

УДК 004.852

DOI 10.52167/1609-1817-2022-121-2-294-301

А.А. Мырзатай¹, Л.Г. Рзаева², Н.Ж. Құрманай¹, М.А. Жакенов³, Г.Абитова⁴

¹Кызылординский университет им. Коркыт-ата, Кызылорда, Казахстан

²Евразийский Национальный Университет им. Л.Н.Гумилева, Нур-Султан, Казахстан

³Гуманитарно-техническая академия, Кокшетау, Казахстан

⁴Astana IT University. Нур-Султан, Казахстан

e-mail: mirzataitegiali@gmail.com

АВТОМАТИЗАЦИЯ СБОРА И ПОДГОТОВКИ ДАННЫХ ИЗ СИСТЕМЫ МОНИТОРИНГА ЛВС ДЛЯ ДАЛЬНЕЙШЕГО ПРИМЕНЕНИЯ В СИСТЕМАХ ПРОГНОЗИРОВАНИЯ БАЗИРУЮЩИХСЯ НА МЕТОДАХ МАШИННОГО ОБУЧЕНИЯ

Аннотация. В статье рассматривается работа по разработке автоматизированной системы сбора и подготовки данных полученных из ПО системы мониторинга ЛВС PRTG Network Monitor. Цель статьи – выявить систематический подход к выявлению проблем, смоделировать бизнес-процессы согласно поставленным задачам и определить функции внедряемых решений. Авторы подчеркивают необходимость наличия систем мониторинга и хранения данных для оптимизации работ служб поддержки ЛВС, а также для дальнейшего внедрения других систем, например, системы сбора и подготовки данных для дальнейшего применения в области машинного прогнозирования. По вышесказанным причинам авторы рассмотрели возможность написания скрипта, который удовлетворял бы поставленным задачам. А именно: экономическая дешевизна при разработке, а не покупке готового софта, возможность работы скрипта с исходящими данными ПО PRTG Network Monitor, пригодный к чтению и анализу исходящих данных системой машинного прогнозирования, объединение максимально возможного числа таблиц в единую таблицу данных, и пр.

Ключевые слова. Прогнозы отказов сетевого оборудования, локальная сеть, системы мониторинга, системы сбора и упорядочивания данных, ЛВС, моделирование бизнес-процессов, скрипт, разработка скрипта, автоматизация сбора данных, автоматизация, объединение данных, dataframe.

Введение.

Из-за роста различных сегментов ЛВС и ее компонентов, таких как, например, коммутаторы и маршрутизаторы, многие фирмы и компании, предоставляющие услуги связи или выхода в сеть «Интернет», применяют системы мониторинга активных оборудований. Эти системы мониторинга показывают оператору показатели тех или иных сенсоров в режиме реального времени: доступность оборудования (ping), показатели температуры процессора сетевого оборудования и серверного помещения, объем курсирующего трафика через маршрутизатор, загруженность процессора или ПЗУ коммутаторов и маршрутизаторов, а также множество других показателей, которые хочет и может регистрировать оператор. [1]

Применение прогнозирующих систем, использующих методы и модели машинного обучения, являются новым этапом развития конкуренции на рынке предоставляемых услуг. Ведь фирма, которая более подготовлена к внештатным ситуациям и знающая срок службы своих активных систем и средств передачи данных, может оказывать более качественные услуги и одновременно сокращать расходы на ремонтные работы. [2] Однако прежде чем применять системы прогнозирования, следует подготовить данные из компонентов ЛВС, которые могут содержать информацию о состоянии активных оборудования. Например, такими системами могут быть ПО систем мониторинга.

Материалы и методы.

Авторы статьи рассмотрели систему мониторинга PRTG Network Monitor, которая была внедрена в одну из столичных фирм. Она предоставляла услуги доступа в сеть «Интернет» своим пользователям. Внедренная система мониторинга осуществляла мониторинг всех сетевых коммутаторов и маршрутизаторов. Информация датчиков от активных сетевых оборудования отправлялся на сервер путем протоколов SMTP, и ПО PRTG Network Monitor выводило на экран оператора необходимые сведения в графическом и информативном виде.

На рисунке 1 продемонстрирована модель процесса мониторинга ЛВС оператором с привлечением ПО PRTG Network Monitor.

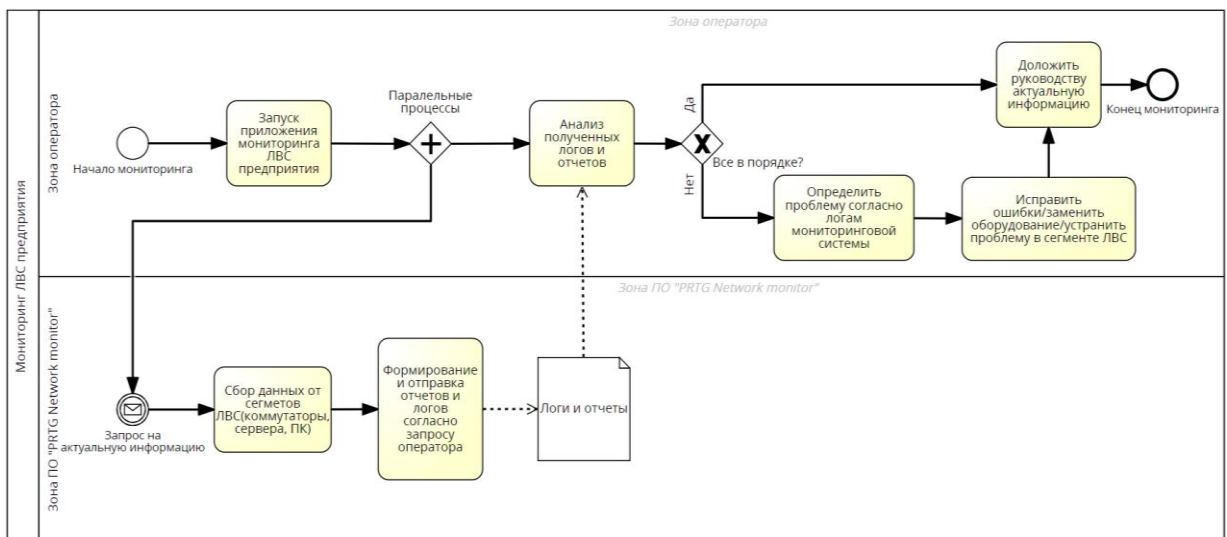


Рисунок 1 – Модель процесса мониторинга ЛВС операторов

В планах развития фирмы был пункт по внедрению систем прогнозирования инцидентов на базе машинного обучения в ближайшие несколько лет (рисунок 2). Этот процесс внедрения подразумевал также и научное исследование в этой области.[3] При исследовании появлялись новые задачи, которые авторы попытались решить, разбив на несколько частей.

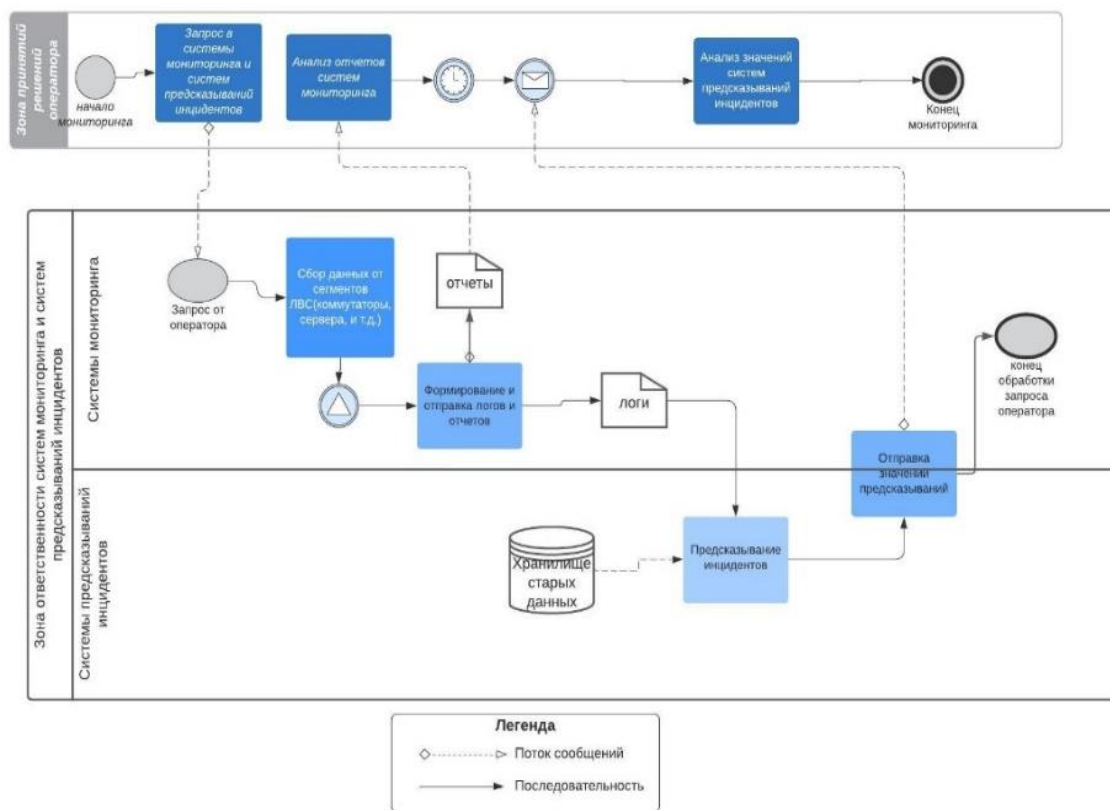


Рисунок 2 - Процесс мониторинга состояния ЛВС после внедрения систем мониторинга и систем прогнозирования

Одной из таких задач было автоматизирование процесса преобразования готовых данных, которых выдает система мониторинга, в подходящий вид, который может использовать система прогнозирования, базирующаяся на машинном обучении. На рисунке 3 показан процесс между событием «Формирование и отправка отчетов и логов» и событием «Предсказывание инцидентов». При автоматизации этого процесса мы столкнулись со следующими явлениями:

1) Главной проблемой использования исходящих данных из ПО PRTG Network Monitor является то, что все данные записаны в формате PDF, а метаданные в формате CSV, но эти данные отображаются некорректно для дальнейшего использования в модели машинного прогнозирования. [4,5]

2) Кроме того, чем больше количество сенсоров нужно вывести в качестве метаданных, тем больше отдельных CSV файлов будет экспортировано из ПО PRTG Network Monitor (рисунок 3). Эта большая проблема, так как использование нескольких баз данных с показаниями сенсоров в машинном обучении трудоемко и нецелесообразно по причине того, что многочисленность dataframe-ов (таблиц) усложняет процесс тренировки и обучения модели машинного обучения. [4,5]

3) Самая важная часть проблемы заключается в том, что подготовка данных человеком/специалистом будет занимать значительно большее время, когда как автоматизация этого процесса является жизненно необходимым для дальнейшего использования в системах машинного прогнозирования.

4) Итоговый продукт должен был быть экономически выгодным для внедрения. Например, если бы сама фирма разработала необходимый скрипт или программу, то необходимость покупки дорогого стороннего лицензионного софта отпала бы сама собой. [4,5,6]

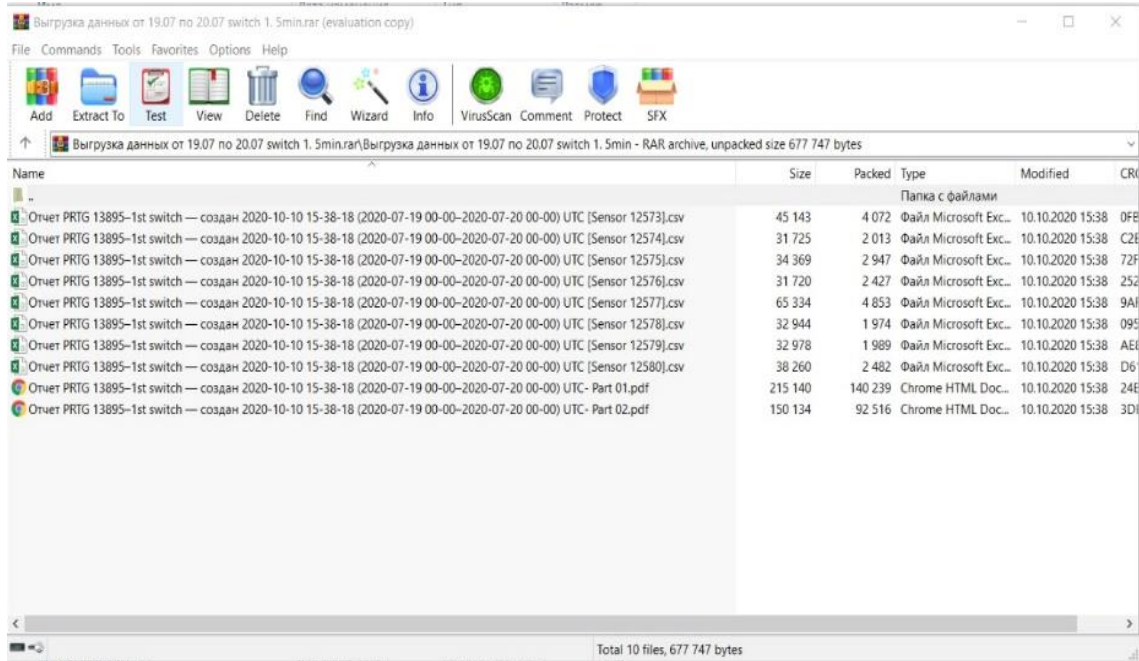


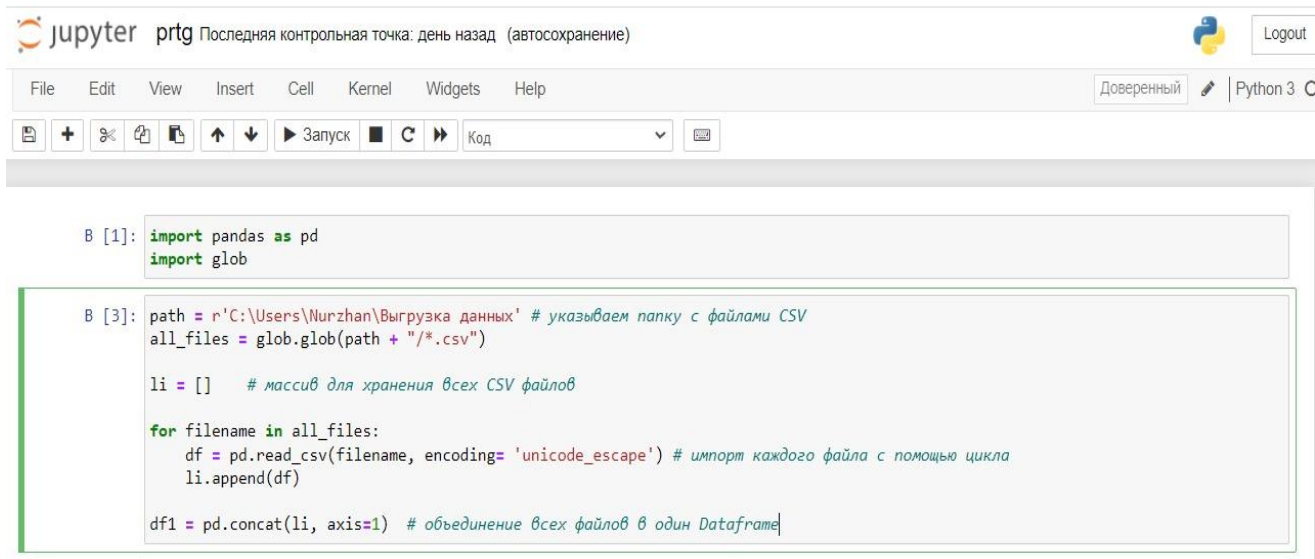
Рисунок 3 - Выгрузка данных от 8 сенсоров одного оборудования

Результаты.

Есть несколько сред разработки для обработки и управления большими данными. Например: Oracle database, MySQL, Python. Во всех вышеприведенных примерах оптимизирован процесс запросов к данным. Однако среда разработки, основанная на языке программирования Python, дает нам возможность пользоваться широким набором инструментов статистических функций. Следует подчеркнуть, что дальнейшие разработки ПО с пользовательскими интерфейсами осуществляется более корректно на объектно-ориентированном программном языке Python.

Решение вышеописанных задач и проблем авторы осуществили с помощью применения встроенной библиотеки языка программирования Python – Pandas, и интерпретатор Jupiter:

Авторы написали скрипт для извлечения из указанной папки несколько CSV dataframe-ов (таблиц) для последующего объединения в один dataframe (таблицу), как продемонстрировано на рисунке 4.



```
In [1]: import pandas as pd
import glob

In [3]: path = r'C:\Users\Nurzhan\Выгрузка данных' # указываем папку с файлами CSV
all_files = glob.glob(path + "/*.csv")

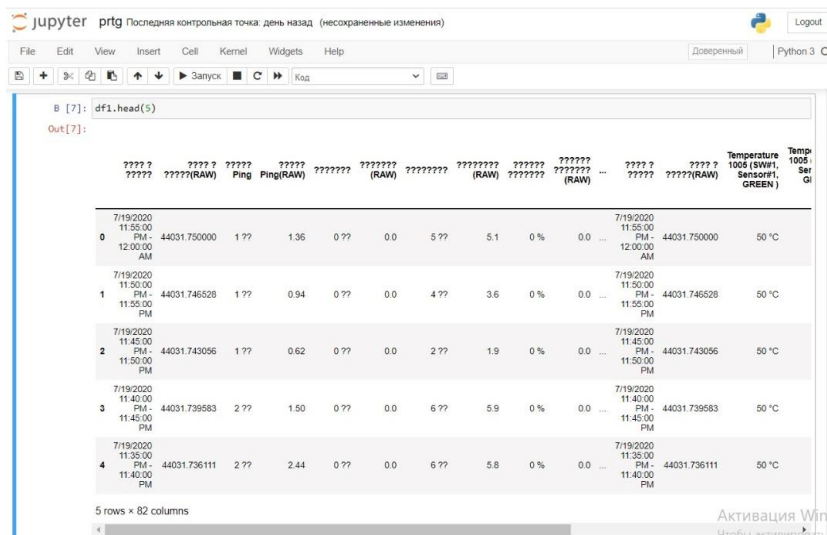
li = [] # массив для хранения всех CSV файлов

for filename in all_files:
    df = pd.read_csv(filename, encoding='unicode_escape') # импорт каждого файла с помощью цикла
    li.append(df)

df1 = pd.concat(li, axis=1) # объединение всех файлов в один Dataframe
```

Рисунок 4 - Код объединения нескольких CSV файлов

После объединения нескольких CSV таблиц можно увидеть, что в итоговом dataframe (таблице) присутствуют дублирующиеся столбцы, которые не несут уникальную информационную нагрузку, как продемонстрировано на рисунке 5.



```
In [7]: df1.head(5)
```

Out[7]:

	???? ?	???? ?	???? ?	???? ?	???????	???????	???????	???????	???????	???????	???????	???????	Temperature 100s (SWIR, Sensor1, GREEN)	Temp 100S Ser GI
0	7/19/2020 11:55:00 PM - 44031.750000	1 ??	1.36	0 ??	0.0	5 ??	5.1	0 %	0.0	...	7/19/2020 11:55:00 PM - 44031.750000	50 °C		
1	7/19/2020 11:50:00 PM - 44031.746528	1 ??	0.94	0 ??	0.0	4 ??	3.6	0 %	0.0	...	7/19/2020 11:50:00 PM - 44031.746528	50 °C		
2	7/19/2020 11:45:00 PM - 44031.743056	1 ??	0.62	0 ??	0.0	2 ??	1.9	0 %	0.0	...	7/19/2020 11:45:00 PM - 44031.743056	50 °C		
3	7/19/2020 11:40:00 PM - 44031.739583	2 ??	1.50	0 ??	0.0	6 ??	5.9	0 %	0.0	...	7/19/2020 11:40:00 PM - 44031.739583	50 °C		
4	7/19/2020 11:35:00 PM - 44031.736111	2 ??	2.44	0 ??	0.0	6 ??	5.8	0 %	0.0	...	7/19/2020 11:35:00 PM - 44031.736111	50 °C		

5 rows x 82 columns

Рисунок 5 - Объединение файлов в один dataframe

Следующим шагом мы очистили дублирующиеся столбцы для дальнейшего анализа и убрали столбцы с сырыми значениями, у которых есть приписка RAW. По итогу, из 82 столбцов у нас остались 20 уникальных. Пример итоговой очистки вы можете увидеть на рисунке 6.

После объединения и очистки dataframe-а от лишних дублирующихся столбцов и сырых значений с припиской RAW следует небольшая лингвистическая корректировка

названий столбцов. После этого работа по подготовке данных для дальнейшего её применения в моделях машинного обучения завершается.

Обсуждение и заключение.

Очевидным результатом этого исследования является автоматизация процесса по сбору и подготовке данных в необходимый вид, который может применяться для дальнейшего процесса машинного прогнозирования, который указывался на рисунке 2.

Авторы с помощью скриптов автоматизировали работу по сбору и объединению CSV файлов в один dataframe. Помимо этого, авторы оптимизировали скрипт функцией очистки объединённого dataframe-а от дублирующихся столбцов и столбцов с сырыми данными. Итоговый продукт, скрипт-программа, может многократно ускорить процесс сбора данных у ПО PRTG Network Monitor, что является неотъемлемой частью обработки, анализа и работы с Big Data.

```
In [7]: for name in df1.columns:
        if 'RAM' in name:
            df1.head(1) = df1.drop(columns = name)

In [10]: df1

Out[10]:
```

	7/19/2020 11:50:00 PM	0%	251 ??	1 ??	0%	0 ??	5 ??	100%	55 ?????	1 ?????	-4.57 ?????	11	10%	100%	Fan 1004 State (Switched, Fan1) (Normal)	F Me
0	7/19/2020 11:50:00 PM	0%	251 ??	1 ??	0%	0 ??	4 ??	100%	55 ?????	1 ?????	-4.57 ?????	11	11%	100%		
1	7/19/2020 11:50:00 PM	0%	251 ??	1 ??	0%	0 ??	2 ??	100%	55 ?????	1 ?????	-4.58 ?????	10	10%	100%		
2	7/19/2020 11:45:00 PM	0%	251 ??	1 ??	0%	0 ??	6 ??	100%	55 ?????	1 ?????	-4.58 ?????	10	10%	100%		
3	7/19/2020 11:40:00 PM	0%	251 ??	2 ??	0%	0 ??	8 ??	100%	55 ?????	1 ?????	-4.58 ?????	10	10%	100%		
4	7/19/2020 11:35:00 PM	0%	251 ??	2 ??	0%	0 ??	8 ??	100%	55 ?????	1 ?????	-4.58 ?????	10	10%	100%		
284	7/19/2020 12:15:00 AM	0%	250 ??	1 ??	0%	0 ??	5 ??	100%	55 ?????	1 ?????	-4.58 ?????	10	10%	100%		
285	7/19/2020 12:10:00 AM	0%	250 ??	1 ??	0%	0 ??	3 ??	100%	55 ?????	1 ?????	-4.58 ?????	10	10%	100%		
286	7/19/2020 12:05:00 AM	0%	250 ??	2 ??	0%	0 ??	6 ??	100%	55 ?????	1 ?????	-4.58 ?????	10	10%	100%		
287	7/19/2020 12:00:00 AM	0%	250 ??	1 ??	0%	0 ??	5 ??	100%	55 ?????	1 ?????	-4.58 ?????	11	10%	100%		
288	7/19/2020 12:00:00 AM	NaN	251 ??	2 ??	0%	0 ??	8 ??	100%	55 ?????	1 ?????	-4.57 ?????	10	10%	100%		

288 rows x 20 columns

Рисунок 6 - Предитоговый вид dataframe-а

Следует отметить, что авторы статьи брали в пример статистику работы одного сетевого оборудования за 24 часа, с интервалом каждые 5 минут. Это означает, что система мониторинга PRTG Network Monitor составляла отчет с информацией с данными от сенсоров раз в сутки одного из множеств своих сетевых коммутаторов. Авторы подчеркивают, что скрипт-программа, разработанный авторами, позволяет работать с неограниченными количествами исходных таблиц и объединить их в одну dataframe (таблицу). Иными словами, скрипт имеет возможность работать не только с данными одного оборудования, а со множеством исходных данных множеств систем которое дает ПО PRTG Network Monitor, и объединить их в одну dataframe (таблицу), настройки которого могут прописать авторы.

Важно отметить что на рынке существуют десятки готовых решений, которые позволяют обрабатывать сырые данные, получаемые из разных систем. Они отличаются различными областями применения, функционалом, наборами инструментов, гибкостью настроек выходных данных и т.д. Но большинство решений не дадут идентичных ожидаемых результатов. К тому же, готовые решения варьируются по итоговой цене

после внедрения в систему мониторинга, что являлось для авторов важной экономической составляющей при разработке системы машинного прогнозирования. Авторы статьи решили применить знания в области программирования для решения поставленных задач и написали скрипт-программу, отвечающую всем их требованиям.

ЛИТЕРАТУРА

[1] С.Ю. Исхаков, А.А. Шелупанов, С.В. Тимченко «Прогнозирование в системе мониторинга локальных сетей» // Доклады ТУСУРа, № 1 (25), часть 2, июнь 2012// С. 100-103. – статья.

[2] Н.И. Шаханов, И.А. Варфоломеев, Е.В. Ершов, О.В. Юдина Прогнозирование отказов оборудования в условиях малого количества поломок. Вестник Череповецкого государственного университета 2016 №6 С.36-41 – статья.

[3] Myrzatay Ali, Rzayeva L.G., Uskenbayeva G.A., Shukirova A.K., Abitova G. “The effect of the amount of data array on the results of forecasting network equipment failures” News of the National academy of sciences of the Republic of Kazakhstan, Physico-mathematical series, Volume 6, Number 340 (2021), 28-36. – статья на английском.

[4] Виктор М. Прогнозная аналитика для эффективного использования оборудования. 2016. URL:

https://filearchive.cnews.ru/files/reviews/2016_03_29/2_Maltsev.pdf - электронный источник.

[5] Oliver M. Predictive maintenance & Service (PdMS) - Outline and Value Proposition. 2014. URL: <https://blogs.saphana.com/wp-content/uploads/2014/11/Predictive-Maintenance-Service-Outline-Value-Proposition.pdf> - электронный источник на английском.

[6] 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.

Али Мырзатай, Қорқыт ата атындағы Қызылорда университеті, оқытушы, Қорқыт ата атындағы Қызылорда университеті, Қызылорда, Қазақстан, mirzataitegiali@gmail.com

Лейла Рзаева, PhD, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Нұр - Сұлтан, Қазақстан, leilarza2@gmail.com

Нұржан Құрманай, магистрант, Қорқыт ата атындағы Қызылорда университеті, Қызылорда, Қазақстан, iamnur739@gmail.com

Мурат Жакенов, т.ғ.к., Гуманитарлық-техникалық академия Көкшетау, Қазақстан, Zhakenov - Murat. @mail.ru

Гульнара Абитова, PhD, Астана ІТ университеті, Нұр-Сұлтан, Қазақстан, abitova.gul@gmail.com

МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНЕ НЕГІЗДЕЛГЕН БОЛЖАУ ЖҮЙЕЛЕРІНДЕ ОДАН ӘРІ ҚОЛДАНУ ҮШІН LAN МОНИТОРИНГ ЖҮЙЕСІНЕН ДЕРЕКТЕРДІ ЖИНАУ МЕН ДАЙЫНДАУДЫ АВТОМАТТАНДЫРУ

Аңдатпа. Мақалада LAN PRTG Network Monitor мониторинг жүйесінен алынған деректерді жинау мен дайындаудың автоматтандырылған жүйесін әзірлеу бойынша жұмыс қарастырылады. Мақаланың мақсаты-проблемаларды анықтауға жүйелі көзқарасты анықтау, қойылған міндеттерге сәйкес бизнес-процестерді модельдеу және енгізілген шешімдердің функцияларын анықтау. Авторлар LAN Қолдау қызметтерінің жұмысын оңтайландыру үшін деректерді бақылау және сақтау жүйелерінің болуын,

сондай-ақ басқа жүйелерді, мысалы, машиналық болжау саласында одан әрі қолдану үшін деректерді жинау және дайындау жүйелерін одан әрі енгізу қажеттілігін атап өтеді. Жоғарыда келтірілген себептерге байланысты авторлар белгілеген талаптарды қанағаттандыратын скрипт жазу мүмкіндігін қарастырды. Атап айтқанда келесі талаптарды орындауға күш жұмсады: дайын бағдарламалық жасақтаманы сатып алудан гөрі скриптті жазу экономикалық тұрғыдан арзан болуы, скрипт PRTG Network Monitor-дан шығатын деректер қорымен жұмыс істеу мүмкіндігі, скрипттен шығатын деректерді машиналық болжау жүйесімен оқуға және талдауға жарамдылығы, көптеген кестелерді бір деректер кестесіне біріктіру және т. б.

Түйінді сөздер. Желілік жабдықтың істен шығуын болжау, жергілікті желі, мониторинг жүйесі, деректерді жинау және реттеу жүйесі, бизнес-процестерді модельдеу, скрипт, скриптті әзірлеу, деректерді жинауды автоматтандыру, автоматтандыру, деректерді біріктіру, dataframe.

Ali Myrzatay, teacher, Korkyt-ata Kyzylorda University Kyzylorda, Kazakhstan, mirzataitegiali@gmail.com

Leila Rzayeva, PhD, L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan, leilarza2@gmail.com

Nurzhan Kurmanai, Master's student, Kyzylorda University im. Korkyt Ata Kyzylorda, Kazakhstan, iamnur739@gmail.com

Murat Zhakenov, Candidate of Technical Sciences, Humanitarian and Technical Academy, Kokshetau, Kazakhstan, Zhakenov-Murat.@mail.ru

Gulnara Abitova, PhD, Astana IT University, Nur-Sultan, Kazakhstan, abitova.gul@gmail.com

AUTOMATION OF DATA COLLECTION AND PREPARATION FROM LAN MONITORING SYSTEM FOR FURTHER APPLICATION IN FORECASTING SYSTEMS BASED ON MACHINE LEARNING METHODS

Abstract. The article discusses the work on the development of an automated system for collecting and preparing data obtained from the software of the PRTG Network Monitor LAN monitoring system. The purpose of the article is to identify a systematic approach to identifying problems, to model business processes according to the tasks set and to determine the functions of the implemented solutions. The authors emphasize the need for monitoring and data storage systems to optimize the work of LAN support services, as well as for the further implementation of other systems, for example, data collection and preparation systems for further use in the field of machine forecasting. For the above reasons, the authors considered the possibility of writing a script that would satisfy the tasks set. Namely: economic cheapness in the development and not the purchase of ready-made software, the ability of a script to work with outgoing data ON PRTG Network Monitor, suitable for reading and analyzing outgoing data by a machine forecasting system, combining the maximum possible number of tables into a single data table, etc.

Keywords. Network equipment failure forecasts, local area network, monitoring systems, data collection and ordering systems, LAN, business process modeling, script, script development, data collection automation, automation, data pooling, data frame.
