Виявлення аномалій роботи веб-серверів методами нейронних мереж та лог-аналітики

Автор : студент 4-го курсу
групи КА-43
Чабанівський Артем
Тарасович

Керівник:
Доцент, к.ф.-м.н.
Каніовська Ірина Юріївна

Актуальність роботи

- Своєчасне виявлення некоректної роботи сервера є необхідним для стабільної роботи веб-застосунку та збереження інформації
- Обробка лог-даних вручну, у пошуках аномалій, є практично неможливою через великі об’єми даних та специфічну структуру лог-запису
- Відсутність Open Source (з відкритою/безкоштовною ліцензією) застосунків подібного типу

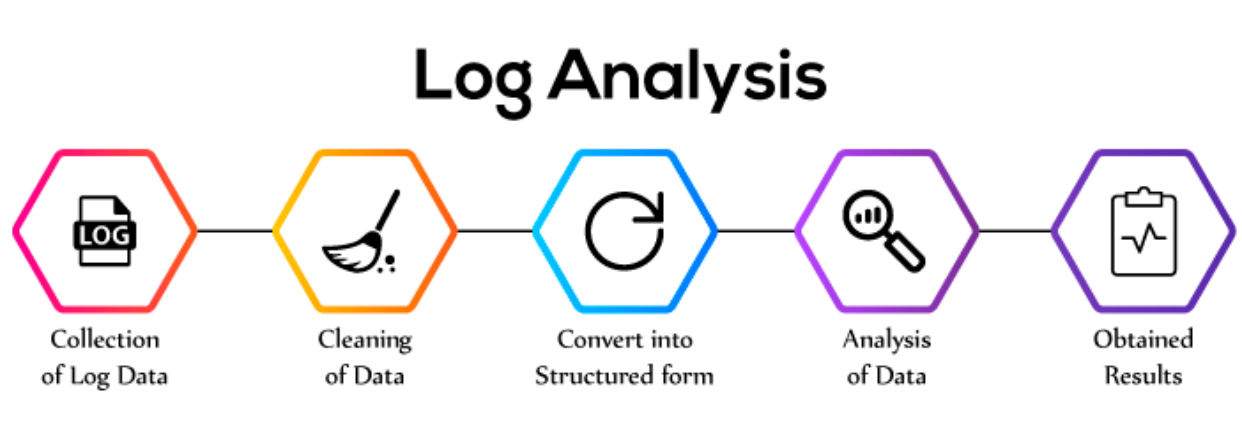
Постановка задачі дипломної роботи

- Проаналізувати існуючі засоби лог-аналізу та їх можливості
- Проаналізувати види аномалій та засоби їх виявлення
- Обрати необхідний інструмент для виявлення аномалій веб-сервера
- Розробити модель нейронної мережі для роботи з лог-даними
- Розробити програмний продукт для розпізнавання стану роботи сервера за даними, що зібрані за певний (заданий) проміжок часу

- **Мета:** розробка ПЗ для виявлення некоректної (аномальної) роботи веб-серверів використовуючи лог-дані
- **Предмет:** Лог-аналітика та методи машинного навчання
- **Об’єкт:** Лог-дані веб-серверу nginx з проекту “Design Contest”

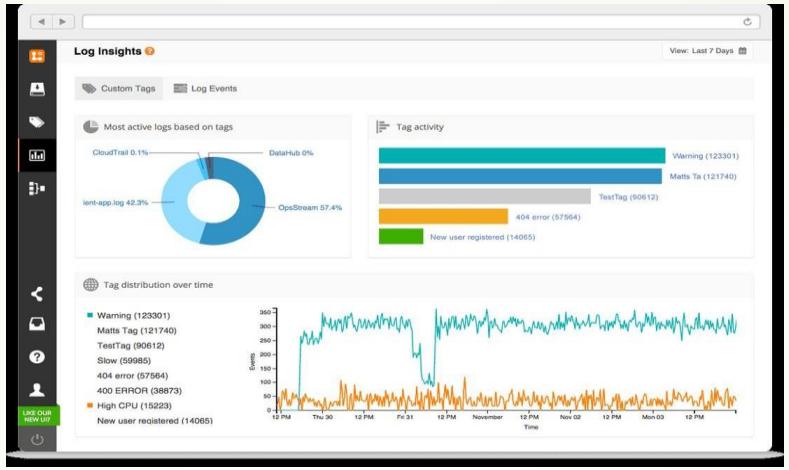
Лог - аналітика

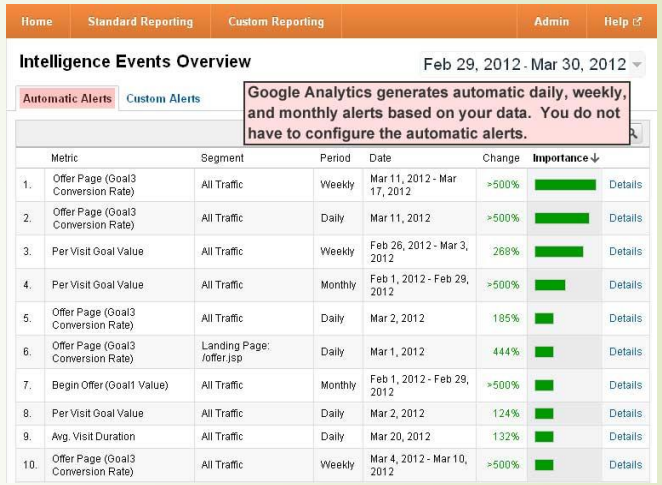
I. Сбір даних
II. Фільтрація даних
III. Структурування даних
IV. Аналіз даних

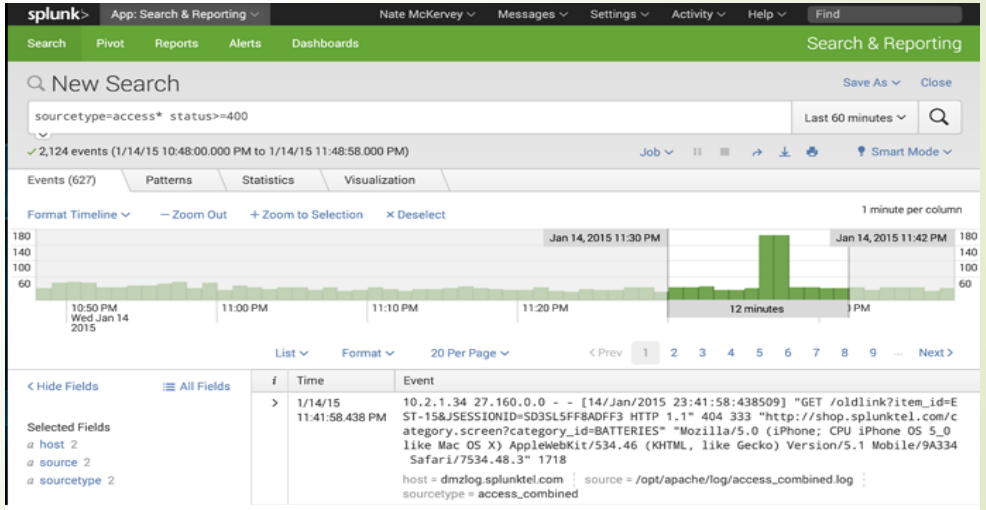


Аналіз існуючих засобів лог-аналітики

1. Logentries
2. GAIE
3. Splunk







Види аномалій

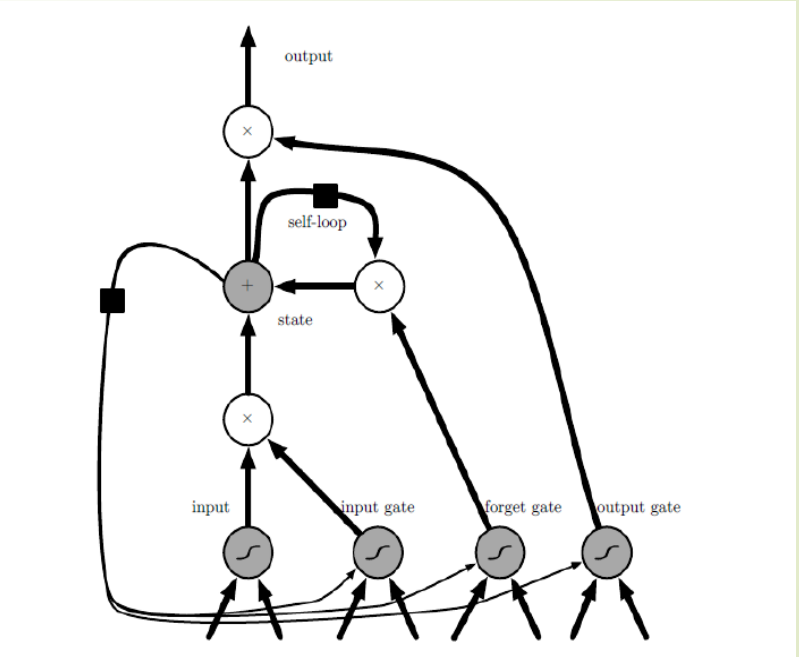
1. **Точкові** : один екземпляр даних розцінюється як точкова аномалія, якщо він знаходиться «далі» від решти даних.
2. **Контекстні** : набір даних є аномальним у специфічних умовах (специфічному контексті), але нормальним в інших випадках.
3. **Колективні** : певний набір даних є аномальним по відношенню до всього набору даних, але окремі одиниці даних не є аномальними.

Методи виявлення аномалій

- Статистичний підхід
  - «сусідство даних»
  - параметричні методи
  - непараметричні методи
- Навчання «без вчителя»
  - лог кластеризація
  - метод головних компонент
  - інваріантний пошук
  - нейронні мережі
- Навчання «з вчителем»
  - логістична регресія
  - дерево рішень
  - метод опорних векторів
  - нейронні мережі

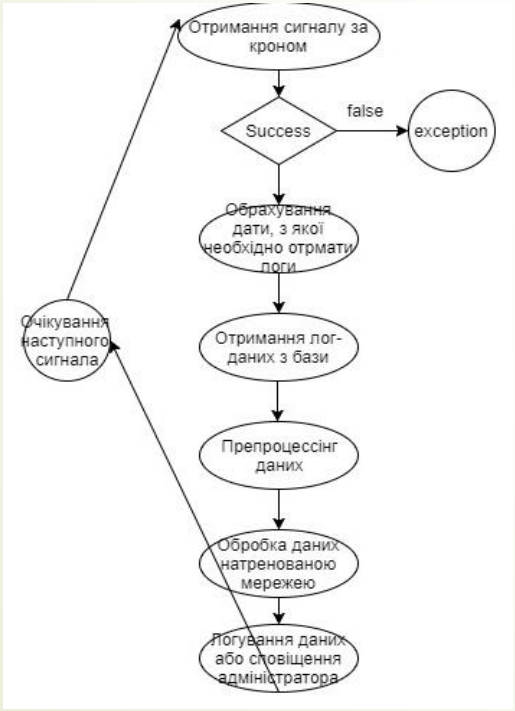
Рекурентні нейронні мережі типу LSTM

- зв'язки між елементами утворюють напрямлену послідовність
- можуть використовувати свою внутрішню пам'ять для обробки послідовностей довільної довжини
- здатні до навчання довготривалим залежностям

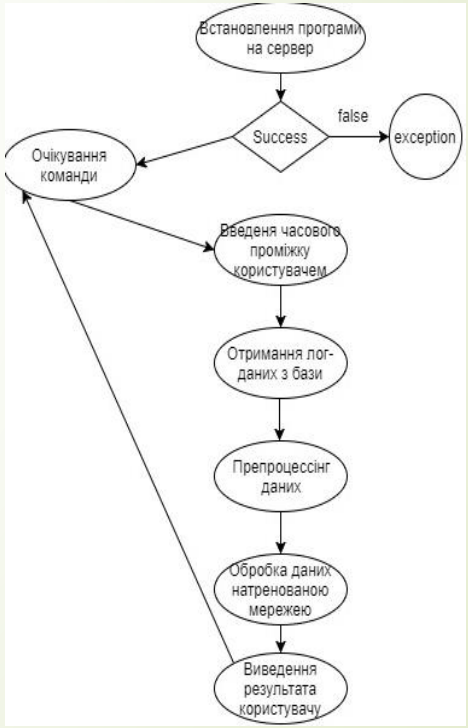


Діаграма діяльності за варіантами використання

За кроном



За запитом користувача



Результати

Тренування мережі

```
Iteration #0
Accuracy : 0.84
Iteration #1
Accuracy : 0.85
Iteration #2
Accuracy : 0.93
```

Робота програми

```
PS D:\Diplom\Project\DiplomaProject\bin\Debug> .\DiplomaProject.exe "1516667744" "1518148846"
Anomaly state of system
PS D:\Diplom\Project\DiplomaProject\bin\Debug> .\DiplomaProject.exe "1515257646" "1516357646"
Normal state of system
```

Аналіз результатів

1. Точність роботи програми є недостатньою для впровадження у проект, де від результатів роботи даного продукту будуть залежати конфіденційна інформація, стан банківських рахунків, тощо, але є достатньою для впровадження у невеликі веб-застосунки
2. Точність роботи програми можливо покращити за допомогою наступних заходів :
    I. Тренувати модель на більш потужній машині
    II. Зібрати більшу базу лог-даних
    III. Запросити експертів задля коректнішого аналізу даних

Висновки

- Спроектовано та розроблено ПЗ, що виявляє аномалії у роботі веб-сервера за лог-даними
- Розроблена система:
  - Працює за запитом користувача або за кроном
  - Виявляє стан роботи система (нормальний/аномальний) за «часовим вікном» з лог-даних
  - Здатна працювати через зазначений проміжок часу, використовуючи засоби, які зазвичай за замовчуванням використовуються у веб-застосунках

Шляхи подальшого розвитку

- Покращити точність роботи
- Розробити зручний інтерфейс користувача
- Розробити можливість розпізнавання конкретного типу атак
- Розробити систему розпізнавання аномалій не тільки веб-серверу, а також інших складових веб-застосунків, таких як, наприклад, база даних

Впровадження результатів роботи

- Зазначене програмне забезпечення впроваджене до веб-системи автоматизації обліку та аналізу потоку продукції “Sales Out Database”(http://sod.smart-mase.com) підприємства CHGROUP LLC.

Дякую за увагу!