Ежегодная международная научно-практическая конференция
«РусКрипто’2023»

Выявление аномалий и классификация компьютерных атак в сети передачи данных на основе фрактального анализа и машинного обучения

Крибель Александр Михайлович, Военная академия связи им. С.М.Буденного
Крибель Ксения Васильевна, Военная академия связи им. С.М.Буденного
Котенко Игорь Витальевич, д.т.н., профессор, СПб ФИЦ РАН

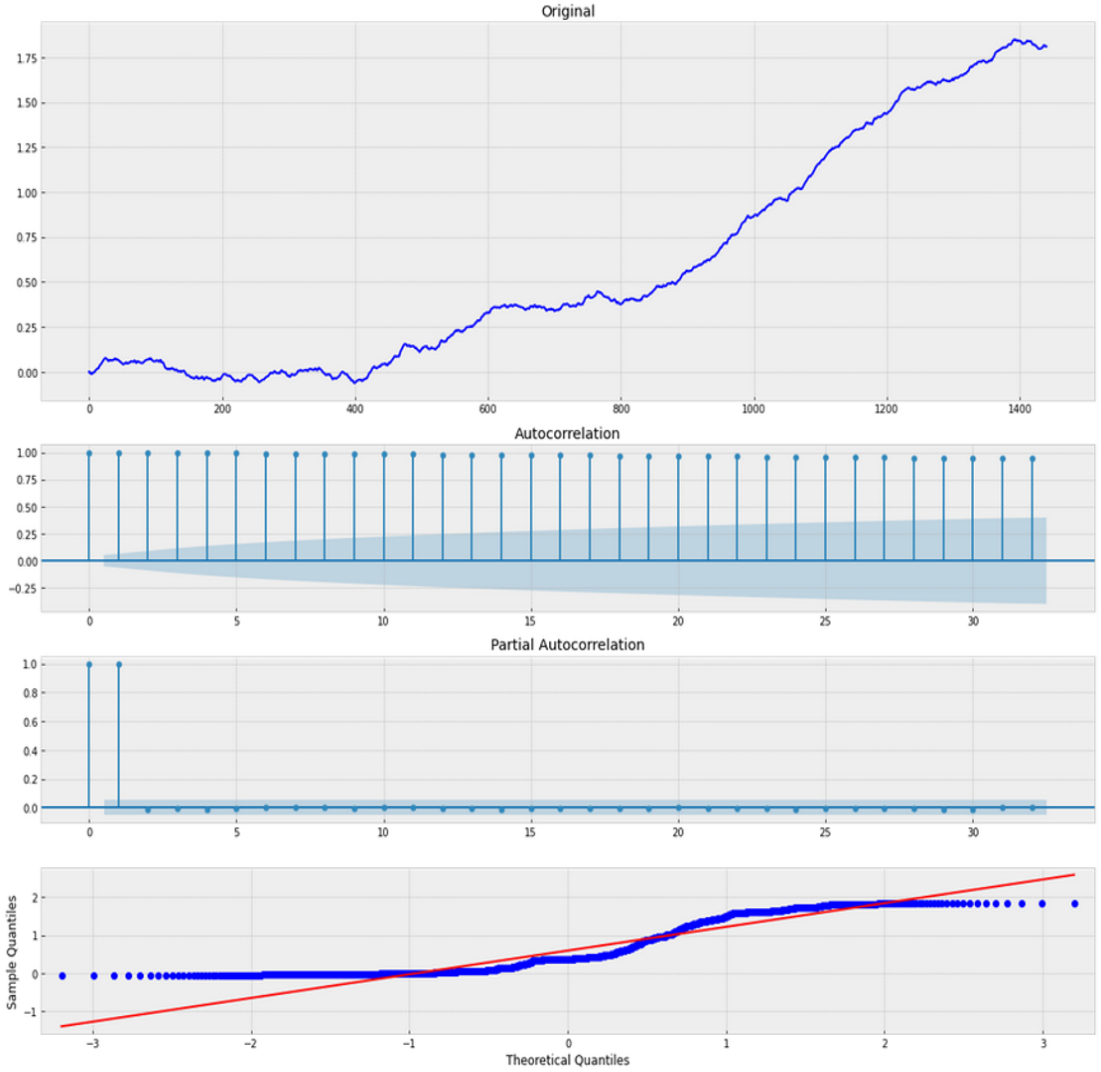


Рисунок 1 — Нестационарный временной ряд

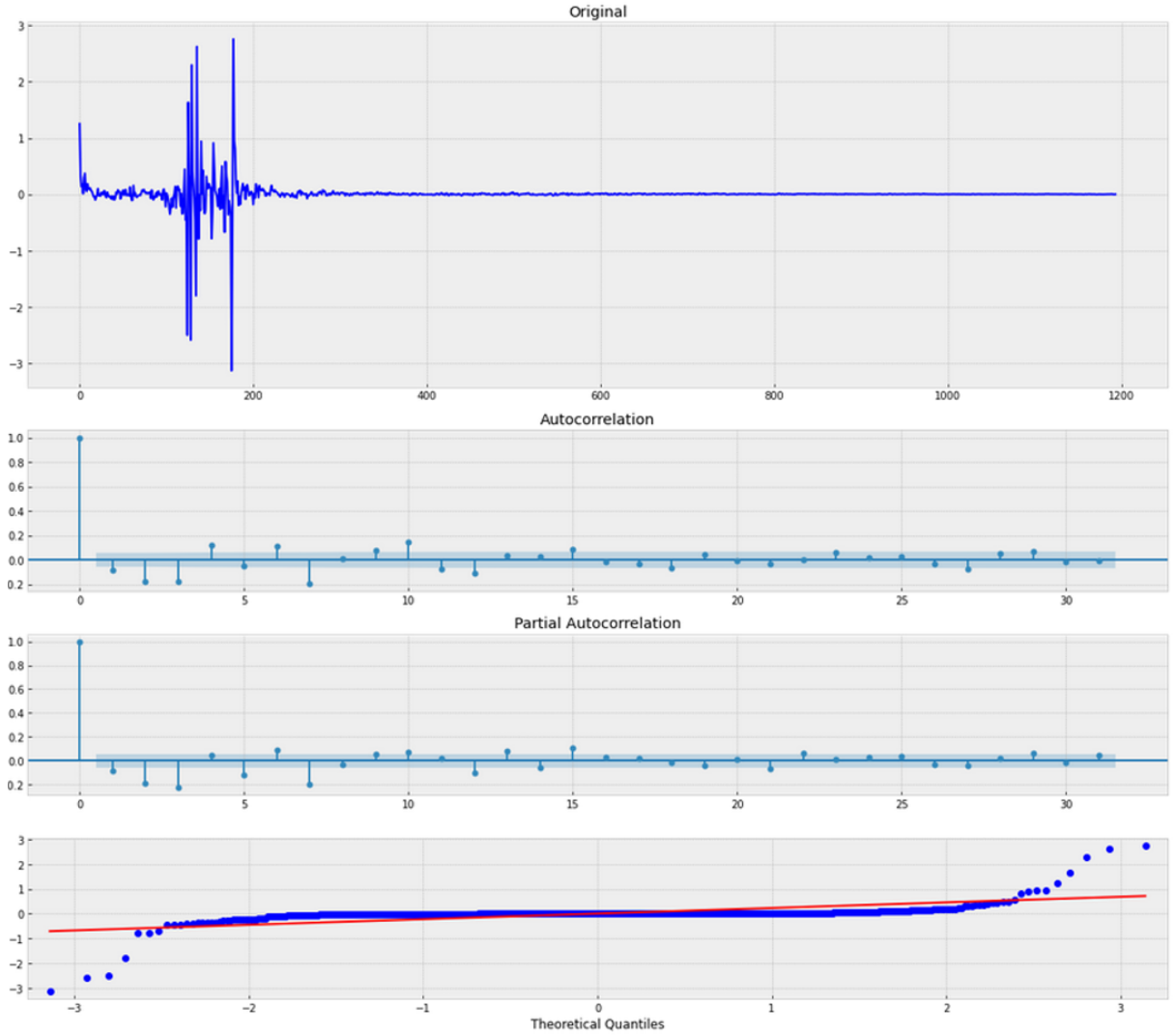
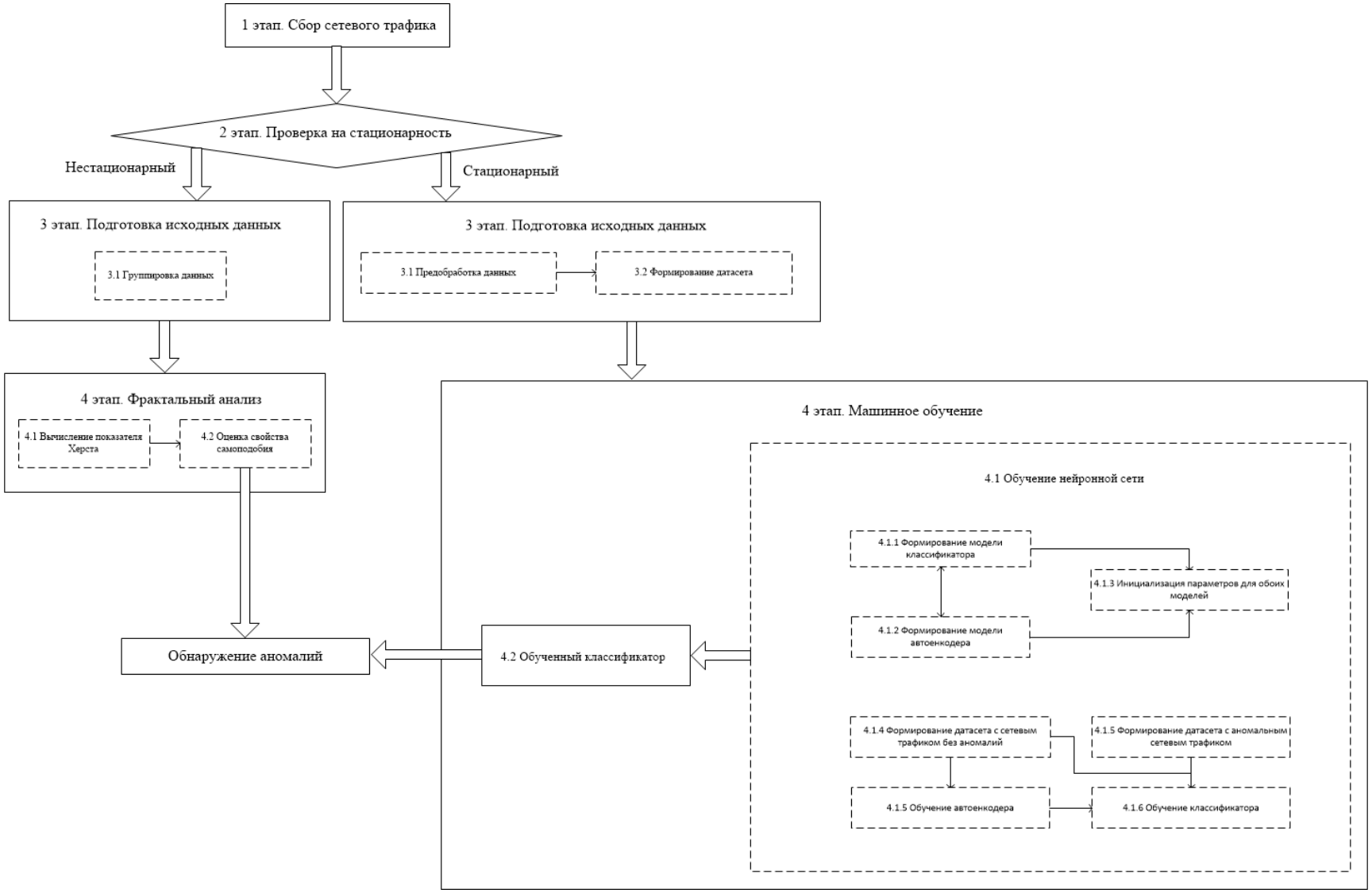


Рисунок 2 — Стационарный временной ряд

Блок-схема методики раннего обнаружения КА в сетевом трафике СПД



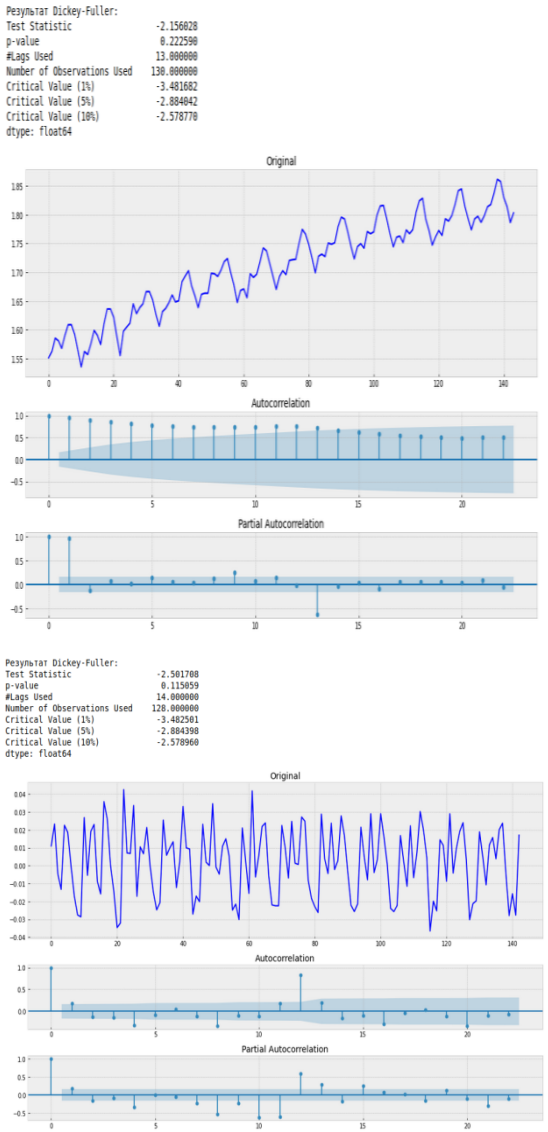


Рисунок 3 — Переход к стационарности

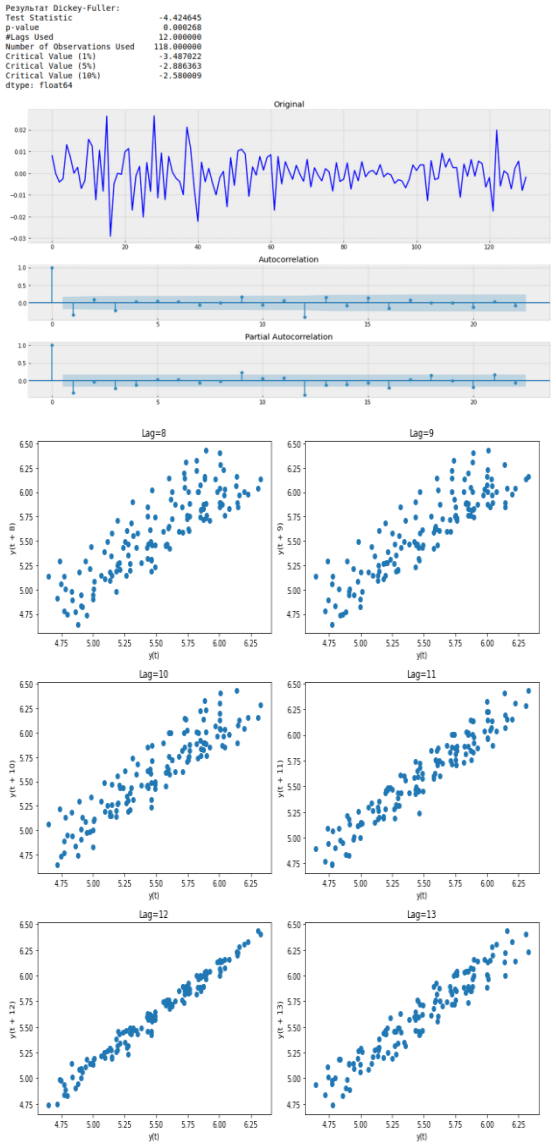


Рисунок 4 — Автокорреляция

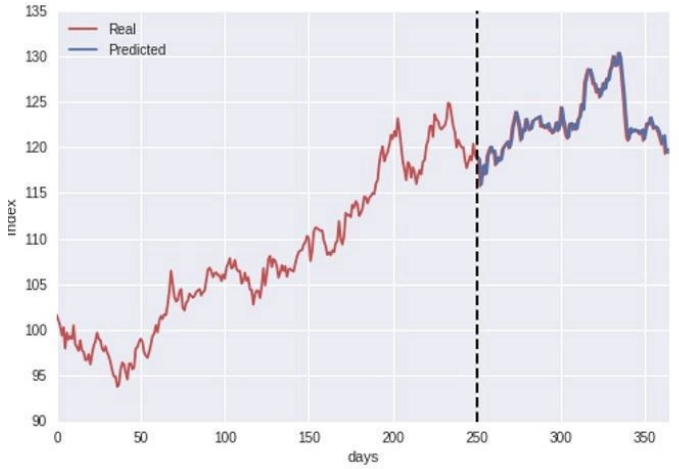


Рисунок 5 — Прогнозирование нестационарного ряда учитывая все временные компоненты

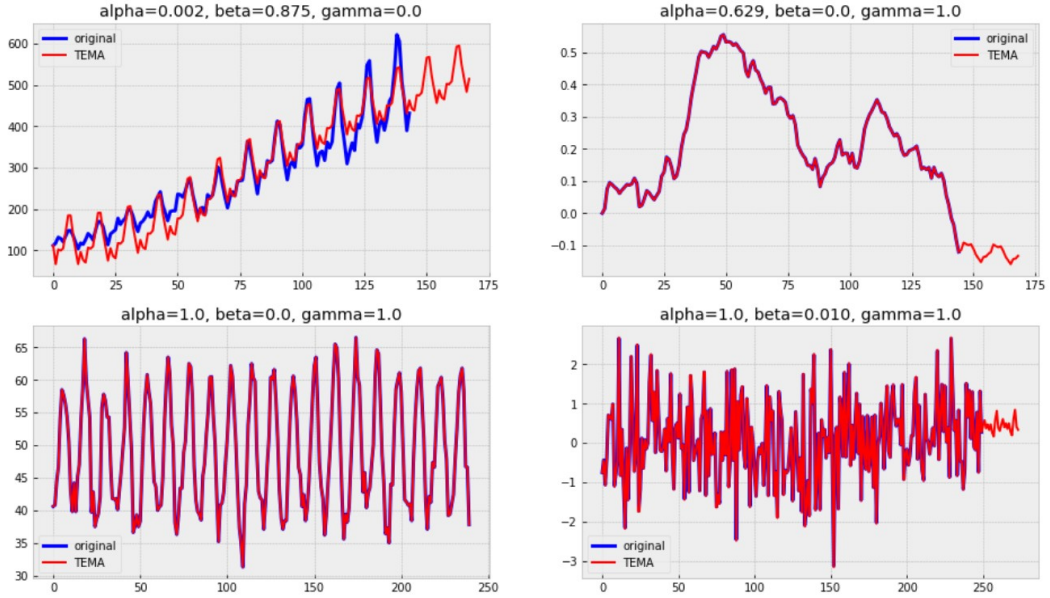


Рисунок 6 — Предсказание нестационарного ряда подбирая временные коэффициенты

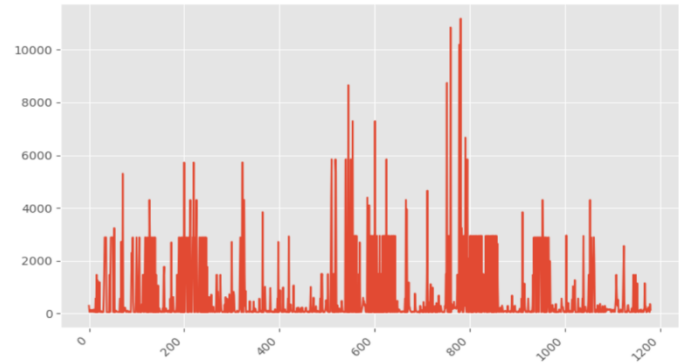


Рисунок 7 — Легитимный трафик

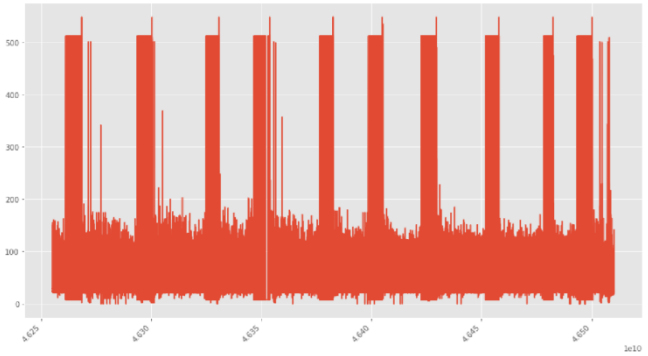


Рисунок 8 — Аномальный трафик

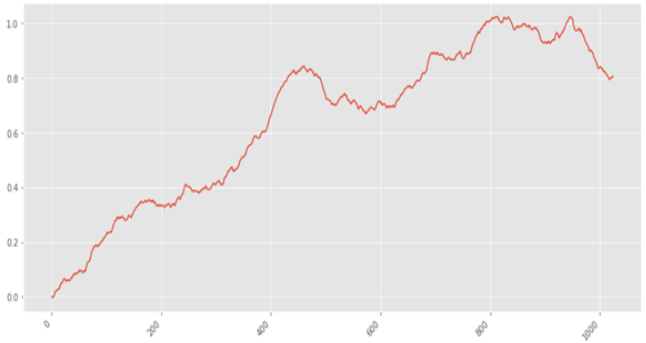


Рисунок 9 — Фрактальное броуновское движение

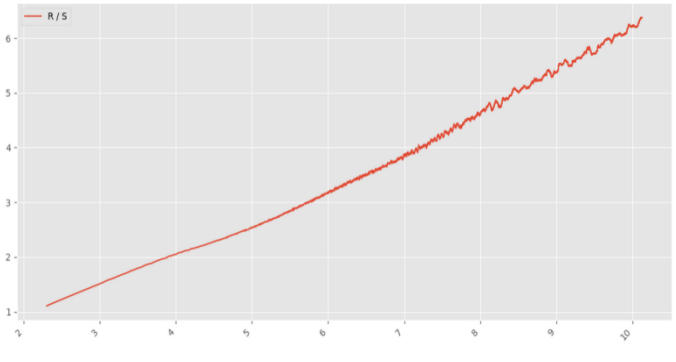


Рисунок 10— R/S анализ (H = 0.762)

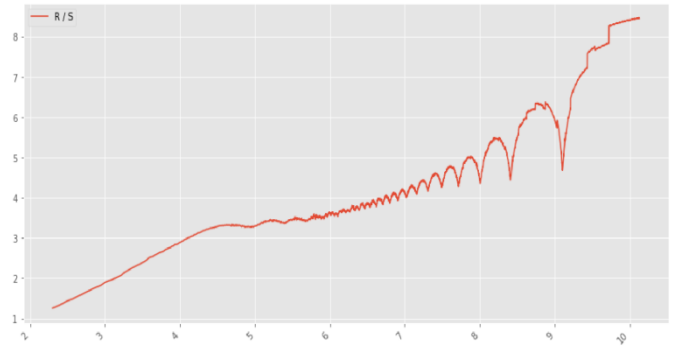


Рисунок 11 — R/S анализ (H = 1.378)

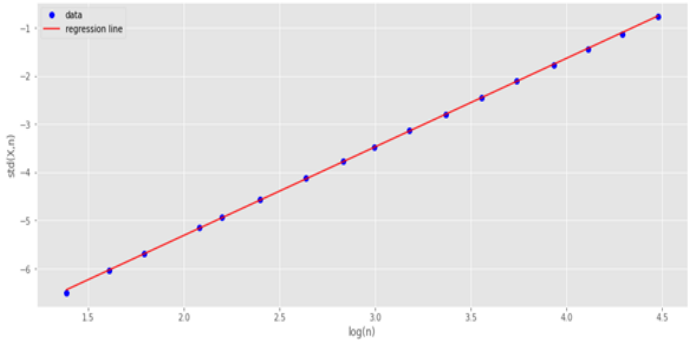


Рисунок 12— DFA анализ (H = 0.837)

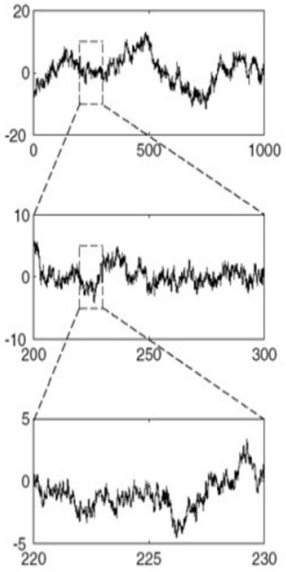


Рисунок 13 — Свойство самоподобия

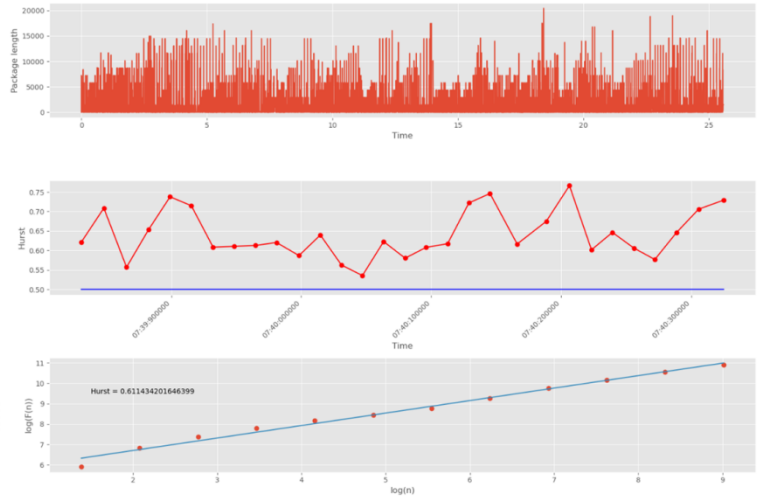


Рисунок 14 — Легитимный трафик

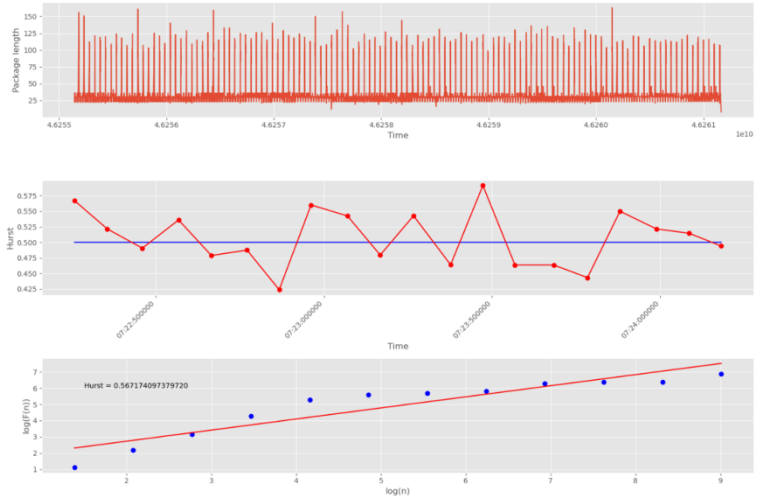


Рисунок 15 — Аномальный трафик

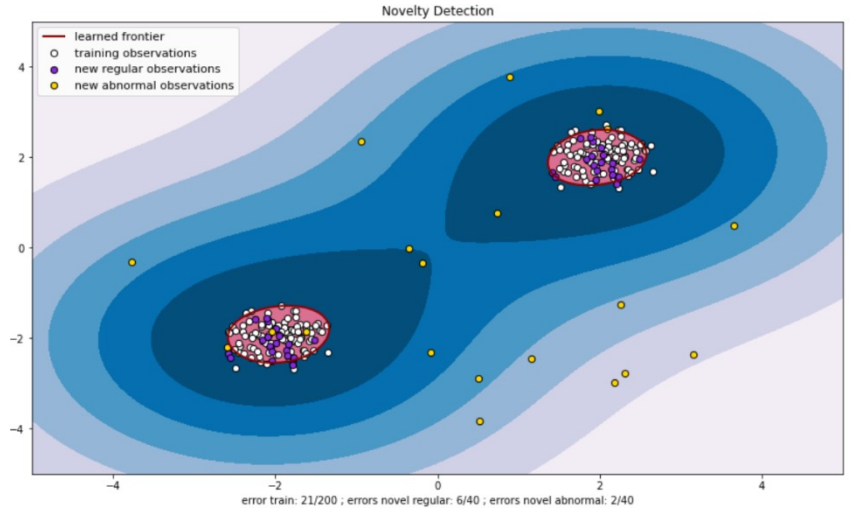


Рисунок 16 — SVM

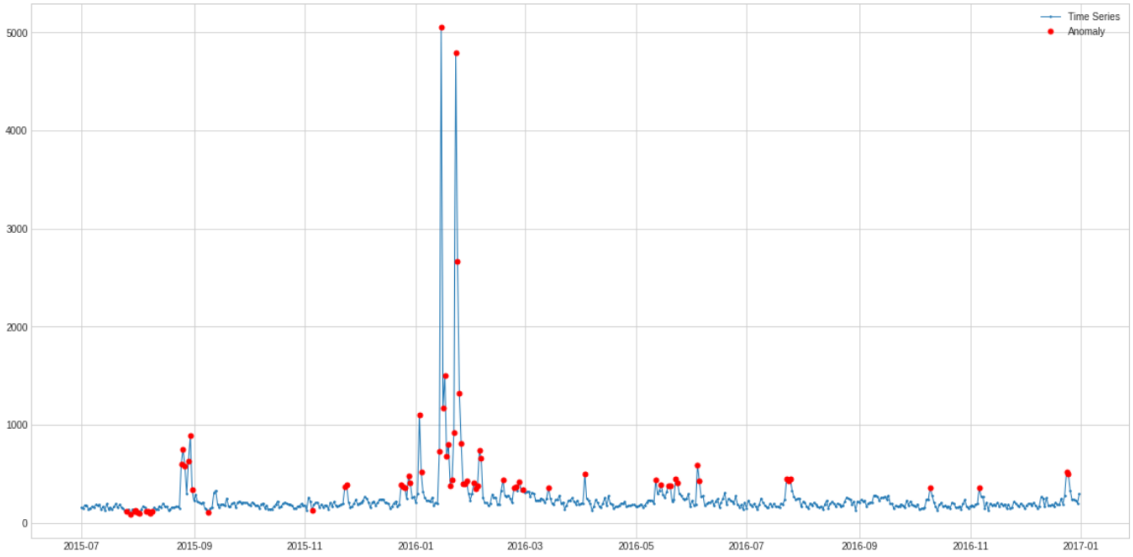


Рисунок 17 — Изолированный лес

```
ta, tai, taf, amp = detect_cusum(serias_2, 5, .05, True, True)
```

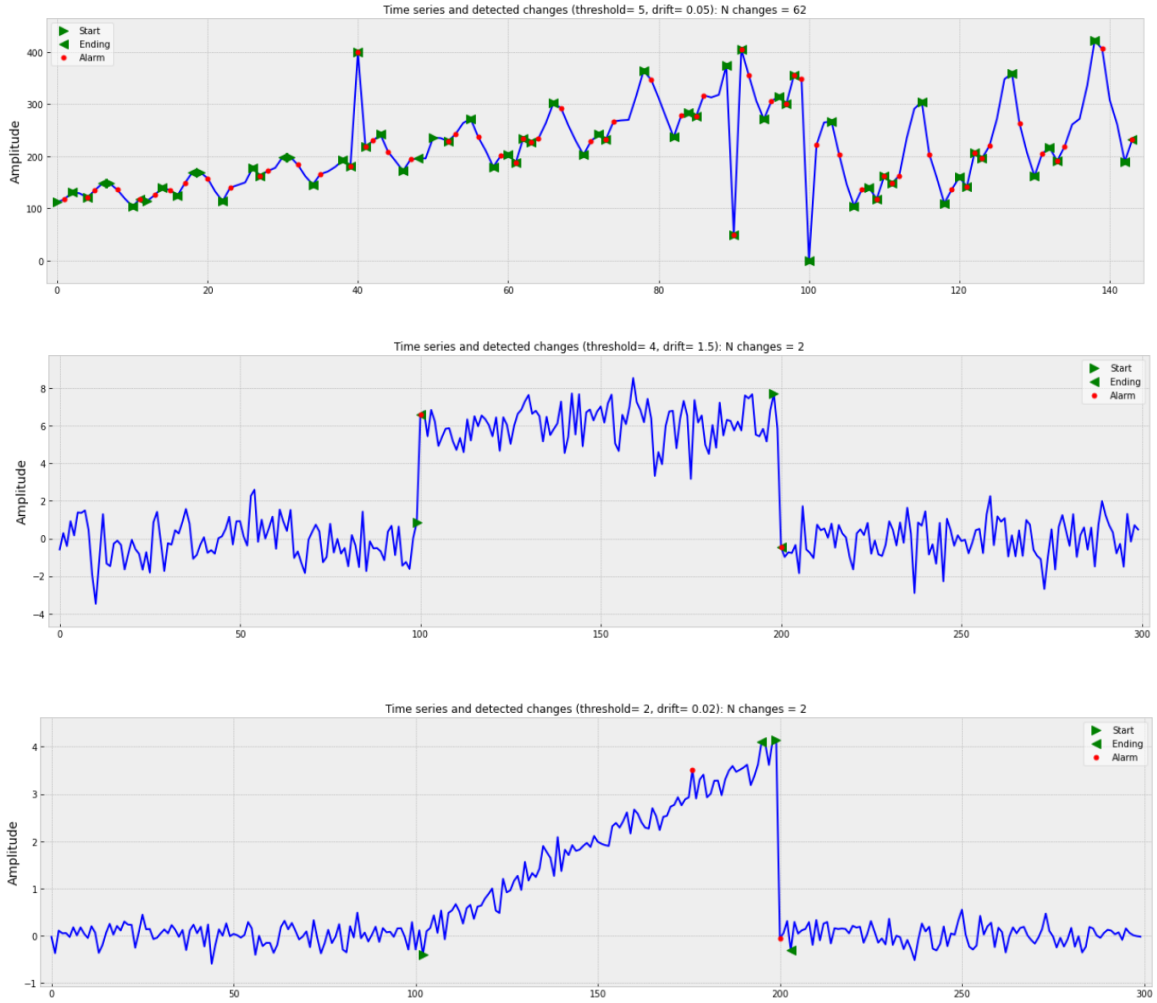
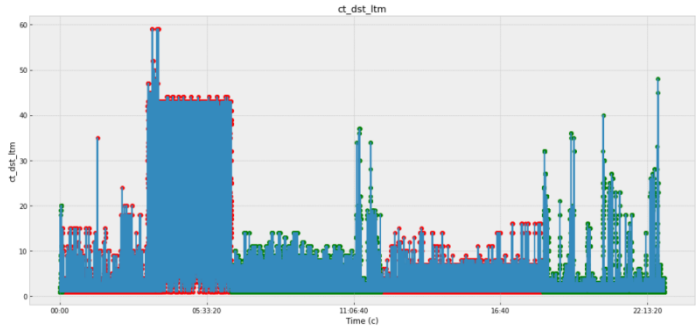
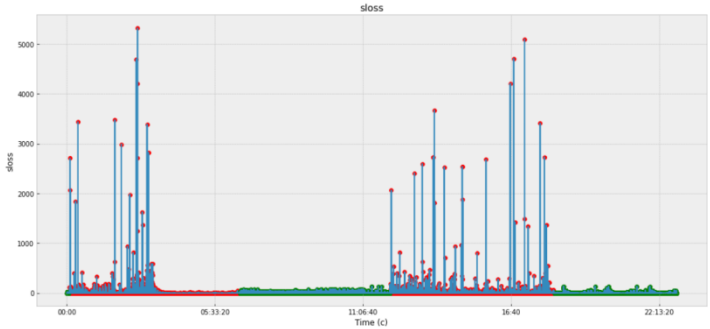


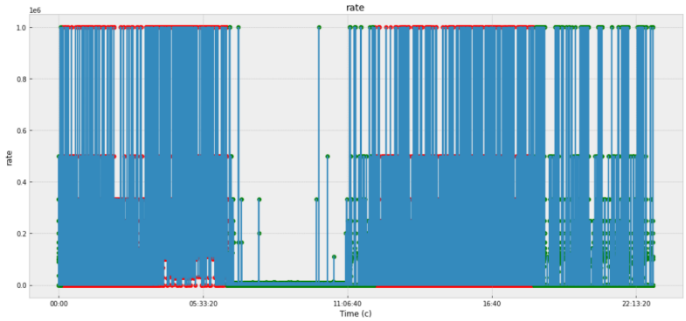
Рисунок 18 — Кумулятивнные суммы

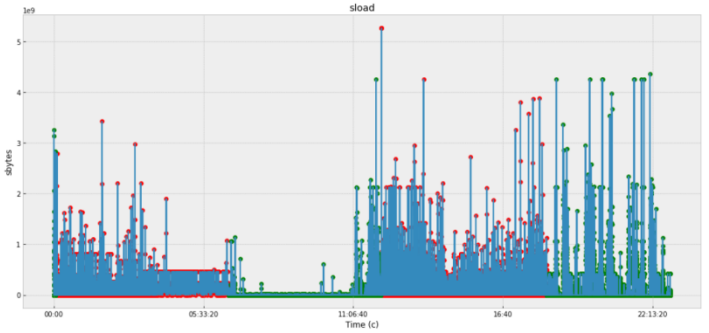
Name	Type	Description
srcip	nominal	Source IP address
sport	integer	Source port number
dstip	nominal	Destination IP address
dsport	integer	Destination port number
proto	nominal	Transaction protocol
state	nominal	Indicates to the state and its dependent proto...
dur	Float	Record total duration
sbytes	Integer	Source to destination transaction bytes
dbytes	Integer	Destination to source transaction bytes
sttl	Integer	Source to destination time to live value
dttl	Integer	Destination to source time to live value
sloss	Integer	Source packets retransmitted or dropped
dloss	Integer	Destination packets retransmitted or dropped
service	nominal	http, ftp, smtp, ssh, dns, ftp-data ,irc and ...
Sload	Float	Source bits per second
Dload	Float	Destination bits per second

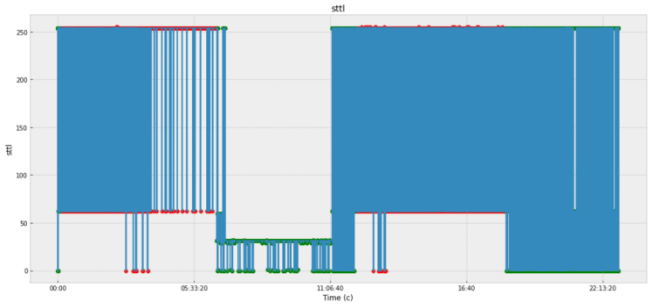
Рисунок 19 — Размеченные параметры

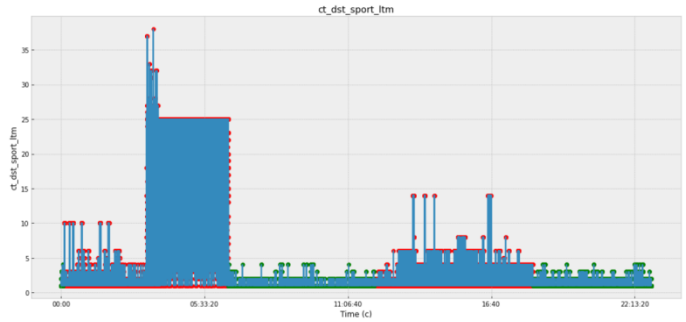


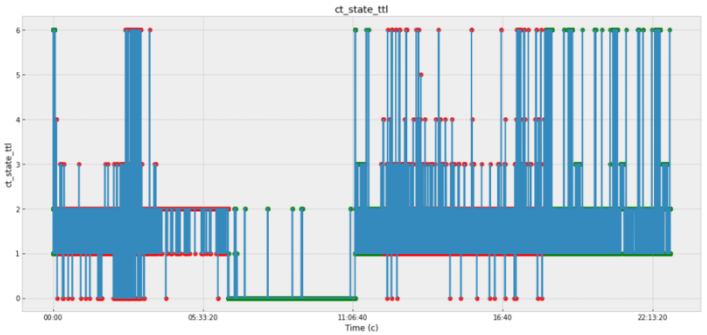












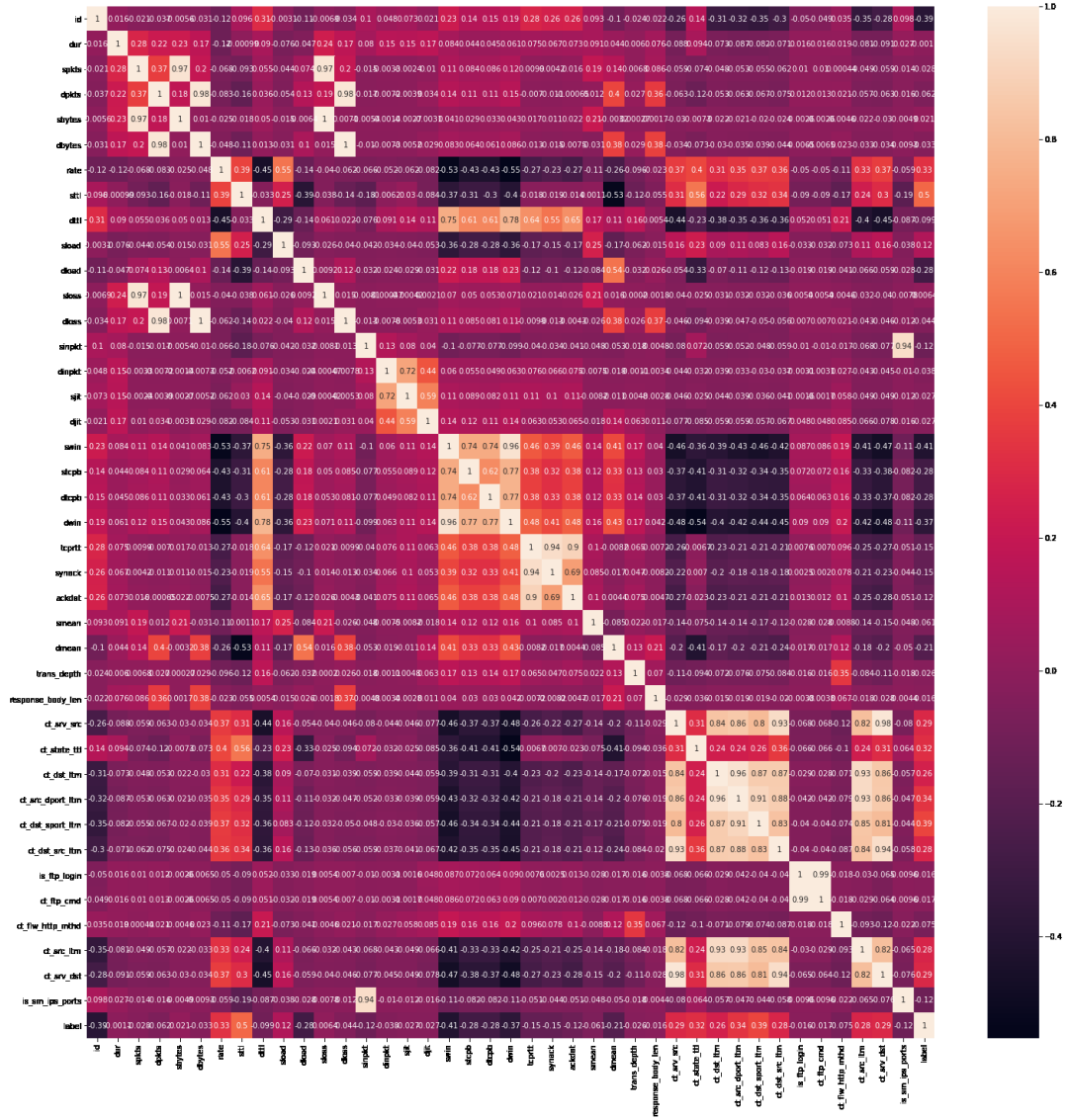


Рисунок 20 — Матрица корреляции

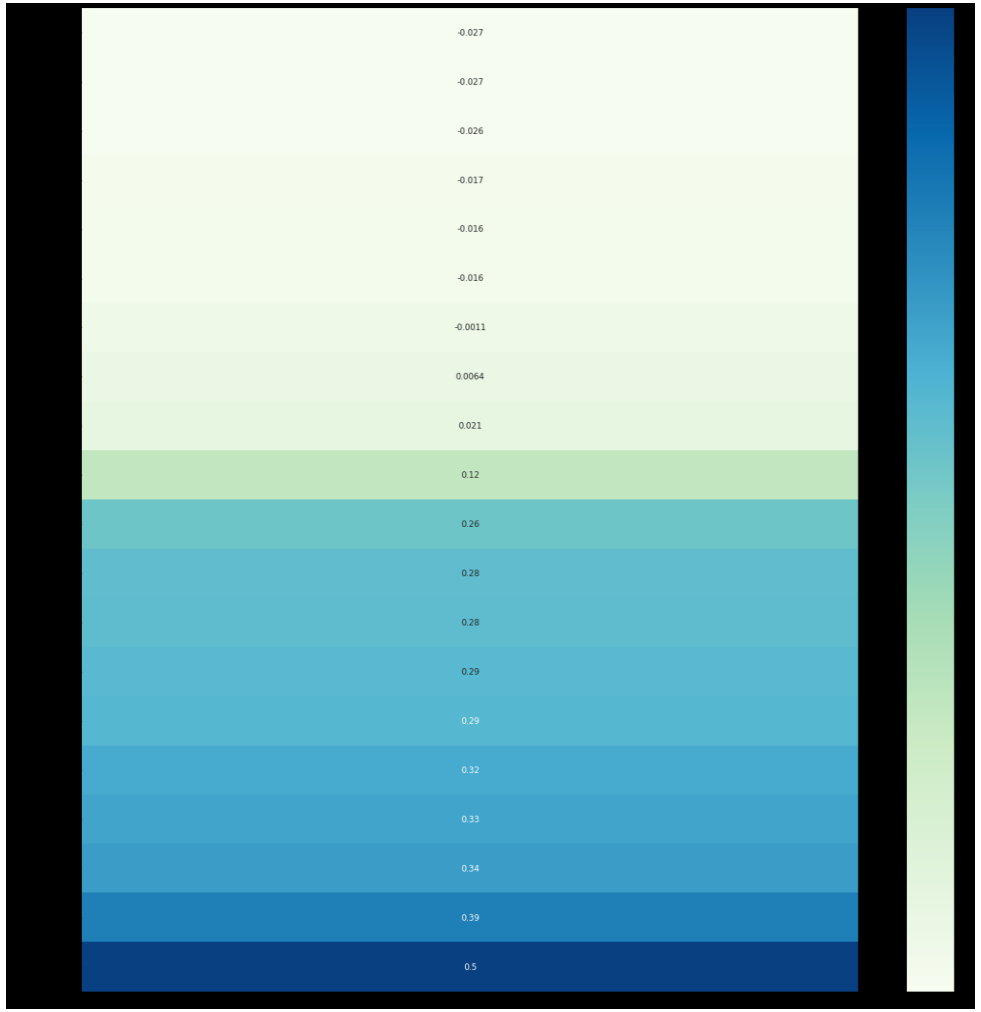


Рисунок 21 — Самые коррелируемые параметры

```
_____________________________________________________
Classificator: LogisticRegression(max_iter=1000, n_jobs=-1)
Best parametrs: {'penalty': 'l2', 'solver': 'lbfgs'}
Best score: 0.8961687849910088
_____________________________________________________
_____________________________________________________
Classificator: KNeighborsClassifier(n_jobs=-1, n_neighbors=1)
Best parametrs: {'n_neighbors': 1, 'weights': 'uniform'}
Best score: 0.999323288700103
_____________________________________________________
_____________________________________________________
Classificator: BaggingClassifier(n_estimators=15)
Best parametrs: {'n_estimators': 15}
Best score: 0.9995141569933788
_____________________________________________________
_____________________________________________________
Classificator: GradientBoostingClassifier(learning_rate=0.5, max_depth=6)
Best parametrs: {'learning_rate': 0.5, 'loss': 'deviance', 'max_depth': 6}
Best score: 0.9999305932110504
_____________________________________________________
_____________________________________________________
Classificator: RandomForestClassifier(max_depth=23, n_estimators=13)
Best parametrs: {'max_depth': 23, 'n_estimators': 13}
Best score: 0.9991324294388775
_____________________________________________________
_____________________________________________________
Classificator: AdaBoostClassifier(learning_rate=1.9000000000000001, n_estimators=51)
Best parametrs: {'learning_rate': 1.9000000000000001, 'n_estimators': 51}
Best score: 0.9999305932110504
```

Рисунок 22 — Самые коррелируемые параметры

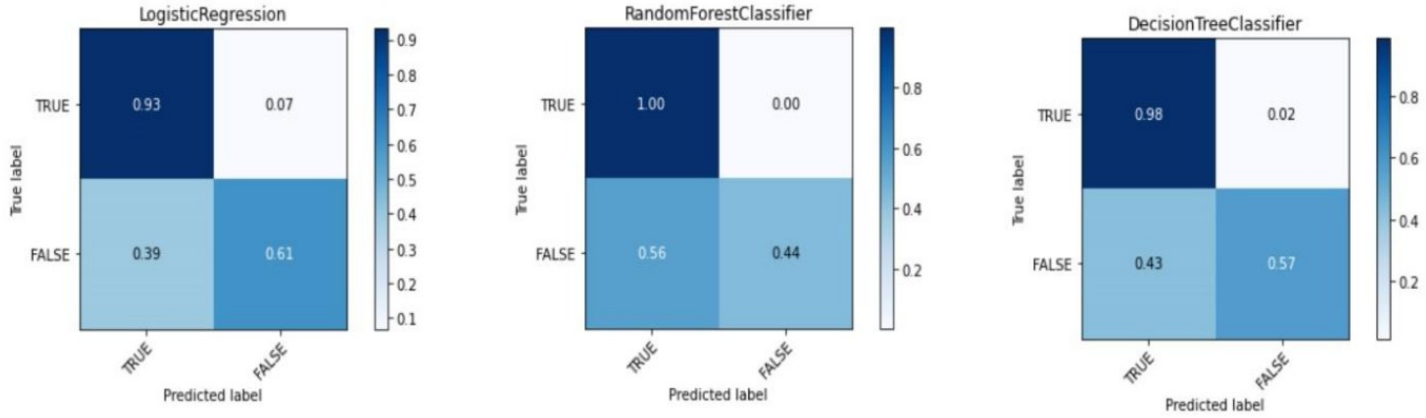


Рисунок 23 — Матрица ошибок первого и второго рода логистической регрессии, случайного леса и дерева решений

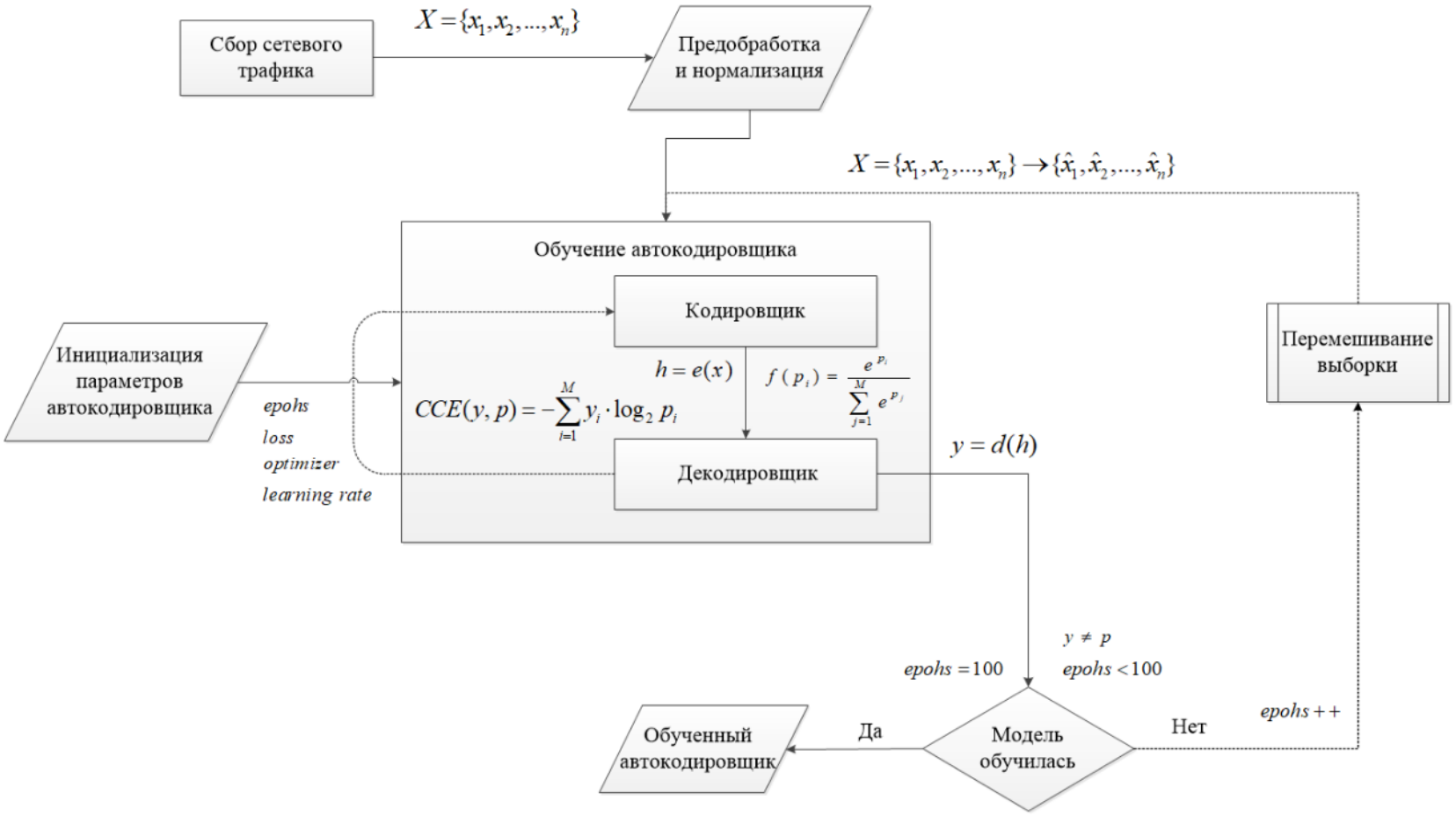


Рисунок 24 — Алгоритм автокодировщика

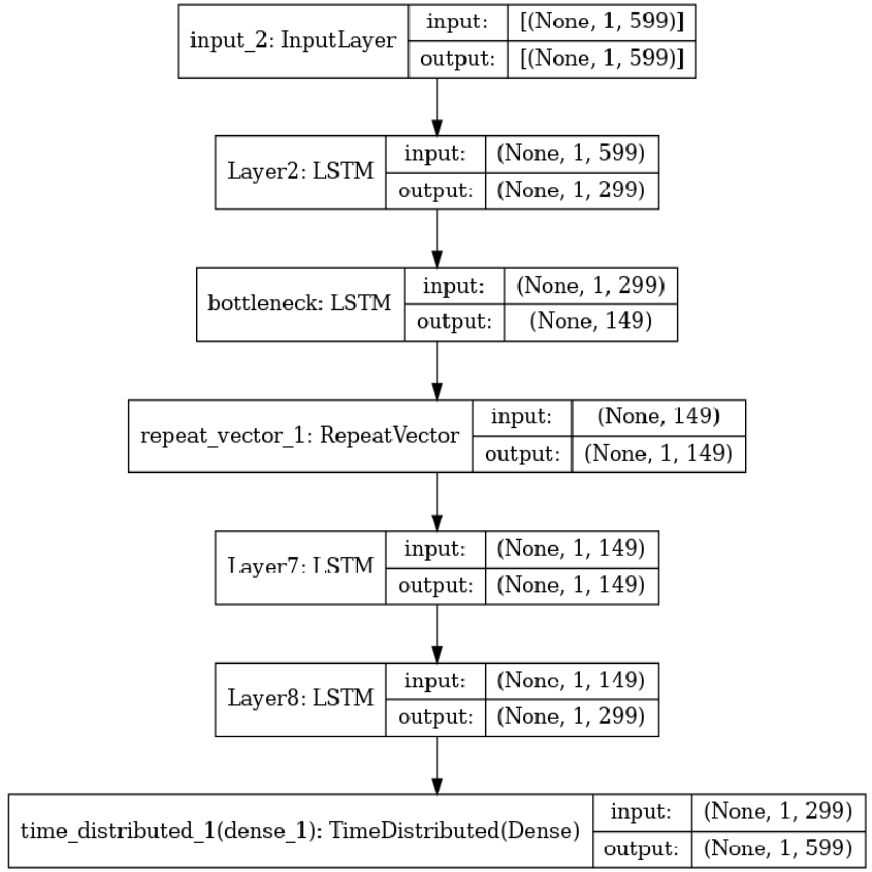


Рисунок 25 — Модель автокодировщика

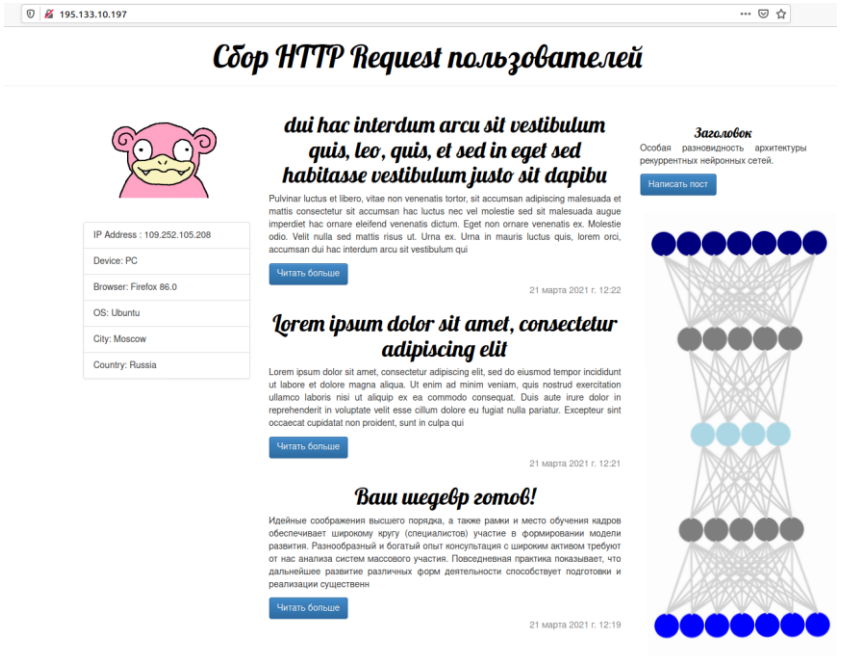


Рисунок 26 — Сбор http-трафика

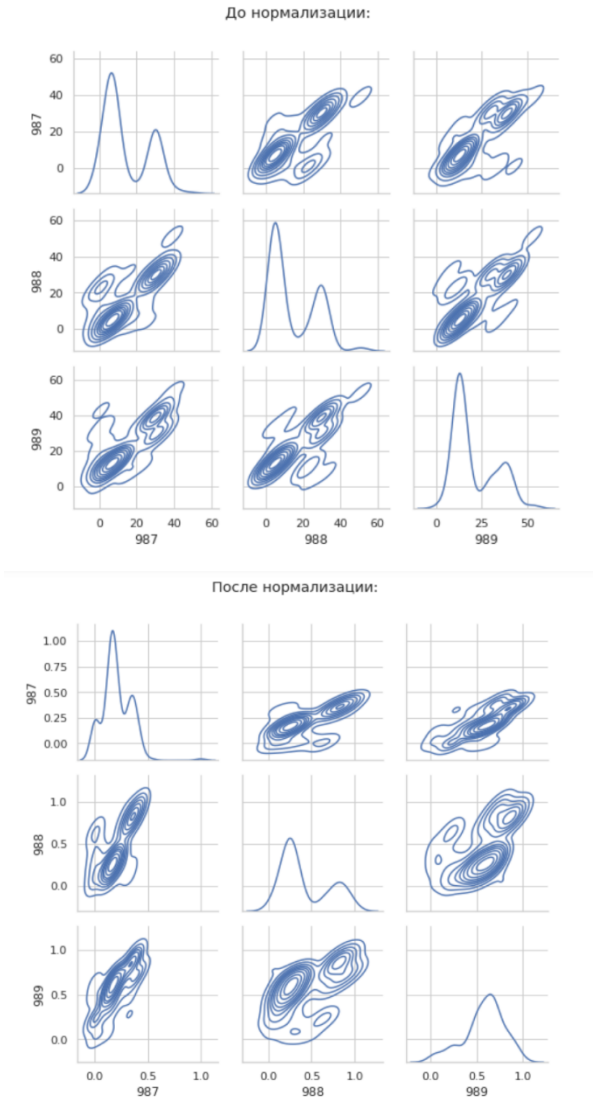


Рисунок 27 —Нормализация данных

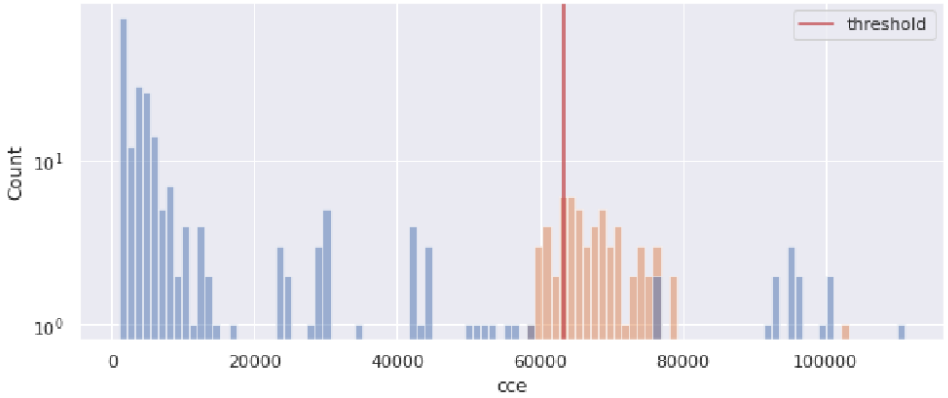


Рисунок 28 — Выбор порогового значения

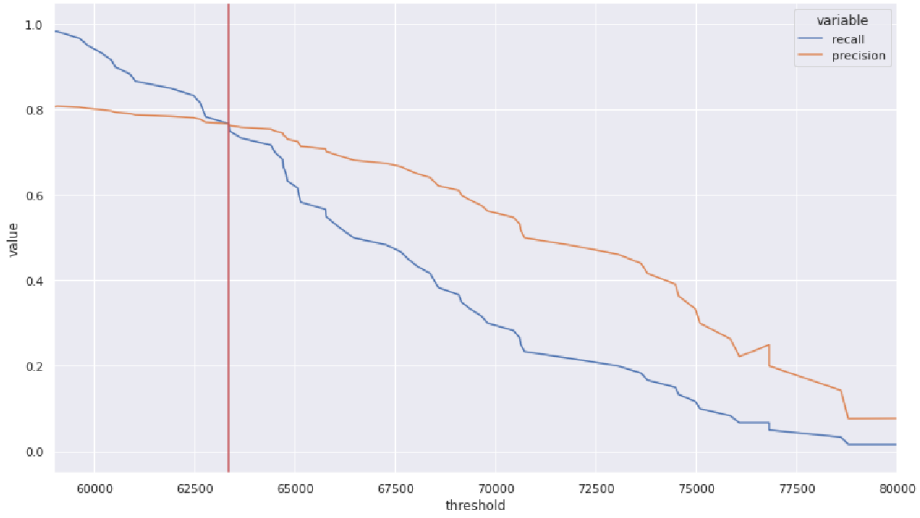


Рисунок 29 — Выбор порогового значения относительно recall и precision

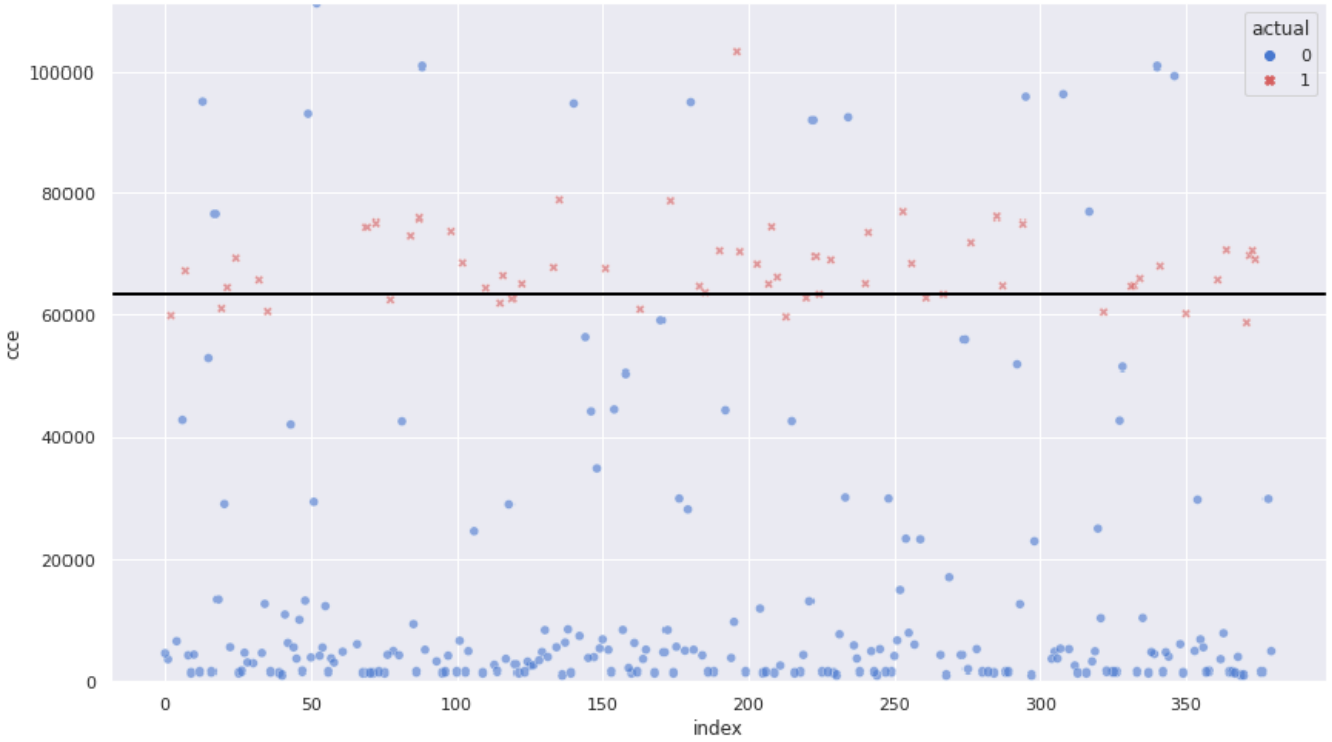


Рисунок 30 — Обнаружение аномалий

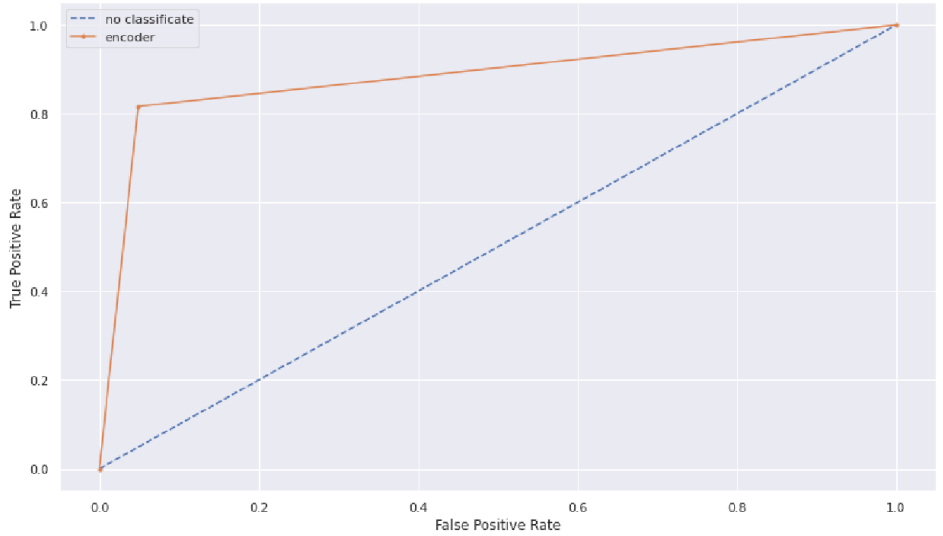


Рисунок 31 — ROC-кривая

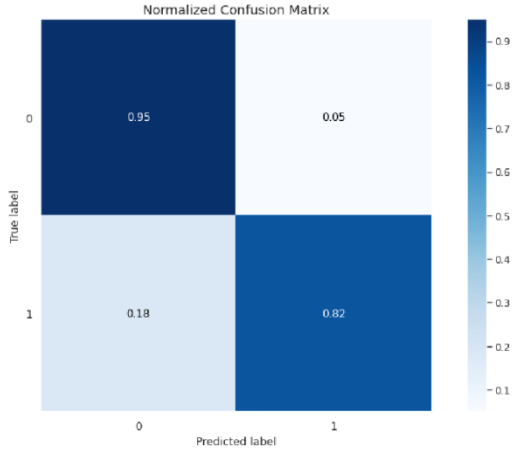


Рисунок 32 — Ошибки первого и второго рода

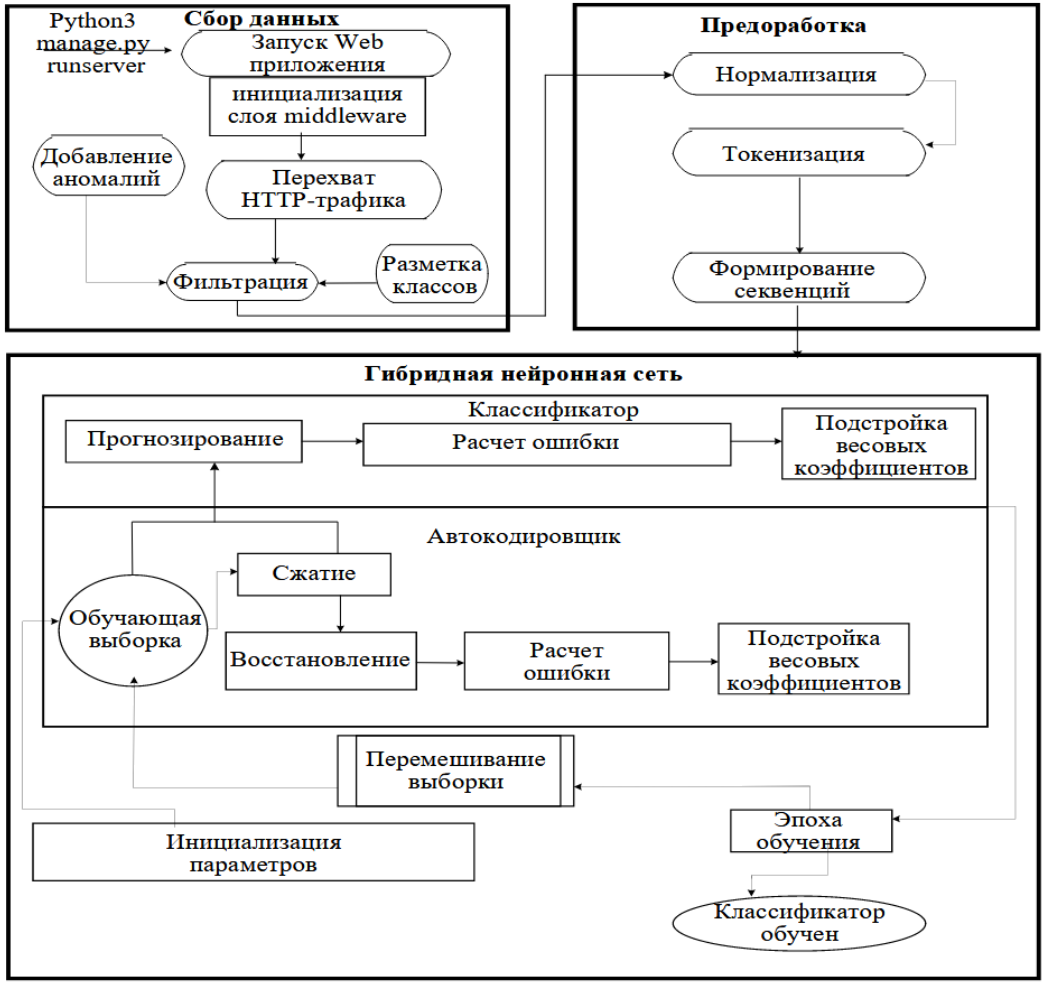


Рисунок 33 — Блок-схема гибридной нейронной сети

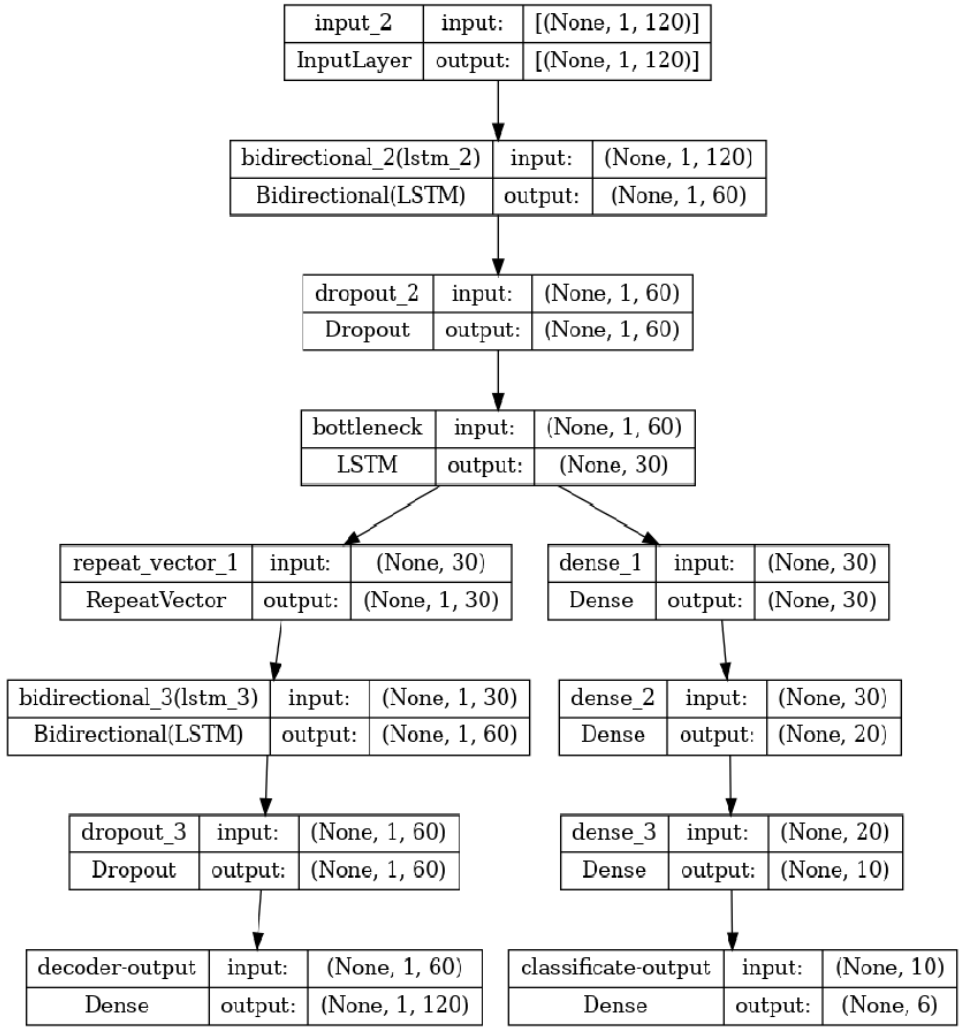


Рисунок 34 — Модель нейронной сети

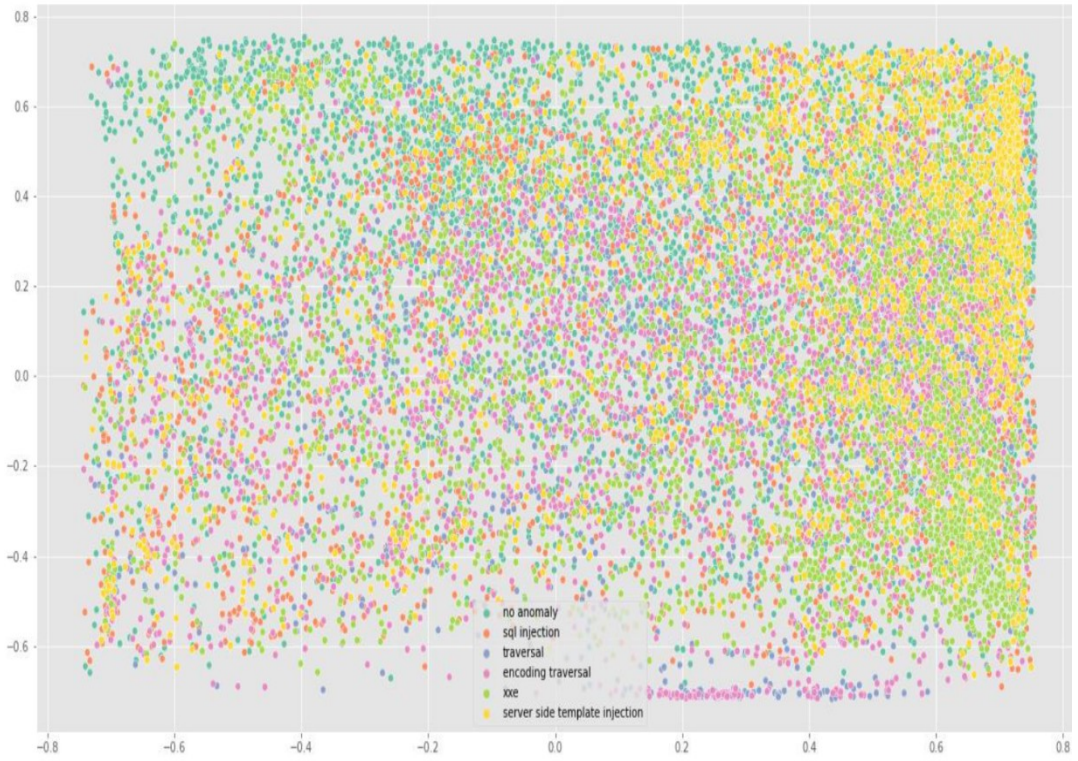


Рисунок 35 — Визуальное представление векторного отображения данных, подаваемых на вход гибридной нейронной сети

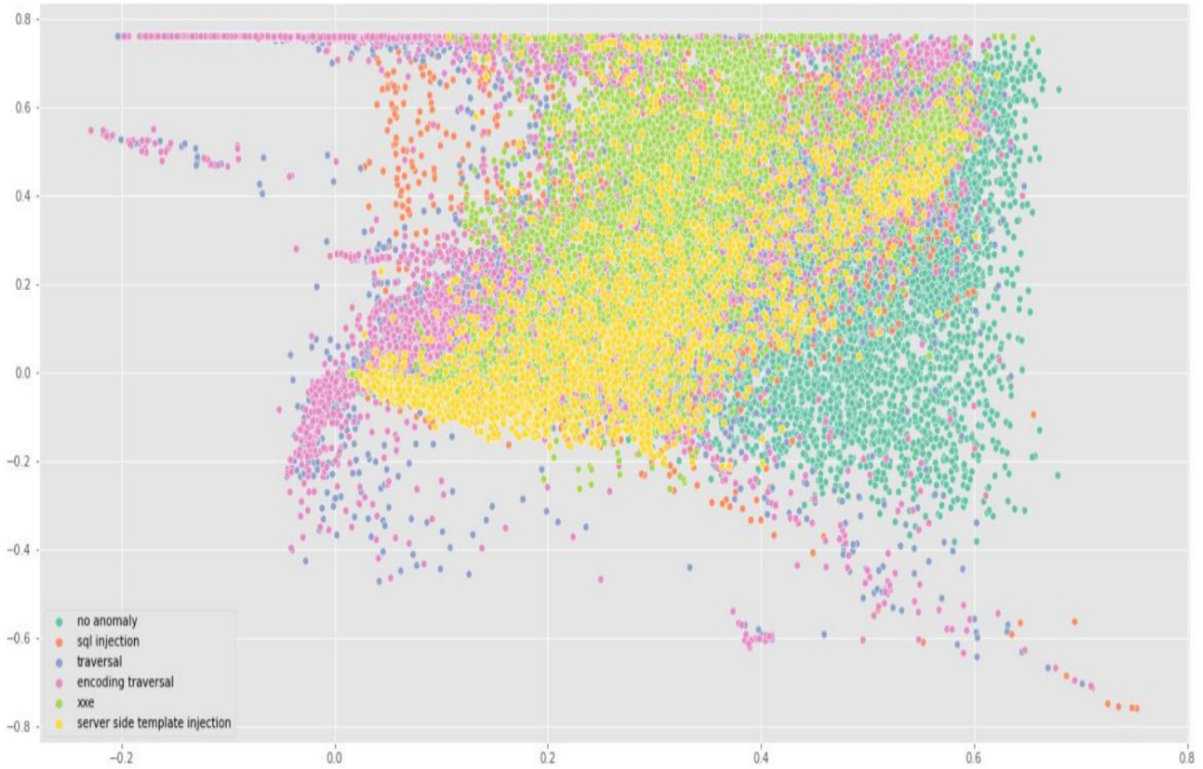


Рисунок 36 — Скрытые латентные представления полученные в результате сжатия информации автокодировщиком

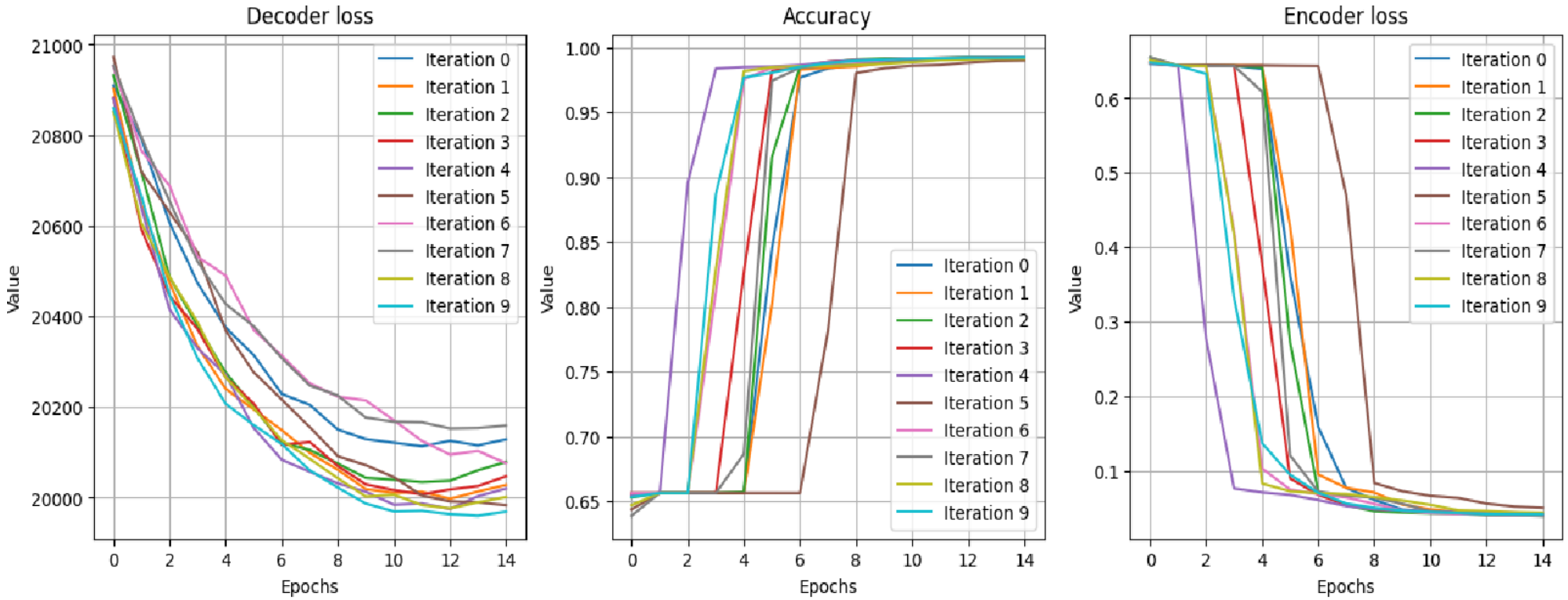


Рисунок 37 — Обучение декодера и классификатора на 15 эпохах в 10 итераций с перемешиванием данных

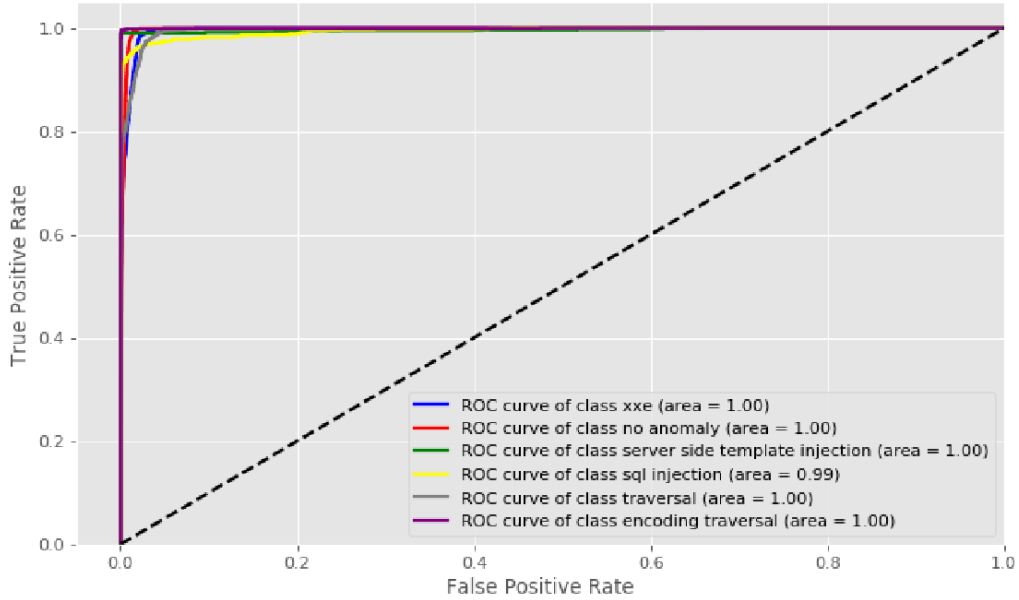


Рисунок 38 — ROC-кривая гибридной нейронной сети

	precision	recall	f1-score	support
xxe	0.83	0.95	0.88	1292
no anomaly	0.99	1.00	0.99	9266
server side template injection	1.00	0.99	0.99	1265
sql injection	0.96	0.93	0.95	1271
traversal	0.94	0.80	0.86	1263
encoding traversal	0.99	0.99	0.99	1305
accuracy			0.97	15662
macro avg	0.95	0.94	0.94	15662
weighted avg	0.97	0.97	0.97	15662

Рисунок 39 — Оценка гибридной нейронной сети

	precision	recall	f1-score	support	classifiers
xxe	0.579972	0.444089	0.503016	939.000000	Logistic Regression
no anomaly	0.903699	0.999436	0.949159	7089.000000	Logistic Regression
server side template injection	0.986715	0.857293	0.917462	953.000000	Logistic Regression
sql injection	0.923077	0.581498	0.713514	908.000000	Logistic Regression
traversal	0.552239	0.558190	0.555198	928.000000	Logistic Regression
encoding traversal	0.980000	0.895699	0.935955	930.000000	Logistic Regression
accuracy	0.868137	0.868137	0.868137	0.868137	Logistic Regression
macro avg	0.820950	0.722701	0.762384	11747.000000	Logistic Regression
weighted avg	0.864330	0.868137	0.860543	11747.000000	Logistic Regression

Рисунок 40 — Оценка logistics regression

	precision	recall	f1-score	support	classifiers
xxe	0.457529	0.252396	0.325326	939.000000	Linear DA
no anomaly	0.959525	0.979828	0.969570	7089.000000	Linear DA
server side template injection	0.834297	0.908709	0.869915	953.000000	Linear DA
sql injection	0.803077	0.574890	0.670090	908.000000	Linear DA
traversal	0.478916	0.685345	0.563830	928.000000	Linear DA
encoding traversal	0.839836	0.879570	0.859244	930.000000	Linear DA
accuracy	0.853409	0.853409	0.853409	0.853409	Linear DA
macro avg	0.728863	0.713456	0.709662	11747.000000	Linear DA
weighted avg	0.849702	0.853409	0.846051	11747.000000	Linear DA

Рисунок 41 — Оценка Linear DA

Спасибо за внимание!

Данные исследования выполняются при финансовой поддержке Гранта РНФ № 21-71-20078 в СПб ФИЦ РАН.