

Список використаної літератури

1. Мохор, В., Богданов, А., Килевой, А. Наставления по кибербезопасности (ISO/IEC 27032:2013). – ООО «Три-К», Киев, 2013.
2. Fathollahi-Fard Mohammad Amir, Hajiaghaei-Keshteli Mostafa, Tavakkoli-Moghaddam Reza. The Social Engineering Optimizer (SEO). Engineering Applications of Artificial Intelligence, 2018. – Vol. 72. – pp. 267-293. doi: 10.1016/j.engappai.2018.04.009
3. Mokhor V.V., Tsurkan O.V., Herasymov R.P., Tsurkan V.V. Information Security Assessment of Computer Systems by Socio-engineering Approach. Selected Papers of the XVII International Scientific and Practical Conference «Information Technologies and Security». Kyiv, 2017. - pp. 92-98. URL: <http://ceur-ws.org/Vol-2067/paper13.pdf>.

ПІДХІД ДО ПРОГНОЗУВАННЯ ДІЄВОСТІ ДЕРЖАВНОГО УПРАВЛІННЯ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ OSINT

Додонов О.Г.¹, Ланде Д.В.^{1,2}, Нестеренко О.В.², Березін Б.О.¹

¹Інститут проблем реєстрації інформації НАН України

²Національний технічний університет України “КПІ ім. Ігоря Сікорського”

1. Постановка проблеми

В період проведення реформ дієвість державного управління країною набуває особливого значення. Одним з важливих елементів державного управління є ефективність законопроектів, що приймаються. Ефективність приймаємих законопроектів та державного управління в цілому знаходить відображення у суспільній думці відносно законопроектів, показником якої можна вважати кількість новинних публікацій відповідної тематики в інформаційному просторі. На сьогодні, державне управління є однією з важливих сфер застосування технологій аналізу відкритих джерел OSINT. При цьому, прогнозування належить до числа невід’ємних складових системи OSINT [1,2].

В даній роботі розглянуто існуючі засоби прогнозування часових рядів кількості тематичних публікацій та запропоновано підхід до прогнозування дієвості державного управління з використанням технологій OSINT.

2. Аналіз прогнозування дієвості державного управління з використанням існуючих засобів

Серед існуючих засобів прогнозування часових рядів, які можуть бути використані для оцінки дієвості державного управління, розглядалися відповідні засоби статистичних пакетів, бібліотек, що входять до складу сучасних мов програмування, а також відповідні онлайн сервіси. З цією метою, за допомогою систем моніторингу інформаційних ресурсів Інтернет було виконано збір даних за тематикою найбільш резонансних у суспільстві законопроектів та законів: про забезпечення функціонування української мови як державної, про розмитнення автомобілів на іноземній реєстрації, про Вищий антикорупційний суд, та деяких інших. Для отриманих часових рядів кількості публікацій відносно резонансних законопроектів було виконано прогнозування за допомогою відповідних бібліотек мов програмування R для статистичних задач та за допомогою засобів пакету Excel.

Серед популярних засобів прогнозування часових рядів відзначають моделі ETS, ARIMA, пакет Prophet та інші [3,4], реалізовані в різних мовах програмування та статистичних пакетах. В основі моделей ETS (Exponential Smoothing) лежить експоненціальне згладжування - метод прогнозування, при якому значення змінної за всі попередні періоди входять в прогноз, експоненціально втрачаючи свою вагу з часом. Це дозволяє моделі з достатнім ступенем гнучко реагувати на новітні зміни в даних, зберігаючи при цьому інформацію про історичну поведінку часового ряду.

Порівняно новий пакет Prophet (розробка 2017 р.), на відміну від раніше існуючих, повинен зменшити трудоемність налагодження параметрів при проведенні прогнозування. В його основі лежить процедура підгонки адитивних регресійних моделей з наступними чотирма основними компонентами: тренд (моделюється за допомогою кусочної лінійної регресії або кусочної логістичної кривої зростання); річна сезонність (моделюється як ряд Фур'є); тижнева сезонність (моделюється з використанням індикаторних змінних); "свята" (наприклад, офіційні святкові та вихідні дні). Оцінювання параметрів підгоняємої моделі виконується з використанням принципів байєсівської статистики. Пакет Prophet реалізовано для мов програмування R та Python.

В якості вихідних даних для прогнозування кількості публікацій по тематикам приймаємих законопроектів та законів, використано дані моніторингу ресурсів Інтернет з відповідними запитам

за період квітень-вересень 2019 р. (закон про забезпечення функціонування української мови як державної) та листопад 2018 р. – вересень 2019 р. (закон про розмитнення автомобілів на іноземній реєстрації). Для зменшення дисперсії вихідних даних було використано логарифмування. Далі, за допомогою методів ETS, Prophet, а також моделі Сорнетте, на основі даних моніторингу було отримано прогностні значення логарифму кількості відповідних публікацій. На рис. 1 - 3 наведено вихідний часовий ряд кількості публікацій відносно закону про забезпечення функціонування української мови як державної. На осі абсцис графіків показано номери днів часового ряду, а на осі ординат – значення логарифму кількості публікацій.

На перші декілька днів часового ряду приходиться пік значення логарифму кількості публікацій, пов'язаний з голосуванням за прийняття закону (3-й день часового ряду). Спадаючий характер подальшої частини ряду пов'язан із зменшенням реакції на цю подію. Перша частина вихідного часового ряду (1 - 134 дні) використовується для навчання моделі прогнозування, друга частина (135 – 167 дні, приблизно 20% довжини ряду) для оцінки точності прогнозування. Отриманий прогноз в цій частині часового ряду порівнюється з вихідними даними (Таблиця 1).

На рис. 1 наведено графік часового ряду логарифму кількості публікацій відносно закону про забезпечення функціонування української мови як державної та графік відповідного прогнозованого часового ряду (пунктирна лінія), отриманого за допомогою метода ETS.

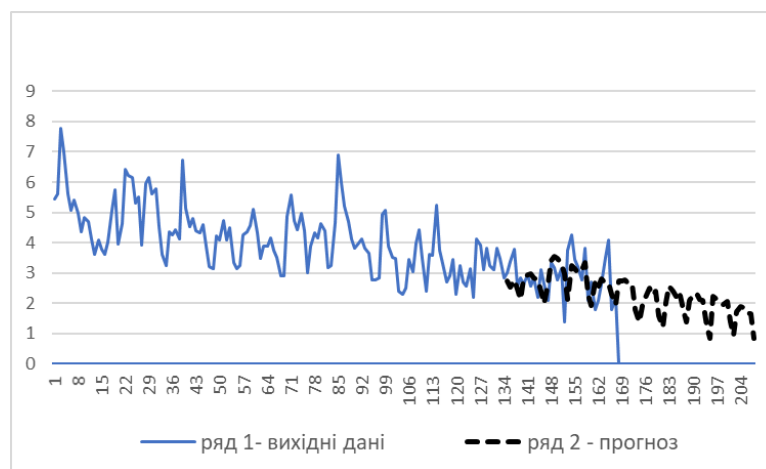


Рис. 1. Прогноз логарифму кількості публікацій відносно закону про забезпечення функціонування української мови як державної, отриманий за допомогою функції ETS. На осі абсцис показано номери днів часового ряду, а на осі ординат – значення логарифму кількості публікацій.

3. Підхід до прогнозування кількості публікацій за допомогою моделі Сорнетте

Серед публікацій, присвячених передбаченню фінансових криз відзначають роботи Д. Сорнетте [5]. Його метод заснований на аналізі закономірності руху ринкових цін на товарних і фондових ринках перед крахом. В роботі відзначається, що перед крахом ціна має степеневе зростання, ускладнене логоперіодичними коливаннями, які сходяться до нескінченності в критичній точці, де ймовірність краху досягає максимальної величини. Для прогнозування можливого краху пропонується аналізувати ринкові дані (часові ряди) з метою виявлення в них коливальних степеневих закономірностей. Для виявлення краху на фінансових ринках, степенева модель, що враховує лінійні логоперіодичні коливання, має наступний вигляд (1).

$$F(t) = A + B(t_c - t)^m \left[1 + C \cos \left(\omega \log \left(\frac{t_c - t}{T} \right) + \varphi \right) \right] \quad (1)$$

У цій моделі t_c – критичний час (час краху на фінансовому ринку). Коефіцієнти моделі визначаються за допомогою процедури підбору. В роботі [6] розглянуто можливості передбачення криз на фінансовому ринку за допомогою методу Д. Сорнетте.

За допомогою використання моделі Сорнетте, на основі даних моніторингу було отримано прогностні значення логарифму кількості відповідних публікацій. Для визначення коефіцієнтів моделі Сорнетте використовувались функції та пакети для підбору рішень у складі мови програмування R та пакету Excel (функція nls – розраховує параметри нелінійної регресійної моделі за методом найменших квадратів; функція solver для пошуку рішень нелінійних та лінійних задач та інші). Крім того, для визначення коефіцієнтів моделі Сорнетте було використано програму на мові Python [7]. На рис. 2 наведено графік часового ряду логарифму кількості публікацій відносно закону про

забезпечення функціонування української мови як державної та графік відповідного прогнозованого часового ряду (пунктирна лінія), отриманий за допомогою використання моделі Сорнетте.

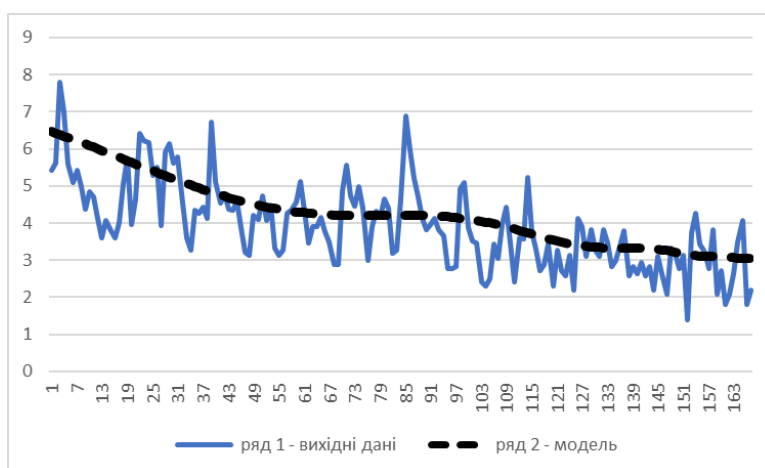


Рис. 2. Прогноз логарифму кількості публікацій відносно закону про забезпечення функціонування української мови як державної, отриманий за допомогою використання моделі Сорнетте. На осі абсцис показано номери днів часового ряду, а на осі ординат – значення логарифму кількості публікацій.

Результати прогнозування кількості публікацій за допомогою вище розглянутих моделей оцінювались на основі відомої метрики, в якій порівнюються відповідні значення вихідного та прогнозованого часових рядів (2).

$$\chi = \sqrt{\frac{1}{n} \sum_{i=1}^n (\tilde{y}(t) - y(t))^2} \quad (2)$$

Для використання метрики у вихідних часових рядах було виділено 20% значень, які порівнювались із відповідними значеннями, отриманими шляхом прогнозування. Результати оцінки для різних засобів прогнозування та двох наборів даних про публікації наведено в Таблиці 1.

Таблиця 1. Оцінка точності прогнозування для різних засобів та вихідних даних.

Засоби, використані для прогнозування	Дані моніторингу про публікації:	
	відносно законів про забезпечення функціонування української мови як державної	відносно законів про розмитнення автомобілів на іноземній реєстрації
Функція ETS	0.83	1.39
Пакет Prophet	0.87	1.45
Модель Сорнетте	0.78	1.33

Для підвищення точності прогнозу з використанням моделі Сорнетте, пропонується враховувати сезонність аналізуємих явищ. У випадку кількості публікацій відносно законопроектів доцільно враховувати вихідні та святкові дні, коли активність публікацій знижується. З цією метою було розраховано сінусоїду, амплітуда та зсув якої підбиралися таким чином, щоб при накладанні її на модель, отриману у відповідності з виразом (1), значення метрики (2) зменшувалось. В розглянутих варіантах параметрів, зменшення метрики (2) складало від одного до кількох відсотків. На рис. 3 наведено графік часового ряду логарифму кількості публікацій відносно закону про забезпечення функціонування української мови як державної та графік відповідного прогнозованого часового ряду (пунктирна лінія), отриманий за допомогою використання моделі Сорнетте з врахуванням фактору сезонності -вихідних днів.

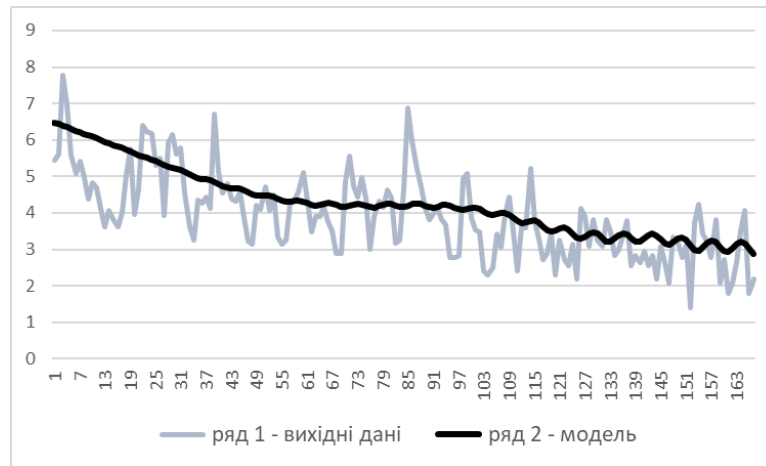


Рис. 3. Прогноз логарифму кількості публікацій відносно закону про забезпечення функціонування української мови як державної, отриманий за допомогою використання моделі Сорнетте з врахуванням сезонності. На осі абсцис показано номери днів часового ряду, а на осі ординат – значення логарифму кількості публікацій.

Отримані значення оцінки точності показують можливість впровадження запропонованого підходу до прогнозування кількості новинних публікацій (та відповідно дієвості державного управління) за допомогою моделі Сорнетте в розробках аналітичних систем для області OSINT [8].

4. Висновки

В роботі розглянуто прогнозування дієвості державного управління на основі оцінки часових рядів публікацій відносно прийомлених законопроектів та законів. На даних моніторингу публікацій в період прийняття та дії законопроектів показано можливість прогнозування часових рядів публікацій за допомогою існуючих програмних засобів. Запропоновано підхід з використанням моделі Сорнетте для прогнозування часових рядів публікацій та показано особливості використання моделі. Отримані результати показують можливість впровадження запропонованого підходу до прогнозування дієвості державного управління в розробках аналітичних систем для області OSINT.

1. Распознавание информационных операций / А.Г. Додонов, Д.В. Ландэ, В.В. Цыганок, О.В. Андрейчук, С.В. Каденко, А.Н. Грайворонская – К.: Инжиниринг, 2017. – 282 с.
2. Додонов А.Г., Ландэ Д.В., Путятин В.Г. Применение OSINT в аналитической деятельности // Реестрация, зберігання і обробка даних. Щорічна підсумкова наукова конференція 17-18 травня 2018 року: збірник / - Київ: ІПРІ НАН України, 2018. - С. 110-112.
3. Jain, G., & Mallick, B. A study of time series models ARIMA and ETS. Available at SSRN 2898968. IJ.Modern Education and Computer Science, 2017, 4, 57-63.
4. Yenidoğan, I., Çayir, A., Kozan, O., Dağ, T., & Arslan, Ç. Bitcoin Forecasting Using ARIMA and PROPHET. In 2018 3rd International Conference on Computer Science and Engineering (UBMK) (pp. 621-624). IEEE.
5. Сорнетте, Д. Как предсказывать крахи финансовых рынков. Критические события в сложных финансовых системах. Litres, 2017. - 394 с.
6. Уренцов, О. В. Проверка возможности предсказания кризисов на финансовом рынке с помощью метода Д. Сорнетте. Труды Института системного анализа Российской академии наук, 2008, 40: 174-191.
7. Nielsen J. Log-Periodic Power Law Singularity (LPPLS) Model for Bubble Detection <https://boulderinvestment.tech/blog/2018/log-periodic-power-law-lppl-model-for-bubble-detection>.
8. Aleksandr Dodonov, Dmitry Lande, Boris Berezin. Semantic Models at Task Monitoring Public Opinions // CEUR Workshop Proceedings (ceur-ws.org). Vol-2318 [http://ceur-ws.org/Vol-2318/paper1.pdf]