

«Сейфуллин окулары-18(2): «XXI ғасыр ғылымы – трансформация дәуірі» Халықаралық ғылыми-практикалық конференция материалдары = Материалы международной научно-практической конференции «Сейфуллинские чтения – 18(2): «Наука XXI века - эпоха трансформации» - 2022.- Т.І, Ч.ІІІ. - С.147-149.

DATA SCIENCE КАК ВАЖНЕЙШАЯ КОМПЕТЕНЦИЯ ВЫПУСКНИКОВ КАЗАХСКОГО АГРОТЕХНИЧЕСКОГО УНИВЕР- СИТЕТА ИМ. С.С. СЕЙФУЛЛИНА

Задаулы Д., к.э.н.
Танкаева С.К., к.ф.-м.н.

Казахский агротехнический университет им. С. Сейфуллина, г. Нур-Султан

Каждый день человечество генерирует квинтиллионы байт различных данных. В 2010-х годах объемы данных стали расти по экспоненте. Свою роль сыграл целый ряд факторов — от повсеместного распространения мобильного интернета и популярности соцсетей до всеобщей оцифровки сервисов и процессов.

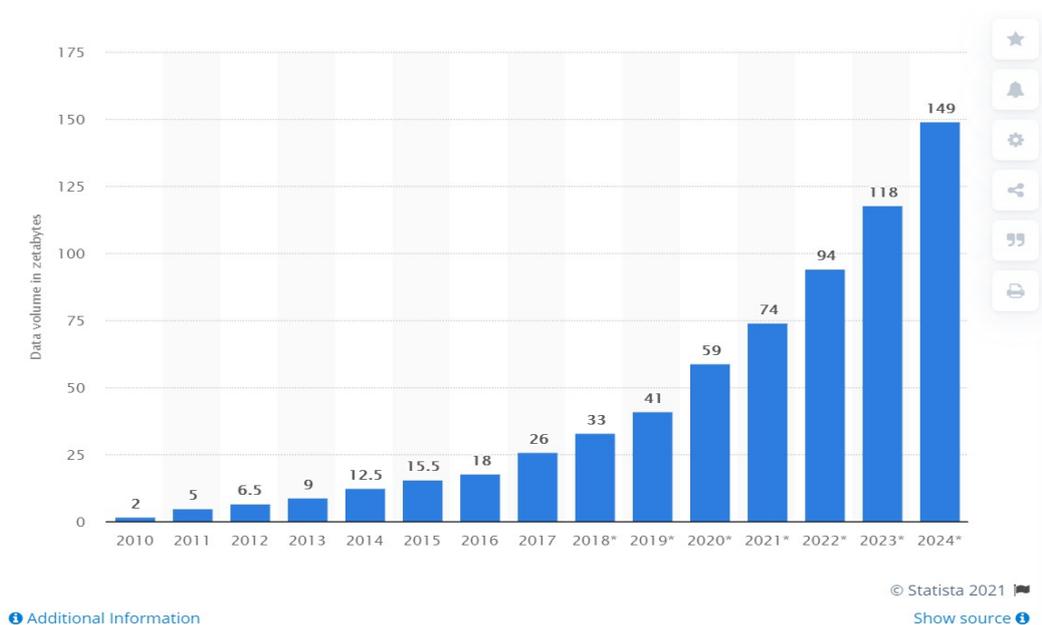


График 1. Объем данных, созданных, собранных и потребленных во всем мире с 2010 по 2024 год (в зеттабайтах)

В итоге профессия дата-сайентиста быстро превратилась в одну из самых популярных и востребованных.

В 2012 году издание Harvard Business Review назвало работу data scientist the sexiest job of the 21 century [1]. McKinsey примерно в то же время

предсказывала нехватку только в США 140–190 тыс. специалистов этого профиля к 2018 году, а к 2024-му - и вовсе 250 тыс.

Информация – один из главных ресурсов современного мира. С помощью правильных инструментов, технологий и алгоритмов, компании и организации могут получать и использовать данные для получения социальных и экономических преимуществ.

Наука о данных (англ. data science; иногда даталогия - datalogy [2]) - раздел информатики, изучающий проблемы анализа, обработки и представления данных в цифровой форме. Объединяет методы по обработке данных в условиях больших объёмов и высокого уровня параллелизма, статистические методы, методы интеллектуального анализа данных и приложения искусственного интеллекта для работы с данными, а также методы проектирования и разработки баз данных.

Data Science (DS) - междисциплинарная область на стыке статистики, математики, системного анализа и машинного обучения, которая охватывает все этапы работы с данными. Она предполагает исследование и анализ сверхбольших массивов информации и ориентирована в первую очередь на получение практических результатов.

Развитие Data Science шло вместе с внедрением технологий Big Data и анализа данных. И хотя эти области часто пересекаются, их не следует путать между собой. Все они предполагают понимание больших массивов информации. Но если аналитика данных отвечает на вопросы о прошлом (например, об изменениях в поведении клиентов какого-либо интернет-сервиса за последние несколько лет), то Data Science в буквальном смысле смотрит в будущее. Специалисты по DS на основе больших данных могут создавать модели, которые предсказывают, что случится завтра. В том числе и предсказывать спрос на те или иные товары и услуги.

Возможности, открывающиеся бизнесу благодаря анализу данных, вызвали к жизни множество университетских программ и курсов Data Science, так что назревающий дефицит стал постепенно заполняться, хотя нехватка качественных спецов до сих пор остаётся. И сегодня работа data scientist'a популярна, престижна и хорошо оплачиваема. Компании продолжают активно нанимать и выстраивают DS-компетенции, определяя Data Science как работу на стыке математических, компьютерных и специализированных (domain knowledge) навыков. В результате сложно найти крупную компанию без DS-подразделения, численность которого может варьироваться от нескольких до нескольких тысяч человек.

Дата-сайентист или специалист по работе с данными работает с Big Data - теми самыми большими массивами структурированных или неструктурированных данных. В частности, он собирает и упорядочивает базы данных, анализирует их, а также ищет в них определенные связи и закономерности. На основе обработанной и полученной информации им создается модель машинного обучения, которая позволяет составить прогноз или предсказать результат в будущем.

Таким образом, суть работы дата-сайентиста - построение полезных для бизнеса предсказательных моделей. Модели Data Science основаны на технологиях искусственного интеллекта (AI), точнее говоря - на алгоритмах машинного обучения (ML).

Например, зарубежом, в связи с развитием блокчейн технологий идет активное внедрение Data Science и искусственного интеллекта в FinTech.

«Financial technology (FinTech) has been playing an increasingly critical role in driving modern economies, society, technology, and many other areas. Smart FinTech is the new-generation FinTech, largely inspired and empowered by data science and artificial intelligence (DSAI) techniques. Smart FinTech synthesizes broad DSAI and transforms finance and economies to drive intelligent, automated, whole-of-business and personalized economic and financial businesses, services and systems».[3]

«Финансовые технологии (FinTech) играют все более важную роль в развитии современной экономики, общества, технологий и многих других областях. Smart FinTech - это FinTech нового поколения, в значительной степени основанный на науке о данных и методах искусственного интеллекта (DSAI). Smart FinTech синтезирует широкий DSAI и трансформирует финансы и экономику для управления интеллектуальными, автоматизированными, комплексными и персонализированными экономическими и финансовыми предприятиями, услугами и системами» - отмечают исследователи Longbing Cao, Qiang Yang и Philip S. Yu в статье «Data science and AI in FinTech: an overview» в журнале «International Journal of Data Science and Analytics»

Поэтому сейчас актуально введение бакалаврских и магистерских программ не только в IT-специальности, но и в финансово-экономические, бизнес-менеджмента и т.п.

Data Science отвечает современным вызовам ИТ-индустрии: подготовка аналитиков данных для инновационной цифровой экономики, обладающих аналитическими, исследовательскими, методическими и технологическими компетенциями в новых областях науки, техники и экономики. Это актуально в современный момент для агротехнологической отрасли.

Образовательное направление по Data Science также включает изучение методов математического моделирования (подробное изложение методов например в [4]) и вычислительных алгоритмов обработки данных, алгоритмов анализа данных как в общей постановке, так и в сфере специализации.

Студенты будут учиться разрабатывать автоматизированные методы анализа больших объемов данных с целью извлечения из них необходимых знаний для принятия решений. Выпускники данной программы будут нацелены проводить оригинальные исследования в выбранных областях машинного обучения и анализа данных и применять полученные результаты в обществе, экономике и бизнесе.

Так же предполагается углубленное изучение фундаментальной информатики и компьютерных наук, статистики, прикладной математики, теоретической физики и теории информации. Каждый учебный курс должен обеспечиваться реальными практическими задачами и примерами их решения.

Выпускники получат все необходимые компетенции, чтобы быть успешными для работы в области вычислительных сетей передачи данных, хранения данных, статистического анализа и быстрого извлечения аналитической информации из неструктурированных массивов разнотипных данных.

После завершения обучения, выпускники будут:

1. Способны эффективно и компетентно отвечать на возрастающие запросы экономики, касающиеся обработки, хранения, передачи данных.

2. Иметь сбалансированные теоретическую подготовку и практические навыки в прикладной математике, статистической физике, теории информации и компьютерных науках, чтобы анализировать многомерные мультимасштабные неупорядоченные наборы разнотипных данных.

3. Способны описывать и трансформировать информацию для того, чтобы открывать взаимосвязи и строение сложных массивов данных.

4. Разрабатывать модели и алгоритмы решения задач, математические методы, методы теории информации и параллельной алгоритмизации.

5. Разрабатывать высокопроизводительное программное обеспечение суперкомпьютерных кластеров, распределенных вычислительных сетей и систем хранения данных.

Что касается должностных обязанностей дата-сайентиста, то к ним относятся:

- поиск связей и закономерностей в наборах больших данных;
- подготовка данных к разработке ML-модели: выборка, очистка, генерация признаков, интеграция и форматирование;
- моделирование и визуализация;
- разработка гипотез по оптимизации экономических и бизнес-показателей при помощи моделей машинного обучения, а также их дальнейшее тестирование.

Безусловно, владение указанными навыками повысит конкурентоспособность наших выпускников на рынке трудовых ресурсов.

Список использованной литературы

1. Davenport T. H., Patil D. J. Data scientist [Text] / Harvard business review. – 2012. – Т. 90. – №. 5. – С. 70-76.
2. <https://www.forbes.com/sites/gilpress/2012/09/27/data-scientists-the-definition-of-sexy/?sh=7183c8ac5f96>

3. Cao L., Yang Q., Yu P. S. Data science and AI in FinTech: An overview [Text] / International Journal of Data Science and Analytics. – 2021. – Т. 12. – №. 2. – С. 81-99.
4. Загоруйко Н. Г. Когнитивный анализ данных [Текст] / Новосибирск: Академическое издательство ГЕО. – 2013. – С. 186.