

# ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ ВИДОБУТКУ ДАНИХ З ІНТЕРНЕТ ПОТОКУ НОВИН ДЛЯ РИНКОВОГО ПРОГНОЗУВАННЯ

Орехов С.В., Черенков І.А.

*Національний технічний університет*

*«Харківський політехнічний інститут», м. Харків*

Якість ринкового прогнозу залежить від точності визначення поточної ситуації на ринку, яку задає черга подій на цільовому та суміжних з ним ринках. Потік новин є відображенням цих подій і використовується в якості вихідних даних для аналізу [1].

Специфіка обробки цього потоку полягає в аналізі новини як відособленого носія інформації про подію, так і в сукупності з іншими новинами. Таким чином, проблематика аналізу потоку новин розділяється на проблему предметного аналізу вмісту новини, що вимагає створення спрощеної моделі ринку для класифікації подій, а також на проблему однозначної ідентифікації події в потоку всіх новин.

Методи машинного навчання дозволяють реалізувати алгоритми автоматичної ідентифікації подій у потоку новин, зокрема, методи класифікації дозволяють одержати алгоритми автоматичної категоризації новини і відображеної події. Методи кластеризації дозволяють порівнювати події в новинах на унікальність і виключити дублювання. Методи машинного навчання як *apriori*, *spearde* дозволяють визначити поточну ситуацію на ринку на основі ретроспективних даних, формуючи при цьому короткостроковий прогноз на майбутнє. Використання таких алгоритмів не вимагає створення якої-небудь моделі функціонування ринку, а тільки наявності достатнього числа статистичних даних.

Таким чином, використовуючи методи машинного навчання можна створити засобу автоматичного видобутку даних з мінімальною експертною участю, який необхідно тільки при створенні класифікатора подій ринку.

**Література:** 1 Орехов С.В., Черенков И.А. Обоснование прогнозирования цен полимеров посредством новостного потока. // Восточно-европейский журнал передовых технологий – 2010. - № 5/7(47). – С. 18-20.