

Ольга Ярославівна КОВАЛЬЧУК

кандидат фізико-математичних наук,
доцент,
кафедра економіко-математичних методів,
Тернопільський національний економічний університет
E-mail: olhakov@gmail.com

Тарас Юрійович ГАЙДА

кандидат економічних наук,
старший викладач,
кафедра міжнародних економічних відносин,
Тернопільський національний економічний університет
E-mail: taras.haida@gmail.com

Сергій Ярославович ЖОНЦА

Тернопільський національний економічний університет
E-mail: proster.ternopil@gmail.com

ТЕХНОЛОГІЇ BIG DATA В ІННОВАЦІЙНОМУ МАРКЕТИНГУ

Ковальчук, О. Я. Технології Big Data в інноваційному маркетингу [Текст] / Ольга Ярославівна Ковальчук, Тарас Юрійович Гайда, Сергій Ярославович Жонца // Український журнал прикладної економіки. – 2018. – Том 3. – № 1. – С. 36-52. – ISSN 2415-8453.

Анотація

Вступ. У сучасному суспільстві інформація та наукові знання є вирішальними факторами розвитку. Новітні інформаційні технології відкривають кардинально нові можливості та способи обробки інформаційних масивів, зокрема у суспільній, економічній та політичній сферах. Інформаційні технології Big Data змінюють підходи до аналізу інформації та способи прийняття рішень. Цей тренд впливає на соціум та політику у міжнародних масштабах.

Мета. Метою дослідження є вивчення можливостей застосування новітніх технологій Big Data та Data Mining для аналізу великих інформаційних масивів з віртуальних мереж з метою прихованого психологічного впливу на процес прийняття рішень цільовою аудиторією

Метод. При проведенні досліджень використано аналітичний та порівняльний методи, контент-аналіз, методи інтелектуального аналізу текстів, асоціативні правила, графічний метод, комп'ютерне моделювання

Результати. У статті проведено дослідження можливостей застосування новітніх технологій Big Data для обробки даних у сфері маркетингу та суспільних комунікацій. Розглянуто нові методики консолідації та аналізу великих масивів інформації з всесвітньої мережі Інтернет для формування індивідуальних профілів користувачів віртуальних мереж з метою вивчення поведінки спільнот та прихованого впливу на прийняття несвідомих рішень цільовою аудиторією. Проаналізовано особливості

використання моделей *Data Mining* у політичному маркетингу, а саме: систем штучного інтелекту розпізнавання образів та інтелектуального аналізу текстів. Проведено дослідження ефективності застосування технологій *Big Data*, використаних у 2016 році для проведення агітаційної *digital*-кампанії претендента на пост Президента США Дональда Трампа. Виконано контент-аналіз 1869 його останніх передвиборчих повідомлень, які він надіслав своєму електорату через соціальну мережу мікроблогів *Twitter*. Виявлено основні асоціативні правила у текстах передвиборчих меседжів Трампа та основне гасло, за допомогою якого він «продав» себе Америці. Побудовано модель *Data Mining*, яку надалі можна використати для аналізу великих масивів неструктурованих текстів та прийняття ефективних маркетингових рішень, зокрема у сфері проведення рекламних та агітаційних кампаній.

Ключові слова: *Big Data*; політичний маркетинг; сервітивне суспільство; мікротаргетинг; інтелектуальний аналіз; контент-аналіз; асоціативні правила.

Olha Yaroslavivna KOVALCHUK

PhD in Physics and Mathematics,
Associate Professor,
Department of Economic and Mathematical Methods,
Ternopil National Economic University

Taras Yuriyovych HAIDA

PhD in Economics,
Senior Lecturer,
Department of International Economic Relations,
Ternopil National Economic University

Serhiy Yaroslavovych ZHONTSA

Ternopil National Economic University

BIG DATA TECHNOLOGIES IN THE INNOVATIVE MARKETING

Abstract

In the article a study of possibilities of application of the newest Big Data technologies is undertaken for processing of data in marketing and public communications field. New methodologies of consolidation and treatment of large arrays of information from the Internet for forming of individual profiles of virtual networks users in order to analyse associations' behaviour and hidden influence on the acceptance of irresponsible decisions by an association are examined. The features of the use of Data Mining models in the political marketing are analysed. They are the intelligence systems for recognition of patterns and intellectual analysis of texts. A study of effective application of Big Data technologies, used in 2016 for realization of agitation digital campaign of pretender on the post of President of the USA, Donald Trump, is undertaken. It is conducted a content-analysis of 1869 latest pre-election Trump's reports, that he sent to the electorate through the Twitter, social micro-blogs network. Basic association rules are determined in texts of his pre-election messages and basic slogan by means of that he "sold" himself to America. Data Mining model is built. It can be used in future for the study of large arrays of unstructured texts and acceptance of effective marketing decisions, in particular in the field of realization of effective advertisement and agitation campaigns.

Keywords: *Big Data; political marketing; servitive society; micro-targeting; Data Mining; content analysis; association rules.*

JEL classification: C55, C53, C63, M31

Вступ

Сучасне суспільство живе в еру цифрових технологій. Інтернет та соціальні мережі зробили світ більш відкритим, а життя більш зручним. Однак є й інший, непередбачуваний, бік новітніх технологій. Зростання кількості користувачів Інтернету та сучасні методи аналізу поведінки в мережі дають можливість з приголомшливою ефективністю впливати на соціум, маніпулювати як окремими особами, так і населенням цілих країн та континентів.

Більшість суспільних комунікацій перейшли у віртуальний простір. Усі найважливіші розмови та «події» відбуваються у мережах і чатах, на форумах та у віртуальних спільнотах. Разом зі стрімким накопиченням інформації швидкими темпами розвиваються і технології аналізу даних.

Особливої актуальності сьогодні набувають технології Big Data (великі дані), які кардинально змінюють підходи до аналізу інформації та способи прийняття рішень у всіх сферах людської діяльності, зокрема в економічній та політичній. Сьогодні можна сміливо говорити не лише про цифрову економіку, а й про цифрову політику.

Експерти прогнозують, що в поточному десятилітті ця область буде прогресувати, генеруючи нові питання, зокрема, як вирішити проблеми безпеки особистих даних, хто саме є суб'єктом розпорядження цими даними, наскільки припустимі окремі способи застосування отриманих даних тощо. На технології роботи з великими даними у світі витрачають мільярди доларів. У цьому секторі створюють мільйони робочих місць. Тільки у США для ефективної обробки великих даних найближчим часом буде потрібно 140-190 тис. аналітиків і понад 1,5 млн менеджерів для управління інформаційними масивами [1].

Нові технології Big Data знаходять застосування в найрізноманітніших галузях: від пошукових сервісів, перекладів та навігації до використання закономірностей, виявлених в Інтернет-середовищі, роздрібних офлайн-продажах і навіть у політичній рекламі.

Питання практичного застосування технологій Big Data на сучасному етапі розглядають здебільшого західні фахівці. Серед них Братосін О. К. (Bratosin O. C.), яка вивчала концепцію великих даних, критерії їх класифікації, архітектуру, інформаційну цінність для бізнесу та вплив Big Data на світові процеси [2]. Бай Ю., Фань Д. і Цай Р. (Bai Ju., Fan Ji., Tsay R.) розглядали великі дані як революційну технологію інновацій, конкуренції та продуктивності економіки, новий ресурс для бізнесу [3]. Титіріска А. (Titirisca A.) вивчала можливості впровадження інноваційних технологій Big Data для розробки бізнес-стратегії, що забезпечить отримання комерційних переваг [4]. Бейл К. А. (Bail C. A.) проводив аналіз методів несвідомого впливу на колективну поведінку учасників неорганізованих груп на сайтах соціальних мереж Facebook і Twitter [5]. Хуст Я. (Worst J.) досліджував питання достовірності, інтеграції, масштабованості та конфіденційності «відкритих» структурованих та неструктурованих даних з соціальних мереж Google, Twitter, Facebook тощо та застосовності інструментів математики та аналітики для їх аналізу [6].

Більшість сучасних компаній збирає, записує й аналізує дані. Проте багато з них стикаються з цілою низкою складних ділових та інформаційно-технологічних проблем, які виникають у зв'язку з необхідністю обробляти величезні масиви даних [1; 2].

ISSN 2415-8453. Український журнал прикладної економіки. 2018 рік. Том 3. № 1.

Інформаційні технології Big Data змінюють підходи до аналізу інформації та способів прийняття рішень і можуть підвищити конкурентоспроможність організації. Провідні світові компанії вже сьогодні отримують реальні стратегічні переваги від накопиченої інформації.

Існує безліч методів аналізу великих масивів даних, в основі яких лежить інструментарій, запозичений із статистики та інформатики (наприклад, машинне навчання). Однак дослідження у сфері створення нових методик обробки значних обсягів інформації та вдосконалення наявних будуть вкрай актуальними ще не одне десятиліття. Крім того, деякі з них застосовні не лише винятково до Big Data. Вони можуть з успіхом використовуватися і для менших за обсягом масивів.

Технології великих даних відкривають широкі можливості не лише для великого бізнесу, а й для політики. Сьогодні тільки розпочинається розробка нових ефективних методик використання інструментів Big Data для підсвідомого впливу на цільову аудиторію, з метою управляти поведінкою, вчинками та навіть вибором спільноти. У багатьох галузях роблять лише перші спроби знайти способи «видобування» та аналізу неявних даних.

Мета статті

Метою статті є дослідження можливостей застосування новітніх технологій Big Data та Data Mining для аналізу великих інформаційних масивів з віртуальних мереж з метою прихованого психологічного впливу на процес прийняття рішень цільовою аудиторією.

Виклад основного матеріалу дослідження

90 % даних у світі було створено протягом останніх кількох років. Згідно з дослідженнями Word Economic Forum, сучасний типовий мешканець західного світу за один день піддається впливу даних, обсяг яких еквівалентний обсягу інформації, яку отримувала за все своє життя пересічна людина у XV ст. У широкому сенсі сучасне суспільство можна визначити як розумне (smart), або сервітивне, в якому цифрові технології широко використовують на всіх рівнях управління. Передбачається, що ринок смарт-технологій до 2020 року становитиме 1,6, а до 2026 року – 3,5 трильйонів доларів [7].

Сьогодні великі обсяги інформації потребують опрацювання майже у всіх галузях. Їх використання відкриває нові можливості для різних сегментів споживачів. Автоматизація обробки великих даних – це друга технологічна революція після механізації процесів. Терміном Big Data позначають набори даних, розмір яких перевершує можливості типових баз даних (БД) для введення, зберігання, обробки та аналізу інформації [1].

Проте великі дані означають дещо більше, ніж просто аналіз великих інформаційних масивів. Проблема не в тому, що сучасні ТНК, суспільні та державні організації обробляють значні обсяги даних, а в тому, що більша їх частина надана у форматі, що не відповідає традиційному структурованому формату БД. Це веб-журнали, відеозаписи, текстові документи, машинний код або, наприклад, геопросторові дані. Усе це зберігається в безлічі різноманітних сховищ, іноді навіть за межами установи. У результаті організації можуть мати доступ до великого обсягу власних даних і не мати необхідних інструментів, щоб встановити взаємозв'язки між цими даними та сформулювати на їх основі вагомні висновки. Крім того, інформація постійно оновлюється і за допомогою традиційних методів аналізу даних вже неможливо ефективно обробляти великі інформаційні масиви.

По суті, поняття Big Data передбачає роботу з даними великих обсягів і різноманітного складу, які отримують з різних джерел, з метою з підвищення ефективності прийняття управлінських рішень та підвищення конкурентоспроможності. Це поняття об'єднує техніки і технології, які «витягують» сенс з даних на екстремальній межі практичності.

У сучасному світі існує безліч джерел великих даних. Ними можуть виступати дані, що безперервно надходять з вимірювальних пристроїв, потоки повідомлень із соціальних мереж, метеорологічні дані, дані супутникового дистанційного спостереження, потоки інформації про місце знаходження абонентів мереж стільникового зв'язку, пристроїв аудіо- та відео реєстрації тощо. Масове поширення перерахованих вище технологій і принципово нових моделей використання різного роду пристроїв та інтернет-сервісів послужило відправною точкою для проникнення великих даних майже в усі сфери діяльності людини. Насамперед – у науково-дослідну діяльність, комерційний сектор і державне управління. Сучасний «цифровий всесвіт» (Digital Universe) містить як цифрові зображення і відео, завантажені з мобільних телефонів, наприклад, на You Tube, так і HD відео, передане по мережах провайдерів. Це як корпоративні дані, що генеруються бізнес-додатками, так і дані, які створює великий андронний коллайдер. Значну частку даних, вироблених до 2020 р., згенерують не люди, а машини в ході взаємодії між собою та іншими мережами даних. Сюди належать, наприклад, сенсори та інтелектуальні пристрої, які можуть обмінюватись потоками інформації з іншими девайсами.

Важко знайти галузь, для якої проблематика великих даних не була б актуальною. Вміння оперувати великими обсягами даних, аналізувати взаємозв'язки між ними і ухвалювати зважені рішення, з одного боку, надає потенціал для компаній з різних вертикалей для збільшення показників прибутковості і підвищення ефективності. З іншого боку, це чудова можливість для додаткового заробітку партнерів вендорів-інтеграторів і консультантів.

Попри на малий термін існування сектору Big Data, на сьогодні відомі оцінки ефективного використання цих технологій, засновані на реальних прикладах. Один з найвищих показників отримано в енергетиці – за оцінками аналітиків, аналітичні технології Big Data здатні на 99 % підвищити точність розподілу потужностей генераторів, а охорона здоров'я США завдяки Big Data може заощадити до \$ 300 млрд. [7].

Інтернет став першою індустрією, що оперує Big Data. Завдяки цій технології вся інформація, поширена у глобальній мережі, та «події», що відбуваються офлайн (offline), залишають цифровий відбиток. Зберігається все – відомості про операції через електронні платіжні системи, дані про запити у пошуковику Google, лайки (likes) у соціальних мережах, розмови у чатах, дзвінки зі Skype та Viber, звичайна прогулянка зі смартфоном тощо.

Абсолютно все, що робить користувач онлайн (on-line), нікуди не зникає. Кожен лайк чи замовлення товару залишає цифровий слід у мережі. Навіть якщо користувач нічого не замовив, а лише переглянув інформацію, кілька наступних днів на різних сайтах, які він буде відвідувати, відобразатимуться рекламні повідомлення з пропозицією придбати річ, якою він цікавився. Такий прийом називається контекстною рекламою. Ще більше інформації про себе, самі того не підозрюючи, користувачі залишають у соціальних мережах. Кожен поставлений лайк – вкрай цінна інформація, на основі якої можна створити точний психологічний портрет конкретної людини. Аналіз кількох десятків чи сотень таких лайків дає можливість отримати інформацію про прибутки користувача, його сімейний стан, релігійні переконання,

ISSN 2415-8453. Український журнал прикладної економіки. 2018 рік. Том 3. № 1.

емоційну стабільність, сексуальну орієнтацію, колір шкіри та політичні погляди. Це терабайти інформації, і не відомо, хто і як може її використати.

Уся зібрана інформація використовується для створення індивідуальних профілів потенційних покупців з метою адресної розсилки відповідної реклами, відслідковування можливих загроз державній безпеці, проведення виборчих кампаній та розробки новітніх технологій маніпулювання суспільною свідомістю. Big Data стали для людства значним досягненням та великою небезпекою.

Технології персоналізованої реклами в мережі Facebook здійснили суттєвий вплив і на результати останніх виборів у США та референдум про вихід Великобританії з ЄС. Про це свідчать результати дослідження швейцарського журналу *Das Magazin* [8].

При використанні технологій Big Data для соціальних досліджень основною проблемою є збір великих обсягів особистої інформації. Однак, з появою Інтернет, а особливо Facebook, Instagram, ВКонтакте, Однокласники та інших соціальних мереж, збір даних відбувається автоматично – користувачі самі роблять загальнодоступними відомості про себе.

Ці ж технології використовують сьогодні і для збору індивідуальних даних. Процес дуже простий. Спочатку користувач отримує перелік питань (онлайн-тест). З відповідей на нього дослідники виявляють особисті цінності піддослідного та вивчають його дії – лайки та репости (reposts) у Facebook, стать, вік та місце проживання. Так формують зв'язки. З простого аналізу даних у мережі можна отримати нетривіальні висновки, зокрема сексуальну орієнтацію користувача, тип характеру, ставлення до політичних рухів, схильність до насильства тощо.

Аналіз 68 лайків у Facebook дає можливість з 95 % ймовірністю встановити колір шкіри піддослідного, з 88 % ймовірністю передбачити його гомосексуальність, з 85 % ймовірністю виявити прихильність до одної з політичних партій [8]. Сучасні технології Big Data здатні визначити інтелектуальний розвиток людини, ставлення до релігії, пристрасть до алкоголю, куріння та наркоманії. За допомогою вдосконалених моделей розробники можуть отримати інформацію про розлучення батьків до повноліття піддослідного. Можливо навіть передбачити відповіді користувача на окремі питання.

Для створення індивідуального психологічного профілю користувача використовують не лише його пости та лайки у Facebook. Аналізують фотокартки у соцмережах, кількість друзів, інтенсивність змін місця розташування, час знаходження у мережі та інші особисті дані, надані їх власником несвідомо. Підключення до мережі відкриває доступ до особистих даних користувача. Наприклад, датчик руху у смартфоні показує, чи розмахує він рукою, як далеко пересувається (ці дії корелюють з емоційною нестабільністю). Смартфон є своєрідним безкоштовним опитувальником, який мимоволі наповнює його власник. За його допомогою можна не лише сформулювати з отриманих даних психологічний портрет, а й відбирати потрібні серед отриманих профілів. Це своєрідна пошукова система по людях, здатна виявити довірливих, схильних до спонтанних дій, злочинних намірів тощо.

Так розробками у сфері прикладного застосування технологій Big Data на основі даних із комп'ютерних мереж займається британська приватна компанія поведінкових досліджень і стратегічних комунікацій «Лабораторія стратегічних комунікацій» SCL (Strategic Communication Laboratories). Вона є глобальною організацією по управлінню передвиборчими кампаніями і спеціалізується на маркетингу, що базується на психології та логіці. У Сполучених Штатах компанія відома як Cambridge Analytica та проводить інтелектуальний аналіз даних і аналіз даних аудиторії користувачів. На основі отриманих результатів розробляються комунікації, конкретно спрямовані на ключові групи аудиторії. Метою є зміна поведінки спільноти відповідно до цілей

замовника та визначення ймовірності отримати небажаний результат. Компанія позиціонує себе як «глобальне агентство з управління виборами». Її основним завданням є вплив на результати виборів [8].

Strategic Communication Laboratories розробила складну корпоративну систему на засадах «податкових гаваней». Вона була використана, зокрема, для організації криз у малорозвинених країнах світу та розробки НАТО методів психологічної маніпуляції жителями Афганістану. У склад SCL входить невелика компанія Cambridge Analytica, яка організовувала цифрові кампанії на підтримку виходу Великобританії з ЄС та виборів Трампа.

З 2015 року ключовою компетенцією цієї фірми є новий тип політичного маркетингу – «мікротаргетинг» (інформація про виборця поєднується з інформацією споживача для моделювання поведінки), який базується на «методі океану» [8]. Міжнародні політтехнологи SCL використовують методологію Big Data для соціальних мереж у великій політиці. Політологи помітили схожість програм Трампа та прихильників Brexit. Встановили і зв'язок новообраного президента США з маловідомою компанією Cambridge Analytica. Сам Трамп ще до оголошення результатів виборів необережно написав у Twitter, що скоро його будуть називати «Містер Brexit».

Вибори президента США 2016 року стали яскравим прикладом використання Big Data для здійснення впливу на несвідомий вибір електорату. До кінця 2014 року передвиборчі кампанії проводились здебільшого на основі аналізу демографічних критеріїв, таких, як стать, раса, соціальний статус тощо. Такий спосіб використала для проведення своєї передвиборчої кампанії і Хіларі Клінтон. Її команда поділила всю спільноту на формально однорідні групи, за приналежністю до яких кожному виборцю були надіслані однакові меседжі. Соціологи визначили, за якими критеріями потрібно виділяти групи, та до останнього передбачали перемогу Клінтон на виборах президента США.

Успіх маркетингових стратегій Cambridge Analytica базується на аналізі психологічної поведінки людини, який використовує «модель океану», та сегментації (оголошення можуть бачити тільки користувачі обраної цільової аудиторії). Це означає персоналізацію реклами та якнайточніше підлаштування рекламного меседжа під тип характеру кожного окремого споживача. Cambridge Analytica постійно купує персональні дані з різноманітних цифрових джерел та проводить консолідацію зібраної інформації зі списками зареєстрованих прихильників Республіканської партії і даними з лайків та репостів у соціальних мережах. Так створюють індивідуальний профіль за «методом океану». Процедура використовує модель на основі IQ-тестів та інших невеликих додатків, щоб отримувати осмислені лайки від користувачів Facebook. Таким чином компанія Cambridge Analytica отримала психограми всіх повнолітніх американців. Це 220 млн чоловік. І кожному з них було надіслано окремий меседж, який здобув прихильність кожного конкретного виборця.

Суперечлива натура Трампа, його безпринципність, які спричинили низку різних повідомлень про нього, випадково надали йому перевагу: кожен окремий виборець отримав свій меседж. На третій день дебатів між Трампом і Клінтон команда Трампа відправила в соцмережі (переважно, Facebook) понад 175 тис. різних варіацій меседжів. Відмінність між цими повідомленнями полягала лише у найдрібніших деталях (різні заголовки і підзаголовки, фон повідомлення, наявність фото чи відеоінформації). Однак усі повідомлення були максимально точно підлаштовані до психотипу кожного конкретного адресата. Така деталізація забезпечила отримання бажаних відгуків на повідомлення навіть у найдрібніших груп населення [8].

Наприклад, у кварталі Маленький Гаїті в Майамі була запущена інформація про відмову Фонду Клінтон брати участь у ліквідації наслідків землетрусу в Гаїті, щоб переконати жителів не віддавати свої голоси за Хілларі. Ще однією метою було утримати електорат Клінтон (зокрема, лівих, що сумніваються, афроамериканців і молодих дівчат) від голосування, «придушити» їх вибір. Були використані і так звані «темні пости» Facebook – платні оголошення посеред стрічки новин, які могли бачити тільки окремі групи осіб. Так афроамериканцям показували пости з відео, на якому Клінтон порівнювала чорношкірих чоловіків з хижакми. Революційний підхід до комунікацій, заснованих на даних, зробив суттєвий внесок у перемогу Дональда Трампа. Хілларі Клінтон стала однією з жертв антиреклами Cambridge Analytica. Важко сказати, якою мірою американське суспільство сьогодні обробляється фахівцями Трампа, адже вони найчастіше використовують соціальні мережі і цифрове ТБ, а не центральне телебачення.

Використання Big Data допомогло «продати» Трампа Америці. Він не придумав нічого кардинально нового, а лише використав технології, які давно застосовують у рекламі. Інші політики також займалися фактично тою ж рекламою, однак, на відміну від бізнесу, використовували технології комунікацій та аналітики 70-х років минулого століття. Успіх реклами Трампа полягає саме в сегментації. Його команда змогла правильно розділити аудиторію на групи, зрозуміти їх бажання і надсилати кожному саме ті повідомлення, які він хотів бачити і чути. Республіканці об'єднали розрізнені дані з різних джерел та застосували до них класичні методи сегментації і таргетування на основі технологій Big Data. Хілларі Клінтон, навпаки, будувала свою передвиборчу кампанію, спираючись на соціологічні дані і стандартний маркетинг, розділивши суспільство лише на формально гомогенні групи (жінки, чоловіки, афроамериканці, латиноамериканці, багаті, бідні тощо).

Окрім «цифрової агітації», кампанія Трампа проводила і традиційні заходи. З липня 2016 року волонтери кампанії Трампа отримали програмне забезпечення, яке містило інформацію про політичні вподобання та особистісні типи жителів кожного будинку. За її допомогою волонтери-агітатори модифікували свою розмову з конкретним виборцем на основі аналізу цих даних. Інформацію про реакцію виборців вони надсилали для подальшої обробки в аналітичний центр компанії Cambridge Analytica.

Аналітики цієї компанії виділили серед американських громадян 32 психотипи, сконцентрувавшись лише на 17 штатах [8]. Вони довели, що шанувальники американських автомобілів достовірно є потенційними прибічниками Трампа. Окрім того, схожі дослідження дали змогу самому Трампу зрозуміти, де і які послання найкраще використовувати. На основі отриманих результатів аналізу передвиборчий штаб прийняв рішення, що доцільніше в останні тижні сконцентруватися на штатах Мічиган і Вісконсин. Виборча кампанія Трампа стала моделлю для застосування аналітичної системи, розробленої Cambridge Analytica.

Важко робити остаточні висновки про вагу впливу психометрії на результат виборів президента США. Cambridge Analytica не розголошує докази успішності своєї кампанії. Однак очевидним є той факт, що завдяки підтримці Cambridge Analytica спостерігалось зростання голосів сільських жителів та скорочення електоральної активності афроамериканців. Навіть той факт, що Трамп витратив на проект лише \$ 15 млн, свідчить про ефективність персоналізованого просування. Три чверті рекламного бюджету своєї передвиборної кампанії він вклав у цифрову сферу [8]. На виборах президента США 2016 року справдились лише прогнози тих статистиків та соціологів, які застосовували новітні методи аналізу великих даних. І Трамп, який за кожної

нагоди критикував «цифрові» технології, отримав перемогу лише завдяки їх цільовому застосуванню.

Науковий аналіз, проведений вченими англійського Кембриджа, підтвердив, що психологічне таргетування, подібно до того, яке використовували в Cambridge Analytica, підвищує кількість кліків (clicks) на рекламу в Facebook на 60 %. Імовірність того, що після перегляду персоналізованої реклами люди перейдуть до дій (куплять розрекламовану річ або проголосують за потрібного кандидата) зростає у 15 разів [8].

У минулому громадську думку висловлювали та аналізували журналісти, соціологи й експерти через засоби масової інформації. Сьогодні це не так. Більшість громадських розмов перейшли в on-line сервіси соціальних медіа та інших цифрових платформ. Визначну роль у формуванні громадської думки відіграють форуми та чати мережі Інтернет.

Досягнення в області Big Data і штучного інтелекту сьогодні дають можливість автоматично класифікувати та аналізувати думки цифрового електорату, пов'язані з виборчими кампаніями. Проблема полягає в тому, що більшість новітніх технологій у всьому світі перебувають під фінансовим тиском і не мають коштів для інвестицій у розвиток Big Data та штучного інтелекту. Окрім того, жодна з них не має ресурсів, щоб конкурувати з технологічними гігантами на зразок Google, Facebook, Amazon та Baidu [7].

Electome, проект Лабораторії соціальних машин у MIT Media Lab, має можливість усунути цю проблему. Більшість її організацій у режимі реального часу вже сьогодні надають аналітику громадської думки, пов'язаної з президентськими виборами в США. Electome – це перша версія машини, призначена для аналізу виборчих кампаній. Вона отримує дані з публічних розмов через традиційні засоби масової інформації, соціальні медіа та цифрові платформи. Особлива увага звертається на громадські розмови на Twitter, що відіграли ключову роль у виборах президента США, оскільки потенційний майбутній президент Трамп постійно спілкувався саме у цій соціальній мережі.

Electome автоматично класифікує розмови з обговорюваних питань і сегментує кожне актуальне питання за демографічними ознаками, тоном розмови та іншими ключовими факторами. Ця машина дає можливість глибше зрозуміти обговорюване питання. Electome здатна будувати якісні прогностичні моделі на основі аналізу інформації з соціальних мереж та інформувати суспільство. Це соціальна комп'ютерна система, яка робить світ кращим через соціально-машинні взаємодії. Доступ до даних в Electome надається сотням журналістів. Тут публікують важливі для суспільства статті. Дані також використовує Комісія з президентських дебатів для аналізу питань, які ставлять кандидатам у президенти США та при обговоренні відповідей на ці питання [9]. Щоб виборці були інформованими, а голос кожного виборця був почутим, потрібно організувати цифрову публічну спільноту, доступну всім користувачам Інтернет.

Facebook та інші соціальні мережі вважають символом нового покоління, епохи без фізичних кордонів. Однак цю потужну пошукову систему по людях можна використати і у злочинних цілях. Технології Big Data для соціальних мереж можуть нести загрозу благополуччю, свободі і навіть життю людей. Тому важливим питанням на сьогодні є дослідження у сфері аналізу контенту у всесвітній павутині засобами новітніх інформаційних технологій, зокрема інтелектуального аналізу даних.

Інструменти Data Mining слугують для виявлення асоціативних зв'язків у масивах даних великих обсягів. Інтелектуальний аналіз даних (Data Mining), або розвідка даних – це процес, метою якого є виявлення у великих інформаційних масивах нових кореляцій (correlations), тенденцій (trends), шаблонів (patterns), зв'язків

(relations) і категорій за допомогою використання засобів розпізнавання зразків (recognition of samples) та інструментів статистики і математики [10].

При проведенні Data Mining виконують багаторазові перетворення та процедури над «неочищеними» даними для знаходження структур, які інтуїтивно зрозумілі людині та краще розкривають суть досліджуваних процесів; для побудови моделей передбачення результатів або характеристик реальних ситуацій на основі аналізу історичних або суб'єктивних даних. Сутність і мета Data Mining – пошук у великих інформаційних масивах прихованих, об'єктивних закономірностей, які мають практичну цінність.

Класична статистика використовує усереднені характеристики аналізованого масиву даних, які часто неточно відображають реальну оцінку явищ та процесів. Тому її методи ефективні здебільшого при перевірці побудованих гіпотез, тоді як формулювання гіпотези в окремих випадках виявляється складним і трудомістким завданням. Сучасні технології Data Mining дають можливість аналізувати інформацію, використовуючи автоматичний пошук шаблонів, притаманних окремим фрагментам неоднорідних багатовимірних даних.

Для знаходження прихованих залежностей та взаємозв'язків у кортежах (наборах) даних великих обсягів застосовують асоціативні правила (association rules). Такий аналіз базується на побудові асоціативних правил зв'язку між спостережуваними процесами [10]. Асоціативні зв'язки – це логічні залежності виду «якщо сталася подія А, то з відповідною ймовірністю відбудеться і подія В». Асоціативні правила дають можливість виявляти закономірності між пов'язаними між собою подіями. Задача генерування асоціативних правил початково була сформульована для пошуку типових шаблонів покупок, зроблених у супермаркетах. Її інша назва – аналіз ринкового кошика (market basket analysis) [11].

Інформація, отримана у результаті побудови асоціативних правил, може бути використана для прийняття рішень щодо маркетингової діяльності, наприклад, при оцінюванні ефективності рекламних кампаній або при виборі місць розташування товарів. Окрім вищезгаданого прикладу з аналізу ринкових кошиків, правила асоціації застосовують сьогодні у багатьох прикладних областях, разом з аналізом web-ресурсів, виявлення несанкціонованих вторгнень у комп'ютерних мережах, підтримку прийняття рішень у сфері діагностики та телекомунікацій, у біоінформатиці тощо. Сфера застосування Data Mining не обмежена.

Одним із яскравих прикладів використання методів інтелектуального аналізу даних для вивчення цифрових масивів великих обсягів у соціальних мережах є нейронна мережа для розпізнавання образів, розроблена з метою передбачення майбутнього президента США. Ця модель для аналізу фотографій знаходить топ-5 найбільш схожих до аналізованого зображення осіб або об'єктів і визначає, що зображено на фото. Навчання мережі проводили на колекції з 14 мільйонів класифікованих зображень, кожне з яких було віднесене до однієї з 21 тисячі категорій (наприклад, професія, посада тощо). Для навчання було використано більше тисячі фотографій різних керівників держав. Мережа «навчилася», як повинен виглядати президент: характерні риси обличчя, одяг, аксесуари, фонові картини тощо. Точність результатів прогнозування у топ 5 – 70 %. Застосована нейронна мережа проаналізувала фотографії кандидатів у президенти США і визначила Дональда Трампа як майбутнього главу держави [7].

Розпізнавання образів на фотографіях на основі нейронних мереж є популярним та ефективним інструментом. Успішно застосовується в Інтернет-компаніях, медицині, економіці тощо. Однак ніхто не пробував передбачити результат президентських

виборів за допомогою цієї технології. Нейронна мережа є математичною моделлю, побудованою на принципі організації нервових клітин головного мозку. Такі моделі можуть складатися з мільйонів нейронів. У процесі навчання мережа, що складається з безлічі з'єднаних між собою шарів нейронів, знаходить складні залежності між їхніми входами і виходами, може працювати з неповною або спотвореною інформацією [11]. Сучасні нейронні мережі широко використовують для обробки зображень як в наукових дослідженнях, так і для ідентифікації зображень, розпізнавання образів, передачі стилю картини (Artisto, Vinci, Prisma).

Data Mining – сукупність великої кількості різноманітних методів виявлення знань. Вибір методу часто залежить від типу наявних даних і характеру результатної інформації. У сучасному світі близько 80 % всієї інформації створюється та зберігається у текстовому форматі. У таких даних зазвичай прихована значна частка інформації, оскільки неструктурованість текстів не дає можливості повною мірою застосовувати алгоритми Data Mining. Для вирішення цієї проблеми використовують методи інтелектуального аналізу неструктурованих текстів. Text Mining, або Text Analytics, – це пошук нових чи наявних фактів шляхом застосування методів обробки природної мови та статистичного навчання [10]. Методи інтелектуального аналізу неструктурованих текстів виникли та розвиваються на перетині декількох прикладних областей: інтелектуальний аналіз даних, обробка природних мов, пошук інформації, витягування інформації й управління знаннями. Виявлення знань у тексті – нетривіальний процес пошуку справді нових, потенційно корисних і зрозумілих шаблонів у неструктурованих текстових даних.

Контент-аналіз передвиборчих меседжів кандидата у президенти США Дональда Трампа, надісланих через соціальну мережу Twitter.

За допомогою інструментів Text Mining проведено аналіз останніх 1869 передвиборчих повідомлень Дональда Трампа, які він надіслав своїм потенційним виборцям через соціальну мережу мікроблогів Twitter [12]. При проведенні досліджень використано безкоштовні (free) версії програмного пакету для статистичного аналізу й управління даними Statistica 10 та відкритої програмної платформи для інтелектуального аналізу даних RapidMiner 8.0.

Виявлено 10 основних предикторів в аналізованих повідомленнях (рис. 1). Словами, які найчастіше використовував Трамп у своїх повідомленнях, були Клінтон, Обама, підтримка (support), ЗМІ – засоби масової інформації (media), безпечний (safe), демократ (democrat), судження (judgment), небезпека (danger), Америка.

Для генерування асоціативних зв'язків, які можуть існувати між словами в аналізованих повідомленнях, до неструктурованих текстових даних послідовно застосовано ряд процесів та операторів (рис. 2).

Process Documents from Files – генерування векторів слів з текстової колекції, що зберігається у різних файлах, за допомогою наступних операторів (рис. 3);

Tokenize – розбиття тексту документа на послідовність токенів (лексем);

Transform Cases – приведення всіх символів у текстовому документі до одного регістру;

Filter Stopwords – видалення з документа стоп-слів (неінформативних);

Filter Tokens (by Length) – фільтрування токенів відповідно до їхньої довжини (кількості символів, які вони містять).

- Numerical Binominal – перетворення числових атрибутів на біноміальні.
- FP-Growth – визначає всі часто використовувані об'єкти.
- Create Associations Rules – генерування кортежу асоціативних правил з аналізованої послідовності частих кортежів.

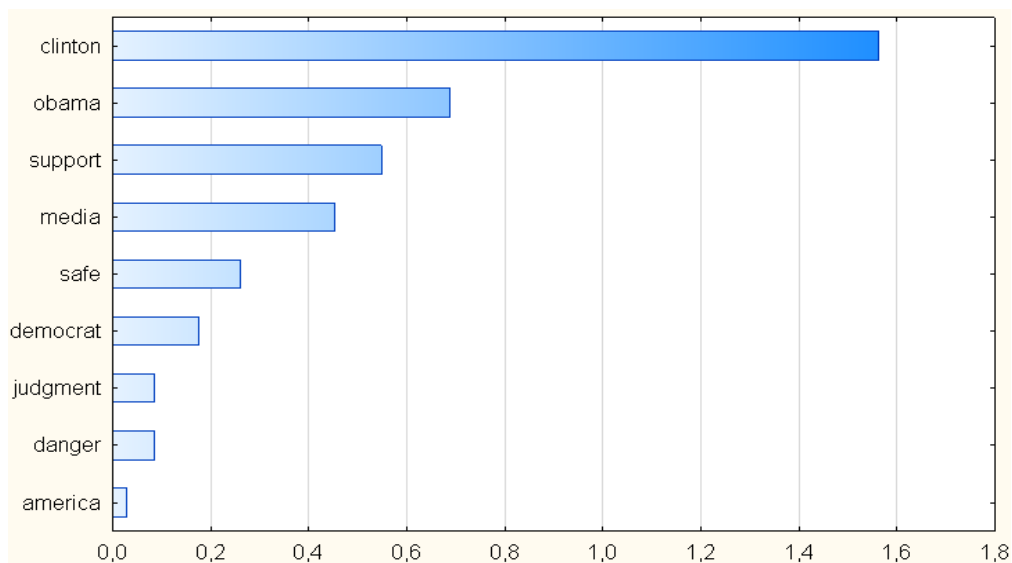


Рис. 1. Основні предиктори передвиборчих меседжів

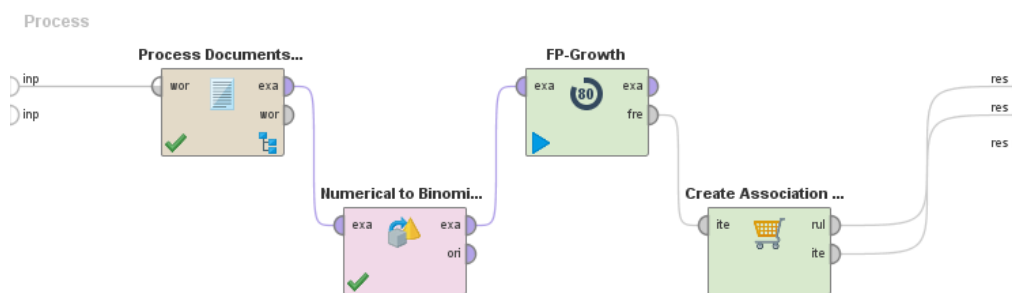


Рис. 2. Оператори, застосовані для генерування асоціативних правил

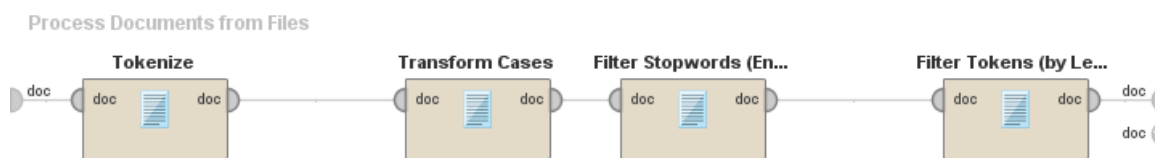


Рис. 3. Оператори процесу Documents from Files

Асоціативні правила є структурами «якщо, то» (if-then). Для кожної з них обчислюються такі оцінки (рис. 4) [10]:

Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
america	make	0.378	0.839	0.950	-0.523	0.203	2.159	3.805
great	make	0.365	0.874	0.963	-0.470	0.203	2.249	4.889
great	america, make	0.365	0.874	0.963	-0.470	0.207	2.314	4.955
great	america	0.371	0.887	0.967	-0.465	0.183	1.971	4.884
make	great	0.365	0.940	0.983	-0.412	0.203	2.249	9.630
make	america, great	0.365	0.940	0.983	-0.412	0.221	2.534	10.408
america, make	great	0.365	0.967	0.991	-0.391	0.207	2.314	17.385
make	america	0.378	0.972	0.992	-0.400	0.203	2.159	19.699
america, great	make	0.365	0.985	0.996	-0.376	0.221	2.534	41.766
great, make	america	0.365	1	1	-0.365	0.201	2.221	∞

Рис. 4. Загальні оцінки моделі для генерації асоціативних правил (фрагмент отриманих результатів)

Підтримка (*Support (F)*) – відношення кількості аналізованих неструктурованих текстових повідомлень, які містять набір (itemset), або кортеж, (слово/словосполучення) *F*, до загальної кількості досліджуваних меседжів. Кортеж називають частим (large itemset), якщо значення його підтримки більше за мінімальне підтримку, задане користувачем: $Support(F) > Support_{min}$. При пошуку асоціативних правил потрібно знайти множину всіх частих кортежів:

$$L = \{Support(X) > Support_{min}\}. \quad (1)$$

Ймовірність (Confidence) – це умовна ймовірність того, що у меседжі буде міститись набір *X* (Premises), якщо у цьому текстовому документі присутній кортеж *Y* (Conclusion), виражена як $\{X \rightarrow Y\}$:

$$Confidence\{X \rightarrow Y\} = \frac{Support(X, Y)}{Support(X)}. \quad (2)$$

Ймовірність Лапласа (LaPlace). Отримані оцінки побудованої нами моделі для генерування асоціативних правил мають високу ймовірність ($> 0,95$).

Корисність (приріст) інформації (Information gain). Зазвичай аналізовані інформаційні масиви містять багато атрибутів, значна кількість з яких можуть бути несуттєвими або навіть надмірними для конкретного дослідження. У теорії інформації, започаткованій Клодом Шенноном [13], кількість інформації вимірюється кількістю бітів (0 та 1), необхідних для зберігання або представлення одного символу в повідомленні. Ця теорія розглядає і поняття значення, або «інформаційного контенту», повідомлень. Такий підхід передбачає мінімізацію кількості атрибутів текстів, необхідних для ідентифікації окремого кортежу. Атрибут з найвищою корисністю інформації (Gain) використовують для генерації правила асоціації. За значенням

приросту інформації цього атрибуту оцінюють корисність інформації інших атрибутів повідомлень [10]:

$$Gain(A) = info(D) - infoA(D), \quad (3)$$

де D – масив текстових даних, $info(D)$ – обсяг масиву D , виміряний у бітах; A – атрибут, $infoA(D)$ – кількість інформації, яку несе атрибут A (у бітах).

- Частка кортежу в інформаційному масиві (sample proportion, p-s).
- Підвищення інтересу (Lift) – відношення частоти появи кортежу X у текстових повідомленнях, які містять обидва кортежі X та Y , до частоти появи кортежу Y загалом у всіх аналізованих текстових повідомленнях:

$$Lift\{X \rightarrow Y\} = \frac{Support(X, Y)}{Support(X) \times Support(Y)}. \quad (4)$$

Значення Lift, більше 1, означає, що кортеж Y з високою ймовірністю буде міститись у текстовому повідомленні, якщо в ньому міститься кортеж X .

- Ступінь наслідування правил (Conviction):
-

$$Conviction\{X \rightarrow Y\} = \frac{1 - Support(Y)}{1 - Confidence\{X \rightarrow Y\}}. \quad (5)$$

За допомогою побудованої моделі отримано наступні асоціативні правила (рис. 5):

```

Association Rules
[america] --> [great, make] (confidence: 0.811)
[america] --> [great] (confidence: 0.823)
[america] --> [make] (confidence: 0.839)
[great] --> [make] (confidence: 0.874)
[great] --> [america, make] (confidence: 0.874)
[great] --> [america] (confidence: 0.887)
[make] --> [great] (confidence: 0.940)
[make] --> [america, great] (confidence: 0.940)
[america, make] --> [great] (confidence: 0.967)
[make] --> [america] (confidence: 0.972)
[america, great] --> [make] (confidence: 0.985)
[great, make] --> [america] (confidence: 1.000)

```

**Рис. 5. Згенеровані асоціативні правила
(фрагмент отриманих результатів)**

Отже, наприклад, з ймовірністю 0,985 можна стверджувати, що при появі у текстовому повідомленні набору {America, great} у цьому меседжі буде наявним і кортеж {make}. А наявність у тексті кортежу {great, make} забезпечує вірогідність (ймовірність = 1) появи у повідомленні слова «America». Отримані асоціативні правила дали змогу виявити у масиві з 1869 неструктурованих текстових повідомлень Трампа у Twitter його основне передвиборче гасло: «Make America great again» («Зробимо Америку величною знову»). Ця фраза ідентифікує всіх американців як єдиний народ,

націю понад націями. Дональд Трамп використав її як дієвий інструмент психологічного нав'язування, що змусило багатьох людей зробити несвідомий вибір.

Граф побудованої моделі надає не лише значення підтримки та ймовірності для кожного із згенерованих асоціативних правил, а й унаочнює, як зв'язки різних атрибутів пов'язані між собою (рис. 6).

Сучасні методи Data Mining дають змогу побудувати модель, яка здатна надати допомогу управлінцям не лише у сфері радикального покращення фінансового та ринкового становища компанії, а й виявити неочевидні закономірності, що можуть змінити розстановку політичних сил в окремо взятій країні чи навіть зміцнити позиції окремої держави на міжнародній арені.

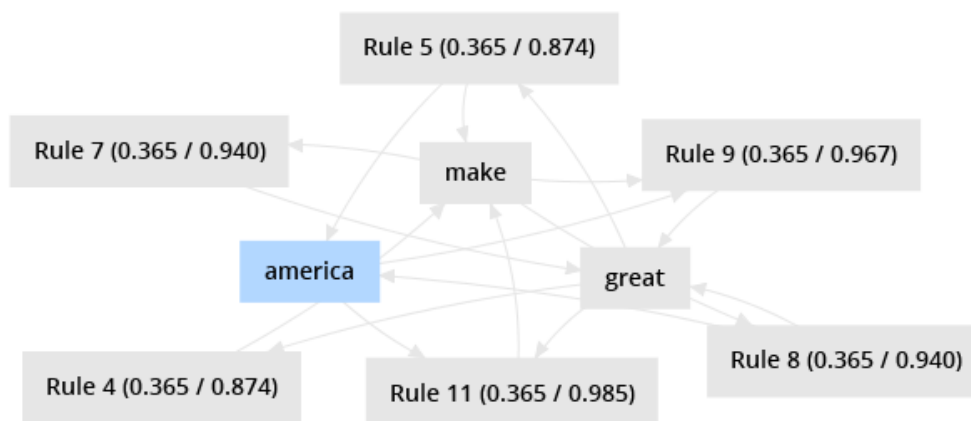


Рис. 6. Графічне представлення зв'язків між словами у передвиборчих меседжах Трампа (фрагмент результатів)

Висновки та перспективи подальших розвідок

У сучасному суспільстві інформація та наукові знання є вирішальними факторами розвитку. Висока мінливість інфосфери (нової інфраструктури планети) спричинена прогресом інформаційних і телекомунікаційних технологій. Інноваційні технології засновані на глобальних телекомунікаційних мережах та інтелектуальних комп'ютерних системах. Сучасні ІТ відкривають кардинально нові можливості та способи обробки інформаційних масивів, зокрема у суспільній, економічній та політичній сферах.

Ми живемо у складний, неспокійний час глобальних загроз та регіональних конфліктів, спричинених бажанням «сильних світу всього» мати щораз більший вплив на громадськість і володіти все більшими територіями та ресурсами. Вибори є ключовим елементом у формуванні сучасного світу, у змінах на краще чи на гірше. Сьогодні суспільні блага багато в чому залежать від досягнень в області аналізу великих даних і штучного інтелекту, зокрема технологій, які використовують для розробки стратегій проведення ефективних виборчих кампаній та забезпечення демократії в нашому сучасному світі.

Тому актуальним на сьогодні є подальше практичне використання методів Big Data і Data Mining для аналізу текстів публічних промов можновладців та інструментів web mining (метод «видобування» інформації з web-ресурсів і виявлення загальних закономірностей у мережі Інтернет) з метою пошуку асоціативних зв'язків та маніпулятивних технологій (наприклад, брехні) політиків та керівників як на регіональному, так і на державному та міжнародному рівнях.

З розвитком інструментів Big Data реально розраховувати на можливість купити голоси, приховати власність чи непривабливі подробиці біографії ніхто з публічних осіб більше не зможе ніколи. Однак виклик тотальної прозорості кинутий не тільки сильним світу цього. Звичайні люди значно вразливіші. Використання великих даних у політиці стало новою зброєю масового ураження. Необмежений доступ до масиву цифрових слідів відкриває необмежені можливості впливати на соціум чи маніпулювати спільнотою. Приватність, інтимність, таємниця, особистий простір – ці поняття втратили сенс у сучасному цифровому суспільстві. Інформаційні технології Big Data сьогодні набувають особливої актуальності, змінюючи підходи до аналізу інформації та способи прийняття рішень. Цей тренд впливає на соціум та політику у міжнародних масштабах.

Список літератури

1. Майер-Шенбергер, В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим: пер. с англ. / В. Майер-Шенбергер, К. Кукьер. – М.: Манн, Иванов и Фербер, 2014. – 240 с.
2. Bratosin, O. C. Boarding to Big data / O. C. Bratosin // Database Systems Journal. – 2016. – vol. 6. – issue 4. – P. 11-17 [Електронний ресурс]. – Режим доступу: <https://EconPapers.repec.org/RePEc:aes:dbjour:v:6:y:2016:i:4:p:11-17>.
3. Bail, C. A. Taming Big Data / C. A. Bail // Sociological Methods & Research. – 2017. – vol. 46. – issue 2. – P. 189-217 [Електронний ресурс]. – Режим доступу: <http://smr.sagepub.com/content/46/2/189>.
4. Titirisca, A. Big Data and the internet of things / A. Titirisca // Management Intercultural. – 2015. – issue 33. – P. 211-215 [Електронний ресурс]. – Режим доступу: http://www.mi.bxb.ro/Articol/MI_33_29.pdf.
5. Bai, Ju. Special Issue on Big Data / Ju. Bai, Ji. Fan and R. Tsay // Journal of Business & Economic Statistics. – 2016. – vol. 34. – issue 4. – P. 487-488 [Електронний ресурс]. – Режим доступу: <http://hdl.handle.net/10.1080/07350015.2016.1197681>.
6. Worst J. The case of the “Big Data” revolution // Working Papers from Maastricht School of Management. – 2014. – No34 [Електронний ресурс]. – Режим доступу: <http://web2.msm.nl/RePEc/msm/wpaper/MSM-WP2014-34.pdf>.
7. Harvard Business Review [Електронний ресурс] // The World Economic Forum. – Режим доступу: <https://www.weforum.org/agenda/2017/10/the-society-of-the-future-looks-nothing-like-you-might-imagine/>.
8. Dehaye, P. -O. The Data That Turned the World Upside Down / P. -O. Dehaye // Public Policy Program. – Stanford, California: Stanford University [Електронний ресурс]. – Режим доступу: <https://publicpolicy.stanford.edu/news/data-turned-world-upside-down>.
9. Stafford, C. MIT Media Lab uses «Electome» to track political discourse / C. Stafford [Електронний ресурс]. – Режим доступу: <http://searchcio.techtarget.com/blog/TotalCIO/MIT-Media-Lab-uses-Electome-to-track-political-discourse>.
10. Барсегян, А. А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. – СПб.: БХВ-Петербург, 2007. – 384 с.
11. Berry, Michael J. A. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. 2nd Edition / Michael J. A. Berry, G. S. Linoff. – Wiley Indianapolis, Indiana: Wiley Publishing, Inc., 2004. – 672 p. [Електронний ресурс]. – Режим доступу: <http://197.14.51>.

-
- 10:81/pmb/GESTION2/MARKETING/Data%20Mining%20Techniques%20For%20Marketing%20Sales%20And%20Customer%20Relationship%20Management%20Ed.pdf.
12. Tweets from @realdonaldtrump [Електронний ресурс]. – Режим доступу: <http://varianceexplained.org/r/trump-tweets/>.
13. Shannon, C. E. The mathematical theory of communication / C. E Shannon. and W. Weaver. – Urbana, IL: University of Illinois Press, 1949.

References

1. Mayer-Schönberger, V. & Cukier, K. (2014). *Bolshye dannye Bolshie dannye. Revolyutsiya, kotoraya izmenit to, kak myi zhivem, rabotaem i myislim [Big Data: A Revolution That Will Transform How We Live, Work, and Think]*. Moscow: Mann, Ivanov & Ferber [in Russian].
2. Bratosin, O. C. (2016). Boarding to Big data. *Database Systems Journal*, 6(4), 11-17. Retrieved from <https://EconPapers.repec.org/RePEc:aes:dbjour:v:6:y:2016:i:4:p:11-17>.
3. Bai Ju. (2016). Special Issue on Big Data. *Journal of Business & Economic Statistics*, 34(4), 487-488. Retrieved from <http://hdl.handle.net/10.1080/07350015.2016.1197681>.
4. Titirisca A. (2015). Big Data and the internet of things. *Management Intercultural*, 33, 211-215. Retrieved from http://www.mi.bxb.ro/Articol/MI_33_29.pdf.
5. Bail, C. A. (2017). Taming Big Data. *Sociological Methods & Research*, 46(2), 189-217. Retrieved from <http://smr.sagepub.com/content/46/2/189>.
6. Worst J. (2014). The case of the “Big Data” revolution. *Working Papers from Maastricht School of Management*, 34. Retrieved from: <http://web2.msm.nl/RePEc/msm/wpaper/MSM-WP2014-34.pdf>.
7. *Harvard Business Review* [Електронний ресурс] // The World Economic Forum. Retrieved from: <https://www.weforum.org/agenda/2017/10/the-society-of-the-future-looks-nothing-like-you-might-imagine/>.
8. Dehaye, P. -O. (2017). The Data That Turned the World Upside Down. *Public Policy Program*. – Stanford, California: Stanford University. Retrieved from: <https://publicpolicy.stanford.edu/news/data-turned-world-upside-down>.
9. Stafford, C. *MIT Media Lab uses «Electome» to track political discourse*. Retrieved from: <http://searchcio.techtarget.com/blog/TotalCIO/MIT-Media-Lab-uses-Electome-to-track-political-discourse>.
10. Barsegjan, A. A., Kupriyanov, M. S., Stepanenko, V. V., & Holod, I. I. (2007). *Tehnologii analiza dannyh. Data Mining, Visual Mining, Text Mining, OLAP [Technologies of data analysis. Data Mining, Visual Mining, Text Mining, OLAP]*. Sankt-Peterburg: BHV-Peterburg [in Russian].
11. Berry, Michael J. A., & Linoff G. S. (2004). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* [2nd ed.]. Wiley Indianapolis, Indiana: Wiley Publishing, Inc. Retrieved from: <http://197.14.51.10:81/pmb/GESTION2/MARKETING/Data%20Mining%20Techniques%20For%20Marketing%20Sales%20And%20Customer%20Relationship%20Management%20Ed.pdf>.
12. *Tweets from @realdonaldtrump* (2016). Retrieved from: <http://varianceexplained.org/r/trump-tweets/>.
13. Shannon, C. E. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

Стаття надійшла до редакції 10.03.2018 р.