

УДК 004.85:004.91

## **ХРОНОЛОГІЧНЕ МОДЕЛЮВАННЯ ТЕМ І АНАЛІЗ ЕМОЦІЙНОГО ЗАБАРВЛЕННЯ ТЕКСТІВ НА ОСНОВІ ML.NET**

Цибань Д.С.

e-mail: dmytro.tsyban@nure.ua

Науковий керівник – к.т.н., ас. Кобилін І. О.

Харківський національний університет радіоелектроніки, каф. СТ  
м. Харків, Україна

Following work is devoted to developing a system for chronological topic representation and sentiment analysis using ML.NET. To structure large-scale text data, the system organizes topics based on keywords, entities, and contextual patterns while analyzing emotional tones. Data is collected from news, blogs, and social media, undergoing preprocessing and feature extraction using named entity recognition and sentiment analysis. Transparency is ensured through SHAP and LIME. The system leverages ML.NET for seamless .NET integration, supporting real-time analysis and scalability.

Експоненціальне зростання цифрових комунікацій спричинило масовий потік текстових даних, що вимагає ефективних методів їх аналізу. Система, яка інтегрує текстові дані для хронологічного представлення тем і аналізу емоційного забарвлення за допомогою ML.NET, пропонує інноваційний підхід до цієї проблеми. Вона структурує події у часі та визначає їх емоційний тон, що спрощує обробку великих текстових масивів.

Основою такої системи є універсальна модель, яка впорядковує дані на основі ключових слів і контекстуальних шаблонів. Додавання аналізу емоційного забарвлення дає змогу оцінювати динаміку настроїв у текстах, що є корисним у журналістиці, аналізі громадської думки та ринкових тенденцій. Журналісти можуть використовувати систему для відстеження змін у наративах, а бізнес-аналітики – для оцінки сприйняття брендів та адаптації маркетингових стратегій.

Ефективність роботи забезпечується ретельним збором і обробкою текстових даних. Інформація з новин, блогів і соціальних мереж проходить очищення, нормалізацію та токенізацію. Методи розпізнавання сутностей, тематичного моделювання та аналізу полярності емоцій підвищують точність моделі. Для прозорості процесу використовуються інструменти SHAP і LIME, що допомагають пояснити отримані результати.

Попри переваги, система має технічні виклики. Робота з великими копусами текстів потребує оптимізації ресурсів, а інтеграція з різними платформами вимагає адаптації до змін у форматах даних. Важливим залишається і питання пояснюваності моделей, адже складні алгоритми можуть давати точні результати, але без належних засобів інтерпретації вони залишаються незрозумілими для кінцевих користувачів.

SHAP (SHapley Additive exPlanations) – це метод для інтерпретації моделей машинного навчання, що показує внесок кожної ознаки у прогноз. Він базується на значеннях Шеплі з теорії ігор, які оцінюють, як кожен елемент впливає на загальний результат. SHAP-значення дозволяють зрозуміти вплив ознак у складних моделях, як-от нейронні мережі, без втрати точності.

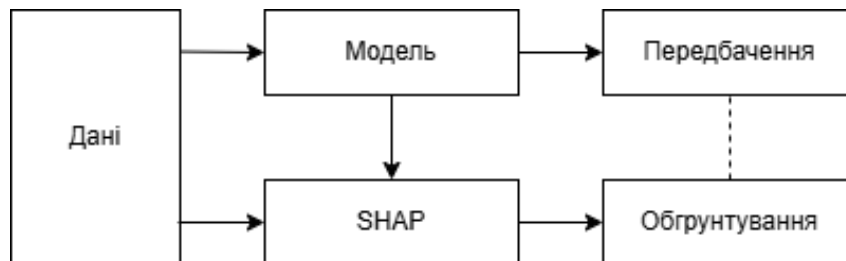


Рисунок 1 – Значення SHAP у практичному використанні

Використання SHAP (рис. 1) дозволяє подолати компроміс між точністю моделі та її інтерпретованістю. Традиційно, складні моделі, такі як глибокі нейронні мережі, забезпечують високу точність, але важко піддаються інтерпретації. Натомість прості моделі, як-от логістична регресія, легше інтерпретувати, але вони можуть поступатися в точності. SHAP-значення надають можливість отримати високу точність, зрозумілість моделі, розкриваючи вплив кожної ознаки на конкретний прогноз.

Практична цінність цієї системи полягає в тому, що завдяки можливості хронологічного представлення тем вона підтримує відстеження наративів, що є необхідним для журналістики під час створення послідовних сюжетів і аналізу потоків інформації. Аналіз емоційного забарвлення додає глибшого розуміння, контекстуалізуючи емоційні відтінки цих наративів, що є безцінним у дослідженнях громадської думки та управління кризами. У поєднанні ці можливості створюють інтегроване рішення, яке є одночасно надійним і адаптованим до змінюваних потреб аналітики.

Список використаних джерел:

1. Esposito D., Esposito F. Programming ML.NET: Train, evaluate, and deploy machine learning models in .NET applications. Microsoft Press, 2022.
1. Liu B. Sentiment analysis and opinion mining. Morgan & Claypool Publishers, 2012.
2. Bouziane K. SHAP Values: The efficient way of interpreting your model. DataDrivenInvestor. URL: <https://medium.datadriveninvestor.com/shap-values-the-efficient-way-of-interpreting-your-model-7de632ed7d2d> (дата звернення: 24.02.2025).