

Математична статистика, машинне навчання та data science – сучасні та перспективні напрямки професійного розвитку



В.П. Зубченко

Київський національний університет імені Тараса Шевченка

Сучасний світ демонструє нові критерії успішності та вимоги до реалізації цілей. Надшвидка динаміка змін, глобалізованість світу, доступність величезної кількості інформаційних ресурсів, нові можливості щодо безперешкодних комунікацій, якісні трансформації у безлічі галузей економіки, цифровізація суспільства, активне намагання впровадження в багатьох сферах здобутків штучного інтелекту та машинного навчання – лише частина трендів, що будуть визначальними для кожного з нас.

Значний вплив зазначені тенденції мають на кожному з етапів професійного становлення – під час навчання, стажування, отримання досвіду роботи, започаткування власної справи. Необхідність запам'ятовування великих обсягів інформації планомірно поступається вмінню швидко орієнтуватись в потоках даних, отримувати доступ до потрібних інформаційних ресурсів, обирати та перевіряти доступні дані, здійснювати їх первинно обробку, поєднувати дані декількох джерел, проводити синтез, формувати та перевіряти гіпотези за допомогою статистичних методів, здійснювати комплексний аналіз досліджуваних процесів. Отримавши певний набір фактичних знань, ми отримуємо конкуренту перевагу на доволі нетривалий проміжок часу, адже технології стрімко ідуть вперед, створюючи необхідність продовження навчання та безперервного подальшого професійного розвитку.

Важливим критерієм успішності для багатьох сфер професійної діяльності стає поєднання знань та навичок. Окрім власне високого рівня компетентності в безпосередній професійній сфері визначальними є вміння будувати дієві процеси взаємодії з колегами, формувати проектні команди, організовувати ефективну комунікацію, використовувати техніки тайм-менеджменту, обирати найбільш дієві методи досягнення поставлених цілей, інвестувати в розвиток учасників команди, покращувати свої професійні та лідерські якості. В той же час, швидкість зміни технологій приводить до необхідності регулярного моніторингу технологічних можливостей та вміння обрати серед них найбільш ефективну для реалізації поставлених задач.

Класична система вертикального кар'єрного розвитку випускника університету стала одним із можливих варіантів професійного становлення. Альтернативою багаторічного розвитку в межах однієї компанії або галузі стало намагання молодих та амбітних студентів отримати досвід роботи в декількох компаніях, які часто можуть належать до різних галузей. Зміна сфери діяльності кожні кілька років дозволяє за певний час отримати досвід в IT,

маркетингу, бізнес-аналітиці, продажах, data science... Тим самим сформувавши унікальний набір професійних знань та навичок, із можливістю подальшого їх поєднання для створення принципово нових моделей та підходів. Успіх – у синтезі досвіду роботи в декількох предметних сферах, здатності динамічно адаптуватись у різних колективах, ефективно взаємодіяти із колегами із абсолютно різними професійними вміннями, стилем мислення, набором особистісних компетенцій, прагненням створювати нове, втілювати, тестувати, випробовувати, помилятись, покращувати, адаптувати, досліджувати, аналізувати, інтегрувати.

Традиційні підходи до організації робочого процесу усе більш активно поступаються впровадженню в різноманітні сфери економіки базових аспектів інноваційного підприємництва, які включають, зокрема: пошук інноваційної бізнес-ідеї, визначення місії, візії та цінності стартапу, пошук та створення команди, методологію дизайну мислення, методи дослідження ринку, побудову бізнес-моделей, методи побудови так званих керованих даними компаній, аналіз конкурентного середовища, способи підготовки ефективної pitch-презентації, основні концепції побудови бренду, основи маркетингу та методи збільшення продажів.

Більшість подібних підходів ґрунтується на активному використанні різноманітних методів аналізу даних. Побудова ефективної стратегії потребує виділення факторів, які мають найбільший вплив на відповідні аспекти проєкту. Аналіз при цьому цілком може бути багатофакторним, із безліччю факторів, частину з яких можна вказати явно й одразу, інші ж можуть бути неявними або ж прихованими. Перетворити цілковиту невизначеність у певну структурованість може допомогти статистичний аналіз, багатий на різноманітні математичні методи аналізу даних. Більшість із них базуються на аналізі попередніх, так званих історичних, даних, які включають в себе різноманітні показники діяльності проєкту – кількість клієнтів, дані про транзакції, найменування, кількість та вартість придбаних товарів, шлях користувача по інтернет-сайту, історія активності в соціальних мережах.

Величезна кількість накопичених даних, або ж так звана big data, відкриває як значні перспективи з точки зору можливого прогнозування майбутнього, так і значну складність виділити із такого масиву даних цінну інформацію. Традиційні методи теорії ймовірностей та математичної статистики усе частіше починають доповнюватись або ж замінюватись сучасними методами машинного навчання, бізнес-аналітики, data science. Вони інтегрують методи математичної обробки та аналізу великого обсягу даних, методи інтелектуального аналізу, елементи штучного інтелекту, найсучасніші програмні інструменти, хмарні технології, використання data-центрів для ефективного збереження та обробки інформації. Відмінність методів data-science від традиційних полягає в наступному. Традиційні підходи передбачають написання чіткого алгоритму, який буде відпрацьовуватись машиною. Однак швидкість сучасного світу приводить до такої ж швидкої зміни факторів, що лежать в основі того чи іншого явища. Саме тому машинне навчання бере за основу підхід, коли комп'ютеру передаються лише певні початкові алгоритми, далі ж машина продовжує самонавчання, коректує чи змінює підходи в процесі роботи на основі нових одержаних даних. Безумовною перевагою підходу є його динамічність, гнучкість, в більшості випадків він дозволяє зробити більш точні висновки щодо досліджуваного процесу або явища.

Втім досить часто методи машинного навчання представляють собою black box («чорну скриню»), яка дає певний, нехай навіть доволі точний, результат, втім набагато слабше за традиційні моделі дозволяє відслідкувати причинно-наслідковий зв'язок між ключовими факторами та результатом відпрацювання математичної моделі. Значна частина підходів науки про дані не є принципово новими – більшість із них були відомі ще в минулому столітті. Однак, обчислювальні потужності того часу просто не дозволяли

ефективно реалізувати відповідні методи, так щоб отримати результати відпрацювання моделі в розумний час. Сучасні персональні комп'ютери вже більш придатні до частини відповідних задач. Водночас більша частина динамічних задач машинного навчання реалізується із використанням хмарних технологій та хмарних дата-центрів.

Динаміка сучасного світу приводить до нових викликів, зміни усталених підходів до вирішення задач, прискорює потреби у нових висококваліфікованих фахівцях по роботі з даними, які вміють поєднувати блискучі математичні та статистичні знання із вмінням розуміти економічну та наукову логіку процесу або явища, прагнуть знайти найбільш ефективну модель для опису процесу, проведення аналізу та підготовки прогнозу, й, зрештою, втілити результат роботи в комплексному продукті, результати роботи якого цілком можуть набути всесвітнього масштабу.