УДК 004.7

Маруняк С.Т., аспірант,
Кирик М.І., д.т.н., професор,
Національний університет «Львівська політехніка»

ІНТЕРПРЕТОВАНИЙ АНАЛІЗ ОЗНАК ДЛЯ ПІДВИЩЕННЯ ТОЧНОСТІ КЛАСИФІКАЦІЇ АНОМАЛІЙ BGP

Протокол BGP забезпечує обмін маршрутизованою інформацією між автономними системами (АС), тому його стабільність є критичною для безперебійної роботи Інтернету. Помилки в конфігурації, підміна префіксів або масові виходи з ладу можуть призводити до дестабілізації глобальних маршрутів. Для зменшення наслідків таких інцидентів важливо не лише виявити факт аномалії, а й визначити її тип – збій у мережі (outage), підміна маршруту (hijack) чи інші форми порушень. Це дозволяє обрати відповідну реакцію: від технічного відновлення до втручання операторів.

Для автоматизованої класифікації BGP-аномалій використовують десятки ознак, сформованих на основі потоків повідомлень BGP. Проте надлишкова кількість або слабоінформативні параметри можуть знижувати якість класифікації. Щоб виявити найбільш релевантні ознаки, застосовують методи пояснення моделей – Explainable AI (XAI). Одним із таких методів є SHapley Additive exPlanations (SHAP) [1], який дозволяє оцінити вплив кожної ознаки на результат класифікації як загалом, так і для окремих класів аномалій.

На рисунку 1 представлено результат глобального аналізу ознак за допомогою SHAP для базової моделі. Найбільший вплив мають характеристики, пов’язані з топологією маршрутів та географічною відстанню. Натомість дві ознаки, пов’язані з довжиною префіксів IPv4, показали найменші значення важливості.

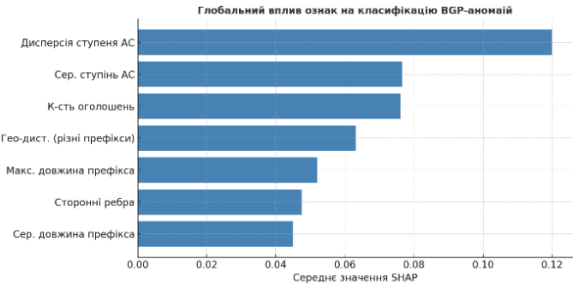

Рисунок 1 – Глобальний вплив ознак на класифікацію аномалій

Для перевірки впливу цих ознак на продуктивність моделі було проведено два вимірювання: з повним набором ознак (табл. 1) та після їх часткового видалення (табл. 2). Модель класифікує три типи аномалій: прямі (direct), непрямі (indirect) та відключення (outage). Ключові метрики: точність (Precision), повнота (Recall), F1-score. Вони характеризують частку правильно класифікованих тестових даних, повноту виявлення об’єктів певного класу та баланс між ними відповідно.

Таблиця 1 – Метрики класифікації вихідної моделі

Class	Precision	Recall	F1-score
Direct	1.00	1.00	1.00
Indirect	0.87	1.00	0.93
Outage	1.00	0.67	0.80

Таблиця 2 – Метрики класифікації моделі після вилучення ознак

Class	Precision	Recall	F1-score
Direct	1.00	1.00	1.00
Indirect	0.91	0.99	0.95
Outage	0.97	0.77	0.86

Після вилучення двох найменш значущих ознак модель почала точніше ідентифікувати випадки відключень: показник повноти для цього класу зріс з 0.67 до 0.77. Це означає, що система стала виявляти більшу частку реальних інцидентів типу збій, не пропускаючи їх. При цьому важливо, що якість класифікації інших типів аномалій залишилася на високому рівні – тобто вилучення ознак не спричинило зменшення точності або зростання хибнопозитивних спрацювань у цих класах. Такий підхід дозволяє не лише оптимізувати модель, а й підвищити її стійкість до шуму у вхідних даних. Подібні інтерпретовані методи можна використовувати і для подальшої оптимізації складніших моделей.

Список використаних джерел:

1. Al-Musawi B., Branch P., Armitage G. BGP anomaly detection techniques: A survey // IEEE Communications Surveys & Tutorials. 2016. Vol. 19, No. 1. P. 377–396. DOI: https://doi.org/10.1109/COMST.2016.2622240

2. Kyryk M., Maruniak S., Tandrukhiv M. SHAP-based Feature Contribution Analysis for Robust BGP Anomaly Classification // Telecommunications and Radio Engineering. 2025. Vol. 84, No. 4. P. 391–400.