

Математичне та програмне забезпечення системи ідентифікації аномалій в даних вимірювання

Студент: Бірук Станіслав, КМ-71

Керівник: старший викладач кафедри ПМА Ладогубець Т.С.

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний технічний університет України
«Київський політехнічний інститут
імені Ігоря Сікорського»

ЗАЛПОВА КНИЖКА № ЗКМ-7103
СТУДЕНТА

Прізвище, ім'я, по батькові Бірук Станіслав Володимирович

Інститут, факультет, відділення Технічної математики
Спеціальність 113 Бірюківська математика
Спеціалізація _____
Рівень вищої освіти бакалавр
Форма навчання денна
Вступи 6 на 1 курс з 04 09 2017 року
Проректор _____
Директор інституту, декан факультету,
завідувач відділення Л.С. Ладогубець
(дата видачі залпкової книжки) (підпис) (підпис та ініціали)

Актуальність теми:

- моніторинг стану обладнання;
- відстежування шахрайських транзакцій;
- моніторинг несподіваних коливань на фінансовому ринку акцій, облігацій та інших цінних паперів;
- облік показників стану здоров'я пацієнтів;
- аналіз телеметричних даних

Постановка задачі

Метою даної дипломної роботи є підвищення ефективності ідентифікації аномалій в даних вимірювання

Розроблена система повинна виконувати наступні функції:

- ітеративно ідентифікувати аномалії (в online режимі);
- виявляти аномалії в багатовимірних часових рядах;
- виявляти не менше 75% наявних аномалій;
- не враховувати додаткову інформацію про структуру даних (розподіл, діапазон значень, наявність або відсутність аномалій);

Розглянуті методи

Методи подібності:

- k-найближчих сусідів
- STORM

Статистичні

Параметричні:

- Змішані гаусівські моделі
- Моделі теорії екстремальних значень
- Моделі станів простору

Непараметричні:

- Моделі негативного відбору
- Моделі ядрової оцінки густини розподілу

Зменшення розмірності:

- Аналіз незалежних компонент
- Взаємо кореляційний аналіз
- Інкрементний аналіз головних компонент

Передбачувальні:

- Рекурентні нейронні мережі
- Згорткові нейронні мережі
- Моделі ієрархічної тимчасової пам'яті

Генеративні:

- Генеративно-змагальні мережі
- Варіаційні автоенкодера

Порівняльна характеристика

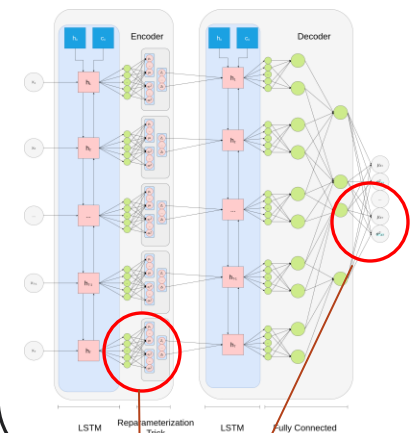
Методи	Залежність між рівнями ряду	Необхідність навчання	Наявність аномалій в навчальній вибірці	Критичне значення
Подібності	Не враховує	Не потрібно	--	Задається вручну
Статистичні	Не враховує	Залежить від типу*	Залежить від типу*	Залежить від типу*
Зменшення розмірності	Враховує	Залежить від типу	Впливає на результат	Не використовується
Передбачувальні	Враховує	Потрібно	Впливає на результат	Задається вручну
Генеративні	Враховує	Потрібно	Не впливає	Задається вручну

Тип*: параметричні та непараметричні методи

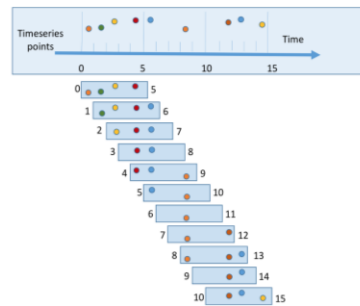
Побудова Математичної Моделі



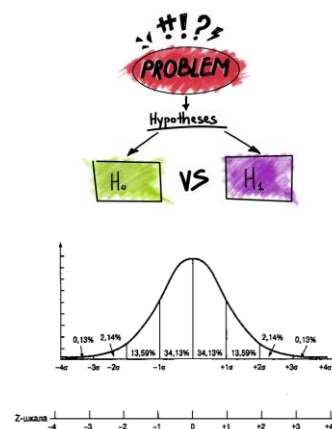
Навчання моделі варіаційного автоенкодера



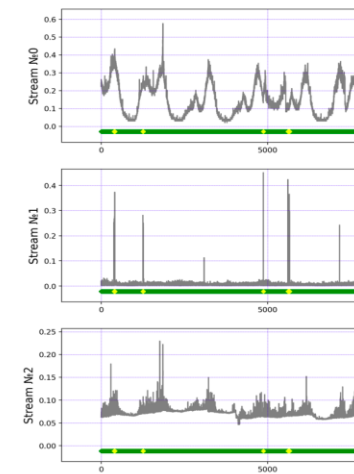
Створення навчальної вибірки



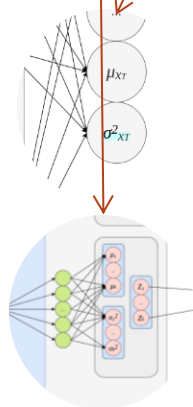
Перевірка на стаціонарність та стандартизація даних



Вхідні дані



Обчислення логарифма ймовірності



$$\log N(x_T | \mu, \sigma^2)$$

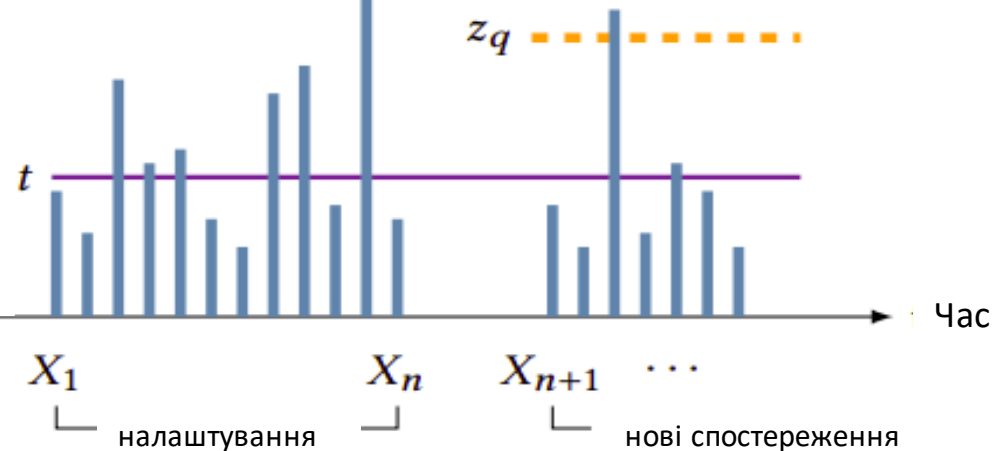
$$\log N(z_T | 0, I)$$

Аномалії

Піки

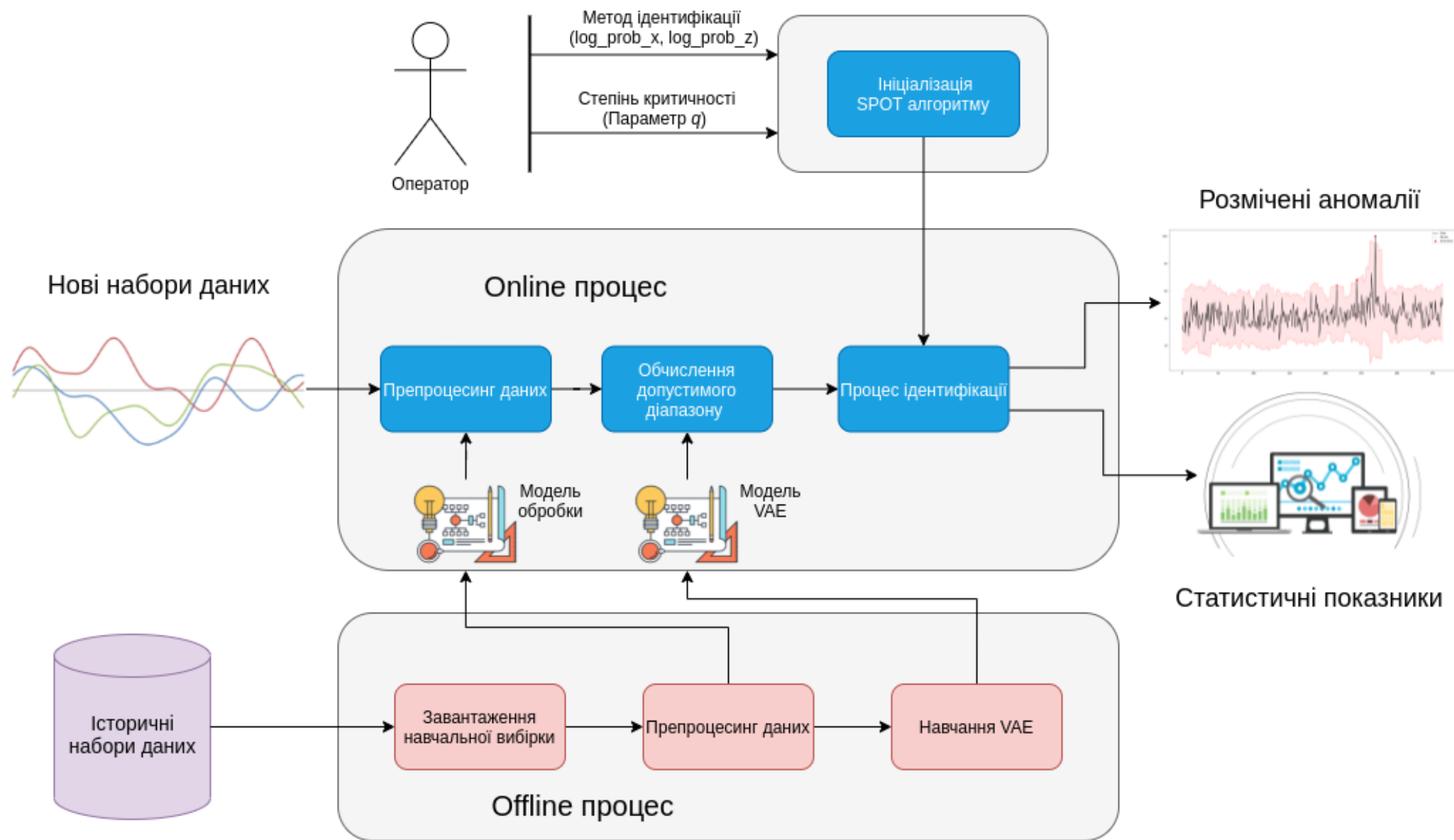
Нормальні спостереження

Критичне значення



Структура системи





Програмна реалізація



Time Series Anomaly Detection



Load Data

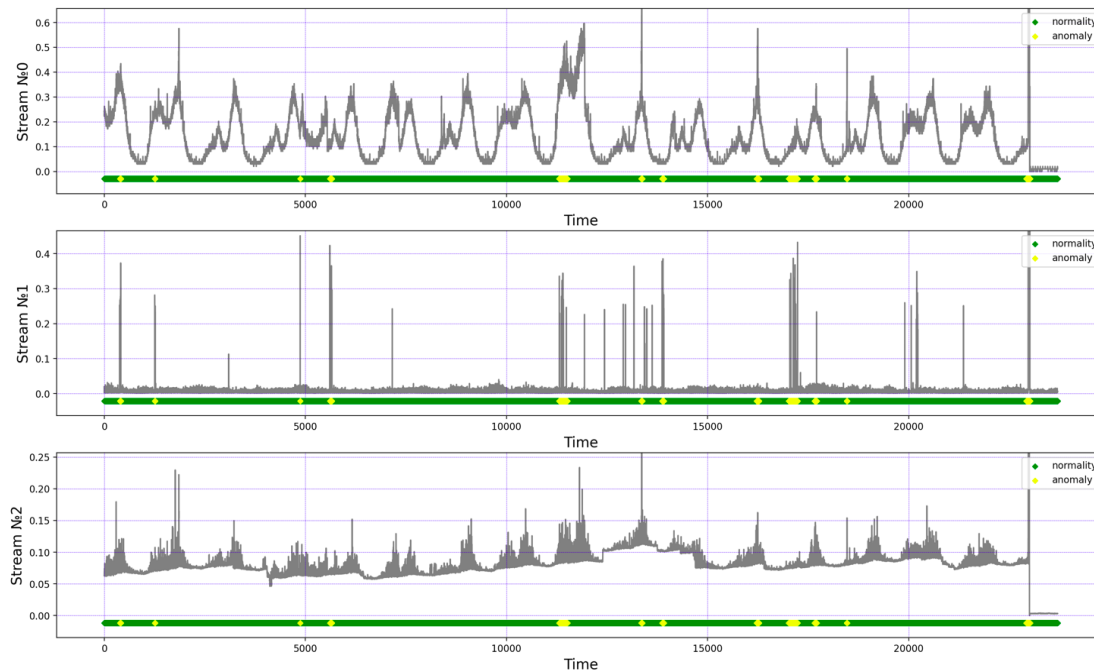
Select one of the datasets

Machine3

Machine1

Machine2

Machine3



Streaming Peaks-Over-Threshold (SPOT) Algorithm

Type of log_prob

- ☐ log_prob_x
- ☒ log_prob_z

Degree of risk parameter



The dimension of the training sample



The dimension of the calibration sample



Run

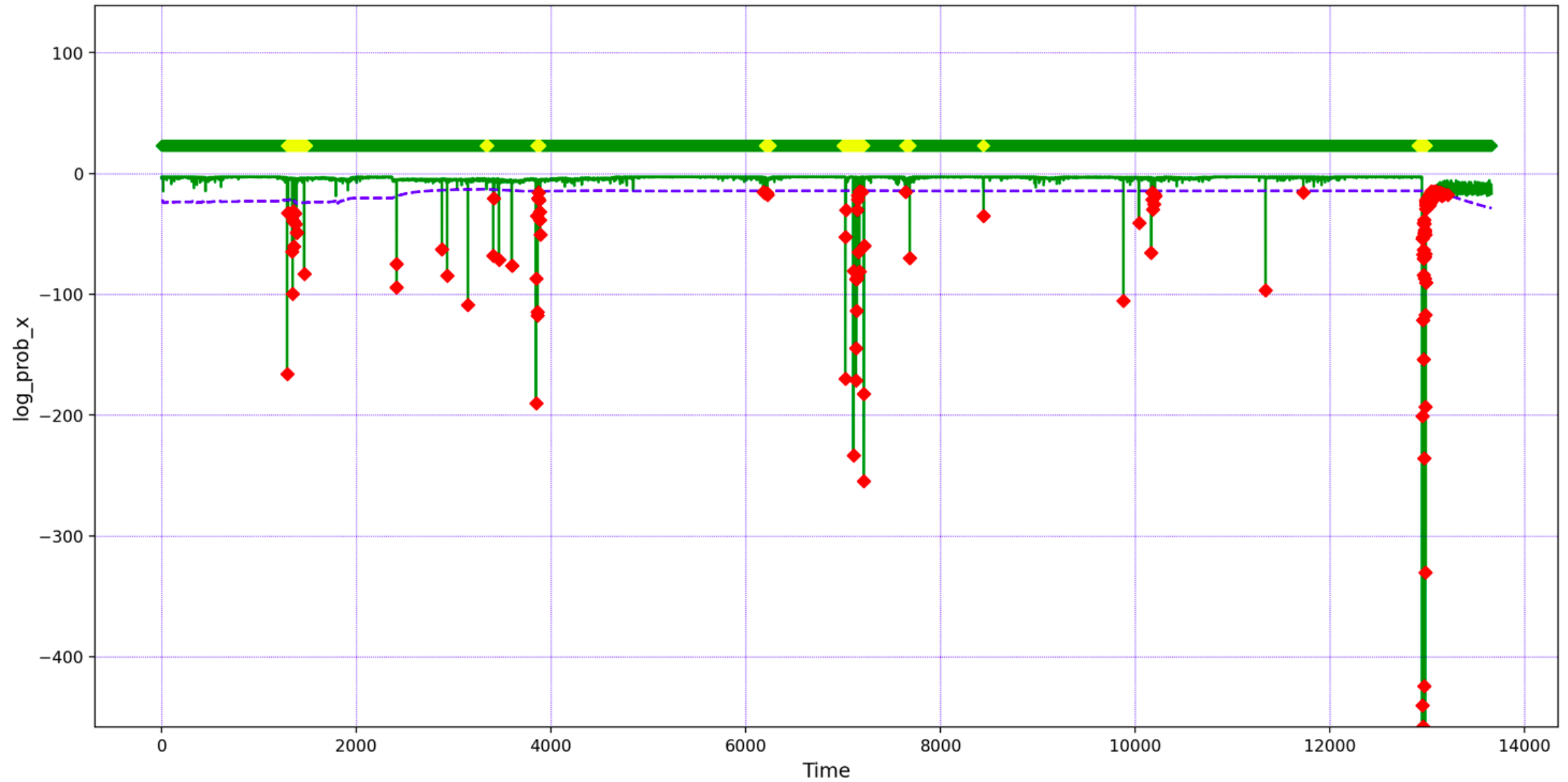
Результати роботи



Результати роботи Варіаційного Автоенкодера



Результати роботи SPOT алгоритму



Результати роботи системи

Метрики якості

Влучність (precision)	Повнота (recall)	F1- оцінка
0.904	0.978	0.939

Матриця невідповідності

		Дійсні класи	
		Аномалія	Не аномалія
Передбачувальні класи	Аномалія	526	56
	Не аномалія	12	13647

Висновки

У рамках дипломного проектування:

- Здійснено аналіз існуючих підходів та методів виявлення аномалій;
- Спроектовано та реалізовано цільову систему;
- Перевірено працездатність системи на реальних даних;
- Відношення виявлених аномалій перевищує 82.6%

Подальші напрямки розвитку та вдосконалення системи:

- Підбір оптимальної архітектури нейронної мережі
- Підбір гіперпараметрів нейронної мережі методами баєсівської оптимізації (BoTorch)
- Заміна гаусівського латентного простору на суміш розподілів
- Оптимізація SPOT алгоритму

Дякую за увагу!