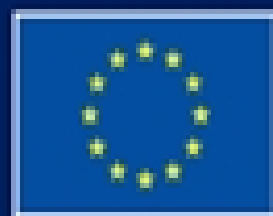


BIDUSA

APPLYING MACHINE LEARNING TO
REAL-WORLD DATA CHALLENGES



Cofinanciado por la
Unión Europea

The BIDUSA project (Big Data Unites the Sciences and the Arts) is co-financed by the European Union. The opinions and viewpoints expressed in all publications are solely those of their authors (Ítaca High School) and do not necessarily reflect those of the European Union or the Spanish Service for the Internationalization of Education (SEPIE). Neither the European Union nor the SEPIE National Agency can be held responsible for them.



Exoplanet Habitability Analysis with Machine Learning Algorithms in Weka

Authors:

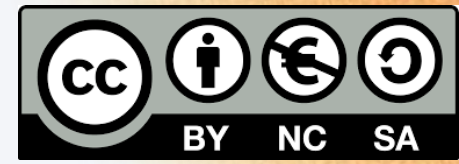
David Jesús Fernández Olmos

Laura Burgos Villegas

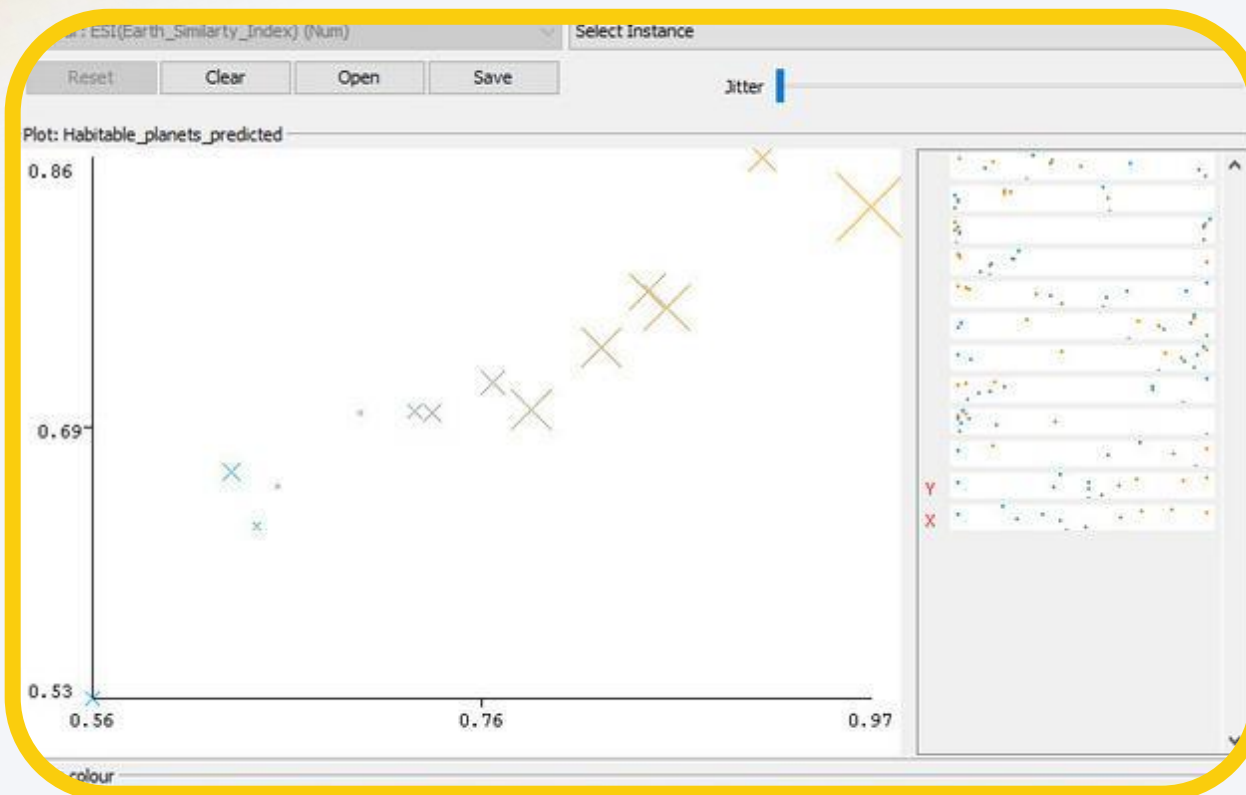
June 10, 2025



Objective of the study

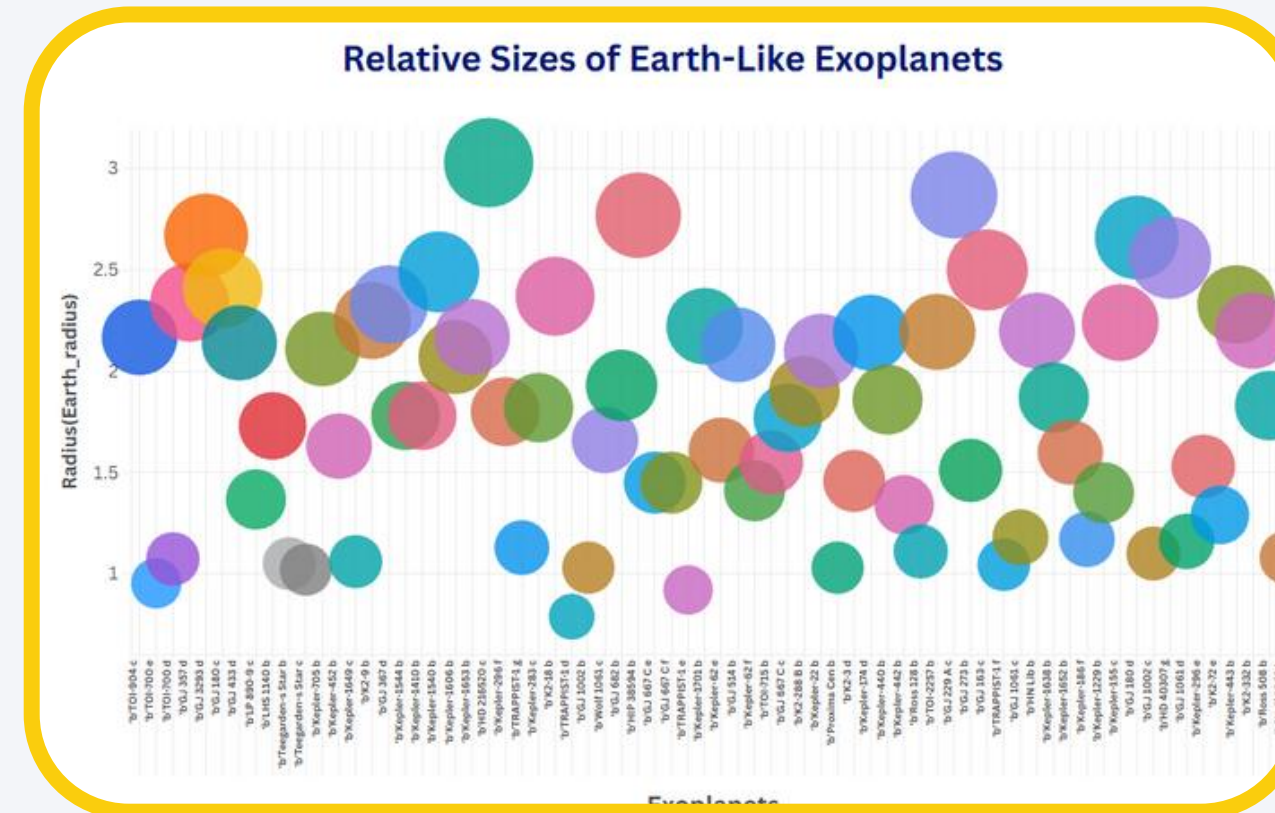


Investigate the habitability potential of exoplanets using machine learning algorithms, supported by the Earth Similarity Index (ESI) and the analysis of key physical variables such as radius, mass and distance to their star.



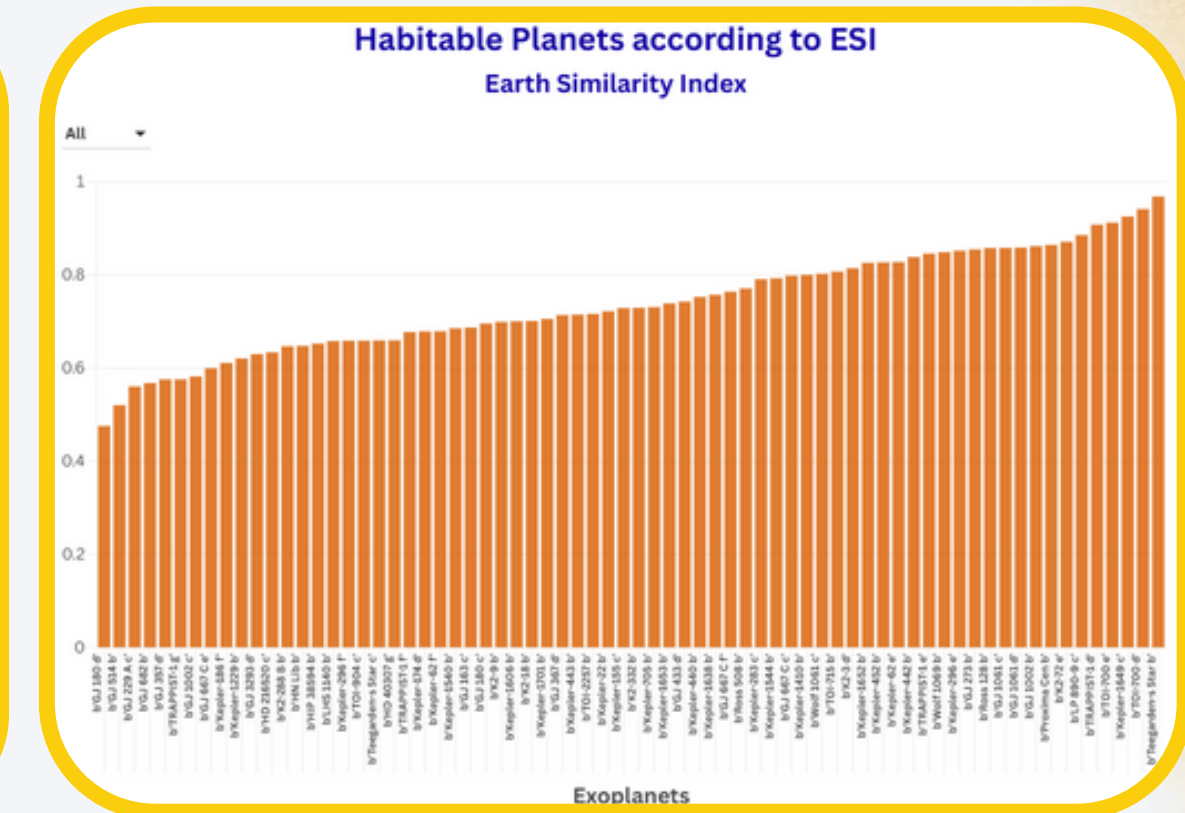
Apply machine learning models in Weka

Algorithms such as Random Forest, SVM, Multilayer Perceptron and KNN were trained to classify exoplanets according to their similarity to Earth.



Evaluate physical and orbital patterns

Analyse characteristics such as mass, radius, orbital distance and temperature to determine their similarity to Earth.



Visualising comparisons with Earth

Use interactive graphs to represent Earth similarity (ESI), showing which planets stand out as potentially habitable.

What is the Earth Similarity Index (ESI)?

The ESI is a mathematical index that assesses how similar an exoplanet is to Earth based on key physical variables.

Its value ranges from 0 (no similarity) to 1 (identical to Earth).

$$ESI = \prod_{i=1}^n \left(1 - \left| \frac{x_i - x_{i0}}{x_i + x_{i0}} \right| \right)^{\frac{w_i}{n}}$$

Processing and methodology



Data Format

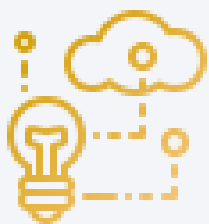
The original data were in CSV format.

They were converted to ARFF format for processing by WEKA.



Software used

- WEKA - For applying machine learning algorithms.
- Microsoft Excel - For cleaning, conversion and pre-processing.
- Flourish Studio - For the creation of interactive visualisations.



Normalisation and data processing

- Variable scaling - Magnitudes (mass, radius, temperature) were adjusted to comparable ranges.
- Elimination of null or incomplete values: records with missing information were cleaned to avoid bias.



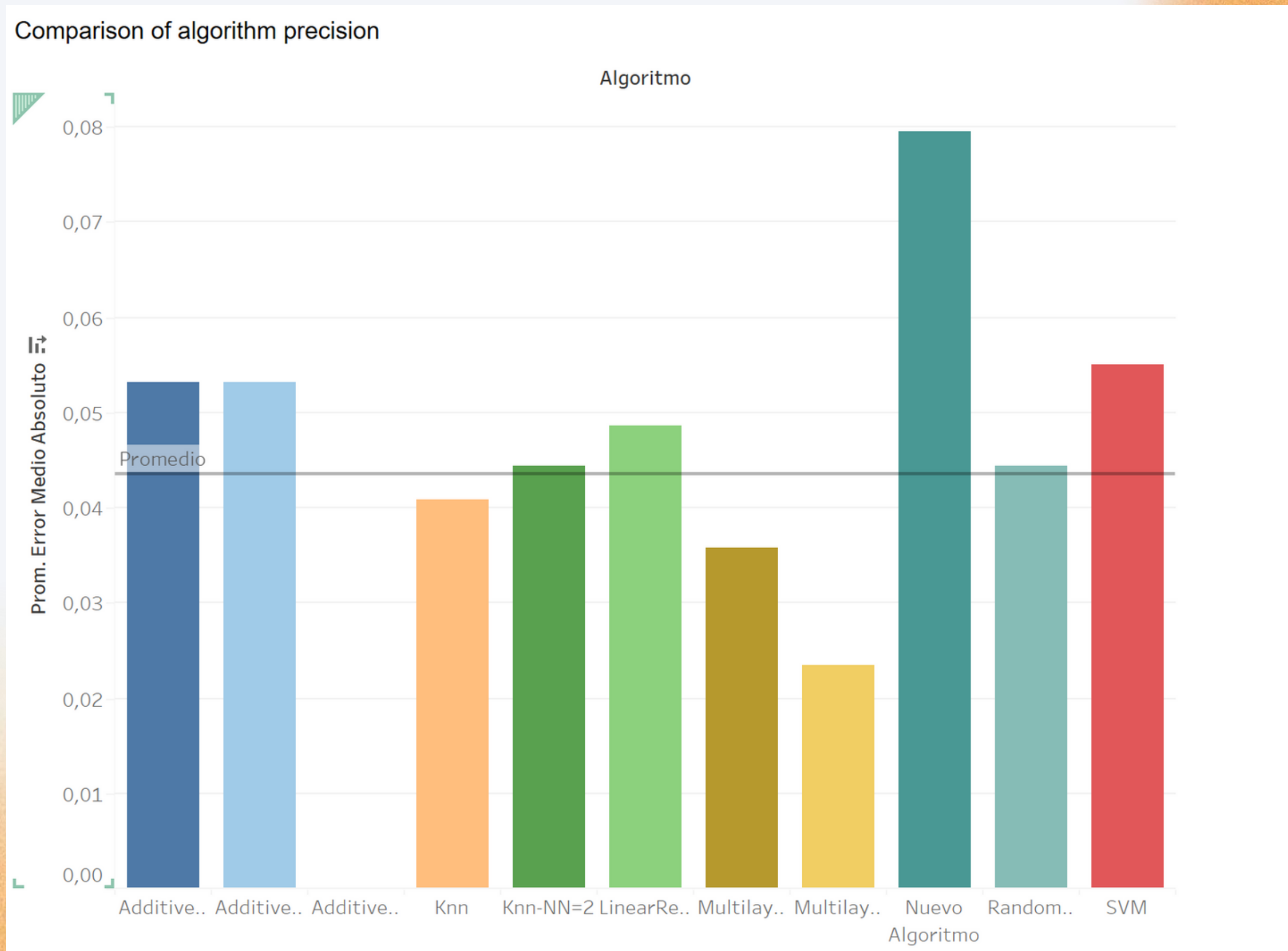
Algorithms used

Algorithm
Multilayer Perceptron (MLP)
Random Forest

Algorithm	Error(%)
JESSI (Java Earth Similarity Index)	1,15%
Additive Regression (RF)	2,41%
MLP (tuned)	2,91%
MLP (untuned)	2,56%
Random Forest	3,21%



Comparison of algorithm precision



JESSI (Java Earth Similarity Index)

The Model

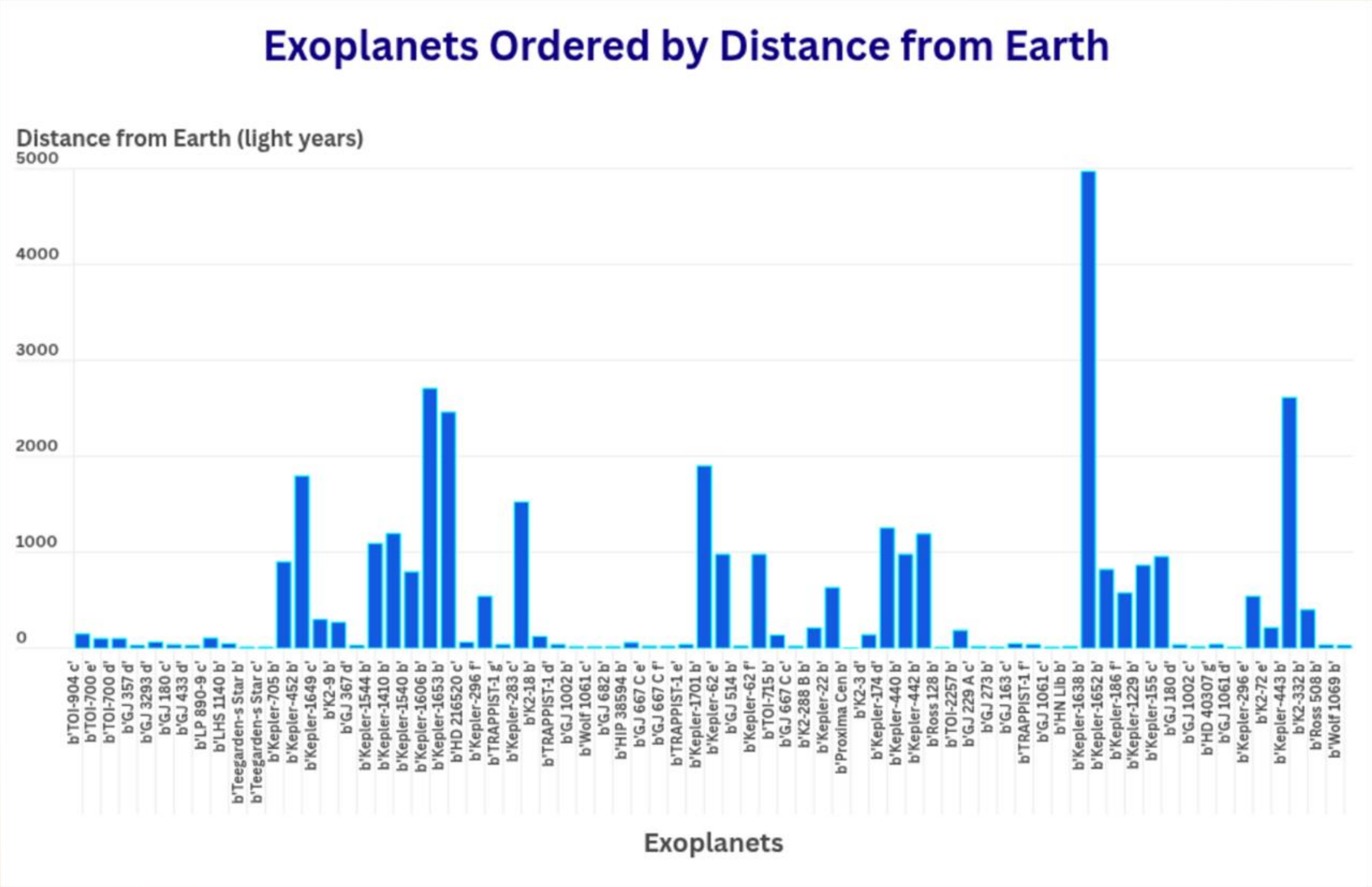
A tooned Stochastic Gradient Descent algorithm made for adjust the weights of the model.

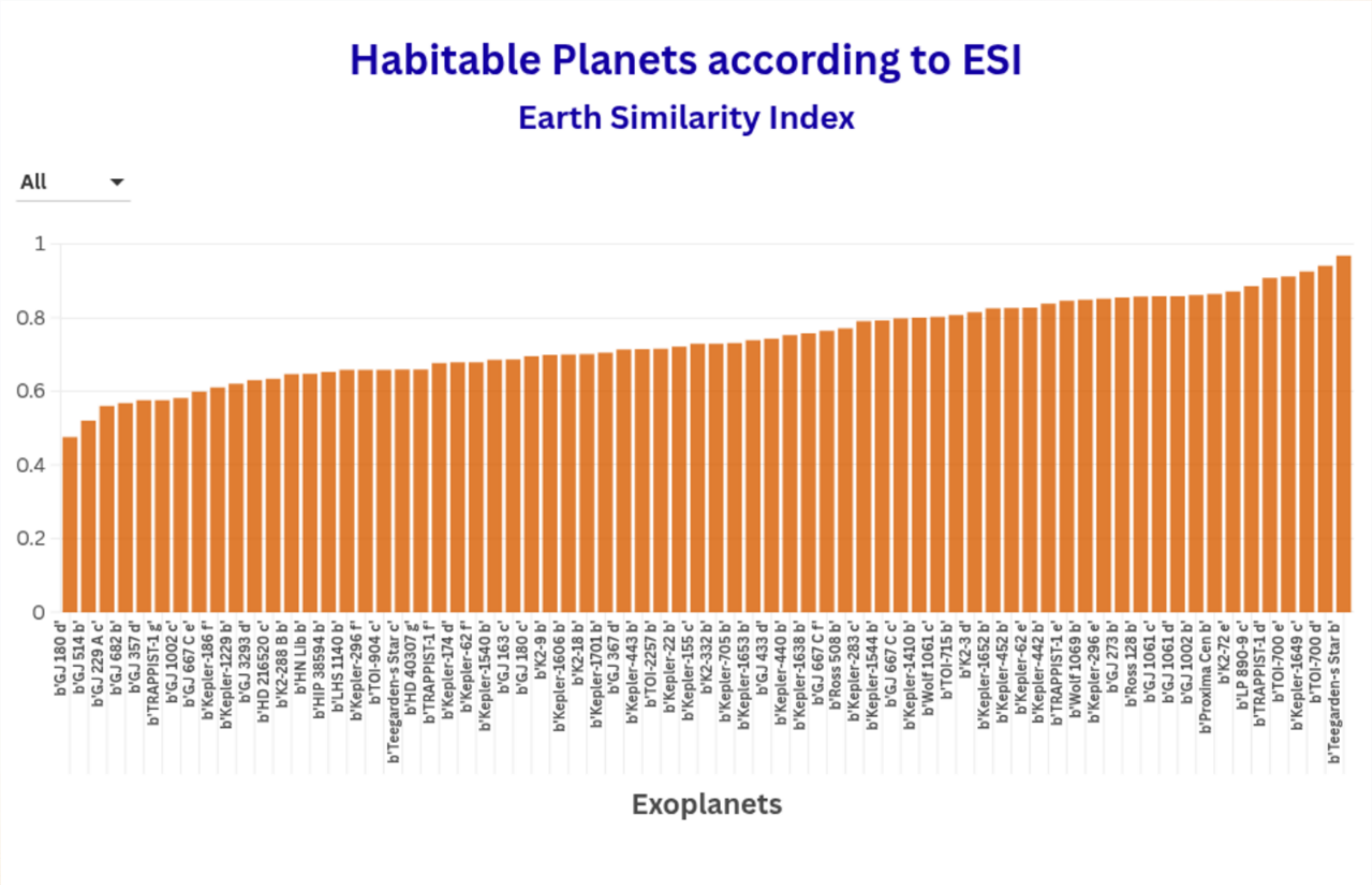
```
/**
 * <h1>esiFormula()</h1>
 * <p>Execute the esi formula</p>
 *
 * @param planet {@link Double[]}
 * @return {@link Double}
 */
private Double esiFormula(Double[] planet){ 1 usage  ⬆ D4vsus
    Double[] earth = modelDT0.getEarth();
    int n = earth.length;
    Double[] weights = modelDT0.getWeights();
    double result = 1.0;

    for (int i = 0; i < n; i++) {
        double parameter_similarity = 1.0 - Math.abs((planet[i] - earth[i]) / (planet[i] + earth[i]));
        double weighted_similarity = Math.pow(parameter_similarity, (weights[i] / n));
        result *= weighted_similarity;
    }
    return result;
}
```

```
64|0,859|0,839|-0,020
65|0,852|0,741|-0,111
66|0,871|0,763|-0,108
67|0,715|0,718|0,004
68|0,729|0,730|0,001
69|0,771|0,795|0,024
70|0,849|0,821|-0,028
```

Quadratic error: 0.011739520137779078



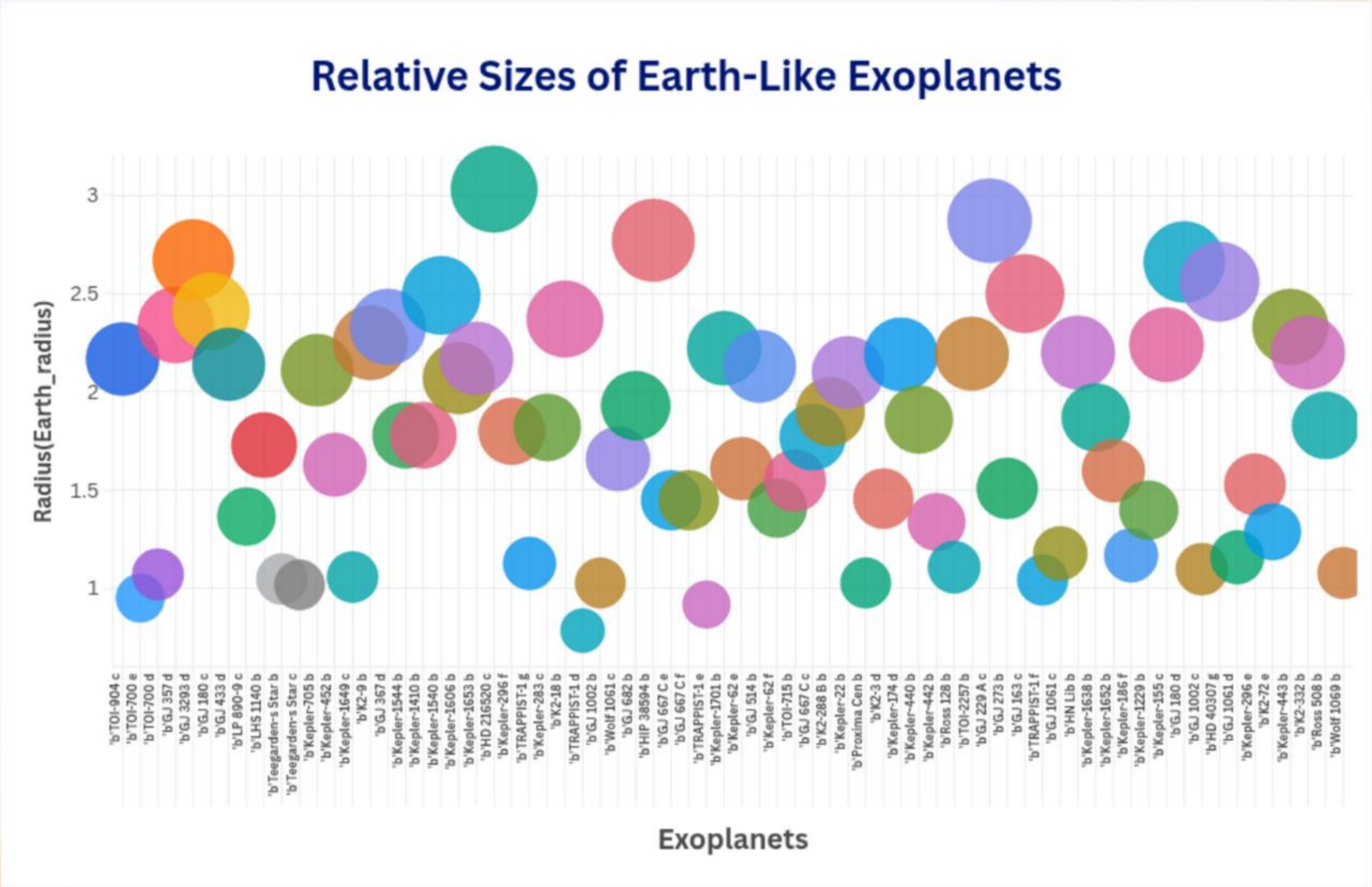


Planetary similarity based on physical properties



Results and analysis

Comparison of sizes with Earth



Results and analysis

Table coloured by ESI

MAIN FINDINGS

🔑 Top 5 Most Habitable Exoplanets (according to ESI)

☒ Habitability Ranking

From most to least habitable, according to the Earth Similarity Index (ESI):

- Teegarden's Star b
- TOI-700 d
- Kepler-1649 c
- TOI-700 e
- TRAPPIST-1 d

Thank you very
much



Resources Page



REFERENCES

Planetary Habitability Laboratory – Habitable World Catalog

University of Puerto Rico at Arecibo

<https://phl.upr.edu/hwc>

NASA Exoplanet Archive – Firefly Atmospheres Tool

California Institute of Technology

<https://exoplanetarchive.ipac.caltech.edu/.../nph-firefly?atmospheres>

Interactive visualisations (own) – Flourish Studio:

- <https://public.flourish.studio/visualisation/22906716/>
- <https://public.flourish.studio/visualisation/23650421/>
- <https://public.flourish.studio/visualisation/23649552/>
- <https://public.flourish.studio/visualisation/22889683/>
- <https://public.flourish.studio/visualisation/23650547/>





Funded by the
Erasmus+ Programme
of the European Union

