

# Извлечение признаков из кривых блеска астрономических источников

А. Д. Лаврухина<sup>1</sup>, К. Л. Маланчев<sup>2,3</sup>

<sup>1</sup>Факультет космических исследований, Московский государственный университет им. М. В. Ломоносова, Ленинские горы, 2-й учебный корпус, Москва, 119991, Россия

<sup>2</sup>Государственный астрономический институт им. П. К. Штернберга, Университетский пр-т. 13, Москва, 119234, Россия

<sup>3</sup>Department of Astronomy, University of Illinois at Urbana-Champaign

## Абстракт

Современные астрономические обзоры содержат информацию о сотнях миллионах кривых блеска переменных астрономических источников, например, релиз данных Zwicky Transient Facility Data Release 3 (ZTF DR3) содержит миллиарды кривых блеска. При решении задач классификации [2] или поиска аномалий [3] в таких больших объемах данных используются методы машинного обучения. Обычно кривые блеска не используются напрямую, вместо чего каждый источник представляется набором признаков, которые наилучшим образом описывают свойства его переменности. В данной работе мы представляем новую библиотеку на языке Python, предназначенную для извлечения признаков из кривых блеска переменных астрономических источников.

## Тестирование библиотеки

В качестве тестовых данных использовались 1855 кривых блеска цефеид и 2394 кривых блеска карликовых новых в красной фотометрической полосе. Кривые блеска взяты из релиза данных ZTF DR3. Цефеиды — пульсирующие звезды-сверхгиганты, периодически изменяющие свою светимость. Карликовые новые — аккрецирующие белые карлики, светимость которых при вспышке за несколько часов возрастает примерно в 100 раз и сохраняется в этом состоянии несколько суток. На Рис. 1 изображены графики кривых блеска для представителей обоих типов.

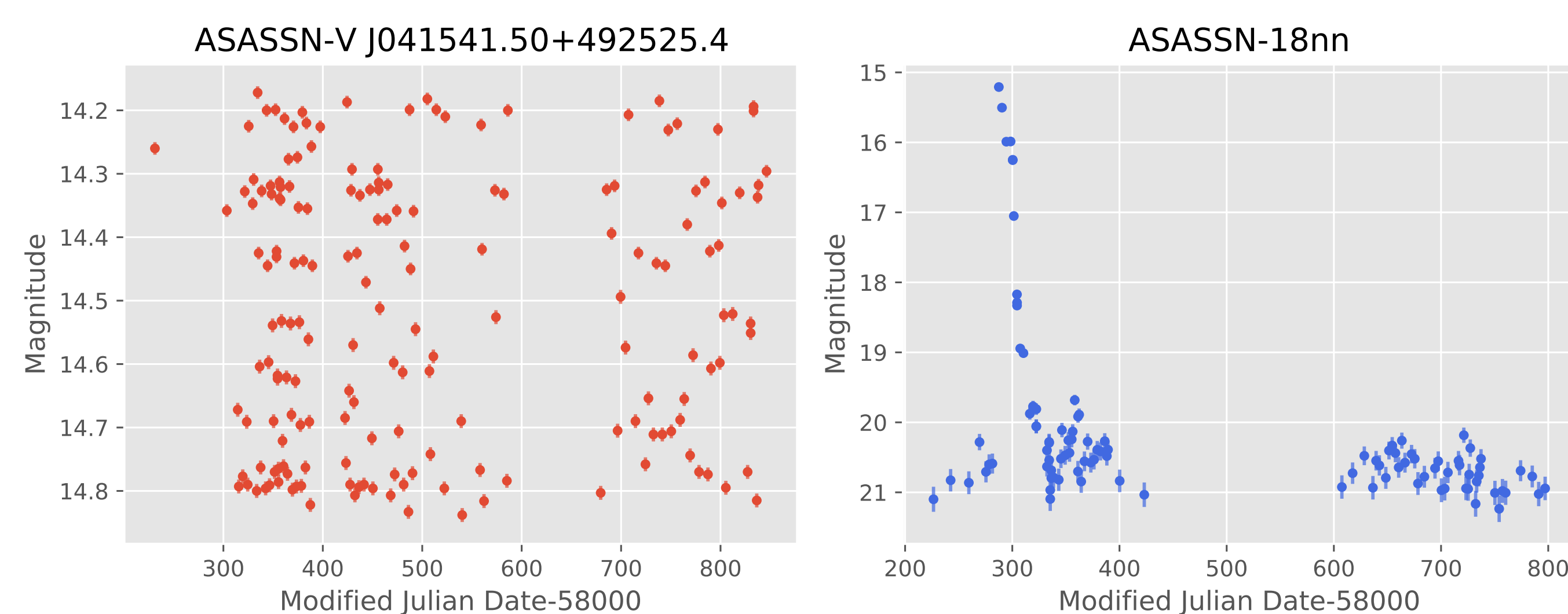


Рис. 1: Графики кривых блеска для цефеиды ASASSN-V J041541.50+492525.4 (красным) и карликовой новой ASASSN-18nn (синим)

Пусть кривая блеска задана набором из  $N$  наблюдений  $\{t_i, m_i, \delta_i\}$ , где  $t_i$  — время наблюдения,  $m_i$  — наблюдаемая звездная величина,  $\delta_i$  — ошибка наблюдения звездной величины. Для построения сравнительных гистограмм, изображенных на Рис. 2, были выбраны следующие признаки:

- 1 **Skew** — коэффициент асимметрии звездной величины

$$\frac{N}{(N-1)(N-2)} \frac{\sum_i (m_i - \langle m \rangle)^3}{\sigma_m^3},$$

где  $\sigma_m$  — среднеквадратическое отклонение звездной величины

- 2 **Inter-percentile range**,  $p = 0.25$  — интерперцентильный размах с  $p = 0.25$  (интерквартильный размах)
- 3 **Beyond  $n$  std** с  $n = 2$  — доля наблюдений, отклонившихся от среднего значения звездной величины более чем на  $n$  среднеквадратических отклонений [1]

- 4 **Median absolute deviation** — медианное значение отклонений звездной величины от ее медианы
- 5 **Amplitude** — полуамплитуда кривой блеска
- 6 **Von Neumann  $\eta$**  [5]

$$\eta \equiv \frac{1}{(N-1)\sigma_m^2} \sum_{i=0}^{N-2} (m_{i+1} - m_i)^2$$

Мы решили тестовую задачу классификации, обучив случайный лес (random forest) с помощью пакета scikit-learn [4] на тренировочном наборе данных (75% выборки, 3186 объектов), составленных из шести представленных признаков кривых блеска. Проверка на тестовом наборе (оставшиеся 25%, 1063 объекта) показала точность работы классификатора в 93%.

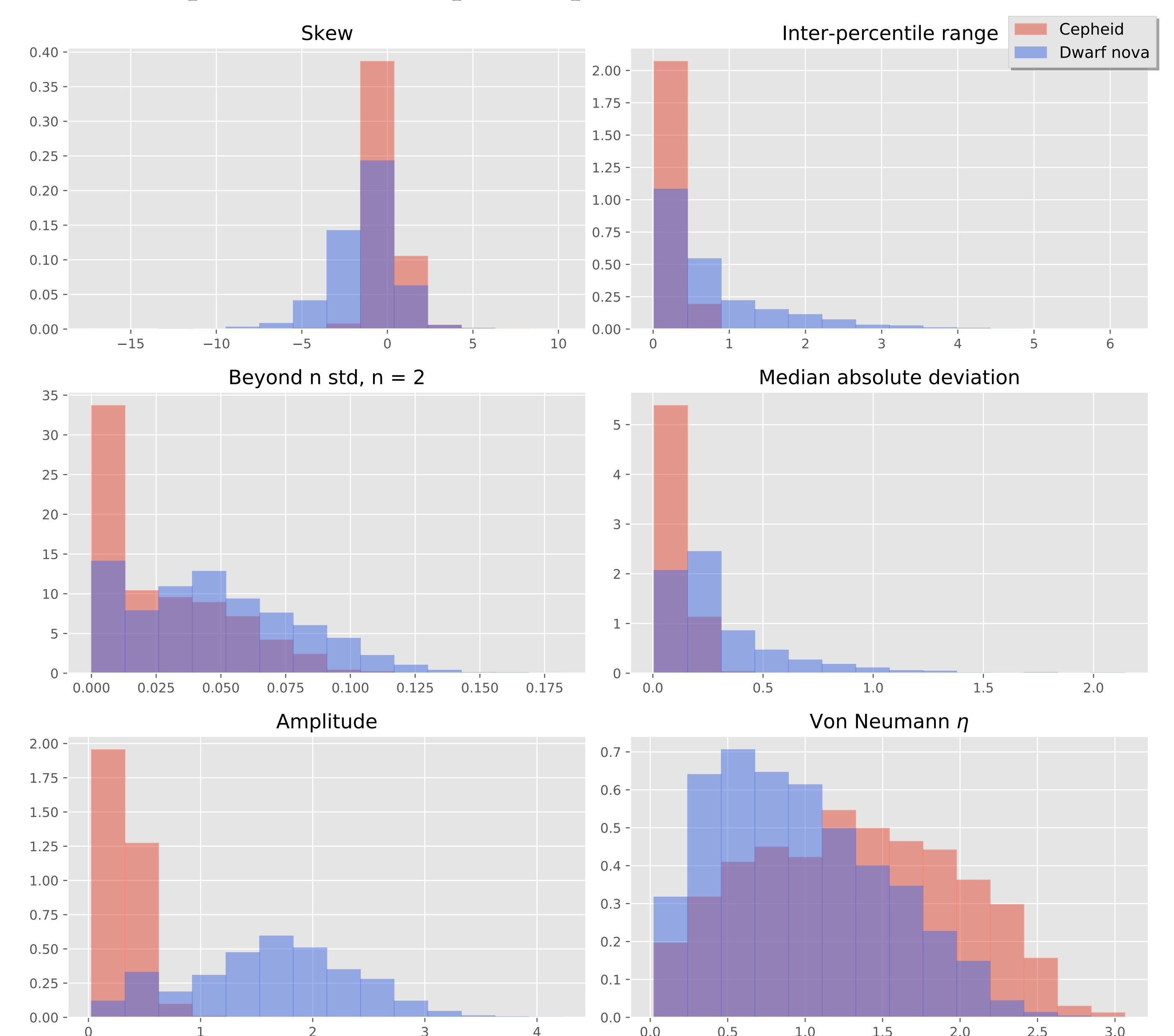


Рис. 2: Гистограммы распределения описанных признаков для цефеид (красным) и карликовых новых (синим)

## Выводы

Использование рассмотренных признаков, вероятно, позволит успешно применять алгоритмы машинного обучения для решения задач классификации или поиска аномалий. В дальнейшем мы планируем реализовать функции для извлечения других признаков, таких как период кривой блеска, а также разработать новые признаки, подходящие для задач классификации. Кроме того, в планах валидация признаков при помощи решения тестовых задач классификации с помощью машинного обучения.

## Ссылки

- [1] A. D'Isanto, S. Cavauoti, M. Brescia, C. Donalek, G. Longo, G. Riccio, and S. G. Djorgovski. An analysis of feature relevance in the classification of astronomical transients with machine learning methods. *Monthly Notices of the Royal Astronomical Society*, 457(3): 3119–3132, 02 2016. ISSN 0035-8711. doi: 10.1093/mnras/stw157. URL <https://doi.org/10.1093/mnras/stw157>.
- [2] D.-W. Kim, P. Protopapas, C. A. L. Bailer-Jones, Y.-I. Byun, S.-W. Chang, J.-B. Marquette, and M.-S. Shin. The epoch project. *Astronomy & Astrophysics*, 566:A43, Jun 2014. ISSN 1432-0746. doi: 10.1051/0004-6361/201323252. URL <http://dx.doi.org/10.1051/0004-6361/201323252>.
- [3] K. L. Malanчев, M. V. Pruzhinskaya, V. S. Korolev, P. D. Aleo, M. V. Kornilov, E. E. O. Ishida, V. V. Krushinsky, F. Mondon, S. Sreejith, A. A. Volnova, A. A. Belinski, A. V. Dodin, A. M. Tatarnikov, and S. G. Zheltoukhov. Anomaly detection in the Zwicky Transient Facility DR3. *arXiv e-prints*, art. arXiv:2012.01419, Dec. 2020.
- [4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [5] J. von Neumann. Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.*, 12(4):367–395, 12 1941. doi: 10.1214/aoms/1177731677. URL <https://doi.org/10.1214/aoms/1177731677>.

## Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-02-00779. Исследование выполнено при поддержке Междисциплинарной научно-образовательной школы Московского университета "Фундаментальные и прикладные исследования космоса"