

В. О. Окунев, Р. М. Хусайнов, Н. Г. Талипов

**КЛАССИФИКАЦИЯ ЭКЗОПЛАНЕТ НА ОСНОВЕ МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ***Ключевые слова: экзопланеты, машинное обучение, классификация, модель машинного обучения, Random Forest.*

*В статье приведено исследование возможности использования методов машинного обучения для классификации экзопланет на основе астрономических данных. Объектом исследования являются экзопланеты, а предметом — подходы к построению и интерпретации моделей классификации экзопланет по их физическим и орбитальным характеристикам. Набор данных Exoplanet Classification Dataset содержит 19761 наблюдение и 16 признаков, включающих параметры звезд, данные фотометрии, а также результирующую метку класса. Разделен набор данных на выборки: обучающая выборка (12646 объектов – 64 %), валидационная выборка (3162 объектов – 16 %), тестовая выборка (3953 объектов – 20 %). Распределено количество объектов по классам: класс 0 — 6311 объектов, класс 1 — 7413 объектов, класс 2 — 6015 объектов, класс 3 — 22 объекта. Выполнена предобработка данных, включающая нормализацию признаков, обработку пропусков и балансировку классов методом SMOTE (Synthetic Minority Over-sampling Technique). Для реализации модели машинного обучения выбран алгоритм Random Forest. Описано сравнение алгоритма Random Forest с другими алгоритмами классификации: логистической регрессией, методом опорных векторов (Support Vector Machine, SVM), градиентным бустингом и простой нейронной сетью (MLP). Проведен сравнительный анализ использования метода SMOTE. Проведена оценка адекватности разработанной модели с использованием метрик точности (Precision), полноты (Recall). Итоговая точность классификации экзопланет на тестовой выборке составила 75 %. На основе полученных моделей определена важность физических признаков, влияющих на принадлежность экзопланет к различным типам, что позволяет интерпретировать результаты не только с точки зрения машинного обучения, но и с позиции астрофизики. Разработанная модель машинного обучения является основой для интеллектуальных систем поддержки научных открытий в современных космических исследованиях.*

V. O. Okunev, R. M. Khusainov, N. G. Talipov

**CLASSIFICATION OF EXOPLANETS BASED ON A MACHINE LEARNING MODEL***Keywords: exoplanets, machine learning, classification, machine learning model, Random Forest.*

*This article presents a study of the feasibility of using machine learning methods to classify exoplanets based on astronomical data. The object of the study is exoplanets, and the subject is approaches to constructing and interpreting exoplanet classification models based on their physical and orbital characteristics. The Exoplanet Classification Dataset contains 19,761 observations and 16 features, including stellar parameters, photometry data, and the resulting class label. The dataset is divided into samples: a training sample (12,646 objects – 64 %), a validation sample (3,162 objects – 16 %), and a test sample (3,953 objects – 20 %). The number of objects is distributed among classes: class 0 — 6,311 objects, class 1 — 7,413 objects, class 2 — 6,015 objects, and class 3 — 22 objects. Data preprocessing was performed, including feature normalization, gap handling, and class balancing using the Synthetic Minority Oversampling Technique (SMOTE). The Random Forest algorithm was selected to implement the machine learning model. A comparison of the Random Forest algorithm with other classification algorithms is described: logistic regression, support vector machine (SVM), gradient boosting, and a simple neural network (MLP). A comparative analysis of the SMOTE method is conducted. The adequacy of the developed model was assessed using the precision and recall metrics. The final accuracy of exoplanet classification on the test set was 75 %. Based on the resulting models, the importance of physical features influencing the classification of exoplanets into different types was determined, which allows the results to be interpreted not only from a machine learning perspective, but also from an astrophysical perspective. The developed machine learning model forms the basis for intelligent systems supporting scientific discoveries in modern space exploration.*

**Введение**

Современная астрофизика характеризуется стремительным ростом объемов данных, получаемых с космических телескопов и наземных обсерваторий. Количество обнаруженных экзопланет увеличивается ежегодно, что требует применения методов автоматического анализа для эффективной обработки и классификации наблюдений [1-3]. Модели машинного обучения зарекомендовали себя как мощный инструмент для распознавания сложных зависимостей в многомерных данных, что делает их особенно перспективными в задачах классификации экзопланет [4-6]. В данной работе рассматривается применение модели машинного обучения для классификации экзопланет на основе данных из

набора Exoplanet Classification Dataset (Kaggle) [7]. Исследуется эффективность различных подходов к построению и обучению модели, а также проводится анализ ее работы на реальных астрофизических характеристиках.

Цель работы заключается в разработке и тестировании модели машинного обучения классификации экзопланет по их физическим и орбитальным характеристикам.

Для достижения поставленной цели решаются задачи:

1. Изучение теоретических основ машинного обучения и их применения в задачах классификации астрономических объектов.

2. Проведение анализа и предобработки данных из Exoplanet Classification Dataset, включая обработку пропусков и нормализацию.

3. Разработка архитектуры и обучение различных видов модели машинного обучения.

4. Проведение сравнительного анализа эффективности обученных моделей.

5. Реализация программного модуля для классификации экзопланет, основанного на полученной модели.

В работе используются методы машинного обучения, статистического анализа данных, методы нормализации и масштабирования признаков, техники регуляризации и кросс-валидации.

Теоретическая значимость работы заключается в применимости подходов машинного обучения к задачам классификации экзопланет и анализе структуры данных, используемых в современной астрономии. Практическая значимость работы заключается в разработке модели, способной автоматически классифицировать экзопланеты, что может быть полезно в научных исследованиях и образовательных целях. Модель машинного обучения целесообразно применять также в качестве инструмента предварительной фильтрации данных при анализе результатов наблюдений, полученных с использованием космических и наземных телескопов, что будет полезно для Российских космических программ по исследованию космоса: «Российская лунная программа» [8], «Федеральная космическая программа» [9].

В случае отсутствия доступа к данным международных архивов, таких как NASA Exoplanet Archive [7], модель может использовать результаты российских обсерваторий, публикуемых в открытых источниках. Разработанная модель не зависит от конкретного зарубежного источника данных, адаптирована для работы в условиях ограниченного доступа к внешним научным ресурсам.

### **Физические и астрофизические характеристики экзопланет и их значимость для классификации**

Экзопланеты представляют собой планеты, вращающиеся вокруг звезд за пределами Солнечной системы. Их изучение позволяет лучше понять процессы формирования планетных систем, архитектуру галактики и потенциальную пригодность планет для жизни. Современные методы наблюдений предоставляют широкий спектр данных о физических свойствах экзопланет, которые используются как признаки для моделей машинного обучения [10, 11].

Ключевые категории характеристик экзопланет:

Орбитальные параметры (определяют характер движения планеты вокруг звезды, включают: период обращения, большая полуось, эксцентриситет, физические параметры планеты).

Период обращения (orbital period) – время одного полного оборота вокруг звезды. Позволяет классифицировать планеты на горячие юпитеры, суперземли, субнептуны и др.

Большая полуось (semi-major axis) – определяет расстояние до звезды, влияющее на климат и температуру.

Эксцентриситет – характеризует форму орбиты; высокий эксцентриситет может указывать на динамическую нестабильность системы [10].

Физические параметры планеты – главные признаки, влияющие на тип объекта: радиус, масса, плотность.

Радиус – один из физических параметров, позволяющих разделять планеты на землеподобные, газовые, ледяные.

Масса – физический параметр, определяющий плотность и внутреннюю структуру планеты.

Плотность – физический параметр, указывающий на состав планеты (каменистая, газовая, водная и др.). Плотность вычисляется из массы и радиуса, является критически важным признаком для классификации.

Свойства звезды-хозяина – параметры, оказывающие прямое влияние на формирование и эволюцию планеты: температура звезды, металличность.

Температура звезды влияет на энергию, получаемую планетой. Радиус и масса звезды связаны с гравитационными процессами в системе. Возраст звезды позволяет определить стабильность планеты и вероятность наличия атмосферы.

Металличность влияет на вероятность формирования крупных планет. Эти признаки важны не меньше самих параметров планет, поскольку косвенно позволяют определить их тип и эволюционную историю.

Фотометрические признаки – это признаки, получаемые из световых кривых, являются важными, поскольку большая часть экзопланет открыта транзитным методом. Световая кривая отражает изменение яркости звезды во времени, и из нее можно извлечь глубину транзита, длительность транзита, наклон орбиты, отношение радиусов планеты и звезды.

Глубина транзита – уменьшение яркости при прохождении планеты перед звездой; используется для оценки радиуса планеты.

Длительность транзита – длительность прохождения, связанная с орбитальной геометрией.

Наклон орбиты определяется по форме транзита и отражает динамику системы.

Отношение радиусов планеты и звезды ( $R_p/R_s$ ) – ключевой параметр в классификации. Машинное обучение позволяет анализировать не только отдельные параметры, но и всю форму световой кривой, что повышает точность классификации [12].

Значимость признаков для модели машинного обучения: каждый параметр представляет собой отдельный признак в обучающем наборе. Модель способна обнаруживать сложные нелинейные зависимости между параметрами, например:

- связь между радиусом планеты и ее плотностью;
- влияние массы звезды на типы планет, формирующиеся вокруг нее;
- зависимость параметров световой кривой от атмосферных особенностей планеты.

Особое внимание уделяется корректной обработке масштабов различных признаков. Например, период обращения измеряется в сутках, а радиус планеты — в радиусах Земли или Юпитера. Такие различия требуют нормализации данных.

Таким образом, понимание астрофизических параметров экзопланет и их взаимодействия является фундаментом для построения высокоточной модели машинного обучения для классификации. Эти характеристики позволяют алгоритму не только различать типы планет, но и выявлять закономерности, связанные с их происхождением и эволюцией [12].

Современные методы машинного обучения занимают ключевое место в автоматизации анализа астрофизических данных, особенно при обработке больших массивов фотометрических наблюдений. Модели машинного обучения позволяют выделять скрытые закономерности в световых кривых, идентифицировать транзитные события и классифицировать экзопланеты по ряду параметров. Их применение значительно ускоряет исследовательские процессы и повышает точность обнаружения слабых сигналов, которые традиционные алгоритмы могут пропустить [13].

#### **Особенности работы с астрономическими данными: дисбаланс классов, шумы и кодирование признаков**

Астрономические данные, используемые для классификации экзопланет, обладают рядом особенностей, которые существенно влияют на качество обучения моделей. В отличие от «земных» выборок, основные трудности здесь связаны с физической природой наблюдений, высокой степенью шума, неполнотой данных и выраженным дисбалансом между реальными экзопланетами и ложными сигналами. Корректная обработка данных является ключевым этапом построения модели классификации.

Одной из значимых проблем является дисбаланс классов. Количество подтвержденных экзопланет значительно меньше числа ложных кандидатов, звездных систем без планет и объектов с недостоверными данными. Например, на тысячу наблюдений может приходиться несколько десятков реальных транзитов. Такой дисбаланс приводит к тому, что модель смещается в сторону предсказания «не планета», исключая редкие, но наиболее значимые случаи [14]. Для компенсации этой проблемы используются методы балансировки. Одним из наиболее эффективных является метод SMOTE (Synthetic Minority Over-sampling Technique) [15], создающий синтетические примеры экзопланетных транзитов путем интерполяции между реально наблюдаемыми кривыми. Это позволяет увеличить представленность класса «экзопланета» без искажения общей структуры данных. В ряде исследований [16, 17] применяется также Random Undersampling – уменьшение количества объектов класса «не планета», что снижает общую нагрузку на модель и повышает точность на редких примерах [15].

Корректная оценка качества моделей, используемых для автоматической идентификации

экзопланет по световым кривым, является важным этапом разработки. В поиске транзитных сигналов особенно критично учитывать влияние различных типов ошибок: ложные срабатывания увеличивают затраты на последующие проверки объекта, а пропущенные планеты уменьшают научную ценность наблюдательных программ [15].

#### **Анализ исходного набора данных и подготовка признаков**

Для обучения модели использован подготовленный набор данных (ML\_ready.csv), содержащий 19761 наблюдение и 16 признаков, включающих параметры звезд, данные фотометрии, а также результирующую метку класса. В частности, датасет включает главные компоненты (PC1–PC3), координаты объекта (RA, DEC), звездные величины в различных диапазонах, характеристики центральной звезды и другие астрофизические параметры [7].

Структура и особенности данных [7] содержат числовые признаки:

- главные компоненты: PC1, PC2, PC3;
- координаты: ra, dec;
- фотометрические параметры: j\_mag, h\_mag, k\_mag, kep\_mag;
- параметры планеты: planet\_radius, orbital\_period;
- параметры звезды: star\_teff, star\_logg, star\_radius, star\_mass;
- временной параметр: lc\_time0.

Целевая переменная label – категориальный признак с четырьмя значениями (0, 1, 2, 3), отражающими тип объекта, где наименование класса 0 – не планета, класс 1 – газовые гиганты, класс 2 – планеты земного и суперземного типа, класс 3 – редкие или атипичные объекты. Наименование классов:

1) класс 0 – не планета (False Positive). Объекты данного класса представляют собой ложные срабатывания транзитного метода, вызванные шумами наблюдений, двойными звездными системами, переменностью звезд или другими астрофизическими эффектами, имитирующими транзит экзопланеты;

2) класс 1 – газовые гиганты. К данному классу относятся экзопланеты с большим радиусом и массой, сопоставимые с Юпитером и Сатурном. Такие планеты, как правило, обладают мощной газовой оболочкой и сравнительно легко обнаруживаются транзитным методом из-за выраженного уменьшения блеска звезды;

3) класс 2 – планеты земного и суперземного типа. Этот класс включает каменные экзопланеты с относительно малым радиусом и массой, близкие по характеристикам к Земле, а также так называемые «суперземли». Их обнаружение является более сложной задачей, поскольку транзитный сигнал имеет меньшую амплитуду;

4) класс 3 – редкие или атипичные объекты. В данный класс входят экзопланеты или кандидаты, встречающиеся в наборе данных в крайне малом количестве. Это могут быть экзотические типы планет, объекты с аномальными орбитальными параметрами или случаи с ограниченным числом подтвержденных наблюдений, что объясняет сильный дисбаланс данного класса.

Первичный просмотр распределения данных выявил дисбаланс классов. Класс 3 представлен всего 22 объектами, что составляет 0,11 % данных. Такой перекося может ухудшать обучение и приводить к исключению редких классов.

Проведена проверка корректности и кодировки данных. На этапе загрузки файл открывался с кодировкой UTF-8, что исключает проблемы несовместимости данных. Также проведен контроль отсутствия некорректных символов, целостности заголовков, корректности типов признаков (все числовые переменные прочитаны как «float» или «int»).

Разделен набор данных на выборки: обучающая выборка (12646 объектов – 64 %), валидационная выборка (3162 объектов – 16 %), тестовая выборка (3953 объектов – 20 %).

Для нормализации диапазонов значений использован StandardScaler ( $x'$ ), приводящий данные к нулевому среднему и единичному отклонению:

$$x' = \frac{x - \mu}{\sigma}, \quad (1)$$

где  $x$  – исходное значение признака;  $\mu$  – среднее значение признака;  $\sigma$  – стандартное отклонение.

Это снижает влияние признаков с большим диапазоном и улучшает сходимость алгоритмов, чувствительных к масштабам. Масштабирование выполнялось только на обучающей выборке, далее преобразование применено к валидационной и тестовой выборкам.

Распределение целевой переменной составило: класс 0 – 6311 объектов; класс 1 – 7413 объектов; класс 2 – 6015 объектов; класс 3 – 22 объекта.

Из полученных данных видно, что отношение минимально представленного класса к наиболее частому составило 0.003, что указывает на крайне серьезный дисбаланс.

### Применение метода SMOTE для устранения дисбаланса классов

Одной из главных проблем, выявленных на этапе предварительного анализа обучающего набора данных, является дисбаланс классов. В обучающей выборке многочисленный класс содержал 4744 объекта, редчайший класс – 14. Подобное соотношение приводит к тому, что модель машинного обучения фактически исключает малочисленные категории, формируя предвзятое решение в пользу доминирующих классов. Для устранения этой проблемы применен метод синтетического порождения примеров SMOTE [14].

Метод SMOTE не копирует существующие объекты, а генерирует новые синтетические примеры, опираясь на структуру многомерного пространства признаков. В многоклассовом режиме алгоритм автоматически выравнивает объем всех классов до наиболее крупного размера. В рассматриваемом эксперименте это привело к тому, что численность каждого класса доведена до 4744 объектов, что позволило существенно увеличить представленность редких наблюдений, особенно для класса 3 [14].

После применения метода SMOTE обучающая выборка стала полностью сбалансированной, что

позволило переобучить модель и повторно оценить ее характеристики. По итогам эксперимента установлено, что:

- точность на основных классах (0, 1 и 2) изменилась незначительно, оставаясь на уровне 0.75-0.78;
- полнота (Recall) редкого класса 3 улучшилась – с 0.25 до 0.50 на тестовой выборке;
- при этом точность (Precision) по классу 3 снизилась, что является ожидаемым эффектом при агрессивном синтетическом расширении данных;
- общая точность модели сохранилась на уровне примерно 75 %, однако сбалансированная метрика (macro average) улучшилась, что подтверждает более корректную работу классификатора по редким категориям.

Полученные результаты показывают, что использование метода SMOTE существенно влияет на распределение вероятностей внутри модели, делая классификацию более устойчивой к дисбалансу данных. Несмотря на определенное снижение точности для искусственно расширенных классов, применение данного подхода оправдано в задачах, где важна способность алгоритма обнаруживать редкие события – например, потенциальные экзопланеты специфических типов [18].

### Обоснование Random Forest, сравнение с другими алгоритмами классификации

Для обоснования выбора алгоритма Random Forest проведено дополнительное экспериментальное сравнение с рядом алгоритмов машинного обучения: логистической регрессией, методом опорных векторов (Support Vector Machine, SVM), градиентным бустингом и простой полносвязной нейронной сетью (MLP). Каждый из алгоритмов обладает различными свойствами и применим к различным типам данных.

Логистическая регрессия является одним из наиболее простых и интерпретируемых методов классификации. Она хорошо работает при наличии линейной зависимости между признаками и целевой переменной. Однако в задачах, где структура данных имеет сложный нелинейный характер, эффективность данного алгоритма может снижаться.

Метод опорных векторов (SVM) способен строить сложные нелинейные разделяющие поверхности с использованием различных ядерных функций. Этот алгоритм часто демонстрирует высокую точность при работе с многомерными данными, однако его вычислительная сложность существенно возрастает при увеличении объема обучающей выборки.

Градиентный бустинг относится к ансамблевым методам и строит последовательность слабых моделей (обычно решающих деревьев), каждая из которых корректирует ошибки предыдущих. Данный подход способен достигать высокой точности, но требует более тщательной настройки гиперпараметров и может быть чувствителен к шумам в данных.

Нейронная сеть типа MLP представляет собой универсальный аппроксиматор нелинейных зависимостей. Она может эффективно выявлять

сложные закономерности в данных, однако требует большего объема обучающей выборки, значительных вычислительных ресурсов и тщательной настройки архитектуры сети [18].

Для корректного сравнения все модели обучались на одинаковой обучающей выборке с использованием одинаковой процедуры нормализации признаков и одинакового разбиения данных на обучающую, валидационную и тестовую выборки. Оценка качества выполнялась по метрике Accuracy. Результаты сравнительного эксперимента представлены в таблице 1.

**Таблица 1 – Результаты сравнения алгоритмов машинного обучения**

**Table 1 – Results of the comparison of machine learning algorithms**

Алгоритм	Accuracy	Особенности
Logistic Regression	0.66	Хорошая интерпретируемость, но ограниченная способность выявлять нелинейные зависимости.
SVM	0.72	Высокая точность, но высокая вычислительная сложность.
Gradient Boosting	0.74	Высокая точность, чувствительность к параметрам.
MLP	0.71	Способность моделировать сложные зависимости, требует больше ресурсов.
Random Forest	0.75	Высокая устойчивость к шумам, хорошая интерпретируемость важности признаков.

Полученные результаты показывают, что алгоритм Random Forest демонстрирует наилучший баланс между точностью классификации, устойчивостью к шумам и вычислительной эффективностью. В отличие от линейных моделей, он способен учитывать нелинейные взаимосвязи между астрофизическими параметрами экзопланет. По сравнению с градиентным бустингом и нейронными сетями алгоритм Random Forest требует меньшей настройки гиперпараметров и является менее чувствительным к переобучению.

Таким образом, проведенное сравнительное исследование подтверждает целесообразность выбора алгоритма Random Forest в качестве основной модели классификации экзопланет в рамках данной работы.

**Анализ результатов и влияние техники балансировки данных**

После построения и обучения модели RandomForestClassifier [19] выполнена оценка качества предсказаний на валидационной и тестовой выборках. Исследование проведено в двух режимах:

без применения метода балансировки классов; с использованием метода SMOTE. Это позволило проанализировать влияние дисбаланса классов на качество классификации, а также оценить целесообразность применения методов искусственного расширения классов.

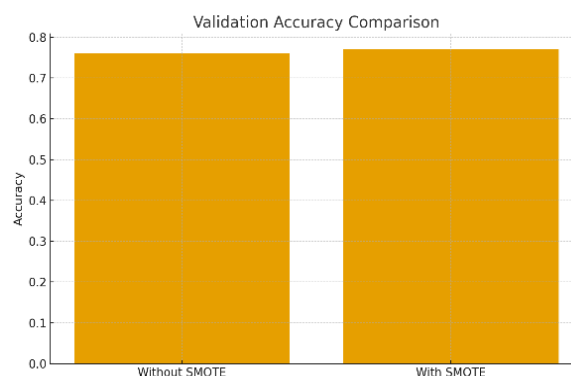
Результаты работы моделей представлены в таблице 2.

**Таблица 2 – Распределение классов до и после применения метода SMOTE**

**Table 2 – Class distribution before and after applying the SMOTE method**

Класс	Количество объектов	
	до применения метода	после применения метода
0	4038	4744
1	4744	4744
2	3850	4744
3	14	4744

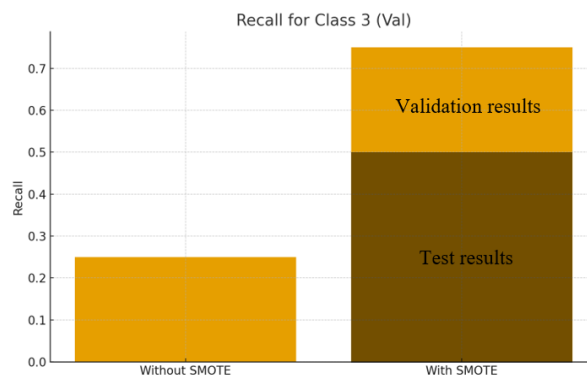
На рисунке 1 представлено сравнение метрики Accuracy до и после применения метода SMOTE.



**Рис. 1 – Результаты сравнения метрики Accuracy до и после применения метода SMOTE**

**Fig. 1 – Comparison of the Accuracy metric before and after applying the SMOTE method**

На рисунке 2 представлены результаты значений метрики Recall для самого редкого класса (3), что особенно важно для оценки способности модели обнаруживать редкие объекты.



**Рис. 2 – Результаты значений метрики Recall для третьего класса**

**Fig. 2 – Recall metric results for the third class**

Исходя из полученного результата получилось, что общая точность модели сохранена примерно на том же уровне (около 0.75), однако изменена структура ошибок.

В тестовой выборке показатель для редкого класса 3 метрика Recall выросла с 0.25 до 0.50, в валидационной выборке до 0.75, т.е. модель стала распознавать редкие случаи значительно лучше. Повышен риск ложных срабатываний, что видно по незначительному снижению показателя достоверности положительных срабатываний Precision. Предпочтение метрики Recall оправданно, так как важно не пропустить редкие объекты при увеличении числа ложных положительных прогнозов. Обучение после применения метода SMOTE потребовало больше времени и памяти, что является недостатком метода при больших данных.

### Заключение

В ходе выполнения работы разработана и исследована эффективность модели машинного обучения для классификации экзопланет на основе астрофизических данных. Для модели выбран алгоритм Random Forest, показавший высокий уровень стабильности на многоклассовой выборке.

Анализ результатов работы модели показал, что после применения метода SMOTE улучшены показатели полноты (Recall) и сбалансированности предсказаний по классам. На тестовой выборке итоговая точность составила 75 %, при этом даже наименьший по размеру класс (3) стал корректнее распознаваться. Таким образом, доказано, что применение методов балансировки данных является ключевым этапом при работе с реальными астрономическими наблюдениями.

Практическая значимость полученных результатов заключается в возможности применения разработанной модели для предварительной фильтрации телескопических наблюдений, ускоренной сортировки кандидатов в экзопланеты, поддержки научных исследований в области астрономии.

Методы машинного обучения способны эффективно решать задачи обнаружения и классификации экзопланет, снижая нагрузку на исследователей и ускоряя процесс анализа наблюдений. В перспективе разработанную модель машинного обучения можно усовершенствовать (добавить дополнительные параметры экзопланет), интегрировать в системы анализа данных астрофизических исследований, используемых в специальной астрофизической обсерватории российской академии наук (САО РАН) [20], а также в системе астрофизических данных SAO/NASA (ADS) - портале цифровой библиотеки для исследователей астрономии и физики [21].

Таким образом, разработанная модель является основой для интеллектуальных систем поддержки научных открытий в современном космическом исследовании.

### Литература

1. А.И. Читайло, *Информационно-коммуникационные технологии в педагогическом образовании*, 6 (93), 23-32 (2024).
2. Г.А. Карасёв, *Тенденции развития науки и образования*, 106-11, 54-58 (2024).
3. А.А. Protko, A.S. Manko, *Recent Achievements and Prospects of Innovations and Technologies*, 4, 226-230 (2025).
4. С.М. Борзов, Е.С. Нежевенко, *Автометрия*, **59**, 3, 52-71 (2023).
5. Т.В. Сафонова, Н.В. Яготинцева, О.Н. Колбина, А.В. Мокряк, А.А. Попович, *Информационные технологии и системы: управление, экономика, транспорт, право*, 4 (48), 126-132 (2023).
6. Ш.И. Хайдаров, *Современные инновации, системы и технологии*, **5**, 2, 5021-5028 (2025).
7. NASA Exoplanet Archive. Confirmed Exoplanet Overview. – URL: <https://exoplanetarchive.ipac.caltech.edu> (дата обращения 12.11.2025).
8. Российская лунная программа. – URL: [https://ru.ruwiki.ru/wiki/Российская\\_лунная\\_программа](https://ru.ruwiki.ru/wiki/Российская_лунная_программа) (дата обращения 12.11.2025).
9. Федеральная космическая программа на 2016 - 2030 гг. – URL: [http://asf.ural.ru/VNKSF/Itogi/re\\_v22/Nauka/Presentazii/11/Ararov.pdf](http://asf.ural.ru/VNKSF/Itogi/re_v22/Nauka/Presentazii/11/Ararov.pdf) (дата обращения 12.11.2025).
10. И. Гудфеллоу, Й. Бенджио, А. Курвиль, *Глубокое обучение*. Москва: ДМК Пресс, 2018. 782 с.
11. M. Perryman, *The Exoplanet Handbook*. Cambridge University Press, 2018. 952 p.
12. R.A. Fisher, *Journal Article*, **7**, 2, 179-188 (1936).
13. C. Bishop, *Pattern Recognition and Machine Learning*, New York: Springer, 2006. 738 p.
14. N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, *Journal of Artificial Intelligence Research*, 16(1), 321-357 (2002).
15. G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning*, Springer, 2021. 616 p.
16. A.P. Ratnasari, *International Journal of Scientific Research and Management*, **12**, 4, 494-501 (2024).
17. T. Wongvorachan, S.He, O.A Bulut, *Information*, 14, 54 (2023).
18. L. Breiman, *Machine Learning*, 45, 5-32 (2006).
19. What is a confusion matrix? – IBM. – URL: <https://www.ibm.com/think/topics/confusion-matrix> – (дата обращения 12.11.2025).
20. Наука - ТАСС. В обсерватории РАН в КЧР планируют создать передвижной планетарий. – URL: <https://nauka.tass.ru/nauka/21704923> – (дата обращения 15.12.2025).
21. WebCatalog. NASA/ADS. – URL: <https://webcatalog.io/ru/apps/ads> – (дата обращения 15.12.2025).

### References

1. A.I. Chitailo, *Information and Communication Technologies in Teacher Education*, 6 (93), 23–32 (2024).
2. G.A. Karasev, *Trends in the Development of Science and Education*, 106–11, 54–58 (2024).
3. A.A. Protko, A.S. Manko, *Recent Achievements and Prospects of Innovations and Technologies*, 4, 226–230 (2025).
4. S.M. Borzov, E.S. Nezhevenko, *Autometry*, **59**, 3, 52–71 (2023).
5. T.V. Safonova, N.V. Yagotintseva, O.N. Kolbina, A.V. Mokryak, A.A. Popovich, *Information Technologies and Systems: Management, Economics, Transport, Law*, 4 (48), 126–132 (2023).

6. Sh.I. Khaidarov, *Modern Innovations, Systems, and Technologies*, 5, 2, 5021–5028 (2025).
7. NASA Exoplanet Archive. Confirmed Exoplanet Overview. – URL: <https://exoplanetarchive.ipac.caltech.edu> (accessed 11/12/2025).
8. Russian Lunar Program. – URL: [https://ru.ruwiki.ru/wiki/Российская\\_лунная\\_программа](https://ru.ruwiki.ru/wiki/Российская_лунная_программа) (accessed 11/12/2025).
9. Federal Space Program for 2016–2030. – URL: [http://asf.ural.ru/VNKSF/Itogi/re\\_v22/Nauka/Presentazii/11/Arapov.pdf](http://asf.ural.ru/VNKSF/Itogi/re_v22/Nauka/Presentazii/11/Arapov.pdf) (accessed November 12, 2025).
10. I. Goodfellow, Y. Bengio, A. Curville, *Deep Learning*. Moscow: DMK Press, 2018. 782 pp.
11. M. Perryman, *The Exoplanet Handbook*. Cambridge University Press, 2018. 952 pp.
12. R.A. Fisher, *Journal Article*, 7, 2, 179–188 (1936).
13. C. Bishop, *Pattern Recognition and Machine Learning*, New York: Springer, 2006. 738 p.
14. N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, *Journal of Artificial Intelligence Research*, 16(1), 321–357 (2002).
15. G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning*, Springer, 2021. 616 p.
16. A.P. Ratnasari, *International Journal of Scientific Research and Management*, 12, 4, 494–501 (2024).
17. T. Wongvorachan, S. He, O.A. Bulut, *Information*, 14, 54 (2023).
18. L. Breiman, *Machine Learning*, 45, 5–32 (2006).
19. What is a confusion matrix? – IBM. – URL: <https://www.ibm.com/think/topics/confusion-matrix> – (accessed 11/12/2025).
20. Science - TASS. The RAS observatory in the KCR plans to create a mobile planetarium. – URL: <https://nauka.tass.ru/nauka/21704923> – (accessed 12/15/2025).
21. WebCatalog. NASA/ADS. – URL: <https://webcatalog.io/ru/apps/ads> – (accessed 12/15/2025).

---

© **В. О. Окунев** – магистрант кафедры Систем информационной безопасности (СИБ), Казанский национальный исследовательский технический университет имени А.Н. Туполева (КНИТУ им. А.Н. Туполева), Казань, Россия, OkunevVO@stud.kai.ru; **Р. М. Хусайнов** – ассистент кафедры СИБ, КНИТУ им. А.Н. Туполева, rumil\_husainov98@mail.ru; **Н. Г. Талипов** – к-т. техн. наук, доцент кафедры СИБ, КНИТУ им. А.Н. Туполева, nafis.talipov@mail.ru.

© **V. O. Okunev** – Master-student of Information Security Systems (ISS) Department, Kazan National Research Technical University named after A.N. Tupolev (KNRTU named after A.N. Tupolev), Kazan, Russia, OkunevVO@stud.kai.ru; **R. M. Khusainov** – Assistant of the ISS Department, KNRTU named after A.N. Tupolev, rumil\_husainov98@mail.ru; **N. G. Talipov** – PhD (Technical Sci.), Associate Professor of the ISS Department, KNRTU named after A.N. Tupolev, nafis.talipov@mail.ru.

Дата поступления рукописи в редакцию – 03.12.25.

Дата принятия рукописи в печать – 08.04.26.