

УДК 519.248:658.518.5  
DOI: 10.15827/0236-235X.136.572-578

Дата подачи статьи: 07.06.21, после доработки: 02.08.21  
2021. Т. 34. № 4. С. 572–578

## **Диагностика состояния технического объекта с помощью классификации методами машинного обучения**

*Н.А. Ломовцева*<sup>1</sup>, магистрант, *natalya.lomovtseva@gmail.com*

*Ю.Е. Кувайскова*<sup>1</sup>, к.т.н., доцент, *u.kuwaiskova@mail.ru*

*В.Н. Клячкин*<sup>1</sup>, д.т.н., профессор, *v\_kl@mail.ru*

<sup>1</sup> *Ульяновский государственный технический университет,  
кафедра прикладной математики и информатики, г. Ульяновск, 432027, Россия*

Для обеспечения безопасности и надежности функционирования сложных технических систем необходима их диагностика. Иногда она сводится лишь к разделению объектов на исправные и неисправные: проводится бинарная классификация методами машинного обучения по прецедентам (с учителем). Однако часто требуется более детальное исследование, когда состояние объекта нельзя отнести к этим двум вариантам. В таком случае проводится многоклассовая классификация состояний объекта. Как и при бинарной классификации, здесь могут эффективно применяться методы машинного обучения.

Полученная по результатам предварительных испытаний выборка разбивается на две части – обучающую и тестовую. Обучающая предназначена для построения моделей, с помощью которых объекты разделяются на заданное количество классов. Предполагается, что есть определенная связь между показателями функционирования объекта и его состояниями. На основе обучающей выборки необходимо построить алгоритм, обеспечивающий для заданного набора показателей функционирования достаточно точную оценку состояния объекта.

Разработана программа многоклассовой классификации для построения модели алгоритма, обеспечивающей надежную диагностику состояния объекта. Для исключения переобучения использована кросс-валидация. Оцениваемые три меры качества построенных моделей позволяют учесть особенности обучающей выборки, при этом применяются различные типы классификаторов. В качестве численного примера рассмотрена навигация робота: по результатам показаний 24 датчиков расстояний определяется одно из четырех направлений его перемещения.

**Ключевые слова:** *техническая диагностика, многоклассовая классификация, кросс-валидация, агрегированный подход, навигация робота.*

Диагностика функционирования сложных технических систем необходима для обеспечения их безопасности и надежности. Иногда она сводится к разделению объектов на исправные и неисправные: проводится бинарная классификация методами машинного обучения по прецедентам (с учителем). По результатам предшествующей эксплуатации объекта имеется набор числовых значений показателей его функционирования и известно, был ли объект исправен при этих значениях. По имеющимся данным строится классификатор, используя который, можно оценить исправность объекта при новом наборе показателей [1, 2].

Часто требуется более детальное исследование состояния объекта, когда его нельзя описать с помощью двух возможных вариантов – исправен объект или нет. Например, подобная задача решалась с применением нейронных сетей при классификации режимов работы авиационного газотурбинного

двигателя [3, 4]. В таком случае проводится многоклассовая классификация состояний объекта. Как и при бинарной классификации, здесь можно эффективно использовать методы машинного обучения.

Выборка, полученная по результатам предварительных испытаний, разбивается на две части – обучающую и тестовую. Обучающая предназначена для построения моделей, с помощью которых объекты разделяются на заданное количество классов. Предполагается, что существует определенная связь между показателями функционирования объекта и его состояниями. На основе обучающей выборки необходимо построить алгоритм, который для заданного набора показателей функционирования обеспечил бы достаточно точный результат, характеризующий состояние объекта.

Для повышения качества классификации и исключения переобучения часто использу-

ется кросс-валидация. Обучающая выборка разбивается на несколько частей. Например, при разбиении на пять частей четыре из них используются непосредственно для обучения, пятая – для контроля. Последовательно перебираются все пять вариантов. Среднее по контрольным выборкам значение критерия качества диагностики характеризует точность алгоритма. Конечное качество классификации с помощью полученных моделей оценивается по тестовой части выборки, не использованной для построения алгоритма классификации.

### Постановка задачи

Для диагностики состояния объекта обучающая выборка представляется в виде матрицы  $X$  показателей функционирования системы, элементы которой  $x_{ij}$  – результат  $i$ -го наблюдения по  $j$ -му показателю,  $i = 1, \dots, l$ ,  $j = 1, \dots, p$  ( $l$  – количество строк, или число наблюдений;  $p$  – количество столбцов, или число показателей), и вектора-столбца ответов  $Y$ , состоящего из номеров классов. Каждой строке  $x_i$  матрицы  $X$  соответствует определенное значение  $y_i$  вектора  $Y$ . Совокупность пар  $(x_i, y_i)$  образует выборку исходных данных – прецедентов.

Задача состоит в построении модели  $a(x, w)$ , которая по заданной строке показателей функционирования  $x_i$  предскажет номер класса  $y_i$ , определяющего состояние рассматриваемого объекта [5, 6].

Чаще всего применяют линейные модели:

$$a(x, w) = w_0 + w_1x_1 + \dots + w_px_p, \quad (1)$$

где  $w = (w_0 \ w_1 \ \dots \ w_p)$  – вектор параметров модели. Процесс подбора параметров  $w_j$  по обучающей части выборки и является обучением алгоритма.

Найденные параметры должны обеспечить оптимальное значение функционала качества. Обычно минимизируется функционал ошибок – среднее количество несовпадений фактического состояния  $y_i$ , соответствующего строке показателей функционирования  $x_i$  и прогнозируемого  $a(x_i)$  по модели (1):

$$\begin{aligned} Q(a, X) &= \frac{1}{l} \sum_{i=1}^l L(a, x_i) = \\ &= \frac{1}{l} \sum_{i=1}^l [a(x_i) - y_i] \rightarrow \min. \end{aligned} \quad (2)$$

Здесь  $L(a, x_i)$  называют функцией потерь, она фиксирует наличие несовпадения опытного значения состояния объекта для задан-

ного множества показателей функционирования  $x_i$  (строки матрицы  $X$ ) со значением, прогнозируемым по построенному алгоритму  $a(x_i)$ .

При несбалансированных классах, когда количество данных об одном из состояний существенно (на порядок) отличается от другого, используется  $F$ -критерий – гармоническое среднее точности и полноты [1].

Полезной характеристикой качества многоклассовой классификации является матрица неточностей. Это матрица размера  $[n \times n]$ , где  $n$  – количество классов. Матрица наглядно демонстрирует, сколько примеров из каждого класса идентифицировано верно (диагональные значения) и сколько наблюдений из каждого класса было неверно принято за  $k$ -й класс. Каждая строка матрицы представляет экземпляры в прогнозируемом классе, а каждый столбец – в фактическом. Сумма всех недиагональных значений матрицы показывает количество допущенных классификатором ошибок.

### Методы многоклассовой классификации

Для многоклассовой классификации может быть использовано множество различных методов; большинство из них используются и при бинарной классификации: наивный байесовский классификатор, логистическая регрессия, метод опорных векторов, нейронные сети, методы ближайших соседей и другие [7].

Специальную группу составляют композиционные методы, представляющие ансамбль отдельных алгоритмов. Опыт показывает, что два основных композиционных метода – бэггинг и бустинг – дают более высокую точность на конкретном наборе данных по сравнению с отдельными алгоритмами. В композиционных методах используется один и тот же метод классификации, или построенный на разных подмножествах выборки, или компенсирующий на каждом шаге ошибки предыдущей итерации [8].

При агрегированном подходе совместно используются различные методы классификации, построенные на обучающей выборке. Для достижения наилучшего результата может быть применен полный перебор наборов из всех базовых методов. Однако проведенные исследования показали [9], что использование в агрегате более чем двух базовых классификаторов не дает существенного по-

вышения точности. При этом и метод агрегирования (по среднему значению, по медиане или по голосованию) также не оказывает значимого влияния на качество классификации. В связи с этим использовалось агрегирование двух классификаторов по среднему значению.

Например, для метода опорных векторов вероятность принадлежности состояния объекта к  $k$ -му классу

$$P(Y = k | X) = \frac{\exp(\sum_{j=1}^p w_{kj} x_j + w_{k0})}{\sum_{i=1}^n (\sum_{j=1}^p w_{ij} x_j + w_{i0})}, \quad (3)$$

где  $w_{ij}$  – параметры метода опорных векторов, определяемые с помощью метода множителей Лагранжа для каждого из  $n$  классов.

Для методов бустинга такая вероятность определяется как

$$P(Y = k | X) = \frac{\exp(\sum_{t=1}^T \alpha_{kt} h_t)}{\sum_{i=1}^n \sum_{t=1}^T \alpha_{it} h_t}, \quad (4)$$

где  $h_t(x)$  – базовые классификаторы бустинга;  $\alpha_{it}$  – коэффициент взвешенного голосования для соответствующего классификатора  $h_t(x)$ .

Математическая модель агрегированного метода классификации по среднему значению для двух базовых методов – опорных векторов и бустинга – примет вид

$$\hat{P}_A(Y = k | X) = \frac{1}{2} \left[ \frac{\exp(\sum_{j=1}^p w_{kj} x_j + w_{k0})}{\sum_{i=1}^n (\sum_{j=1}^p w_{ij} x_j + w_{i0})} + \frac{\exp(\sum_{t=1}^T \alpha_{kt} h_t)}{\sum_{i=1}^n \sum_{t=1}^T \alpha_{it} h_t} \right]. \quad (5)$$

Целесообразность использования агрегированных классификаторов обусловлена тем, что при проведении бинарной классификации такой подход привел к существенному повышению качества диагностики во всех рассмотренных технических системах [1, 9].

### Программа многоклассовой классификации

Программа разработана на языке программирования Python. Используется библиотека Sklearn (scikit-learn), из которой импортируются готовые конструкторы для моделей-классификаторов:

- LogisticRegression – логистическая регрессия;
- LinearDiscriminantAnalysis – линейный дискриминантный анализ;
- KNeighborsClassifier – метод  $k$ -ближайших соседей;
- DecisionTreeClassifier – деревья решений;
- GaussianNB – наивный байесовский классификатор;
- SVC – метод опорных векторов;
- RandomForestClassifier – случайный лес;
- AdaBoostClassifier – адаптивный бустинг.

Также из библиотеки Sklearn импортируются готовые модули для вычисления метрик многоклассовой классификации:

- accuracy\_score – функционал ошибок;
- f1\_score – F-мера;
- confusion\_matrix – матрица неточностей и модуль для метода кросс-валидации – cross\_val\_score.

Метод агрегирования по среднему значению программировался вручную, но с использованием функции Python для нахождения вероятностей принадлежности каждого объекта к соответствующему классу: predict\_proba.

Программа обеспечивает ввод файла исходных данных, разбиение его в заданном соотношении на обучающую и тестовую части случайным образом, проведение кросс-валидации с заданным пользователем количеством блоков.

Обеспечивается также отбор значимых показателей функционирования объекта двумя методами: по коррелированности показателей и по значимости регрессоров в линейной регрессионной модели зависимости номера класса объекта от показателей.

Результат работы программы – расчет трех мер качества классификации по тестовой выборке для базовых, комбинированных и агрегированных классификаторов. В зависимости от характера исходных данных пользователь выбирает наилучший классификатор по тому или иному критерию. Этот классификатор в дальнейшем используется для прогнозирования состояния объекта. По вновь найденным показателям функционирования объекта прогнозируется, к какому классу относится его состояние.

**Численное исследование:  
навигация робота**

Прогнозируются направления перемещения робота по результатам анализа показаний ультразвуковых датчиков для измерения расстояний [10]. По показаниям датчиков, расположенных на роботе, необходимо провести классификацию по направлениям его перемещения. Полученная информация о состоянии робота – набор исходных данных: множество ситуаций с заданными показателями и множество возможных состояний объекта. Предполагается наличие неизвестной зависимости между показаниями датчиков и перемещениями робота. По имеющимся данным необходимо построить алгоритм для восстановления этой зависимости, с помощью которой для любого объекта по новым показаниям датчиков можно выдать достаточно точный ответ о его перемещении (см. рисунок). Для решения используются методы многоклассовой классификации с учителем.

Обозначения показаний датчиков:  $x_1$  – датчик в передней части робота, опорный угол  $-180^\circ$ ;  $x_2$  – опорный угол  $-165^\circ$ ;  $x_3$  – опорный угол  $-150^\circ$ ;  $x_4$  – опорный угол  $-135^\circ$ ; ...,  $x_{12}$  – опорный угол  $-15^\circ$ ;  $x_{13}$  – датчик в задней части робота, опорный угол  $0^\circ$ ;  $x_{14}$  – опорный угол  $+15^\circ$ ; ...,  $x_{23}$  – опорный угол  $+150^\circ$ ;  $x_{24}$  – опорный угол  $+165^\circ$ .

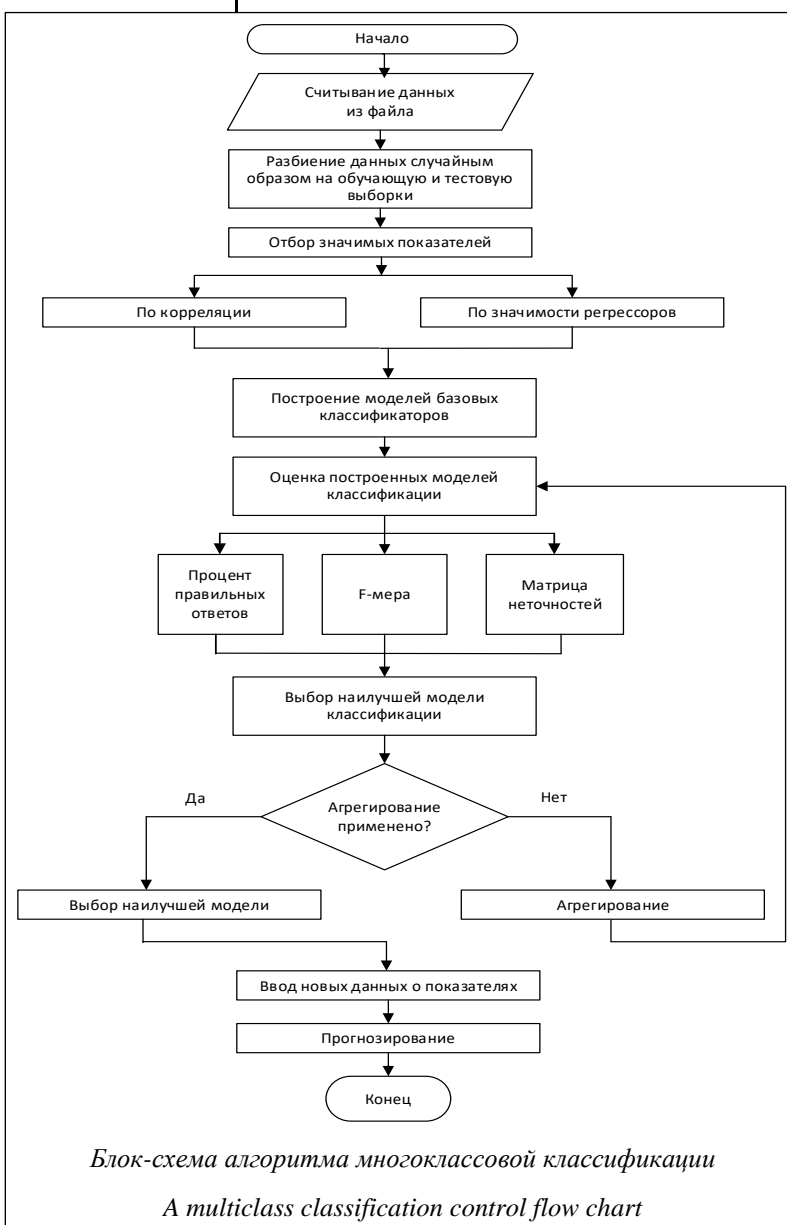
Заданы четыре класса, соответствующие направлениям перемещения робота: 1 – движение вперед, 2 – плавный правый поворот, 3 – резкий поворот направо, 4 – плавный левый поворот.

Всего получено 5 456 наблюдений. Процент правильных ответов моделей базовых классификаторов на тестовой выборке:

- логистическая регрессия – 68,13 %;
- линейный дискриминантный анализ – 65,75 %;
- алгоритм k-ближайших соседей – 89,01 %;
- дерево решений – 99,27 %;
- наивный байесовский классификатор – 53,48 %;
- метод опорных векторов – 90,48 %;
- случайный лес – 99,63 %;
- адаптивный бустинг – 78,75 %.

F-мера моделей базовых классификаторов на тестовой выборке:

- логистическая регрессия – 0,675;
- линейный дискриминантный анализ – 0,649;



Блок-схема алгоритма многоклассовой классификации

A multiclass classification control flow chart

- алгоритм k-ближайших соседей – 0,890;
- дерево решений – 0,993;
- наивный байесовский классификатор – 0,535;
- метод опорных векторов – 0,904;
- случайный лес – 0,996;
- адаптивный бустинг – 0,705.

Матрица неточностей случайного леса:

$$\begin{bmatrix} 212 & 0 & 1 & 0 \\ 0 & 73 & 0 & 0 \\ 0 & 0 & 219 & 0 \\ 1 & 0 & 0 & 40 \end{bmatrix}.$$

Процент правильных ответов на тестовой выборке (не использованной в расчетах) оказался наилучшим при использовании случайного леса – 99,63 %; близкий результат показало дерево решений – 99,27 %. Аналогичные результаты получены и по F-мере. Матрица неточностей случайного леса показывает, что наилучший классификатор ошибся два раза на тестовой выборке: присвоил метку 1-го класса для наблюдения из 3-го класса и присвоил метку 4-го класса для наблюдения из 1-го класса. Худшие результаты показал байесовский классификатор.

Заметим, далеко не всегда получается так, что результаты по различным мерам близки. В таком случае решение о наилучшем классификаторе приходится принимать пользователю.

Учитывая некоторую несбалансированность данных (к 4-му классу относится значительно меньше наблюдений, чем к первым трем), в качестве основной характеристики была принята F-мера. По этой мере оценивались агрегированные классификаторы. Результаты агрегирования по среднему значению моделей базовых классификаторов (F-мера на тестовой выборке):

- метод опорных векторов + алгоритм k-ближайших соседей – 0,896;
- метод опорных векторов + адаптивный бустинг – 0,996;
- метод опорных векторов + логистическая регрессия – 0,776;
- метод опорных векторов + линейный дискриминантный анализ – 0,733;
- алгоритм k-ближайших соседей + адаптивный бустинг – 0,978;
- алгоритм k-ближайших соседей + логистическая регрессия – 0,875;
- алгоритм k-ближайших соседей + линейный дискриминантный анализ – 0,859;
- адаптивный бустинг + логистическая регрессия – 0,976;
- адаптивный бустинг + линейный дискриминантный анализ – 0,971;
- логистическая регрессия + линейный дискриминантный анализ – 0,657.

Заметим, что в отличие от бинарной классификации, когда агрегирование повышало качество диагностики, в этой задаче значение F-меры при использовании агрегированных классификаторов не увеличилось: лучший результат 0,996 показало агрегирование метода опорных векторов с адаптивным бустингом, при этом случайный лес также показал значение 0,996.

### Заключение

Разработанная программа многоклассовой классификации для диагностики состояния технического объекта методами машинного обучения обеспечивает достаточно высокую точность классификации по выборке исходных данных, предварительно сформированной по результатам предшествующей работы объекта. Наилучшая модель классификатора позволяет достоверно прогнозировать состояние системы по заданным показателям ее функционирования.

### Литература

1. Биргер И.А. Техническая диагностика. М.: Машиностроение, 1978. 240 с.
2. Клячкин В.Н., Жуков Д.А. Прогнозирование состояния технического объекта с применением методов машинного обучения // Программные продукты и системы. 2019. Т. 32. № 2. С. 244–250. DOI: 10.15827/0236-235X.126.244-250.
3. Васильев В.И., Жернаков С.В. Классификация режимов работы ГТД с использованием технологии нейронных сетей // Вестн. УГАТУ. 2009. Т. 12. № 1. С. 53–60.
4. Жернаков С.В., Гильманшин А.Т. Применение интеллектуальных алгоритмов на основе нечеткой логики и нейронных сетей для решения задач диагностики отказов авиационного ГТД // Сб. докл. II Междунар. конф. ИПРМ. 2014. С. 112–115.
5. Witten I.H., Frank E. Data Mining: Practical Machine Learning Tools and Techniques. San Francisco: Morgan Kaufmann Publ., 2005, 525 p.

6. Воронина В.В., Михеев А.В., Ярушкина Н.Г., Святов К.В. Теория и практика машинного обучения. Ульяновск: Изд-во УлГТУ, 2017. 290 с.
7. Klyachkin V.N., Kuvayskova J.E., Zhukov D.A. Aggregated classifiers for state diagnostics of the technical object. Proc. Int. Multi-Conf. FarEastCon, 2019, art. 8934362. DOI: 10.1109/FarEastCon.2019.8934362.
8. Vijayarani S., Dhayanand S. Liver disease prediction using SVM and Naïve Bayes algorithms. Int. J. of Science, Engineering and Technology Research, 2015, vol. 4, no. 4, pp. 816–820.
9. Wyner A.J., Olson M., Bleich J. Explaining the success of AdaBoost and random forests as interpolating classifiers. The J. of Machine Learning Research, 2017, vol. 18, no. 1, pp. 1558–1590.
10. Freire A.L., Barreto G.A., Veloso M., Varela A.T. Short-term memory mechanisms in neural network learning of robot navigation tasks: A case study. Proc. VI LARS, 2009, pp. 1–6. DOI: 10.1109/LARS.2009.5418323.

Software & Systems  
DOI: 10.15827/0236-235X.136.572-578

Received 07.06.21, Revised 02.08.21  
2021, vol. 34, no. 4, pp. 572–578

### Diagnosing the condition of a technical object using machine learning classification

*N.A. Lomovtseva*<sup>1</sup>, Graduate Student, [natalya.lomovtseva@gmail.com](mailto:natalya.lomovtseva@gmail.com)

*Yu.E. Kuvayskova*<sup>1</sup>, Ph.D. (Engineering), Associate Professor, [u.kuvayskova@mail.ru](mailto:u.kuvayskova@mail.ru)

*V.N. Klyachkin*<sup>1</sup>, Dr.Sc. (Engineering), Professor, [v\\_kl@mail.ru](mailto:v_kl@mail.ru)

<sup>1</sup> Ulyanovsk State Technical University, Ulyanovsk, 432027, Russian Federation

**Abstract.** Diagnosing the functioning of complex technical systems is necessary to ensure their safety and reliability. Sometimes the diagnosis is reduced to the division of objects into healthy and faulty: there is a binary classification of machine learning methods according to precedents (with the teacher). However, when there is a need to describe an object's state with several possible options (not just two: a healthy object or a faulty object), a more detailed study is often needed. In this case, a multi-class classification of the object's states is carried out. Machine learning techniques can be used effectively as for binary classification.

The sample obtained from the preliminary tests is divided into two parts: training and test. The training part is for building models that help to divided objects into a given number of classes. It is assumed that there is some connection between the object's performance indicators and states. Based on the training sample, it is necessary to build an algorithm that provides a sufficiently accurate object's state assessment for a given set of performance indicators.

The paper presents a developed multi-class classification program allowing building an algorithm model for reliable diagnosis of the object's condition. At the same time, cross-validation is used to eliminate retraining. The three quality measures of the built models are used to take into account the specifics of the training sample applying different types of classifiers. As a numerical example, the authors consider the robot's navigation: according to the results of 24 distance sensors, one of the four directions of its movement is determined.

**Keywords:** technical diagnostics, multi-class classification, cross-validation, aggregated approach, robot navigation.

### References

1. Birger I.A. *Technical Diagnostics*. Moscow, 1978, 240 p. (in Russ.).
2. Klyachkin V.N., Zhukov D.A. Forecasting the state of a technical object using machine learning methods. *Software and Systems*, 2019, vol. 32, no. 2, pp. 244–250. DOI: 10.15827/0236-235X.126.244-250.
3. Vasilyev V.I., Zhernakov S.V. Classification of GTE operation modes with use of neural network technology. *Vestnik USATU*, 2009, vol. 12, no. 1, pp. 53–60 (in Russ.).

4. Zhernakov S.V., Gilymanshin A.T. Applying intelligent algorithms based on fuzzy logic and neural networks to solve diagnostics problems of aviation GTE failures. *Proc. II Int. Conf. ITIPM*, 2014, pp. 112–115 (in Russ.).
5. Witten I.H., Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco: Morgan Kaufmann Publ., 2005, 525 p.
6. Voronina V.V., Mikheev A.V., Yarushkina N.G., Svyatov K.V. *Machine Learning Theory and Practice*. Ulyanovsk, 2017, 290 p. (in Russ.).
7. Klyachkin V.N., Kuvayskova J.E., Zhukov D.A. Aggregated classifiers for state diagnostics of the technical object. *Proc. Int. Multi-Conf. FarEastCon*, 2019, art. 8934362. DOI: 10.1109/FarEastCon.2019.8934362.
8. Vijayarani S., Dhayanand S. Liver disease prediction using SVM and Naïve Bayes algorithms. *Int. J. of Science, Engineering and Technology Research*, 2015, vol. 4, no. 4, pp. 816–820.
9. Wyner A.J., Olson M., Bleich J. Explaining the success of AdaBoost and random forests as interpolating classifiers. *The J. of Machine Learning Research*, 2017, vol. 18, no. 1, pp. 1558–1590.
10. Freire A.L., Barreto G.A., Veloso M., Varela A.T. Short-term memory mechanisms in neural network learning of robot navigation tasks: A case study. *Proc. VI LARS*, 2009, pp. 1–6. DOI: 10.1109/LARS.2009.5418323.

#### Для цитирования

Ломовцева Н.А., Кувайскова Ю.Е., Клячкин В.Н. Диагностика состояния технического объекта с помощью классификации методами машинного обучения // Программные продукты и системы. 2021. Т. 34. № 4. С. 572–578. DOI: 10.15827/0236-235X.136.572-578.

#### For citation

Lomovtseva N.A., Kuvayskova Yu.E., Klyachkin V.N. Diagnosing the condition of a technical object using machine learning classification. *Software & Systems*, 2021, vol. 34, no. 4, pp. 572–578 (in Russ.). DOI: 10.15827/0236-235X.136.572-578.