Hazardous asteroids forecast via Markov Random Fields: A case study for Explainable Artificial Intelligence (XAI)

Introduction

Research issue

Can machine learning algorithms be safely used in physics?

Literature

"As machine learning is incorporated into the physicist's toolbox, it is reasonable to expect that physicist may shed light on some of the notoriously difficult questions machine learning is facing. Specifically, **physicists are already contributing to issues of interpretability**, *techniques to validate or guarantee the results, and principle ways to choose the various parameters of the neural networks architectures"* Carleo et al. Machine learning and the physical sciences. Reviews of Modern Physics, 2019 [8]

Definitions

Machine Learning

"Machine learning (ML) is a branch of artificial intelligence (AI) and computer science that focuses on the using data and algorithms to enable AI to imitate the way that humans learn, gradually improving its accuracy" IBM [1]

Intepretability

"The language of S can be translated into the language of T in such a way that T proves the translation of every theorem of S" (Tarski et al.) [16]

Explainable artificial intelligence (XAI)

"Explainable artificial intelligence (XAI) is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms" IBM [2]

Approaches for intepretability

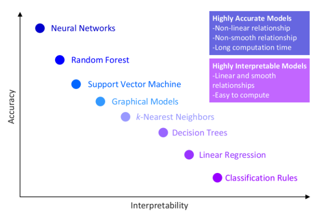
Intrinsecally intepretable

- Regressions (Linear, Quadratic, Logistic, etc.)
- Graphical Methods: Markov Random field or Bayesian Networks

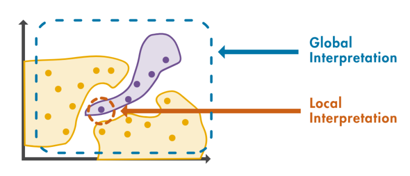
Ex-Post

- Global Methods (e.g. Partial dependence plots PDP)
- Local Methods (e.g. Local interpretable model-agnostic explanations - LIME)

Trade-off [13]



Global vs. local [3]



Graphical models: Conditional dep. into a graph

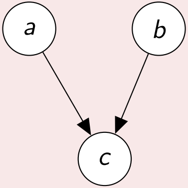
Independence [15, 14]

A is independent of event B if $P(A, B) = P(A)P(B)$

Conditional Independence

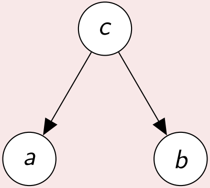
A and B are conditionally independent given event C if $P(A, B|C) = P(A|C)P(B|C)$ or $A \perp B|C$

Collider



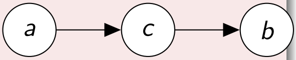
$p(a, b, c) =$
$p(a)p(b)p(c|a, b)$

Fork



$p(a, b, c) =$
$p(a|c)p(b|c)p(c)$

Chain



$p(a, b, c) =$
$p(a)p(c|a)p(b|c)$

Dataset

Source

The asteroid dataset was retrieved from Kaggle [4], which reports into a more machine-readable form the dataset of The Center for Near-Earth Object Studies (CNEOS) [5], a NASA research centre.

Description

- 3552 asteroids (the proportion hazardous/not hazardous was 1:5)
- Among the 40 features, the ones connected only to the other name of the asteroid or connected only to the name of the orbit and the one connected with the orbiting planet were discarded

Definition [6]

- NEAs whose Minimum Orbit Intersection Distance (MOID) with the Earth is 0.05 au or less and whose absolute magnitude (M) is 22.0 or brighter

Methods

Algorithms

We considered the random forest, neural networks, support vector machines, and Markov random fields. The random forest is interpretable but does not provide the relation between the variables, the neural networks and the SVM are not interpretable, and the Markov random field is interpretable and provides the conditional dependence between the variables.

Computational details

Except for the Markov Random Fields, all the algorithms were trained with the caret R package (train/test splitting 0.6 / 0.4). A 5-fold cross-validation was considered.

Intepretation

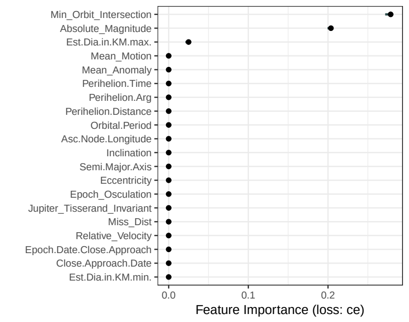
Except for the Markov Random Fields, all the algorithm's final models were interpreted with variable importance and partial dependence plots as implemented in the *iml* R package [12]

Random Forest (ranger)

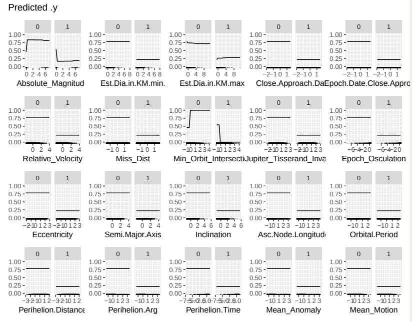
Description

Algorithm: Random forest as implemented in the *ranger* R package [17]. The Matthews correlation coefficient ϕ was 0.98.

Variable Importance



Partial Dependence Plots

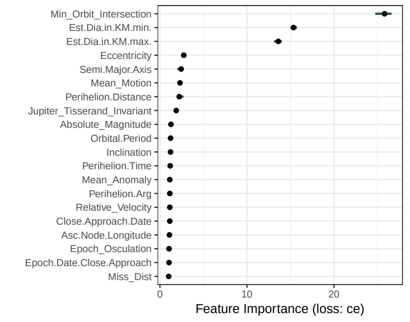


Multilayer perceptron

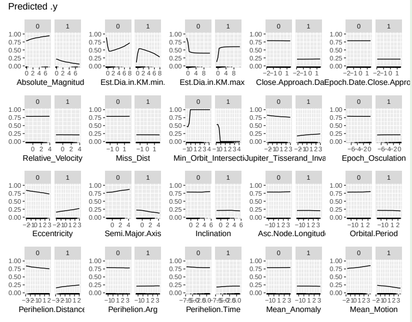
Description

Algorithm: Multilayer perceptron [7] with five hidden layers as implemented in the *mlp* R package [17]. The Matthews correlation coefficient ϕ was 0.93.

Variable Importance



Partial Dependence Plots

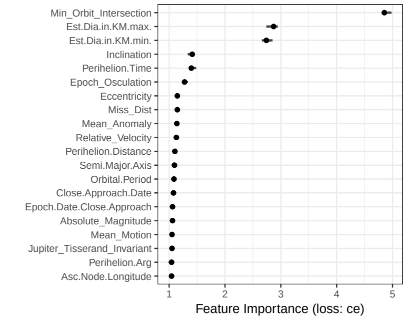


Support Vector Machine

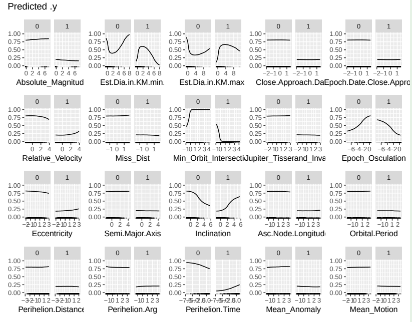
Description

Algorithm: Support Vector machines with a linear kernel [10]. The linear kernel was considered to have an algorithm that works with a lower performance than the previous ones. This was done to have a counterfactual for the interpretation. The Matthews correlation coefficient ϕ was 0.79.

Variable Importance

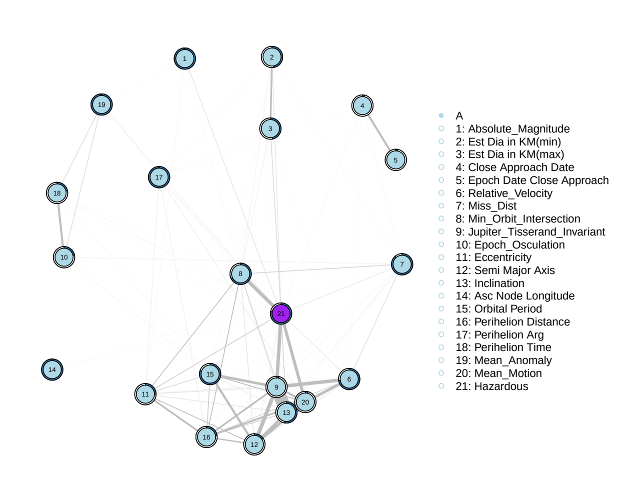


Partial Dependence Plots



Markov Random Field

Plot

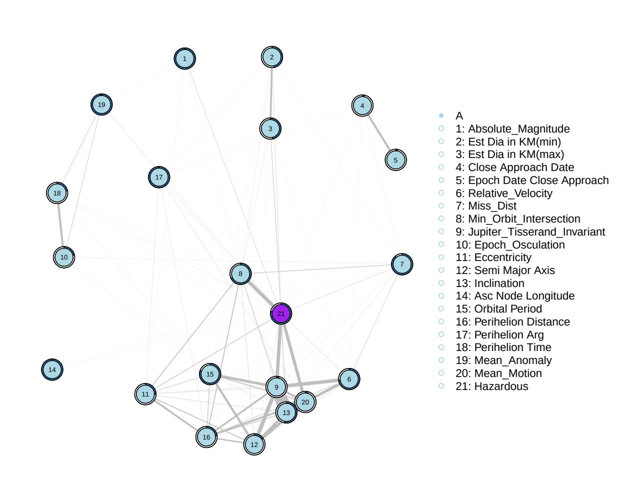


Description

Algorithm: Mixed graphical model [9]. The Matthews correlation coefficient ϕ was $\phi = 0.58$

Markov Random Field

Plot



Outcome

The algorithm correctly captured the definition. Some links between features are correct, others not. The critical point is that the algorithm's decision process is fully interpretable and then verifiable.

Conclusions

- Markov Random Fields (MRF) correctly captured the definition for hazardous asteroids.
- With respect to other algorithms which can be interpreted with PDP and variable importance (such as the Random Forest, Neural Network or the SVM), MRF provides a highly interpretable model and a graph which connects or does not the input features.
- The graphs of MRF can provide a helpful aid for cases (datasets) in which the underlying physics is not already clear.
- Since mistakes may be present, **the connections provided by the graphs should be validated by deriving them from a consolidated theory**.

Outlook & Contact

Outcome

We aim to extend this approach to more complicated datasets in which the underlying physics is only partially explored. We are interested in collaborations.

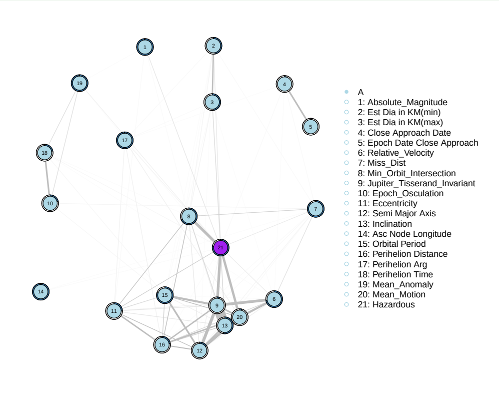
Contact Information

dr. Marzio De Corato (Dep. Computer Science UNIMI) marzio.decorato@unimi.it
prof. Alfio Ferrara (Dep. Computer Science UNIMI) alfio.ferrara@unimi.it
prof. Silvia Salini (Dep. of Economics, Management and Quantitative Methods UNIMI) silvia.salini@unimi.it
dr. Claudia Vello (Dep. of Physics UNIMI) claudia.vello@studenti.unimi.it

Supporting Information

Discussion on MRV model

Plot

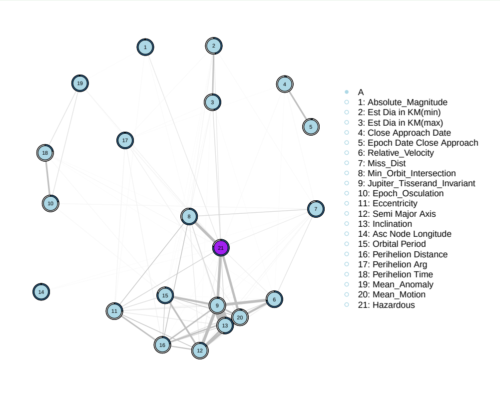


8 - 21, Min Orbit Intersection - Hazardous

The Minimum Orbit Intersection Distance (MOID) is an astronomical measure used to evaluate the risk of close encounters and potential collisions between celestial objects. It represents the smallest distance between the nearest points of the osculating orbits of two bodies.

Discussion on MRV model

Plot

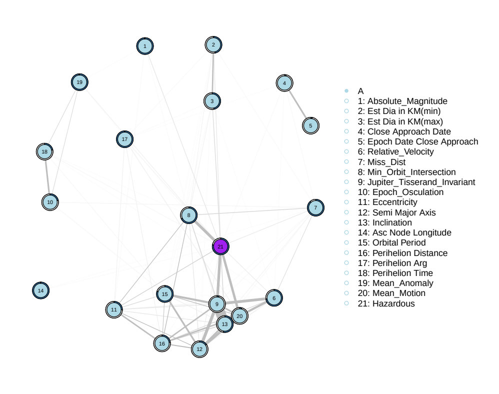


13 - 21, Inclination - Hazardous

Orbital inclination *i* measures the tilt of an asteroid's orbit relative to the plane of Earth's orbit (the ecliptic). Asteroids with high inclinations (orbits significantly tilted away from the ecliptic) are less likely to intersect Earth's path, reducing their overall hazard.

Discussion on MRV model

Plot



11 - 15, Eccentricity - Orbital Period

This connection has no physical background. Indeed $T = (2\pi a^{3/2})/\sqrt{(}Gm_s)$, where m_s is the mass of hosting planet/star and a the semimajor axis. The good point is that such spurious connection used by the model is clear (and in principle can be corrected)

Discussion on MRV model

1 - 21, Absolute Magnitude - Hazardous

The connection between absolute magnitude and hazardousness is weakly highlighted in the graph. However, for an asteroid to be classified as potentially hazardous, it must be large enough to cause significant damage upon impact with Earth, generally having an absolute magnitude H that corresponds to a diameter of about 140m or more. While absolute magnitude itself does not directly measure hazardousness, it offers critical insights into the asteroid's size and brightness, which are important for evaluating impact risk.

Discussion on MRV model

9 - 21, Jupiter Tisserand Invariant - Hazardous

The Tisserand parameter is used to assess whether an object's orbit is primarily controlled by Jupiter or if it may be shifting between different orbital classifications, such as from a comet-like orbit to an asteroid-like orbit. It is given by: $T_J = \frac{a_J}{a} + 2\sqrt{\frac{a}{a_J}(1 - e^2)}\cos i$ where a_J is the semi-major axis of Jupiter's orbit, a is the semi-major axis of the small body's orbit, e is the eccentricity of the small body's orbit and I is the inclination of the small body's orbit relative to Jupiter's orbit. From this equation, we can also confirm the relationships between the Tisserand parameter and the semi-major axis (13), as well as its relationship with the inclination (12). Asteroids with $T_J > 3$ are typically stable and stay within the asteroidbelt, while those with $T_J < 3$ (e.g., comets and Near-Earth Objects) oftenhave unstable, eccentric orbits that can cross Earth's path.

Methods - Performance metrics

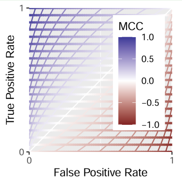
Performance metrics

The Matthew Correlation Coefficient ϕ was considered to assess the algorithms' performances. This is defined as

$$\phi = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where TP/FP is the number of true/false positives, and TN/FN is the number of true/false negatives. The definition can be generalized for a multi-class case

Graphical interpretation [11]



Bibliography I

[1] https://it.mathworks.com/discovery/interpretability.html.

[2] https://www.ibm.com/topics/explainable-ai.

[3] https://www.ibm.com/topics/machine-learning.

[4] https://www.kaggle.com/shrutimehta/nasa-asteroids-classification.

[5] https://cneos.jpl.nasa.gov/.

[6] https://cneos.jpl.nasa.gov/about/neo_groups.html.

[7] Christoph Bergmeir e José M. Benítez. "Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS". In: *Journal of Statistical Software* 46.7 (2012), pp. 1–26.

[8] Giuseppe Carleo et al. "Machine learning and the physical sciences". In: *Reviews of Modern Physics* 91.4 (2019), p. 045002.

Bibliography II

[9] Jonas M. B. Haslbeck e Lourens J. Waldorp. “mgm: Estimating Time-Varying Mixed Graphical Models in High-Dimensional Data”. In: *Journal of Statistical Software* 93.8 (2020), pp. 1–46. DOI: 10.18637/jss.v093.i08.

[10] Alexandros Karatzoglou et al. “kernlab – An S4 Package for Kernel Methods in R”. In: *Journal of Statistical Software* 11.9 (2004), pp. 1–20. DOI: 10.18637/jss.v011.i09.

[11] David Lovell et al. “Never mind the metrics-what about the uncertainty? Visualising binary confusion matrix metric distributions to put performance in perspective”. In: *International Conference on Machine Learning*. PMLR. 2023, pp. 22702–22757.

Bibliography III

[12] Christoph Molnar, Bernd Bischl e Giuseppe Casalicchio. "iml: An R package for Interpretable Machine Learning". In: *JOSS* 3.26 (2018), p. 786. DOI: 10.21105/joss.00786. URL: https://joss.theoj.org/papers/10.21105/joss.00786.

[13] Manuel Eugenio Morocho-Cayamcela, Haeyoung Lee e Wansu Lim. "Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions". In: *IEEE access* 7 (2019), pp. 137184–137206.

[14] Kevin P. Murphy. *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023. URL: http://probml.github.io/book2.

[15] Kevin P. Murphy. *Probabilistic Machine Learning: An introduction*. MIT Press, 2022. URL: probml.ai.

Bibliography IV

[16] A. Tarski, A. Mostowski e R.M. Robinson. *Undecidable Theories*. Studies in logic and the foundations of mathematics. North-Holland, 1953. ISBN: 9780444533784. URL: https://books.google.it/books?id=XtLbjZjB1B8C.

[17] Marvin N. Wright e Andreas Ziegler. "ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R". In: *Journal of Statistical Software* 77.1 (2017), pp. 1–17. DOI: 10.18637/jss.v077.i01.