were used in the input layer (see Figure 1). The most accurate results, as mentioned above, were produced when Absolute Magnitude (H) and eccentricity (e) were used in the input vector.

The algorithm was trained over the entire data set. Accuracy was also measured as the deviation of all data points from the classifier. The results are from the testing period. The system had an accuracy of 90% when it was trained over 160 sets of data from the above mentioned sources. With an increase in the size of the training set, the accuracy of the algorithm gradually increased. With the large amount of data available in the IAU database, the accuracy can further be increased. The false positive and false negative rates as well as the inconclusive rates were low.

The results have been summarised in Table 4.

<table>
<tr><td colspan="5">Overall Performance</td></tr>
<tr><td colspan="2">Positive</td><td colspan="2">Negative</td><td>Sensitivity : 84.7%</td></tr>
<tr><td>TP : 50</td><td>FP : 7</td><td>TN : 90</td><td>FN : 9</td><td>Specificity : 92.7%</td></tr>
<tr><td colspan="2">Positive Prediction Rate : 88%</td><td colspan="2">Negative Prediction Rate : 91.0%</td><td rowspan="2">Matthews Correlation Coefficient : 0.780</td></tr>
<tr><td colspan="4">Inconclusive : 0</td></tr>
<tr><td colspan="5">Overall Accuracy : 90%</td></tr>
</table>

Table 4 : Summary of the results

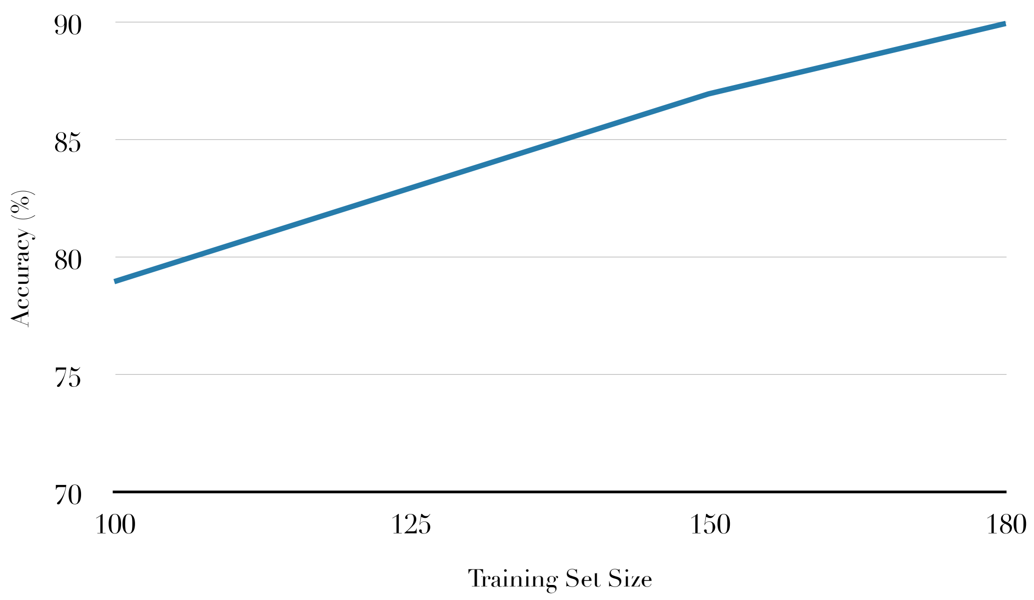

Figure 5 : Comparison of the accuracy of the fminunc and the training set size

4 Minimum Distance of Approach

4.1 Artificial Neural Network

A sample architecture of a multi layer, multi neuron artificial neural network is figure 1. A similar structure (different number of neurons in the hidden layer) was used for predicting the minimum distance upto which an NEO could approach planet Earth.

4.1.1 Feature Selection

From the Minor Planet Center Database provided by The IAU and the Caltech's JPL Small-Body Database Browser, the features listed in table 1 were used by the algorithm. The Absolute Magnitude (H) and eccentricity (e) produced the most accurate results.

4.1.2 ANN Development

The neural network was developed using a the Neural Network Fitting tool in MATLAB. The two features listed above were used as inputs. The ANN was three layered wherein the hidden layer contained ten neurons. Figure 1 illustrates a three layered, 4 hidden layer neuron artificial neural network architecture.

The Levenberg - Marquardt Algorithm was used to train the Neural Network.

4.1.3 Results

The Artificial Neural Network used over 250 sets of data to train itself on predicting the minimum distance upto which an NEO could approach the planet Earth. In a single test run, 2 features were used in the input layer (see Figure 1). The most accurate results, as mentioned above, were produced when Absolute Magnitude (H) and eccentricity (e) were used in the input vector.

A breakup of 70 : 15 : 15 was used in training, validating and testing of the algorithm on the dataset. The results are from the testing period of the ANN.

The algorithm was highly accurate. The mean squared error, that is the average squared distance between the output and the target, was 0.00090. A value of 0 of the MSE means no error. The regression value, that is the correlation between targets and outputs, was 0.79. A value of 1 means a close relation.

The results are have been summarised below (Table 5).

Overall Accuracy	
Mean Squared Error : $9.0 * 10^{-4}$	Regression R value : 0.79

Table 5 : Result Summary

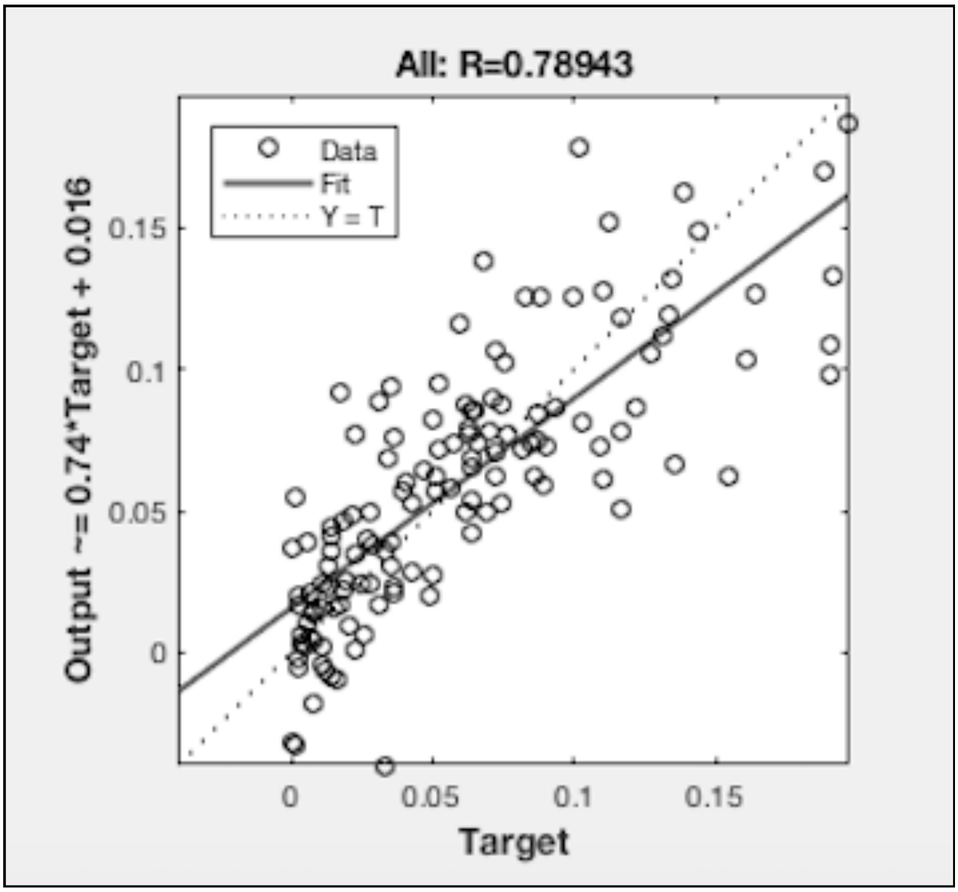

Figure 6 : Regression Fit for the Training Algorithm

5 Conclusion

The algorithms in the research were highly accurate. The artificial neural network based on the backpropagation algorithm achieved an accuracy of 92%. The logistic classification based on unconstrained minimisation using the fminunc function had an accuracy of 90%. An artificial neural network based on the Levenberg - Marquardt Training Algorithm was used to predict the minimum distance of approach. It achieved a regression value of 0.79 and a mean square error value of 0.00090. Moreover, the algorithms addressed the high number of false positives and false negatives in the conventional NEO hunting techniques by achieving very low rates.

As shown in figure 2 and figure 5, the accuracy of the machine learning algorithms increases gradually with increase in the training size. Thus, with more input data, these algorithms can achieve near perfect accuracies.

Data was collected from The Minor Planet Center Database provided by The International Astronomical Union (IAU) and from the California Institute of Technology's Jet Propulsion Laboratory Small-Body Database Browser as they are extremely authentic, reliable and reputed sites of data. Moreover, the algorithms were able to achieve high levels of accuracy due to proper techniques and exhaustive research that went into feature selection. Each algorithm was trained several times with different combination of features and different number of neurons in the hidden layer of the ANN (algorithms used in section 4.1 an 5.1 used ANN) before the features were finalised.

All algorithms were either custom coded on MATLAB and Octave or used via the toolboxes provided by MATLAB. An unconstrained minimisation system used to classify NEOs is also being developed in python. The python coded web application would allow professionals to gather data on new NEOs and take effective measures beforehand.

Since software systems have hardly made their entry in the field of NEO detection at the moment, the algorithmic system developed in this report would come as a boon.

I intend to convert the entire algorithmic system into a web application. Moreover, I am also working on exploring the applications of machine learning in other areas of space exploration and research.

6 Acknowledgements

The project would not have been possible without the constant support from friends and family.

I would also like to thank my maths teacher for his help in elucidating the math behind the machine learning algorithms. I would also like to take this opportunity to thank Andrew Ng for his inspiring course on 'Machine Learning' on Coursera and Herbert Lee for the course on 'Bayesian Statistics : From Concept to Data Analysis ' on Coursera. These courses gave a solid foundation in statistics and machine learning and inspired me to undertake the project.

Finally, I would like to thank IAU and California Institute of Technology for their open source and available to all database on the near earth objects.

7 References

[1] The International Astronomical Union, Minor Planet Center - Minor Planet, Asteroid and Comet database http://www.minorplanetcenter.net/iau/lists/CloseApp.html Web

[3] California Institute of Technology Jet Propulsion Laboratory Small-Body Database Browser http://ssd.jpl.nasa.gov/ Web

[3] Mohnish Chakravarty and Tanay Kothari. 2015. A Comprehensive Study on the Applications of Machine Learning for Diagnosis of Cancer. arXiv.

[4] Nasa Space Apps 2016 https://2016.spaceappschallenge.org/ Facts and Figures Web

[5] K. P. Bennett and O. L. Mangasarian. 1992 .Robust linear programming discrimination of two linearly inseparable sets. Optimization Methods and Software 1:23-34.

[6] Sonali. B. Mind and Priyanka Wankar. 2014. Research Paper on Basic of Artificial Neural Network. ISSN 2(1):96-100.

[7] D.F. Specht. 2002. A General Regression Neural Network. IEEE Transactions on Neural Networks 2(6):568-576.

[8] V.N. Vapnik. 2002. An overview of Statistical Learning Theory. IEEE Transactions on Neural Networks 10(5):988-999.

[9] M.T. Hagan and M.B. Menhaj. 2002. Training Feedforward Networks with the Marquardt Algorithm. IEEE Transactions on Neural Networks 5(6):989-993.

[10] C. R. Nugent, D. Tholen and R. Walker. 2012. Physical Parameters of Asteroids Estimated from the WISE 3 Band Data and NEOWISE Post-Cryogenic Survey.

[11] https://en.wikipedia.org/wiki/Near-Earth_object Facts Web

[12] Shital Sonlanki and H.B. Jethwa. 2013. A Review on Backpropagation Algorithms for Feedforward Networks. ISSN 2(1):73-75

[13] Dr. David R. Williams. NASA Goddard Space Flight Center. Asteroid Fact Sheet. Web

[14] Pedro Domingos. A Few Useful Things to Know about Machine Learning.