

Использование методов интеллектуального анализа данных для прогнозирования поломок устройств сети передачи данных

¹*МЫРЗАТАЙ Али Алтынбайұлы, докторант, mirzataitegiali@gmail.com,

²РЗАЕВА Лейла Гумметовна, PhD, и.о. доцента, leilarza2@gmail.com,

³УСКЕНБАЕВА Гульжан Амангазыевна, PhD, и.о. доцента, gulzhum_01@mail.ru,

³ШУКИРОВА Алия Косаналиевна, PhD, и.о. доцента, aliya.shukirova@mail.ru,

²АБИТОВА Гульнара Аскеровна, PhD, и.о. доцента, gulya.abitova@gmail.com,

¹НАО «Кызылординский университет имени Коркыт Ата», Казахстан, Кызылорда, ул. Айтеке би, 29А,

²Astana IT University, Казахстан, Нур-Султан, пр. Мәңгілік Ел, Бизнес-центр EXPO, блок С1,

³Евразийский национальный университет имени Л.Н. Гумилева, Казахстан, Нур-Султан, ул. Сатпаева, 2,

*автор-корреспондент.

Аннотация. Рассматриваются два метода прогнозирования отказов сетевого оборудования. Цель статьи – раскрыть актуальность предложенного авторами подхода к использованию данных большого объема в выбранном методе машинного обучения и произвести сравнительный анализ итоговых значений с работами мировых исследователей. В первом разделе авторы анализируют работу ученых из Пекинского университета почты и телекоммуникаций, отмечая сильные и слабые стороны их метода. В разделе 2 авторы применили метод случайных деревьев при моделировании методов машинного обучения студии RapidMiner. Авторы провели работу с большим объемом данных и провели сравнительный анализ результатов моделирования методов. Доказана важность использования модели и методов прогнозирования отказов сетевого оборудования. В заключительном разделе авторы подчеркивают необходимость совершенствования моделей прогнозирования для их дальнейшего внедрения в рабочую среду. Также авторы подчеркивают, что выбранный метод прогнозирования сбоя в системе локальной сети ученых из Пекинского университета почты и телекоммуникаций является частным случаем, так же как и метод, предложенный авторами этой статьи.

Ключевые слова: методы машинного обучения, моделирование метода машинного обучения, прогнозы отказов сетевого оборудования, локальная сеть.

Введение

Постоянное развитие информационных технологий является фактором роста числа используемого активного оборудования и размеров локальных сетей предприятий, вследствие чего неизбежно растут и проблемы, связанные с грамотным построением ЛВС и её мониторингом. В этом примере мы бы хотели взять за основу компании, оказывающие услуги в сфере информационных технологий, в частности услуги, связанные с предоставлением систем ЛВС. Данные ИТ-фирмы предоставляют конечным пользователям услуги доступа к интернету, организацию, настройку и сопровождение ЛВС и активного оборудования, таких как коммутаторы, маршрутизаторы, хабы, серверы и прочее. Из-за масштабов развертываемых сетей такими сервисными компаниями очень сложно предоставлять услуги высокого качества и одновременно с этим поддерживать как технические, так и экономические показатели системы в

оптимальном состоянии. Существующие автоматизированные системы управления предприятием (EMS) должны соответствовать требованиям в рамках постоянного роста и степени сложности задач:

а) неуклонный рост числа предоставляемых услуг связи;

б) вследствие пункта «а», развертывание все большей и сложной системы маршрутных сетей передачи данных;

с) управление логистикой для обеспечения отказоустойчивости активных систем передачи данных;

д) вследствие пунктов «а» и «б», своевременная реакция на возникающие инциденты для обеспечения бесперебойной работы среды передачи данных конечным клиентам связи.

Чаще всего организации, предоставляющие такие услуги, используют системы оперативно-го мониторинга. Эти системы фиксируют инци-

денты, связанные с поломками оборудования, перегревом или выходом из строя из-за внешних факторов, задержками или потерями трафика, и множество других событий, которые могут быть выявлены в ходе активного мониторинга среды передачи данных. Инциденты регистрируются с устройств (коммутаторов) организации через операционные или коммуникационные каналы, а также и через конечных пользователей услуг. Общепринятая практика – обрабатывать инциденты на уровне диспетчерских служб или центров мониторинга – распространена во многих развитых организациях. Тем не менее, некоторые категории событий требуют немедленного решения, например: все неисправности или выход из строя активного оборудования. Для повышения эффективности работы организации можно применить прогнозирование стабильности работы ряда активного оборудования.

Одной из актуальных тем исследований поиска данных является успешное применение данных алгоритмов в сфере обслуживания, в частности, в работе банков, а также торгового и страхового бизнеса. Публикации на данную тему достаточно широко охватывают сферы применения этих моделей. Например, Ирина Чучуева в своей диссертации описала работу подобной модели, результатом которой является прогноз изменения цен на энергорынке. В модели использовалась экстраполяция касательно выборки максимального сходства для прогнозирования показателей энергетического рынка Российской Федерации [1].

Основным ожидаемым бизнес-эффектом от использования инструментов прогнозирования является снижение финансовых затрат за счет повышения технической готовности оборудования. Эта цель должна быть достигнута через сравнение изученных методов прогнозирования и составления собственного метода прогнозирования отказоустойчивости сетевого оборудования передачи данных [5].

В этой статье рассматривается вопрос применения методов машинного обучения и анализа данных в области информационных технологий, путем сравнения метода SVM-DES, выбранного китайскими исследователями, с выбранным авторами методом прогнозирования случайных деревьев [2]. В статье описаны частные случаи отказа оборудования и связанные с этим сложности, для предотвращения которых разработана модель прогноза, а также принципы построения такой модели. Проводится сравнение эффективности интеллектуальных способностей человека и искусственного интеллекта.

На сегодняшний день существует более двухсот различных методов прогнозирования, но на практике наиболее распространенными являются около двадцати. Число методов и моделей постоянно растет, добавляются новые, наиболее эффективные для определенных предметных областей или определенных задач [5].

Очевидно, что разработать единый принцип работы для всех возникающих случаев не представляется возможным, так как каждая модель прогнозирования имеет свои собственные требования к данным для входа, и не каждые входные данные соответствуют этим требованиям или разным ситуациям, которые могут быть спрогнозированы. Несоблюдение этих правил может повлиять на конечные результаты прогнозов. Особенное внимание уделяется пониманию работы модели, ее способу обработки входных данных, так как одна определенная модель может быть крайне эффективна для одних задач, а для других моделей разрешение этих задач может быть затруднено вследствие несоответствия, даже в случае соответствия данных задаваемым параметрам.

1. Обзор и анализ статьи «Failure prediction using machine learning and time series in optical network» [2].

В статье «Failure prediction using machine learning and time series in optical network», написанной исследователями из «Пекинского университета почты и телекоммуникаций» совместно с «Китайской корпорацией мобильной связи», предлагается метод прогнозирования отказа оборудования оптической сети, основанный на методе, сочетающем двойное экспоненциальное сглаживание (DES) и специально выбранный алгоритм опорных векторов машинного обучения (SVM). Предлагаемый метод DES-SVM может применяться в программно определяемой сети. Основным алгоритмом является несложным, поэтому представленный метод легко установить в сетевом контроллере. Информация об оборудовании из записи мониторинга системы управления сетью в реальной сети с мультиплексированием по длине волны (WDM) использовалась для подтверждения достоверности выбранного метода.

Типичный алгоритм машинного обучения, как правило, сталкивается с проблемой огромного объема предварительных данных (например, обычно более 100 000 элементов), необходимых для обучения модели принятия решений. При таком масштабе контроллер должен обладать большой емкостью памяти и мощными вычислительными возможностями. В реальной сети определение доступности узла выполняется множеством различных видов оборудования. Если контроллер будет отслеживать и прогнозировать состояние каждой единицы оборудования с такой шкалой, на него ляжет серьезная нагрузка. Необходимо выбрать алгоритм машинного обучения, который может обучить высокоточную модель с меньшим количеством данных. Алгоритм SVM обладает высокой эффективностью и высокой точностью при объеме данных менее 5000, что подходит для практического применения в оптической сети WDM.

Основным сценарием применения предлагаемого метода является программно-определяемая городская сеть (SDMAN/ПОГС), как показано на

рисунке 1.

На физическом плане узлы WDM образуют сеть мегаполиса в ячеистой топологии. На плоскости управления центральный контроллер собирает данные об эксплуатации и техническом обслуживании со всех узлов WDM, а затем анализирует данные перед предоставлением инструкций.

Предлагаемый китайскими исследователями метод работает в контроллере, используя данные операции для обучения модели прогнозирования и прогнозирования отказа оборудования. Если контроллер идентифицирует потенциальный сбой оборудования с помощью метода прогнозирования DES-SVM, он заранее инициирует меры защиты, чтобы рассчитать наилучший подход к защите услуг. Затем контроллер отправляет управляющие сообщения всем узлам WDM. Эти узлы переключают службы на безопасный путь, чтобы предотвратить потерю данных. Основная процедура метода прогнозирования DES-SVM описана ниже:

1. Контроллер собирает данные и выбирает

индикаторы. В алгоритме SVM мы используем функции для формирования вектора x в каждом элементе данных. Эти особенности и являются значением показателей. На рисунке 2 показано, что мы можем использовать точность классификации, обученную одному показателю, чтобы получить взаимосвязь между этим показателем и отказом оборудования. Если показатель тесно связан с отказом оборудования, изменение этого показателя, очевидно, повлияет на состояние отказа оборудования.

2. Далее идет обучение лучшей модели диагностики сбоев. При использовании алгоритма SVM два параметра – функция ядра и коэффициент наказания – были выбраны в соответствии с характеристиками проблемных данных. Выбор функции ядра и фактор наказания напрямую влияют на точность модели.

3. Третьим шагом прогнозирования был прогноз значения индикатора. На этом этапе они используют DES для каждой функции, тем самым устанавливая и обновляя кривую прогнози-

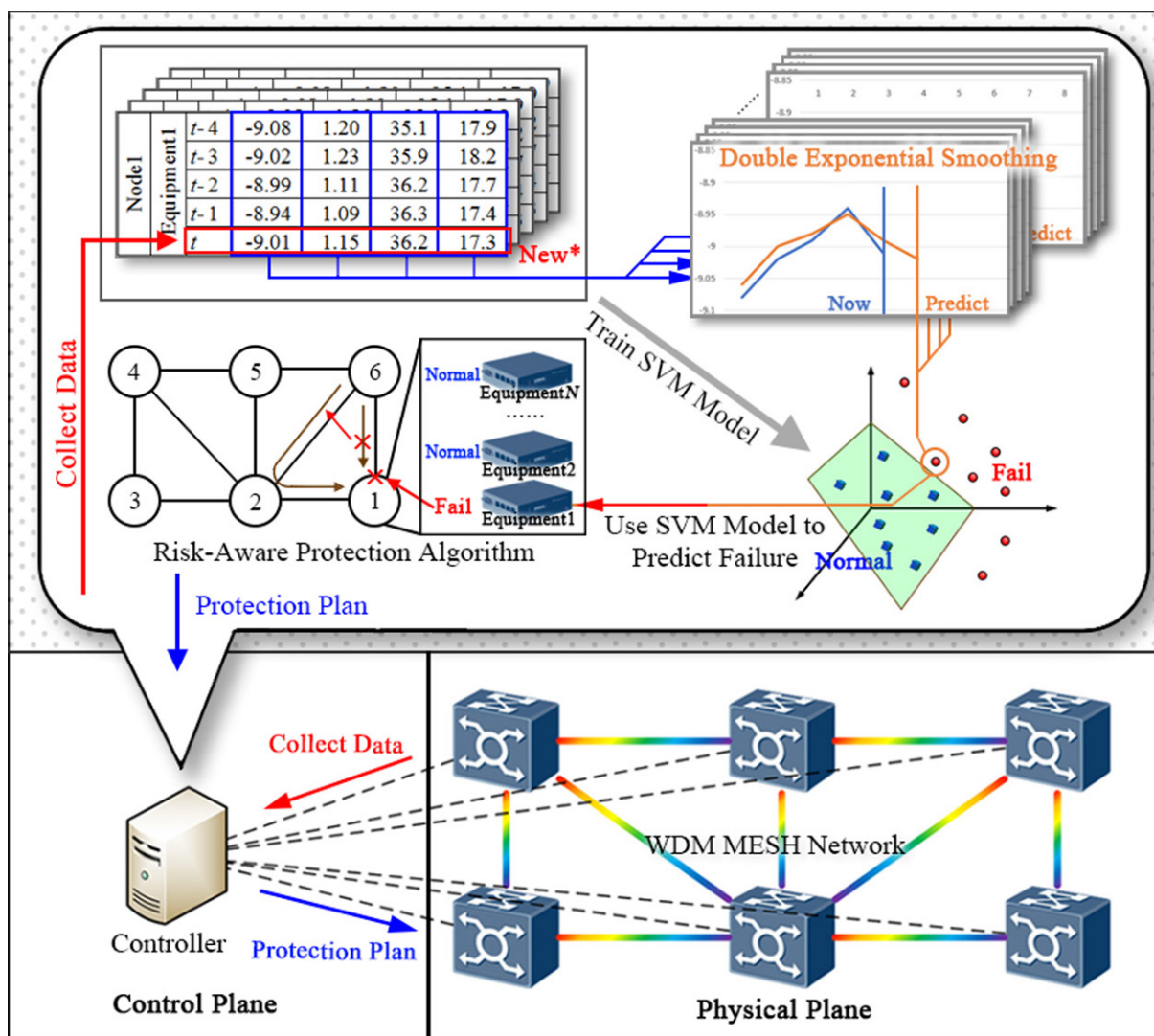


Рисунок 1 – Метод DES-SVM, предсказывающий отказ оборудования, и алгоритм защиты с учетом рисков, предотвращающий потерю данных в SDMAN/ПОГС

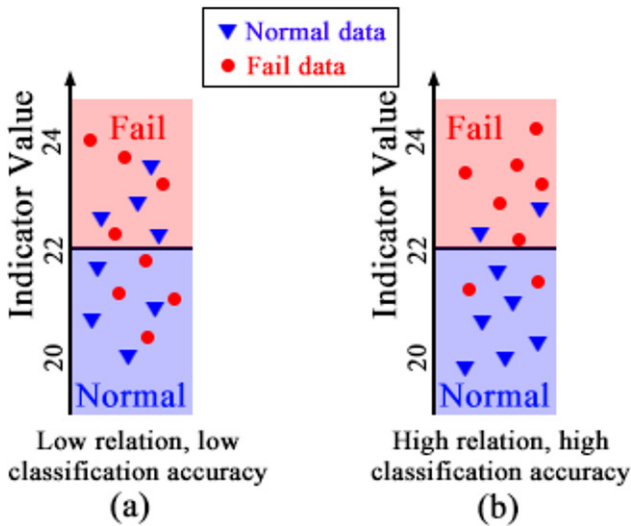


Рисунок 2 – Точность классификации по одному показателю отражает степень взаимосвязи между этим показателем и отказом оборудования

рования в соответствии с последними данными мониторинга.

4. Затем идет попытка прогноза сбоя оборудования. Тем самым, они получают модель диагностики неисправностей оборудования на этапах 1 и 2 и выводят прогнозное значение каждой функции на этапе 3.

Согласно данным эксперимента, показанным на рисунке 3, точность метода прогнозирования DES-SVM в среднем составляла 95,59%, что означало, что состояние отказа 95% плат можно было правильно предсказать. Более того, они выбрали для прогнозирования примерно десять плат, которые обычно выходили из строя. Хотя их показатели претерпевали очень большие колебания, точность прогнозирования достигла

86,37%. Эти результаты показывают, что их метод чаще всего правильно предсказывал состояние оборудования.

Согласно результатам эксперимента, метод прогнозирования DES-SVM предлагает следующие ключевые преимущества. Он может предсказать сбой платы в сети WDM, что означает, что службы могут быть защищены от потери данных до того, как произойдет сбой сети. Кроме того, SVM и DES не являются сложными алгоритмами. Таким образом, предлагаемый комбинированный метод легко устанавливается в системе контроллера.

Тем не менее, из-за плохой реакции DES в период, когда тенденция данных заметно изменилась, их метод не смог точно предсказать точку перегиба. Если бы данные были более точными (например, данные записывались ежечасно, а не за день), можно было бы определить более эффективные средства прогнозирования значений показателей.

2. Обзор метода прогнозирования сбоев методом случайных деревьев решения машинного обучения и его сравнение с вышеописанным методом зарубежных исследователей в данном конкретном случае

Современное оборудование оснащено большим количеством датчиков и контроллеров, которые позволяют контролировать и производить мониторинг его состояния работы. Нужно всего лишь выбрать наиболее информативный параметр при прогнозировании сбоев. Данную задачу можно решить, используя методы машинного обучения [3].

Для начала, в сравнении с первым методом, нам следует определиться с параметрами, которые будут подаваться на вход контроллера. Ав-

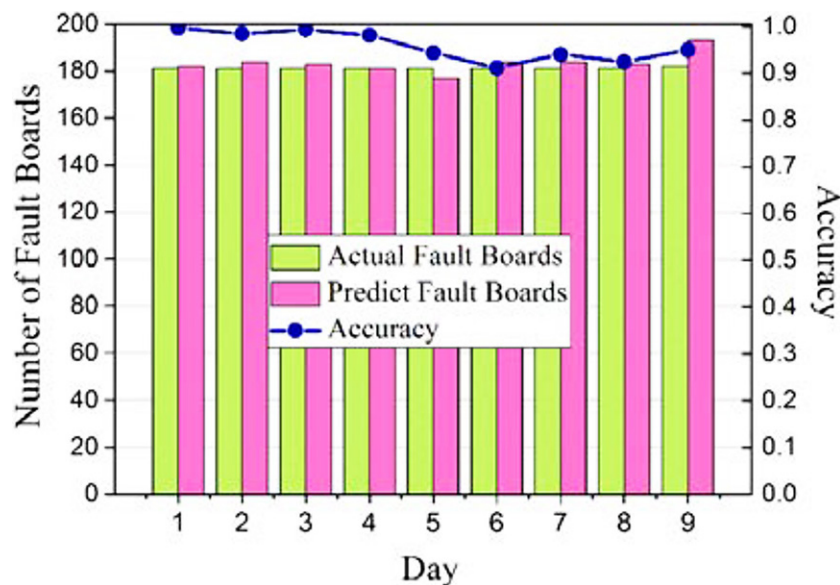


Рисунок 3 – Количество прогнозируемых плат неисправностей, по сравнению с фактическими платами неисправностей, и точность прогнозирования DES-SVM

торы статьи выбрали параметры, которые им удалось извлечь из системы мониторинга PRTG Network Monitor исходя из статистических данных об оборудовании:

- статистика поломок и инцидентов выхода из строя с учетом характера поломок и его периодичность;

- задействованные компоненты коммутатора;
- потеря пакетов между коммутаторами;
- время отклика коммутатора;
- индекс трафика на входе и выходе;
- число тревог и ошибок на коммутаторе;
- время пингов;
- и другие возможные параметры.

Следующим шагом является построение модели на основе метода машинного обучения (machine learning) – метод случайных деревьев с перекрестной проверкой: «Random forest with cross validation».

Следует отметить факт анализа и предварительной подготовки входных данных и определения их взаимосвязей непосредственно перед построением модели.

2.1 Выбор параметров и критериев для определения точности построения прогноза

В настоящее время широко используется такой критерий для определения точности построения прогноза, как сумма квадратов ошибок:

$$SSE = \sum_{t=1}^n e_t^2, \quad (1)$$

где e_t – разность между реальным значением и прогнозируемым; n – число значений.

Данный критерий наиболее часто используется при выборе оптимальной модели прогно-

зирования. В критерии каждое значение ошибки прогноза возводится в квадрат, что подчеркивает большие ошибки прогноза. Данная особенность очень важна, так как модель прогнозирования, постоянно дающая средние по величине ошибки, в некоторых случаях может быть предпочтительнее другой модели, имеющей малые ошибки, но периодически дающей неприемлемые по абсолютной величине выбросы [5].

Критерий (1) хорошо подходит для моделей, которые возвращают результаты в виде вероятности наступления события. Но в данном конкретном случае были использованы модели, возвращающие дискретные значения. В связи с этим был выбран критерий определения точности модели, учитывающий число верных «попаданий» [5]:

$$A = \frac{a+d}{a+b+c+d}, \quad (2)$$

где a – сумма верно предсказанных событий первого типа («поломка»); b – сумма неверно предсказанных событий первого типа («поломка»); c – сумма неверно предсказанных событий второго типа («рабочее состояние»); d – сумма верно предсказанных событий второго типа («рабочее состояние»).

Суть модели на основе применения метода случайных деревьев при решении машинного обучения:

- входные данные состояли из записей в количестве от 1700 до 2600 строк, а количество параметров каждой строки была не менее 14 (рисунок 4).

- связи между параметрами и их веса назначались программным обеспечением, тем самым создавались, так называемые, модели деревьев решений (рисунок 5) [4].

Breaking	Время безн...	Downtime	CPU Load	CPU 1	Available Me...	Percent Ava...	Available Me...	Percent Ava...	Available Me...	Response TI...	CPU Load In...	Traffic Index	Alarms
No	182	0	8.570	8.640	?	?	?	?	?	?	?	?	?
No	181	0	8.270	8.430	?	?	?	?	?	?	?	?	?
No	180	0	8.520	8.530	?	?	?	?	?	?	?	?	?
No	179	0	8.770	8.770	?	?	?	?	?	?	?	?	?
No	178	0	8.780	8.760	?	?	?	?	?	?	?	?	?
No	177	0	8.770	8.770	?	?	?	?	?	?	?	?	?
No	176	0	8.810	8.780	?	?	?	?	?	?	?	?	?
No	175	0	8.600	8.630	?	?	?	?	?	?	?	?	?
No	174	0	8.120	8.340	?	?	?	?	?	?	?	?	?
No	173	0	8.280	8.380	?	?	?	?	?	?	?	?	?
No	172	0	8.600	8.640	?	?	?	?	?	?	?	?	?
No	171	0	8.670	8.660	?	?	?	?	?	?	?	?	?
No	170	0	8.670	8.730	?	?	?	?	?	?	?	?	?
No	169	0	8.330	8.470	?	?	?	?	?	?	?	?	?
No	168	0	8.590	8.660	?	?	?	?	?	?	?	?	?
No	167	0	8.090	8.310	?	?	?	?	?	?	?	?	?
No	166	0	8.290	8.360	?	?	?	?	?	?	?	?	?
No	165	0	8.610	8.680	?	?	?	?	?	?	?	?	?

Рисунок 4 – Образец входных данных

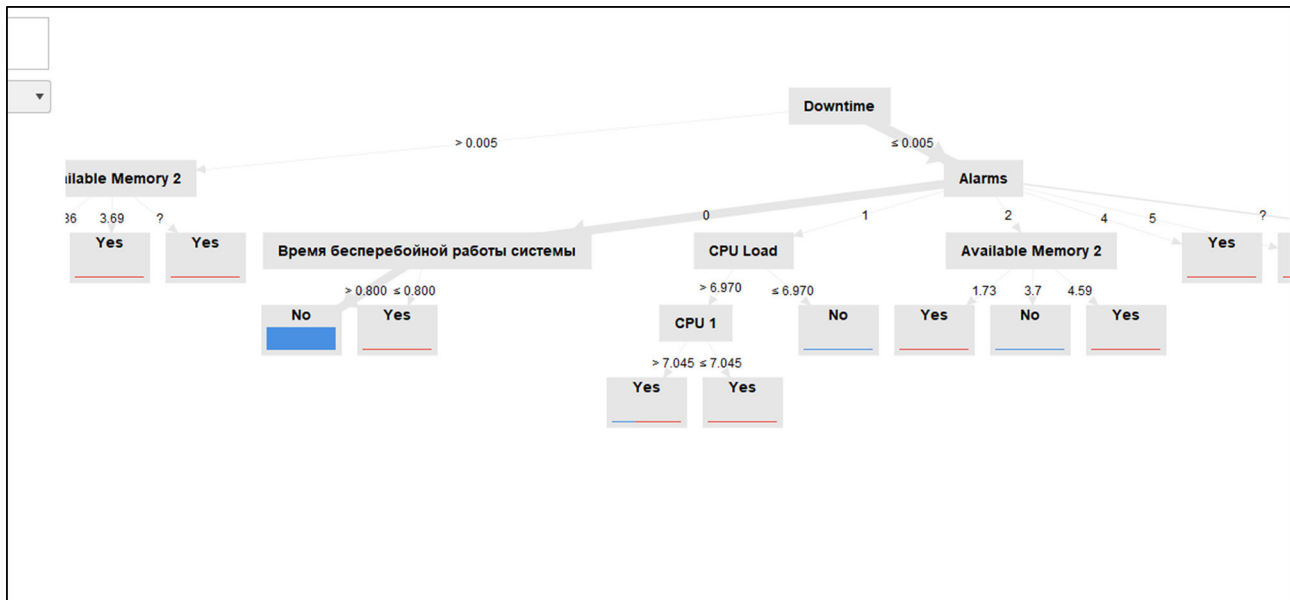


Рисунок 5 – Одно из 1000 деревьев решений в модели

• Однако некоторые параметры были либо упрощены, либо не учитывались построенной моделью. При этом при оценке модели имеющиеся в наличии данные разбиваются на «n» частей. Далее, на «n-1» частях данных производится обучение модели, а оставшаяся часть данных используется для тестирования.

• Процесс повторяется n раз; в итоге каждая из n частей данных используется для тестирования. В итоге получается оценка эффективности выбранной модели с наиболее равномерным использованием имеющихся данных.

Самый важный момент при построении модели машинного обучения – это подготовка данных. Это означает, что надо провести комплексные работы по очистке данных от пустых значений, объединение, добавление значений и полей, а также другие работы, которые можно делать как вручную, так и в автоматизированном виде. Метод автоматизированной работы предпочтителен ввиду того, что данных для обработки может быть очень много, и передать эту работу компьютеру является эффективным и рациональным использованием ресурсов. В этом случае основная подготовка данных проводилась с помощью инструментов системы мониторинга PRTGMonitor и Excel, но было выполнено дополнительное уравнивание данных [5].

2.2 Подготовка выборки данных, программа реализации и оптимизация

Выборка данных для построения модели прогнозирования была получена на основе данных за один год работы пяти единиц сетевого оборудования. Все собранные данные были разделены на три части: реальная выборка данных с малым количеством поломок одного оборудования за один год, реальная выборка данных с малым количе-

ством поломок пяти единиц оборудования, объединенных в одну базу данных, и сгенерированная выборка данных с количеством поломок в два раза больше семи единиц оборудования, объединенных в одну базу данных. Все три базы данных были поочередно использованы в обеих моделях.

Анализ данных с помощью прогнозирующей модели производился на обычном компьютере (Intel(R) Core(TM) i5-7500, 3.40GHz, ОЗУ 16 Гб), вместо кластера, как это обычно происходит в промышленных масштабах.

Модель (рисунок 6) была построена с помощью инструмента RapidMiner Studio version 9.8, который является решением с открытым исходным кодом и при этом обладает мощным функционалом для разработки и интуитивно-понятным графическим интерфейсом.

Результаты исследования

1. С увеличением общего числа поломок, количества выборок и параметров (14 критериев) полученный показатель улучшился до 72% на основе метода машинного обучения (рисунок 7).

2. После увеличения общих чисел выборок около 2500 записей при тех же 14 параметрах на запись, точность прогноза возросла до 100%, при этом точность прогнозов поломок снизилась до 96% (рисунки 8 и 9).

Выводы

Проведенное в статье сравнение двух исследований показало, что использование методов интеллектуального анализа данных для прогнозирования поломок устройств сети передачи данных позволяет спрогнозировать потенциальные неисправности. Однако следует разделить два метода прогнозирования по их характеристикам входных данных на входе регуляторов и количеству обраба-

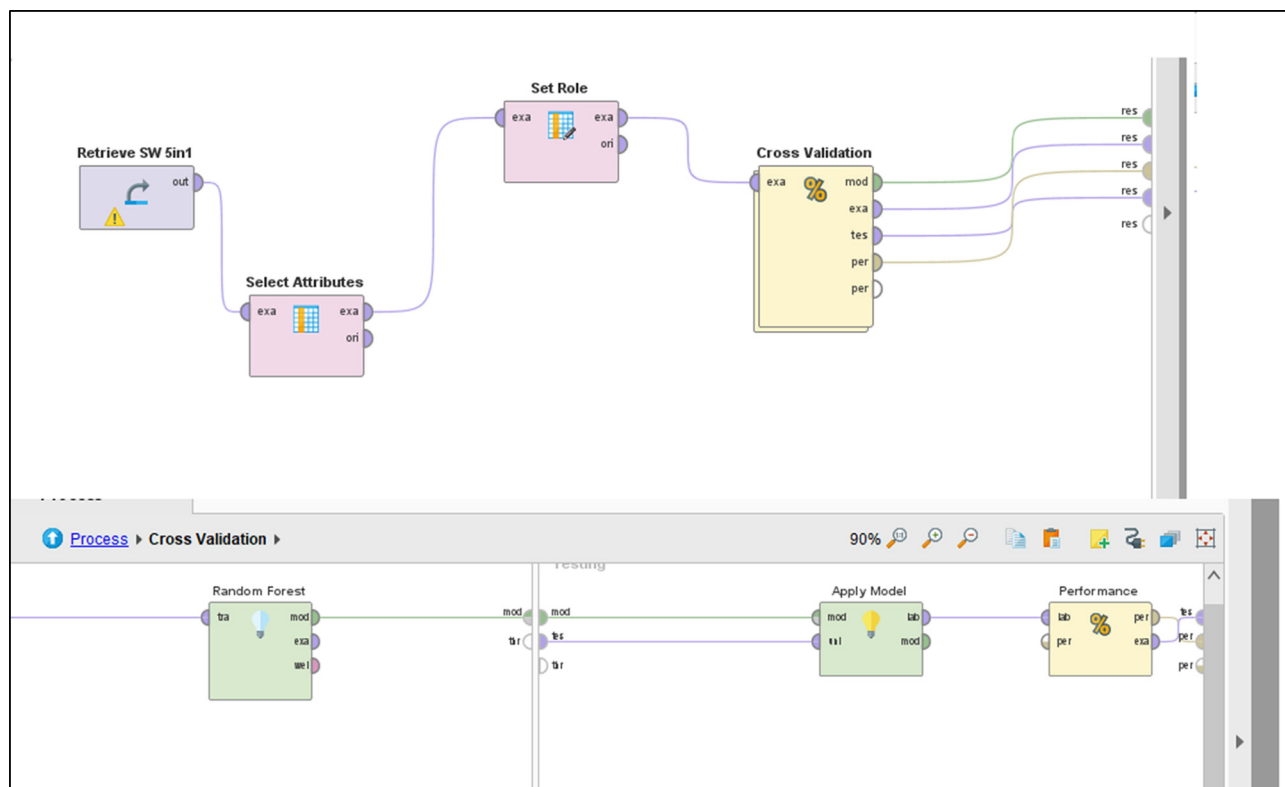


Рисунок 6 – Random forest model with cross validation

Table View Plot View

accuracy: 99.84% +/- 0.26% (micro average: 99.84%)

	true No	true Yes	class precision
pred. No	1814	3	99.83%
pred. Yes	0	8	100.00%
class recall	100.00%	72.73%	

Рисунок 7 – Показатель прогноза МО при обработке 1800 записей

Table View Plot View

accuracy: 99.97% +/- 0.11% (micro average: 99.97%)

	true No	true Yes	class precision
pred. No	2889	0	100.00%
pred. Yes	1	32	96.97%
class recall	99.97%	100.00%	

Рисунок 8 – Показатель прогноза МО при обработке 2500 записей

тываемых значений в единице времени. Так, если вспомнить метод исследователей из Пекинского университета почты и телекоммуникаций, метод DES-SVM, то недостатком этого метода является трудность обработки входных данных больше

5000 за единицу времени, когда как наш метод мог позволить обрабатывать более $2500 \cdot 14 = 35000$ единиц данных за единицу времени. Этот показатель крайне важен, так как прогноз поломок будет более точными, если данные на входе контроллеров

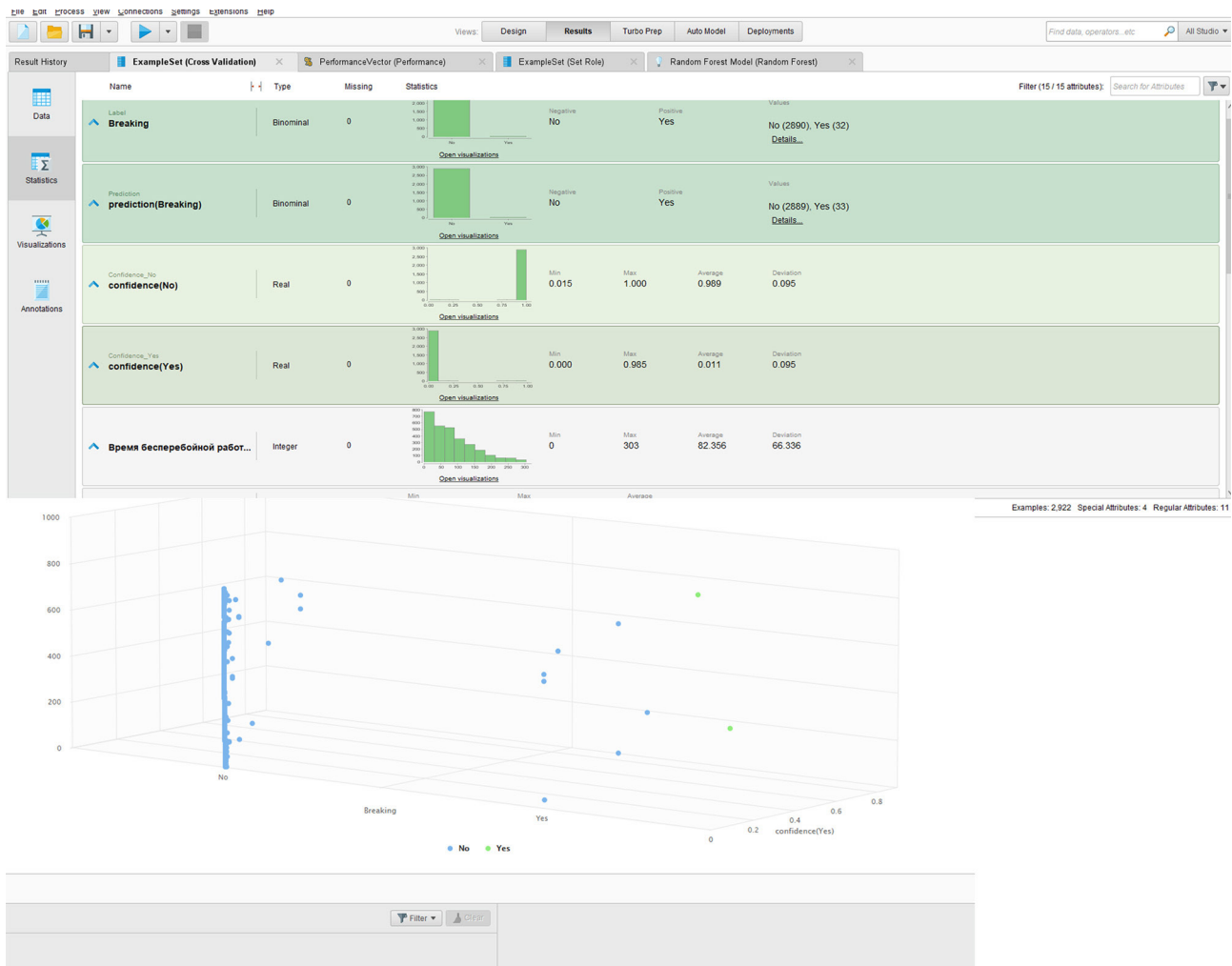


Рисунок 9 – Визуализация результатов модели в виде статистики и графов

будут более вариативными и объемными. Иными словами, чем больше данных, тем меньше шансов на ложный прогноз моделью.

Также следует отметить, что из-за плохой реакции DES в период, когда тенденция данных заметно изменилась, их метод не смог точно предсказать точку перегиба. Если бы данные были более точными (например, данные записывались каждую секунду, а не за день), можно было бы определить более эффективные средства прогнозирования значений показателей. В то же время, метод составления работы авторами статьи, использова-

ние дополнительных ПО (PRTGMonitor) для сбора больших значений сенсоров, показали незначительную, но более эффективную работу метода алгоритма машинного обучения.

Однако следует подчеркнуть, что исследователи из Пекинского университета почты и телекоммуникаций доказали на практике свою теорию, и их данные по проверке метода DES-SVM являются более живыми и достоверными, в отличие от протестированного авторами статьи метода Random Forest with Cross Validation на ПО RapidMiner, хоть и с живыми метаданными.

СПИСОК ЛИТЕРАТУРЫ

1. Чучуева И.А. Модель прогнозирования временных рядов по выборке максимального подобию. – Москва, 2012. – 155 с.
2. Zhilong Wang, Min Zhang, Danshi Wang, Chuang Song, Min Liu, Jin Li, Liqi Lou, and Zhuo Liu «Failure prediction using machine learning and time series in optical network» Optics Express Vol. 25, No. 16, 18553-18565 pages, 7 Aug 2017 – статья на английском языке.
3. Alestra S., Brand C., Burnaev E., Erofeev P., Papanov A., Bordry C., Silveira-Freixo C. Rare event anticipation and degradation trending for aircraft predictive maintenance // 11th World Congress on Computational Mechanics, WCCM 2014, 5th European Conference on Computational Mechanics, ECCM 2014 and 6th European Conference on Computational Fluid Dynamics, ECFD 2014 11, 2014. С. 6571-6582. – статья на английском языке.

4. Чистяков С.П. Случайные леса: обзор // Труды Карельского научного центра РАН. 2013. № 2. С. 117-136.
5. Пивоварова Н.В., Порошина В.И. Интеллектуальный анализ данных для пригородных пассажирских перевозчиков // Интернет-журнал «Науковедение». Май-июнь 2017. Том 9. № 3.

Деректер желісінің құрылғыларының бұзылуын болжау үшін деректерді интеллектуалды талдау әдістерін қолдану

¹***МЫРЗАТАЙ Али Алтынбайұлы**, докторант, mirzataitegiali@gmail.com,

²**РЗАЕВА Лейла Гумметовна**, PhD, доцент м.а., leilarza2@gmail.com,

³**УСКЕНБАЕВА Гульжан Амангазыевна**, PhD, доцент м.а., gulzhum_01@mail.ru,

³**ШУКИРОВА Алиа Косаналиевна**, PhD, доцент м.а., aliya.shukirova@mail.ru,

²**АБИТОВА Гульнара Аскеровна**, PhD, доцент м.а., gulya.abitova@gmail.com,

¹«Қорқыт Ата атындағы Қызылорда университеті» КеАҚ, Қазақстан, Қызылорда, Әйтеке би көшесі, 29А,

²Astana IT University, Қазақстан, Нұр-Сұлтан, Мәңгілік Ел даңғылы, Бизнес-центр EXPO, блок С1,

³Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Қазақстан, Нұр-Сұлтан, Сәтпаев көшесі, 2,

*автор-корреспондент.

Аңдатпа. Желілік жабдықтың істен шығуын болжаудың екі әдісі қарастырылған. Мақаланың мақсаты – машиналық оқытудың таңдалған әдісінде үлкен көлемдегі деректерді қолдануға авторлар ұсынған тәсілдің өзектілігін ашу және қорытынды мәндерді әлемдік зерттеушілердің жұмыстарымен салыстырмалы талдау жасау. Бірінші бөлімде авторлар Пекин пошта және телекоммуникация университетінің ғалымдарының жұмысын талдайды, олардың әдісінің күшті және әлсіз жақтарын атап өтеді. 2-бөлімде авторлар RapidMiner студиясының машиналық оқыту әдістерін модельдеуде кездейсоқ ағаш әдісін қолданды. Авторлар деректердің үлкен көлемімен жұмыс жасады және әдіспен модельдеу нәтижелеріне салыстырмалы талдау жасады. Желілік жабдықтың істен шығуын болжаудың моделі мен әдістерін қолданудың маңыздылығы дәлелденді. Қорытынды бөлімде авторлар оларды жұмыс ортасына одан әрі енгізу үшін болжау модельдерін жетілдіру қажеттілігін атап өтеді. Сондай-ақ, авторлар Пекин пошта және телекоммуникация университетінің ғалымдарының таңдаған жергілікті желі жүйесіндегі сәтсіздіктерді болжау әдісі осы мақаланың авторлары ұсынған әдіс сияқты ерекше жағдай болып табылады.

Кілт сөздер: машиналық оқыту әдістері, машиналық оқыту әдісін модельдеу, желілік жабдықтың істен шығуын болжау, жергілікті желі.

The Use of Data Mining Methods for Predicting Breakdowns of Data Transmission Network Devices

¹***MYRZATAY Ali**, doctoral student, mirzataitegiali@gmail.com,

²**RZAYEVA Leila**, PhD, Acting Associate Professor, leilarza2@gmail.com,

³**USKENBAYEVA Gulzhan**, PhD, Acting Associate Professor, gulzhum_01@mail.ru,

³**SHUKIROVA Alia**, PhD, Acting Associate Professor, aliya.shukirova@mail.ru,

²**ABITOVA Gulnara**, PhD, Acting Associate Professor, gulya.abitova@gmail.com,

¹NCJSC «Korkyt Ata Kyzylorda University», Kazakhstan, Kyzylorda, Aiteke bi Street, 29A,

²Astana IT University, Kazakhstan, Nur-Sultan, Mangilik El Avenue, EXPO Business-Center, Block C1,

³L.N. Gumilyov Eurasian National University, Kazakhstan, Nur-Sultan, Satpayev Street, 2,

*corresponding author.

Abstract. Discusses two methods for predicting network equipment failures. The purpose of the article is to reveal the relevance of the approach proposed by the authors to the use of large-volume data in the chosen method of machine learning and to make a comparative analysis of the final values with the works of world researchers. In the first section, the authors analyze the work of scientists from the Beijing University of Post and Telecommunications, noting the strengths and weaknesses of their method. In section 2, the authors applied the random tree method in modeling machine learning methods of RapidMiner studio. The authors carried out work with a large amount of data and conducted a comparative analysis of the results of modeling by the method. The importance of using models and methods for predicting network equipment failures is proved. In the final section, the authors emphasize the need to improve forecasting models for their further implementation in the working environment. The authors also emphasize that the chosen method for predicting failures in the local network system of scientists from Beijing University of Posts and Telecommunications is a special case, as well as the method proposed by the authors of this article.

Keywords: machine learning methods, modeling of machine learning method, network equipment failure forecasts, LAN.

REFERENCES

1. Chuchueva I.A. Model' prognozirovaniya vremennyh rjadov po vyborke maksimal'nogo podobija. – Moscow, 2012. – 155 p.
2. Zhilong Wang, Min Zhang, Danshi Wang, Chuang Song, Min Liu, Jin Li, Liqi Lou, and Zhuo Liu «Failure prediction using machine learning and time series in optical network» Optics Express Vol. 25, No. 16, 18553-18565 pages, 7 Aug 2017. – article
3. Alestra S., Brand C., Burnaev E., Erofeev P., Papanov A., Bordry C., Silveira-Freixo C. Rare event anticipation and degradation trending for aircraft predictive maintenance // 11th World Congress on Computational Mechanics, WCCM 2014, 5th European Conference on Computational Mechanics, ECCM 2014 and 6th European Conference on Computational Fluid Dynamics, ECFD 2014 11, 2014. pp. 6571-6582. – article
4. Chistjakov S.P. Sluchajnye lesa: obzor // Trudy Karel'skogo nauchnogo centra RAN. 2013. No. 2. pp. 117-136.
5. Pivovarova N.V., Poroshina V.I. Data mining for suburban passenger carriers // Online journal «Naukovedeniye». May-June 2017. Volume 9. No. 3.