СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ

УДК 004.032.2

МОДЕЛЬ КЛАССИФИКАТОРА БОРТОВЫХ СОСТОЯНИЙ ИЗДЕЛИЯ РАКЕТНО-КОСМИЧЕСКОЙ ТЕХНИКИ НА ОСНОВЕ НЕЯВНЫХ ПРИЗНАКОВ ТРАЕКТОРНЫХ ОТКЛОНЕНИЙ

С. В. Спицын, инженер-конструктор 1 категории АО «РКЦ «Прогресс» – филиал ОКБ «Спектр», Рязань, Россия;
orcid.org/0000-0001-6468-5043, e-mail: spitsyn62@gmail.com
С. А. Тихомиров, доцент кафедры САПР ВС РГРТУ, главный конструктор проекта АО «РКЦ «Прогресс» – филиал ОКБ «Спектр», Рязань, Россия;
orcid.org/0000-0002-3174-1536, e-mail: serge-tikhomirov@yandex.ru

Рассматривается модель классификатора бортовых состояний сложного многопараметрического изделия ракетно-космической техники, реализованная на основе анализа неявных признаков траекторных отклонений объекта от расчётной и априорной траектории при проведении ракетных пусков (полигонных испытаний). ***Целью работы*** *является нахождение обобщенного алгоритма определения возникновений нештатных и аварийных ситуаций на борту сложного многопараметрического объекта на основе траекторных данных, регистрируемых в ходе проведения лётного эксперимента полигонными измерительными средствами в составе общего потока бортовой телеметрической информации. Представлен поэтапный процесс построения эффективной модели классификатора бортовых состояний, основанной на современных и перспективных методах интеллектуального анализа данных и машинного обучения, в настоящий момент не получивших широкого применения в технологических процессах ракетно-космической отрасли.*

Ключевые слова: *информационно-измерительное обеспечение ракетных пусков, траекторные измерения, машинное обучение, системный анализ, телеметрическая информация.*

DOI: 10.21667/1995-4565-2021-75-133-141

Введение

Анализ используемых в настоящее время подходов к сбору, обработке, анализу и представлению телеметрической информации, содержащей траекторные измерения сложного изделия ракетно-космической техники в ходе проведения полигонных испытаний, показывает необходимость модернизации информационно-аналитической и вычислительной инфраструктуры полигонных измерительных комплексов в части применения современных и перспективных информационных технологий. В связи с этим в качестве одного из таких направлений, предназначенных для усовершенствования технологических процессов ракетных пусков, предлагается реализовать эффективную модель классификатора состояний изделия, основанную на интеллектуальном анализе неявных признаков траекторных отклонений изделия. Разработка подобной модели может выступить выгодным дополнением к существующим методам контроля и сопровождений натурных работ и иметь высокую практическую значимость, что в конечном счёте способствует повышению информативности и объективности оценок результатов пуска, упрощению информационно-аналитического обеспечения полигонных испытаний и снижению объёма контролируемых параметров телеметрии, характеризующих режимы функционирования бортовых систем многопараметрических объектов.

Теоретическая часть

Определение подзадач в области исследований обработки и анализа траекторных измерений объекта

В ходе проведения этапа лётного эксперимента полигонных испытаний сложных многопараметрических изделий ракетно-космической техники итоговыми результатами обработки траекторной измерительной информации являются некоторые «зависимые переменные». В качестве таких «зависимых переменных» могут выступать как численные значения географических координат объекта, характеризующие его текущее местоположение и баллистическую траекторию, так и факторные номинативные переменные, характеризующие определённое состояние сложного многопараметрического объекта или признак такого состояния. Исходя из этого, для исследований предметной области сбора, обработки и анализа траекторной телеметрии в рамках современных и перспективных методов предикативной аналитики и интеллектуального анализа данных можно выделить два направления исследования.

1. Задачи регрессионного анализа, результатом решения которых являются прогнозы величин, как правило, вещественного типа. Для области анализа траекторных измерений при проведении ракетных пусков вопрос решения подобных задач ранее рассматривался в [1, 2].

2. Задачи классификации, результатом решения которых в простейшем случае будут являться бинарные метки класса, позволяющие классифицировать различные состояния объекта. Под задачу бинарной классификации по принципу «норма/ненорма» наилучшим образом подходит вопрос оперативного определения текущих режимов функционирования бортовых систем объекта – штатных или нештатных. Для ракетно-космической отрасли вопрос эффективной и быстрой классификации таких состояний является достаточно актуальным, так как оперативное выявление нештатных режимов на борту изделия даёт время на их парирование и предотвращение перехода в необратимую аварийную ситуацию [3].

Таким образом, процесс разделения (классификации) направлений исследований по данной тематике на подзадачи в зависимости от результирующих переменных можно представить в качестве алгоритма, показанного на рисунке 1.

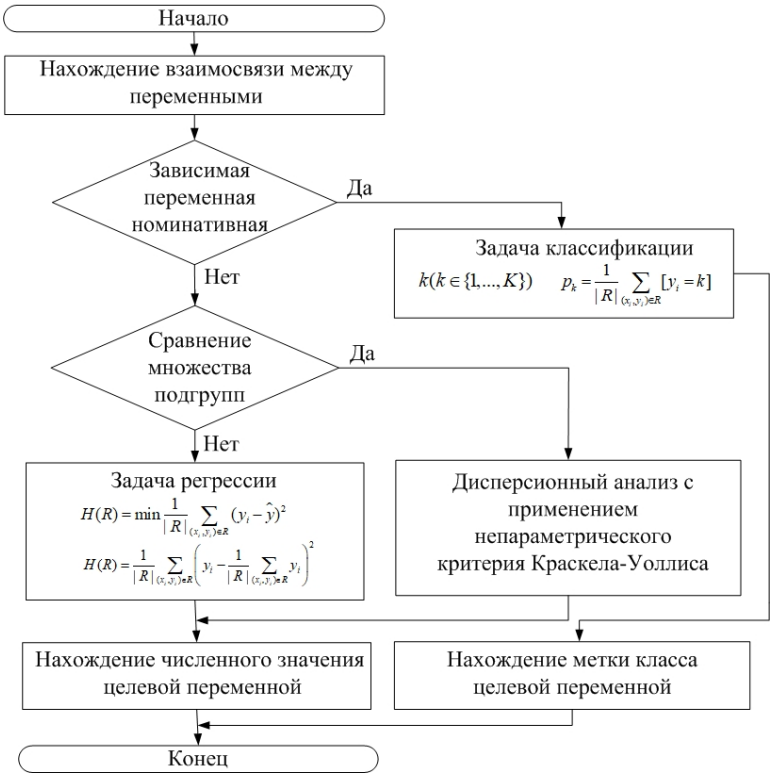

Рисунок 1 – Классификация подзадач направлений исследования
Figure 1 – Classification of research subtasks

В рамках решения задачи по второму направлению исследований предлагается разработать модель бинарного классификатора штатных и нештатных (аварийных) состояний бортовых систем сложного многопараметрического объекта на основе неявных признаков его траекторных отклонений от расчётной и априорной траектории в ходе проведения лётного эксперимента. Разработка подобной модели может стать эффективным дополнением к существующим методам контроля бортовых состояний на основе явных признаков (анализа массива цифровых параметров бортовой телеметрии), но также она способна выступить и в качестве полноценного самостоятельного решения.

Экспериментальные исследования

В качестве исходного набора траекторных измерений берётся набор данных, ранее полученный в ходе натурных работ с многоступенчатым изделием при проведении лётных испытаний и используемый в исследованиях возможностей прогнозирования географических координат отделяющихся частей изделия с целью формирования признаков попадания и промаха с допустимым уровнем точности [1]. В исходный набор, фрагмент которого представлен в таблице 1, вносятся искусственные изменения в значения географической долготы не характерные для штатного прохождения объекта по априорной траектории (неявные признаки траекторных отклонений). На основании внесённых изменений (выделены в таблице курсивом) создаётся синтетическая колонка бинарного признака аварийной ситуации, в которой признаку «1» соответствует возникновение нештатной ситуации, а признаку «0» соответствует отсутствие такой ситуации, т.е. штатное функционирование.

Таблица 1 – Фрагмент исходного набора траекторных данных объекта
Table 1 – Fragment of object original trajectory dataset

index	Время	Широта	Долгота	Признак аварии
…		…	…	…
2399	53946,5	47,5384	46,7895	0
2400	53946,6	47,5487	*46,7900*	1
2401	53946,7	47,5487	*46,7905*	1
2402	53946,8	47,5487	*46,7911*	1
2403	53946,9	47,5487	*46,7916*	1
2404	53947,0	47,5487	*46,7921*	1
2405	53947,1	47,5482	46,7949	0
…	…	…	…	…
3999	53999,4	47,5246	46,9721	0
4000	53999,5	47,5246	*46,9553*	1
4001	53999,6	47,5246	*46,9553*	1
4002	53999,7	47,5246	*46,9553*	1
4003	53999,8	47,5246	*46,9443*	1
4004	54000,9	47,5246	*46,9443*	1
4005	54001,0	47,5246	46,9863	0
…	…	…	…	…

В качестве критерия качества (критерия разбиения классификатором набора данных на подвыборки) для построения модели классификатора состояний объекта выбирается энтропийный критерий Шеннона для системы с n возможными состояниями:

$$E = -\sum_{i=1}^{n} p_i \log_2 p_i , \quad (1)$$

в котором p_i – это частотная вероятность для класса i.

Следовательно, для бинарной классификации частотных вероятностей $p_{(+)}$ и $p_{(-)}$ критерий (1) принимает вид (2):

$$E = -p_{(+)} \log_2 p_{(+)} - p_{(-)} \log_2 p_{(-)}. \quad (2)$$

Все возможные ошибки модели классификатора при определении нештатных режимов функционирования бортовых систем можно описать, приняв в терминологии статистического анализа данных за нулевую гипотезу абсолютное отсутствие на борту нештатных состояний. Таким образом, можно выделить:

– ложноотрицательные классификации (*FN*), когда нештатные состояния будут классифицированы как штатные;

– ложноположительные классификации (*FP*), когда штатные состояния будут ошибочно классифицированы как нештатные.

Аналогично формулируются и критерии эффективных классификаций:

– отрицательные (*TN*), т.е. верные классификации штатных состояний;

– положительные (*TP*), т.е верные классификации нештатных состояний.

При этом обобщённую score-метрику модели, характеризующую количество верных классификаций, можно представить как отношение суммы истинных классификаций к сумме всех возможных, учитывая и ложные:

$$score = \frac{TP + TN}{TP + TN + FP + FN}.$$

С помощью инструментальных средств языка программирования высокого уровня Python, а также входящих в его состав библиотек анализа больших и неструктурированных массивов данных ScikitLearn и Pandas была построена первичная модель классификатора бортовых состояний (экземпляр класса DecisionTreeClassifer), основанная на предложенном Л. Брейманом алгоритме решающих деревьев [4, 5].

По результатам циклов подбора оптимальной глубины ветвления первичной модели классификатора на исходном наборе траекторных данных (таблица 1) посредством его перекрёстной проверки (кроссвалидации) и разбиения на обучающую и тестовую выборки сформирована база знаний score-метрик, графическое представление которой показано на рисунке 2.

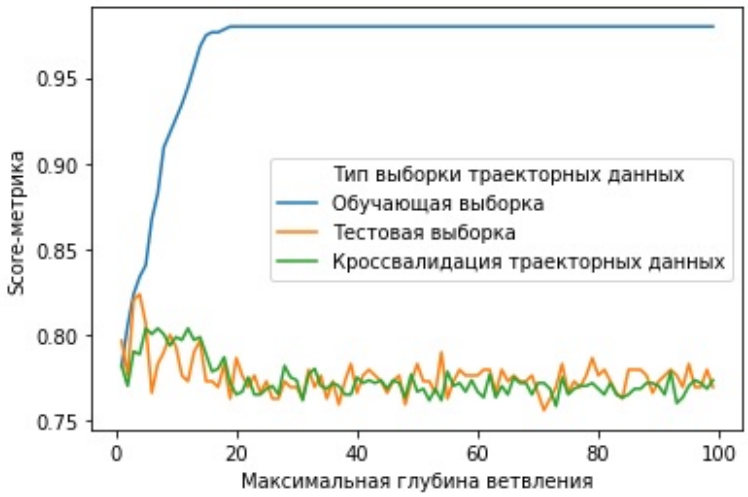

Рисунок 2 – База знаний score-метрик модели классификатора состояний
Figure 2 – Score-metrics knowledge base of state classifier model

Фрагмент базы знаний представлен в таблице 2. Анализ score-метрик базы позволяет определить числовое значение оптимальной глубины ветвления классификатора: 6, что соответствует значению кроссвалидационной метрики 0,818031, а также 0,865453 и 0,769452 для обучающей и тестовой выборки соответственно.

В связи с тем, что возникновение нештатных и аварийных ситуаций при проведении лётных испытаний сложных изделий ракетно-космической техники является достаточно редким, но чреватым серьезными последствиями явлением, в вопросе анализа набора траекторной измерительной информации, при формировании обучающего и тестового поднаборов

данных будет иметь место выраженный дисбаланс классов. При построении модели классификатора наличие несбалансированных классов говорит о необходимости введения дополнительных метрик качества и полноты классификаций, реализуемых моделью. В терминологии современных феноменов Big Data и машинного обучения (Machine learning) такими метриками выступают «precision» и «recall», характеризующие долю истинно положительных состояний объекта, определенных классификатором как положительные (precision – способность различать моделью классы), и долю истинных положительных классификаций из всех положительных классификаций модели (recall – способность модели к распознаванию определённого класса).

Таблица 2 – Фрагмент базы знаний score-метрик модели классификатора
Table 2 – Fragment of knowledge base of state classifier model

Максимальная глубина ветвления классификатора	Score-метрика		
	Обучающая выборка	Тестовая выборка	Кроссвалидация исходных данных
1	0,761876	0,796611	0,781961
2	0,803367	0,773271	0,771196
3	0,821825	0,821339	0,790222
4	0,831891	0,821729	0,788327
5	0,840401	0,810159	0,797041
6	0,865453	0,769452	0,818031
7	0,880552	0,773271	0,744434
8	0,905713	0,796101	0,785322
9	0,915781	0,813231	0,803045
10	0,956322	0,752513	0,730124
11	0,932568	0,765202	0,786989
12	0,942632	0,773881	0,805410

С помощью построения графика ROC-кривой был произведен подбор оптимального соотношения показателей полноты и точности классификаций первичной модели. При построении графика кривой бинарная классификация определялась путём варьирования специфического порога классификации, в зависимости от которого определялись значения статистических ошибок 1 и 2 рода: для каждого значения порога в интервале [0;1] с шагом $d_x = 0{,}01$ рассчитывались значения метрик точности и полноты с построением графика их взаимозависимости.

Оценка полученной ROC-кривой производится с помощью определения площади фигуры под ней согласно таблице 3. Малоэффективным классификаторам соответствуют значения до 0,5, наиболее эффективным моделям – значения в интервале [0,5;1,0]. Показатель площади *AUC* (Area Under Curve) рассчитывается с помощью численного метода трапеций:

$$AUC = \int f(x)dx = \sum_i \left[\frac{x_{i+1} + x_i}{2} \right] (y_{i+1} - y_i),$$

где x и y – это численные значения метрик точности и полноты на осях абсцисс и ординат соответственно.

Таблица 3 – Таблица оценки модели классификатора по показателю AUC
Table 3 – Classifier model assessment table for the AUC indicator

Численное значение метрики AUC	Качество модели классификатора
0,9-1,0	Максимальное
0,8-0,9	Высокое
0,7-0,8	Выше среднего
0,6-0,7	Среднее
0,5-0,6	Неудовлетворительное

Полученная ROC-кривая результатов расчётов метрик точности и полноты классификаций первичной модели с показателем AUC = 0,82 показана на рисунке 3.

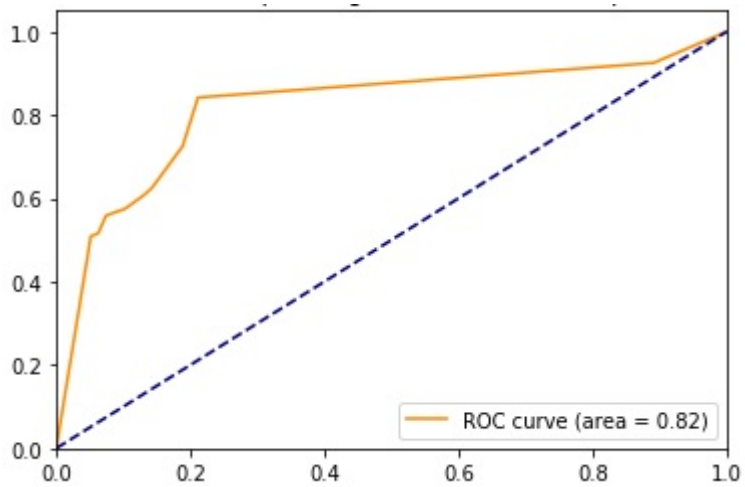

Рисунок 3 – ROC-кривая первичной модели
Figure 3 – ROC curve of primary model

По результатам анализа ROC-кривой первичной модели классификатора бортовых состояний объекта получены значения параметров и соответствующих метрик модели, наиболее подходящих для реализации качественной классификации (показатель точности классификаций 80 % при показателе полноты 86 %). Полученные значения параметров первичной модели представлены в таблице 4. Оптимальная первичная модель классификатора штатных и аварийных состояний показана на рисунке 4. Многочисленные исследования в области анализа больших объемов данных показывают, что одним из наиболее эффективных алгоритмов классификации неструктурированной информации является алгоритм случайного леса (random forest) [6]. В связи с этим, завершающим этапом исследования является переход от первичной к итоговой модели классификатора состояний, основанной на алгоритме случайного леса, основанном на использовании ансамблей (комитетов) первичных моделей классификаторов по типу решающих деревьев. Для этого с помощью инструмента автоматического подбора параметров GridSearchCV из программной библиотеки ScikitLearn был построен оптимальный экземпляр класса RandomForestClassifer, в качестве поддеревьев которого выступает полученная ранее первичная модель. Параметры полученной итоговой модели на основе алгоритма решающего леса из 27 поддеревьев также представлены в таблице 4.

Таблица 4 – Параметры оптимальных моделей классификатора состояний
Table 4 – Parameters of optimal state classifier models

Параметр классификатора	Первичная модель	Итоговая модель
Критерий	Энтропийный	Энтропийный
Глубина ветвления	7	7
Score-метрика	0,80341	0,87112
Recall-метрика	0,86135	0,89133
Precission-метрика	0,80532	0,85253

Заключение

Результатом проведённого исследования является эффективная модель классификатора бортовых состояний изделия ракетно-космической техники на основе неявных признаков траекторных отклонений, построенная на современных методах классификации и интеллектуального анализа данных. Было показано, что в результате применения разработанной модели достигаются показатели точности классификации 80 % и полноты классификации 86 %.

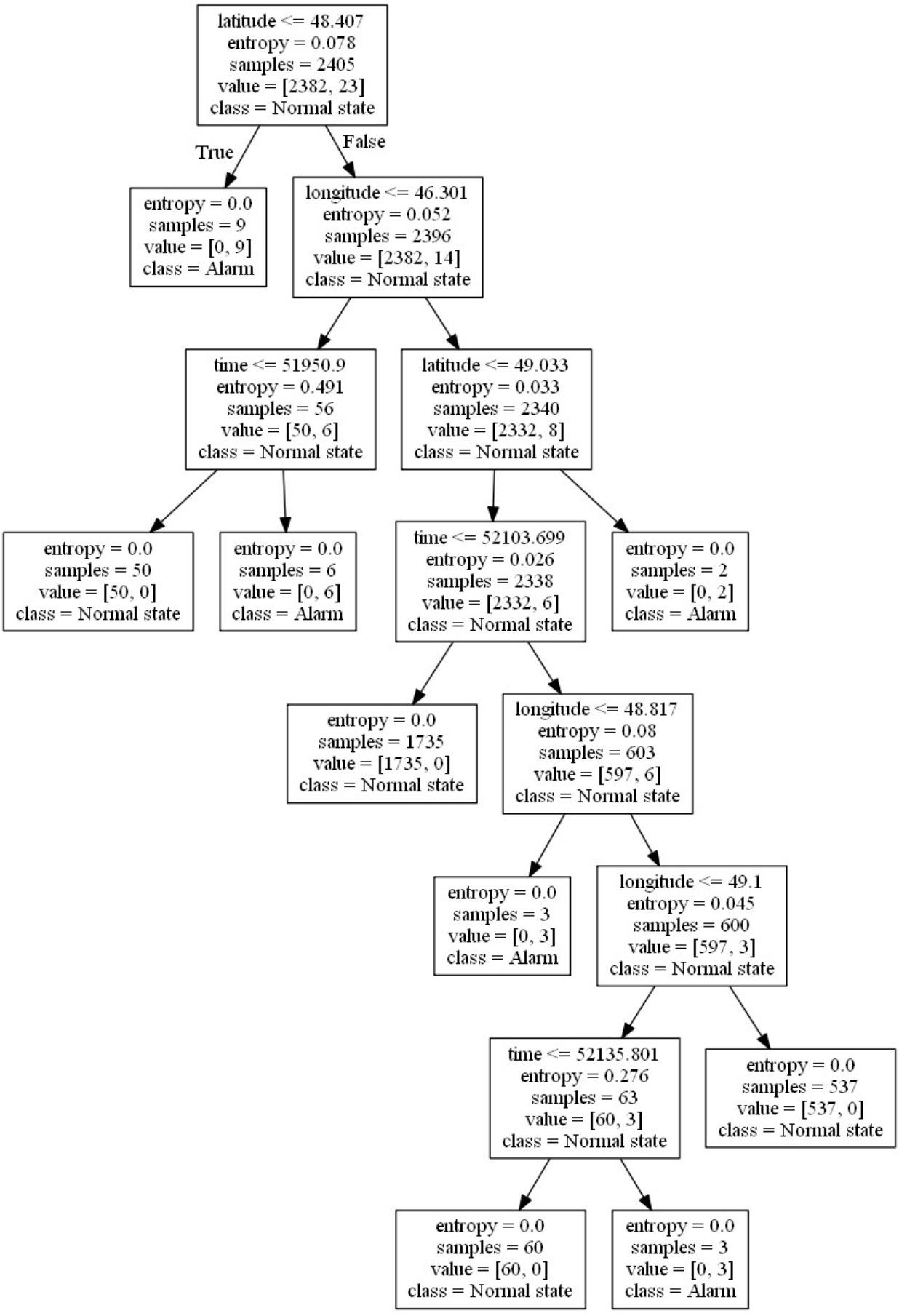

Рисунок 4 – Оптимальная первичная модель классификатора бортовых состояний
Figure 4 – Optimal primary model of onboard states classifier

Положительный эффект, возникающий в результате практического применения полученных в рамках исследования результатов заключается в возможности оперативного определения нештатных и аварийных ситуаций на борту сложного многопараметрического объекта ракетно-космической техники при проведении лётного эксперимента во время полигонных испытаний. Показатели эффективности полученной модели классификатора подтверждают возможность её использования в качестве интеллектуального средства парирования нештатных ситуаций и запуска аварийных циклограмм.

Интеграция разработанной модели в информационно-аналитическую инфраструктуру полигонного измерительного комплекса позволяет увеличить эффективность автоматизированных систем обработки и анализа бортовой телеметрии в части:

– сокращения времени принятия решений при значительном объёме контролируемых телеметрических параметров;

– повышения количества верно идентифицируемых нештатных состояний на борту изделия;

– разработки высокопроизводительных автоматизированных систем анализа телеметрии на базе перспективных компьютерных технологий.

Библиографический список

1. **Богданов А. А., Спицын С. В., Тихомиров С. А.** Методы регрессионного анализа в обработке траекторной информации изделий ракетно-космической техники // Вестник Рязанского государственного радиотехнического университета. 2020. № 74. С. 80-88. DOI:10.21667/1995-4565-2020-74-80-88.

2. **Спицын С. В., Товпеко А. В., Тихомиров С. А**. Алгоритм обработки траекторных измерений при проведении ракетных пусков // Вестник Рязанского государственного радиотехнического университета. 2019. № 70. С. 118-126. DOI: 10.21667/1995-4565-2019-70-118-126.

3. **Балашов С. Ю., Спицын С. В.** Особенности обработки, анализа и представления траекторных измерений при проведении пусков изделий РКТ // Вестник Рязанского государственного радиотехнического университета. 2019. № 69. С. 175-184. DOI:10.21667/1995-4565-2019-69-175-184.

4. **Richert W., Coelho L. P.** Building Machine Learning Systems with Python. Packt Publishing. 2013, 290 p.

5. **Breiman L.** Random Forests. Machine Learning Journal. 2001, 45(1), 5-32.

6. **Ali J., Khan R., Ahmad N., Maqsood I.** Random Forests and Decision Trees Machine Learning // IJCSI International Journal of Computer Science Issues. September, 2010, vol. 9, issue 5, pp. 272-278.

UDC 004.032.2

ON-BOARD DATA CLASSIFIER MODEL OBTAINED FROM ROCKET AND SPACE EQUIPMENT VEHICLE ON THE BASIS OF IMPLICIT FEATURES OF TRAJECTORAL DEVIATIONS

S. V. Spitsyn, post-graduate student, RSREU, design engineer, JSC «SRC «Progress» – department of special design bureau «Spectrum», Ryazan, Russia;
orcid.org/0000-0001-6468-5043, e-mail: spitsyn62@gmail.com
S. A. Tikhomirov, assistant professor, RSREU, senior specialist, JSC «SRC «Progress» – department of special design bureau «Spectrum», Ryazan, Russia;
orcid.org/0000-0002-3174-1536, e-mail: serge-tikhomirov@yandex.ru

The paper considers on-board data classifier model obtained from rocket and space equipment vehicle on the basis of implicit features of trajectoral deviations. The proposed classification model is based on specific binary classification algorithm in terms of modern and perspective machine learning methodology algorithms such as decision tree and random forest. ***The aim of the work*** *is to find a generalized algorithm for determining the occurrence of abnormal and emergency onboard situations in complex multiparametric objects (rocket and space equipment vehicle) based on trajectory data analysis. The trajectory data is obtained*

by polygon measuring instruments as part of general flow of onboard telemetry information during flight experiment. The paper presents a step-by-step process of constructing effective classifier model for normal and abnormal on-board states. The model is based on modern methods of data mining and machine learning, which are currently not widely used in technological processes of rocket and space industry.

Key words: *information and measurement support of rocket launches, trajectory measurements, machine learning, system analysis, telemetry information.*

DOI: 10.21667/1995-4565-2021-75-132-141

References

1. **Bogdanov A. A., Spitsyn S. V., Tikhomirov S. A.** Metody regressionnogo analiza v obrabotke traektornoj informacii izdelij raketno-kosmicheskoj tekhniki. *Vestnik Rjazanskogo gosudarstvennogo radiotehnicheskogo universiteta*. 2020, no. 74, pp. 80-88. (in Russian). DOI:10.21667/1995-4565-2020-74-80-88

2. **Spitsyn S. V., Tovpeko A. V., Tikhomirov S. A.** Algoritm obrabotki traektornyh izmerenij pri provedenii raketnyh puskov. *Vestnik Rjazanskogo gosudarstvennogo radiotehnicheskogo universiteta*. 2019, no. 70, pp. 118-126. (in Russian). DOI:10.21667/1995-4565-2019-70-118-126

3. **Balashov S. V., Spitsyn S. V.** Osobennosti obrabotki, analiza i predstavleniya traektornyh izmerenij pri provedenii puskov izdelij RKT. *Vestnik Rjazanskogo gosudarstvennogo radiotehnicheskogo universiteta*. 2019, no. 69, pp. 175-184. (in Russian). DOI:10.21667/1995-4565-2019-69-175-184

4. **Richert W., Coelho L. P.** *Building Machine Learning Systems with Python*. Packt Publishing. 2013, 290 p.

5. **Breiman L.** Random Forests. *Machine Learning Journal*. 2001, 45(1), 5-32.

6. **Ali J., Khan R., Ahmad N., Maqsood I.** Random Forests and Decision Trees Machine Learning. *IJCSI International Journal of Computer Science Issues*. September, 2010, vol. 9, issue 5, pp. 272-278.