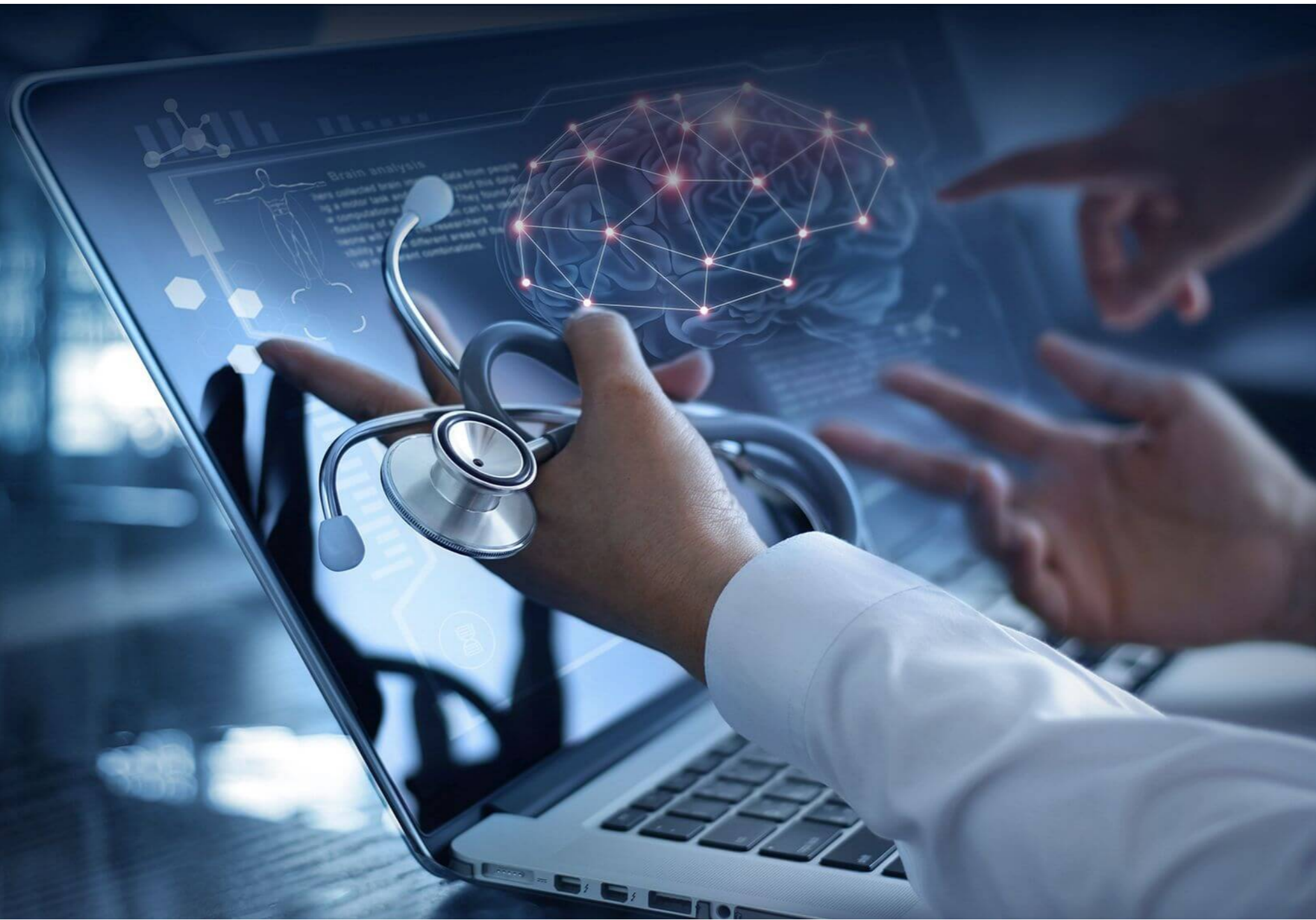


Data Science in Medicine

365 



ШКОЛА
СИСТЕМНОГО
АНАЛИЗА



Мира Карлаш

- Более 5 лет в анализе данных
- Дополнительное образование от Stanford в Data analysis
- Повышение квалификации от Google & Samsung
- Дополнительное образование от Duke в области нейронаук
- В роли тренера более 4 лет

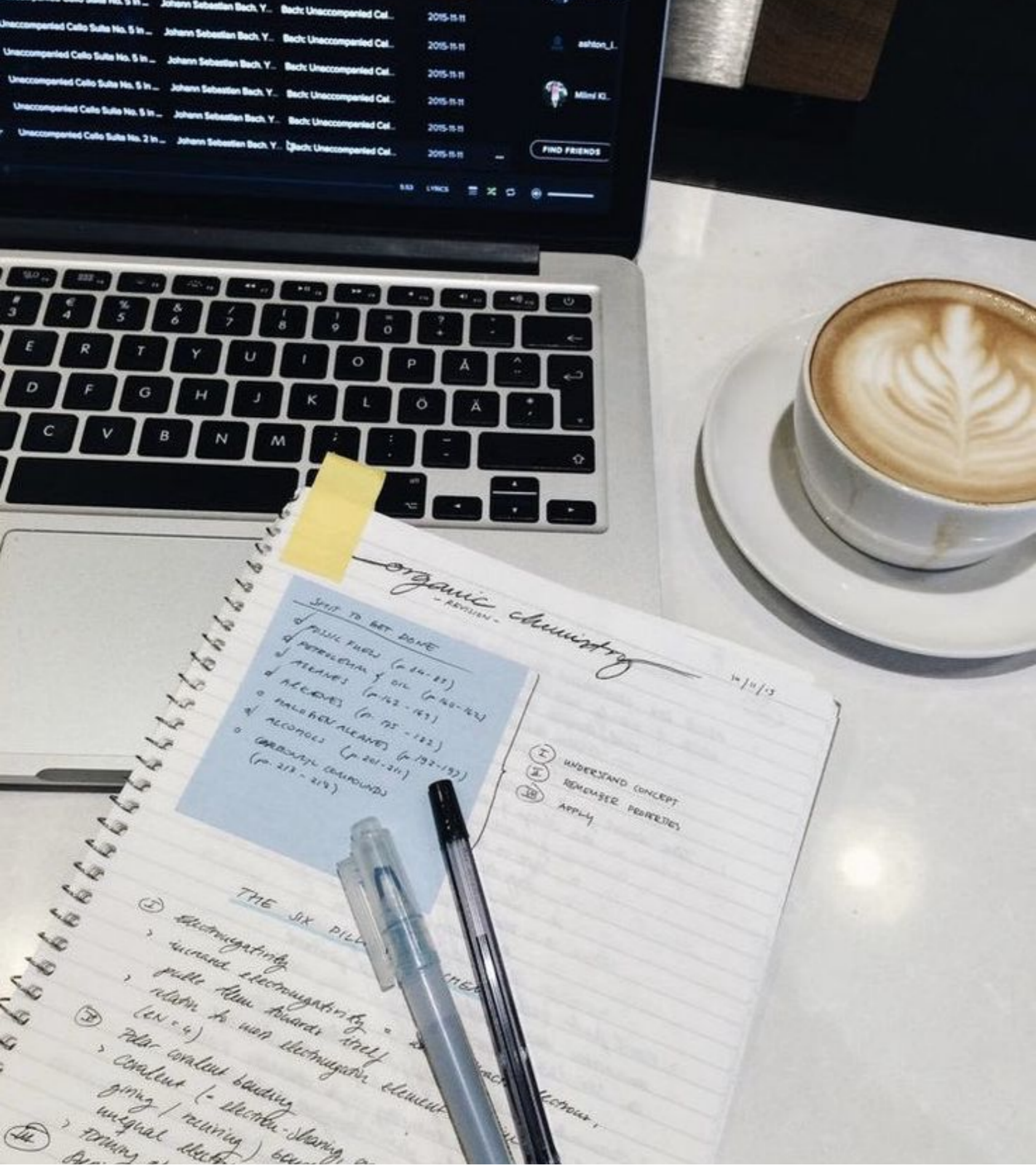
 @mira_karlash

ШКОЛА
СИСТЕМНОГО
АНАЛИЗА

Дмитрий Андреев



- 10+ лет в банках
- “войти в айти” - 2017 год
- аналитик “всего”
- проклятие бизнес-аналитика



План

- I. Введение
 - A. Приветствие и представление
 - B. Введение в тему конференции: "Data Science in Medicine"

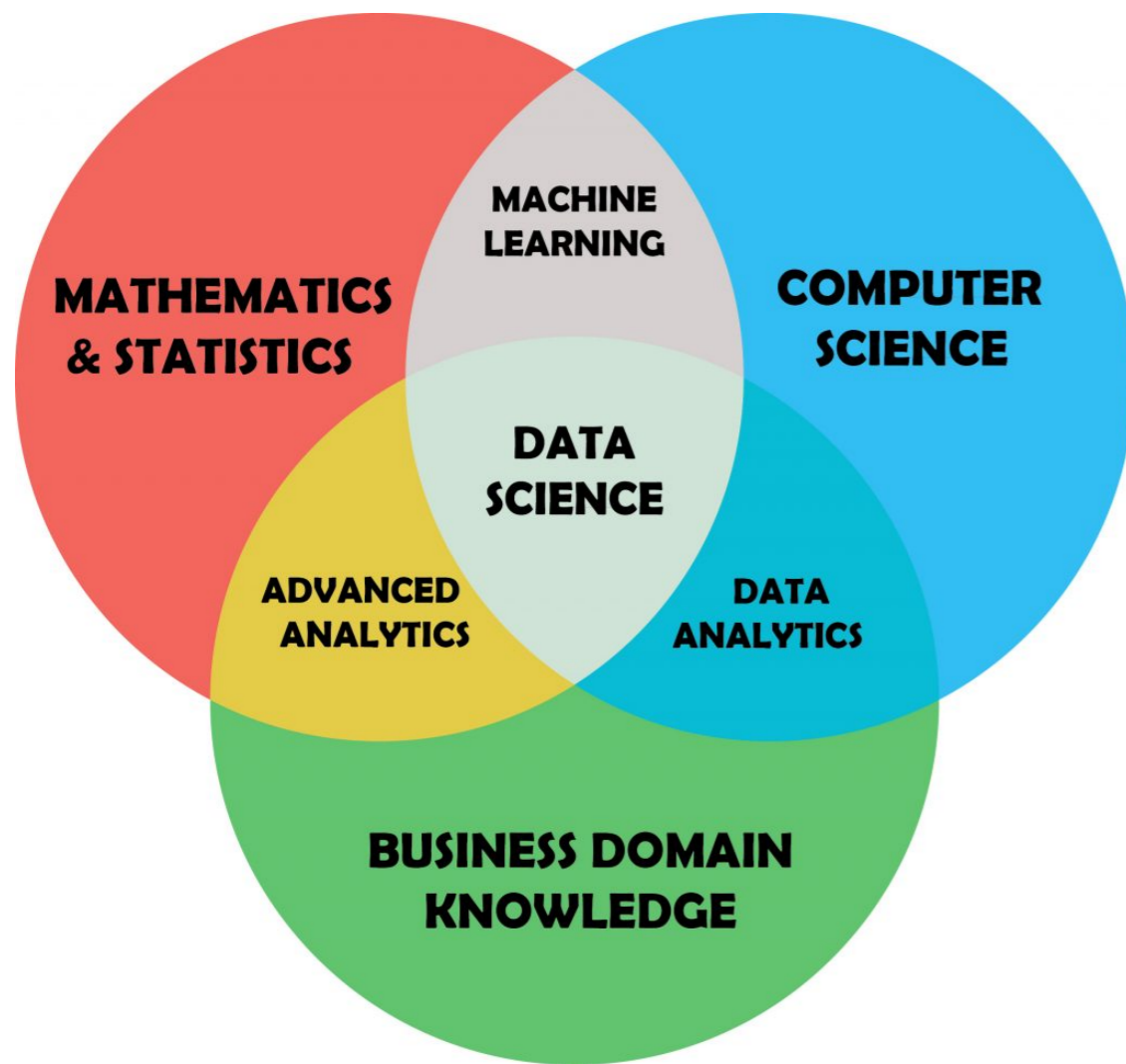
- II. Кто такие Data Scientist и их роль в медицине
 - A. Определение Data Scientist и их ключевые навыки
 - B. Роль Data Scientist в медицине и их значимость для современной практики
 - C. Примеры успешного использования Data Science в медицине

- III. Возможности и перспективы применения Data Science в медицине
 - A. Текущие возможности и преимущества применения Data Science в медицине
 - B. Перспективы развития и новые направления исследований
 - C. Примеры успешного внедрения Data Science в других предметных областях (например, банки, агрегаторы и др.)

- IV. Сбор требований для системного аналитика с учетом потребностей Data Scientist
 - A. Важность сотрудничества между Data Scientist и системным аналитиком
 - B. Методы сбора требований для проектов, связанных с Data Science в медицине
 - C. Примеры процесса сбора требований для проектов в других предметных областях

- V. Заключение
 - A. Подведение итогов и основные выводы
 - B. Значимость и перспективы развития Data Science в медицине
 - C. Приглашение к дальнейшему обсуждению и вопросам

system analyst vs data analyst vs data scientist



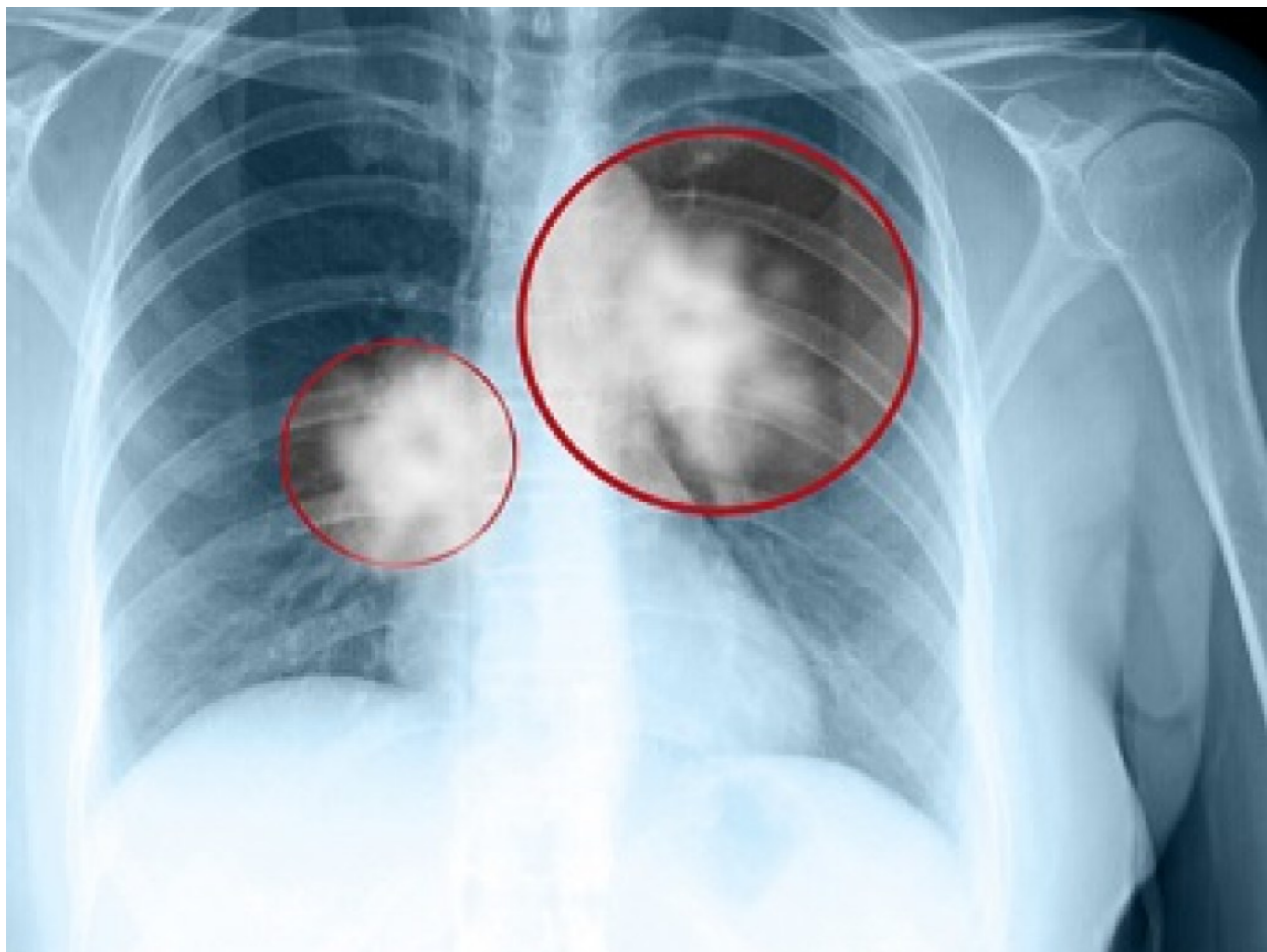
	Data analyst	Data scientist
Mathematics	Foundational math, statistics	Advanced statistics, predictive analytics
Programming	Basic fluency in R, Python, SQL	Advanced object-oriented programming
Software and tools	SAS, Excel, business intelligence software	Hadoop, MySQL, TensorFlow, Spark
Other skills	Analytical thinking, data visualization	Machine learning, data modeling

Определение Data Scientist и их ключевые навыки



- Статистика и анализ данных: Понимание статистических методов и умение работать с различными видами данных для извлечения информации.
- Машинное обучение и искусственный интеллект: Знание методов обучения на основе данных для создания моделей прогнозирования и классификации.
- Программирование: Опыт работы с языками программирования, такими как Python или R, для обработки данных и создания алгоритмов.
- Базы данных и хранение данных: Умение работать с базами данных, структурировать и хранить большие объемы информации.
- Визуализация данных: Создание информативных и наглядных визуализаций для представления результатов анализа данных.

Примеры успешного использования Data Science в медицине



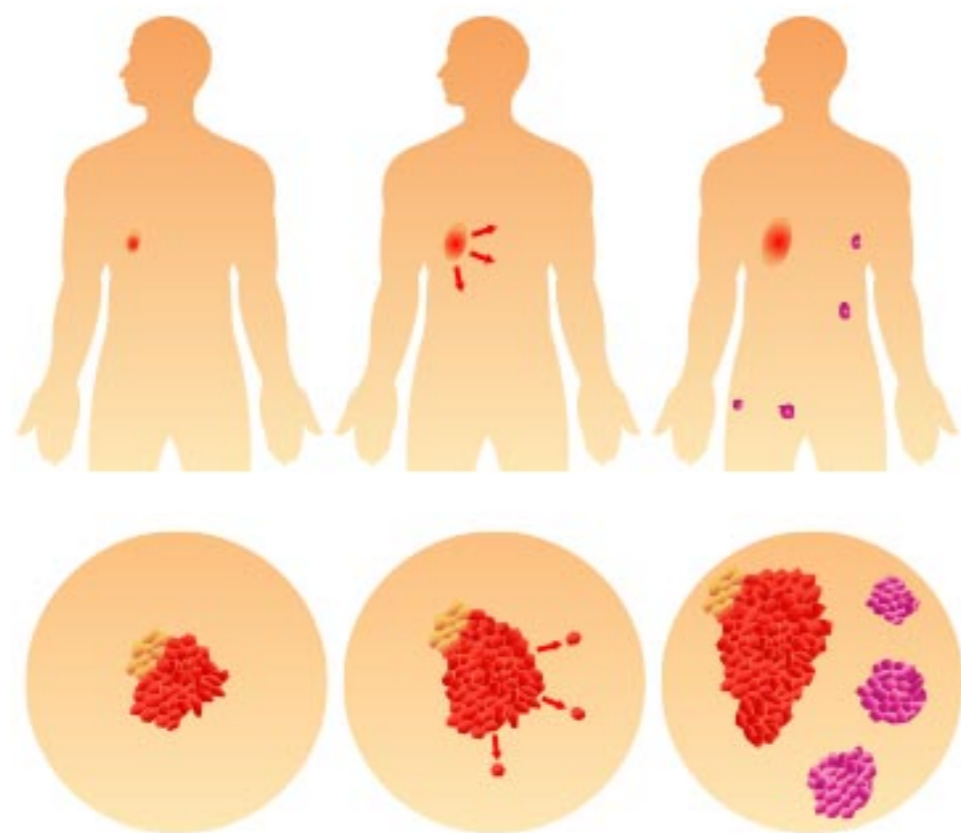
Дано:

В наличии снимки КТ легких.
Тяжелая эпидемиологическая обстановка в
связи с COVID-19.

Задача:

Диагностировать поражение легких на
уровне 5-10% и выше в 90+% случаев.

Примеры успешного использования Data Science в медицине



Дано:

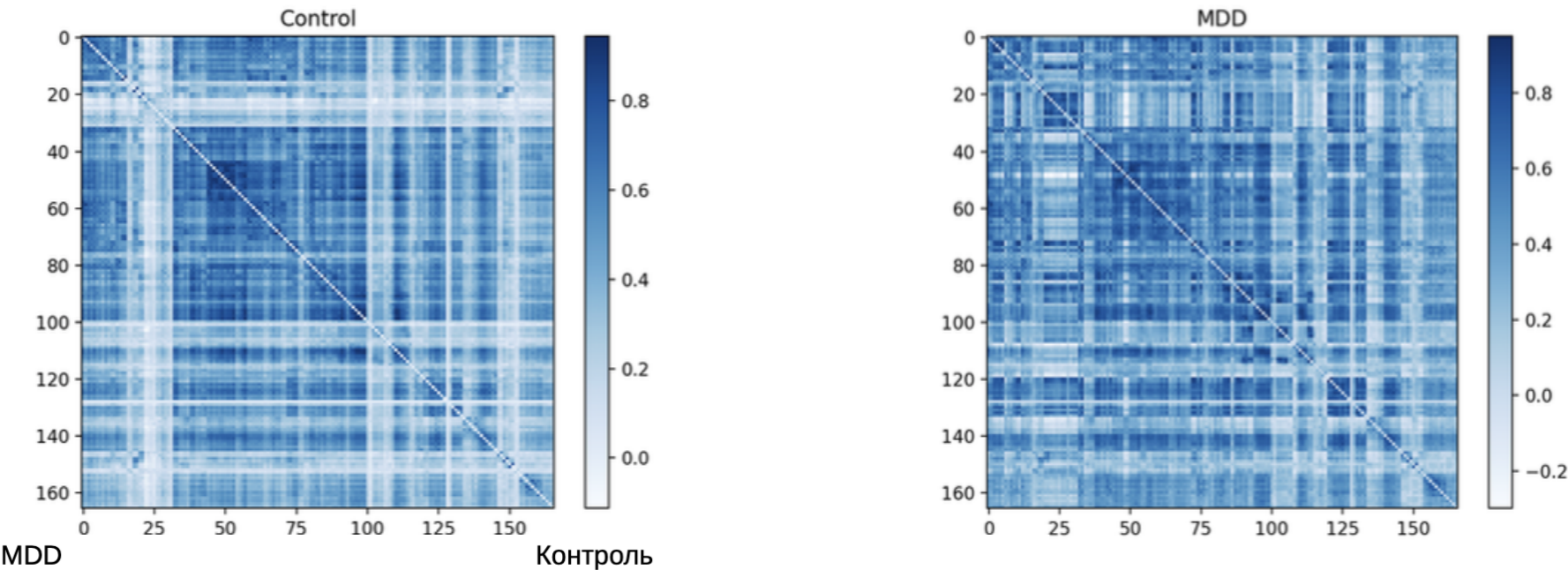
Материалы предположительно пораженного участка раком.

Рак входит в топ 10 смертельных болезней из-за поздней диагностики.

Задача:

Диагностировать рак или его рецидива на первых стадиях в 85+% случаях.

Примеры успешного использования Data Science в медицине



Дано:

Функционально-МРТ головного мозга

Каждый 8й человек страдает психическими расстройствами

Нет точного метода диагностики*

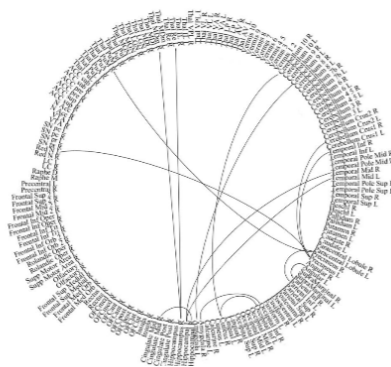
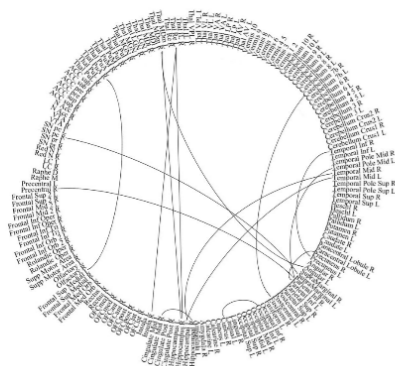
Задача:

Диагностика Большого Депрессивного расстройства точными методами (92%)

ШКОЛА СИСТЕМНОГО АНАЛИЗА

Связь	Позиция	Название	Название
0.949929	(54, 47)	Occipital Inf L	Cuneus R
0.949390	(56, 101)	Fusiform L	Cerebellum 7b R
0.946578	(64, 157)	SupraMarginal L	SN pc R
0.938073	(63, 86)	Parietal Inf R	Temporal Pole Mid L
0.935851	(63, 1)	Parietal Inf R	Precentral R
0.935393	(68, 89)	Precuneus L	Temporal Inf R
0.934752	(39, 128)	Hippocampus R	Thal Re L
0.930222	(66, 131)	Angular L	Thal Mdm R
0.930167	(69, 62)	Precuneus R	Parietal Inf L
0.929899	(18, 143)	Frontal Sup Medial L	Thal PuA R
0.927467	(41, 35)	ParaHippocampal R	Cingulate Mid R
0.927361	(40, 84)	ParaHippocampal L	Temporal Mid L
0.924283	(40, 85)	ParaHippocampal L	Temporal Mid R
0.920424	(33, 128)	Insula R	Thal Re L
0.920232	(39, 131)	Hippocampus R	Thal Mdm R

Связь	Позиция	Название	Название
0.949929	(54, 47)	Occipital Inf L	Cuneus R
0.949390	(56, 101)	Fusiform L	Cerebellum 7b R
0.946578	(64, 157)	SupraMarginal L	SN pc R
0.938073	(63, 86)	Parietal Inf R	Temporal Pole Mid L
0.935851	(63, 1)	Parietal Inf R	Precentral R
0.935393	(68, 89)	Precuneus L	Temporal Inf R
0.934752	(39, 128)	Hippocampus R	Thal Re L
0.930222	(66, 131)	Angular L	Thal Mdm R
0.930167	(69, 62)	Precuneus R	Parietal Inf L
0.929899	(18, 143)	Frontal Sup Medial L	Thal PuA R
0.927467	(41, 35)	ParaHippocampal R	Cingulate Mid R
0.927361	(40, 84)	ParaHippocampal L	Temporal Mid L
0.924283	(40, 85)	ParaHippocampal L	Temporal Mid R
0.920424	(33, 128)	Insula R	Thal Re L
0.920232	(39, 131)	Hippocampus R	Thal Mdm R



Итог успешного использования Data Science в медицине

Преимущества использования Data Science в диагностике :

- **Раннее обнаружение:** Анализировать большие объемы данных, включая медицинские изображения, биомаркеры и клинические показатели, для выявления ранних признаков изменений. Повышает шансы на успешное лечение.
- **Точность диагноза:** Могут обучаться на больших наборах данных и выявлять тонкие и характерные паттерны, которые могут быть неочевидны для человеческого глаза. Это способствует более точному и надежному диагнозу.
- **Интеграция различных данных:** Объединять разнообразные типы данных, включая генетические, клинические, молекулярные и изображения, для создания комплексного и всестороннего обзора состояния пациента. Это помогает выявлять более полную картину заболевания.
- **Прогнозирование и стратификация:** Data Science может использоваться для создания моделей прогнозирования, которые оценивают вероятность распространения, развития осложнений заболевания и эффективности различных методов лечения. Это позволяет проводить более индивидуализированный и подходящий лечебный план.



Итог успешного использования Data Science в медицине

- **Поддержка принятия решений:** Алгоритмы Data Science могут предоставлять врачам рекомендации и советы на основе анализа данных. Это помогает врачам принимать более обоснованные решения о диагнозе и лечении.
- **Минимизация субъективности:** Использование Data Science сокращает влияние субъективных факторов на диагностику, так как алгоритмы оперируют на основе объективных данных и анализируют их независимо от личных предпочтений или восприятия.
- **Оптимизация лечения и мониторинга:** Data Science может использоваться для отслеживания эффективности лечения и мониторинга динамики изменений состояния пациента. Это позволяет корректировать терапию и подходы к уходу, основываясь на реальных данных.
- **Быстрота анализа:** Автоматизированные методы Data Science позволяют быстро анализировать большие объемы данных и генерировать результаты. Это важно для оперативности диагностики и начала лечения.



Текущие возможности и преимущества применения Data Science в медицине

- **Ранняя диагностика и прогнозирование заболеваний:** Data Science позволяет анализировать медицинские данные, включая изображения, генетическую информацию и клинические записи, для выявления рисков и предсказания развития заболеваний на ранних стадиях.
- **Персонализированная медицина:** Анализ данных о пациентах позволяет разрабатывать индивидуальные подходы к лечению и диагностике, учитывая особенности каждого пациента.
- **Оптимизация лечения:** Data Science помогает определить наиболее эффективные методы лечения, предсказать реакцию на лекарства и даже разрабатывать новые лекарственные препараты.
- **Анализ клинических испытаний:** Методы Data Science помогают анализировать результаты клинических испытаний, оптимизировать дизайн и выбор пациентов, ускоряя процесс разработки новых лекарств.
- **Здравоохранительная аналитика:** Data Science используется для оптимизации управления здравоохранением, прогнозирования потребности в медицинских ресурсах и планирования медицинских услуг.
- **Медицинская визуализация:** Создание наглядных визуализаций медицинских данных помогает врачам и исследователям лучше понимать информацию и принимать обоснованные решения.



В каких еще областях можно встретить успешное внедрение Data Science

- 1. Анализ финансовых данных:** Data Science используется для анализа больших объемов финансовых данных, включая исторические котировки ценных бумаг, данные о торговле, долги компаний и другие финансовые показатели. Алгоритмы анализа данных помогают выявлять паттерны, тренды и особенности в финансовых данных, что позволяет инвесторам и аналитикам принимать более обоснованные решения.
- 2. Прогнозирование рисков:** Data Science позволяет создавать прогнозные модели для оценки рисков в финансовых операциях. Например, можно прогнозировать вероятность дефолта заемщика, риски инвестиций, изменения валютных курсов и другие факторы, которые могут повлиять на финансовую стабильность.
- 3. Обнаружение мошенничества:** Алгоритмы машинного обучения позволяют обнаруживать аномалии и необычные паттерны в финансовых транзакциях, что помогает банкам и финансовым учреждениям выявлять случаи мошенничества, включая кражу личных данных, фальсификацию транзакций и другие виды финансовых преступлений.



Финансы и банковское дело

В каких еще областях можно встретить успешное внедрение Data Science

1. **Анализ поведения пользователей:** Собранные данные о поведении пользователей на веб-сайтах или платформах электронной коммерции могут быть анализированы с помощью Data Science для выявления тенденций, паттернов и предпочтений. Это позволяет понять, какие продукты или услуги наиболее популярны, как пользователи взаимодействуют с сайтом и какие шаги они предпринимают перед совершением покупки.
2. **Персонализированные рекомендации продуктов:** Алгоритмы рекомендаций используют данные о предыдущих покупках, просмотрах товаров и поведении пользователей, чтобы предложить им наиболее релевантные продукты или услуги. Это повышает вероятность совершения покупки и улучшает пользовательский опыт.
3. **Прогнозирование спроса:** Data Science позволяет создавать модели прогнозирования спроса на продукты. Это помогает эффективно управлять запасами, избегать дефицита товаров и избытка товаров на складе.
4. **Анализ эффективности маркетинговых кампаний:** Data Science позволяет оценивать эффективность различных маркетинговых стратегий и кампаний. Анализ данных о конверсиях, кликах, откликах и других метриках помогает определить, какие маркетинговые действия наиболее успешны.



Электронной коммерция

В каких еще областях можно встретить успешное внедрение Data Science

- 1. Оптимизация маршрутов:** Data Science используется для разработки алгоритмов оптимального планирования маршрутов транспортных средств. Это позволяет сократить расходы на топливо и время, улучшить обслуживание клиентов и снизить негативное воздействие на окружающую среду.
- 2. Управление транспортными сетями:** Data Science помогает анализировать данные о движении транспортных средств, трафике и перевозках для эффективного управления транспортными сетями. Это включает в себя управление движением, контроль за состоянием дорог и инфраструктуры, а также распределение ресурсов.
- 3. Предсказание спроса на услуги:** С использованием алгоритмов прогнозирования Data Science предсказывает спрос на транспортные услуги. Это позволяет компаниям адаптироваться к изменениям спроса, эффективно планировать ресурсы и предоставлять высококачественное обслуживание клиентам.
- 4. Управление запасами и складами:** Data Science используется для оптимизации управления запасами и складскими запасами. Анализ данных о потребности и поставках помогает минимизировать затраты на хранение и обеспечить своевременную поставку товаров.



Транспорт и логистика

Важность сотрудничества между Data Scientist и системным аналитиком

Сотрудничество между ними позволяет:

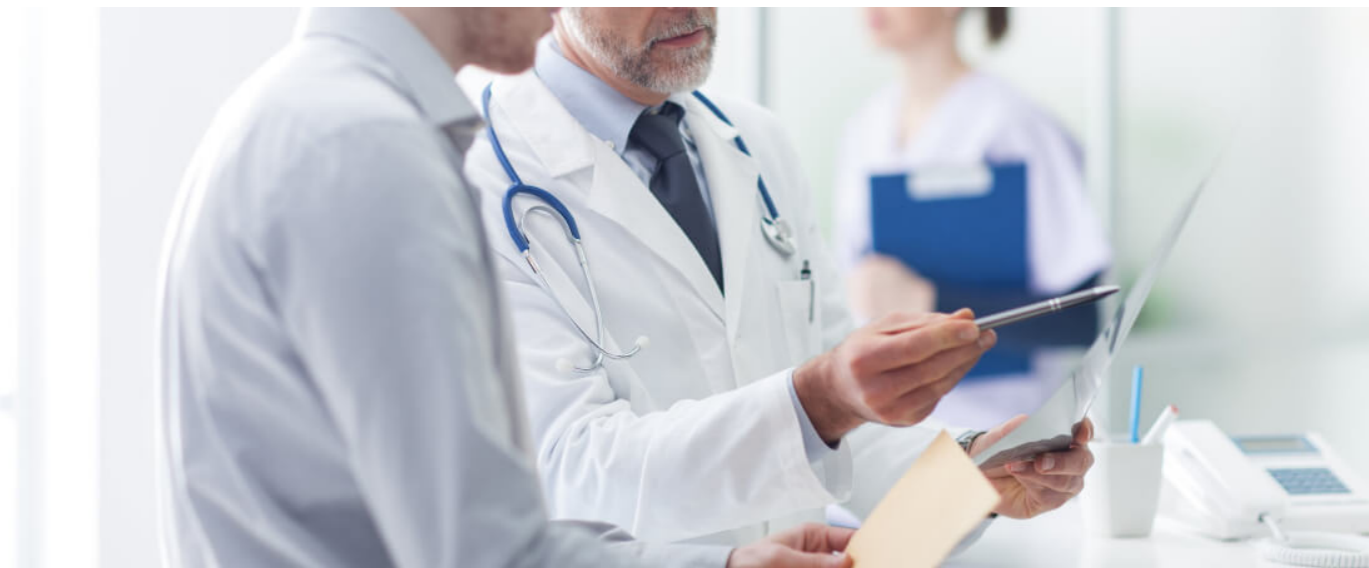
- Понимание потребностей: Системный аналитик помогает Data Scientist понять цели и потребности бизнеса или медицинской практики, что важно для нацеливания анализа данных на решение конкретных задач.
- Определение требований: Системный аналитик помогает формулировать требования к данным, методам анализа и представлению результатов, что облегчает работу Data Scientist.
- Дизайн системы: Совместная работа позволяет разработать архитектуру системы, включая базы данных, инфраструктуру и интерфейсы, учитывая специфику задач Data Science.
- Интеграция: Системный аналитик помогает обеспечить интеграцию разработанных моделей и алгоритмов в рабочую среду, что важно для их практического применения.
- Оценка рисков: Системный аналитик может помочь идентифицировать потенциальные риски и сложности в реализации проекта, что позволяет более эффективно планировать и управлять им.



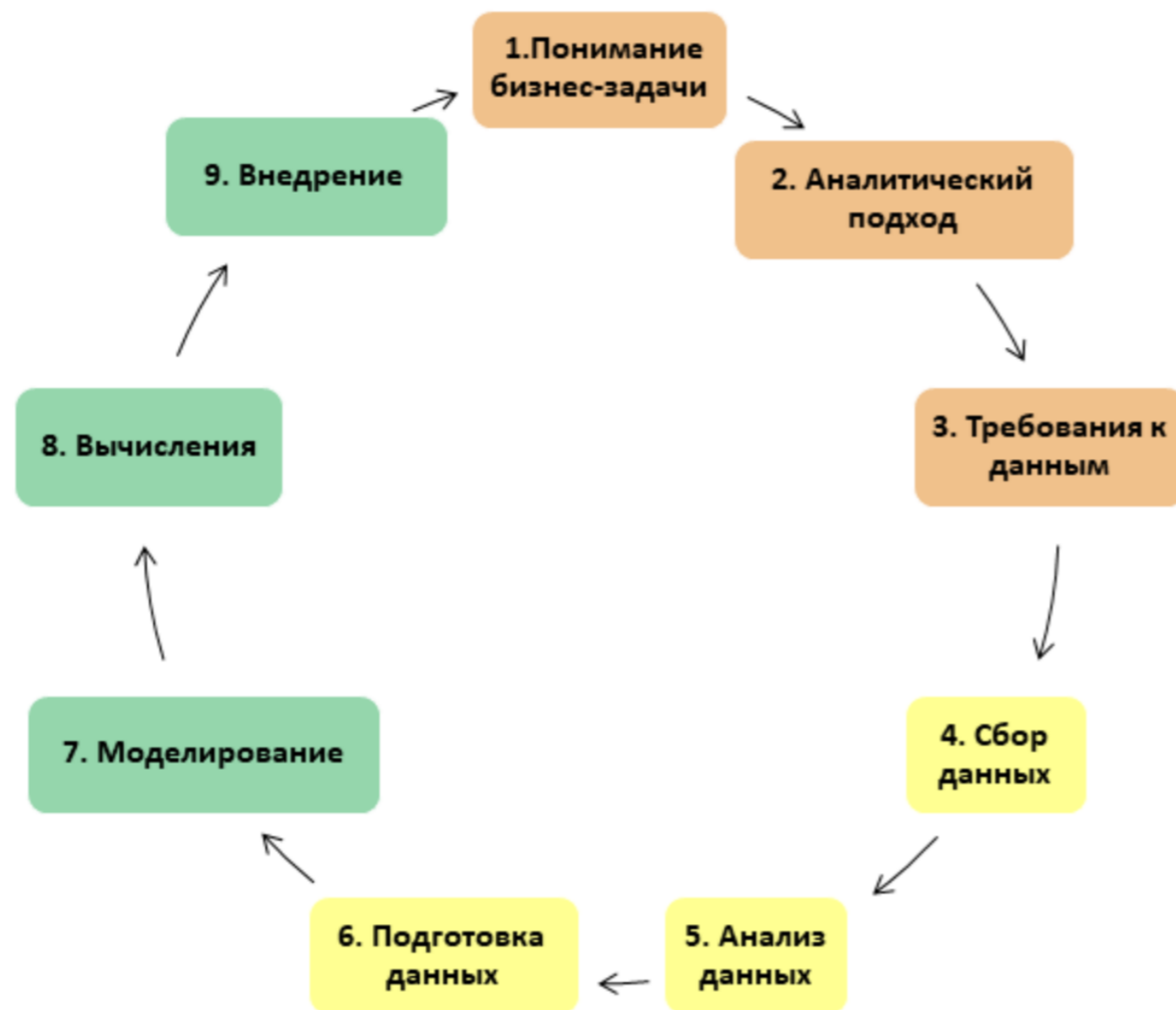
Методы сбора требований для проектов, связанных с Data Science в медицине

Методы сбора требований для проектов, связанных с Data Science в медицине:

1. **Интервью с заинтересованными сторонами:** Беседы с врачами, медицинским персоналом и администрацией позволяют понять бизнес-потребности и ожидания от решения.
2. **Анализ документации:** Изучение медицинских протоколов, клинических данных и стандартов помогает определить требования к данным и методам анализа.
3. **Рабочие мастер-классы:** Организация совместных мастер-классов или воркшопов с Data Scientist и медицинскими специалистами способствует обмену знаниями и определению ключевых потребностей.
4. **Проектирование прототипов:** Создание прототипов аналитических моделей и визуализаций позволяет более конкретно обсудить и уточнить требования.



Примеры процесса сбора требований для проектов в других предметных областях



Любая успешная информационная система состоит из трех элементов:

- полезность;
- удобство использования;
- использование.

Чек-лист основных шагов

Сбор требований системным аналитиком с учетом потребностей Data Scientist может включать в себя следующие шаги и аспекты:

1. Понимание бизнес-потребностей:

- Взаимодействуйте с Data Scientist, чтобы получить понимание их целей и задач, связанных с анализом данных и разработкой моделей.
- Установите, какие именно бизнес-проблемы или вопросы Data Scientist пытаются решить через анализ данных.

2. Определение требований к данным:

- Обсудите с Data Scientist виды данных, которые им потребуются для анализа, включая источники, форматы, объемы и частоту обновления данных.
- Установите необходимые метаданные для эффективной обработки и анализа данных.

3. Требования к инструментам и платформам:

- Узнайте о предпочтениях Data Scientist в отношении инструментов и платформ для анализа данных, программирования и разработки моделей.
- Обсудите требования к вычислительным ресурсам, необходимым для обработки больших объемов данных или выполнения сложных вычислений.

4. Архитектурные требования:

- Разработайте архитектурные решения для интеграции аналитических моделей Data Scientist с существующими системами и базами данных.
- Учтите потребности масштабирования и обработки данных в режиме реального времени, если это необходимо.

5. Требования к безопасности и конфиденциальности:

- Обсудите требования по обеспечению безопасности и конфиденциальности данных, особенно если анализируются медицинские или чувствительные данные.

6. Визуализация результатов:

- Обсудите с Data Scientist, какие типы визуализаций и отчетов им потребуются для представления результатов анализа данных.
- Уточните требования к интерфейсам и инструментам визуализации.

7. Требования к обучению и развертыванию моделей:

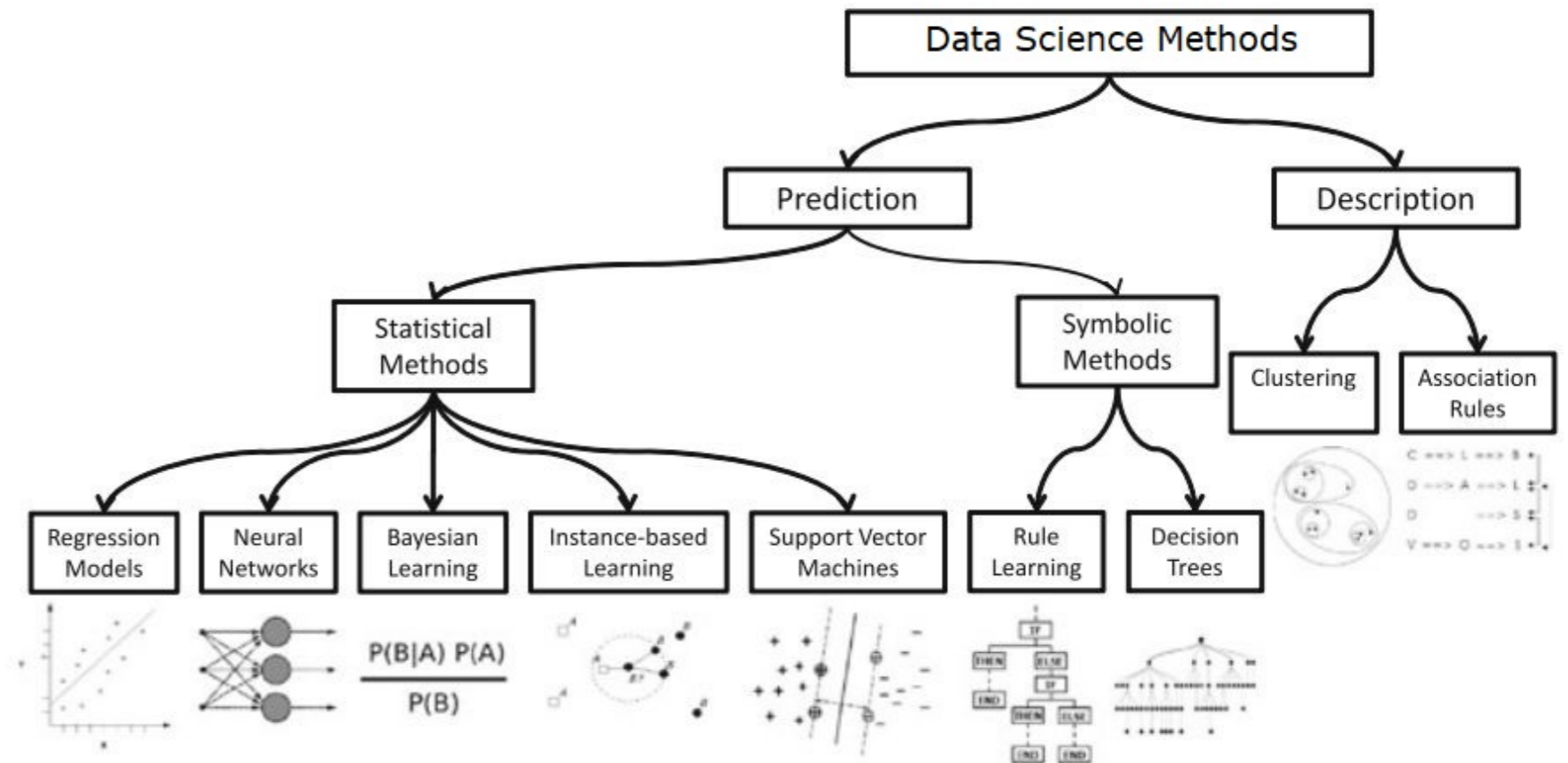
- Узнайте о потребностях Data Scientist в области обучения моделей, валидации, тестирования и развертывания их в боевые условия.
- Разработайте процессы и инфраструктуру для управления жизненным циклом моделей.

8. Взаимодействие с другими системами:

- Рассмотрите, какие другие системы или приложения могут использовать результаты анализа данных Data Scientist, и разработайте требования для их интеграции.

Методы в Data Science

1. Кластеризация (Clustering)
2. Классификация (Classification)
3. Регрессия (Regression)
4. Ассоциативные правила (Association Rules)
5. Рекомендательные системы (Recommendation Systems)
6. Обработка естественного языка (Natural Language Processing, NLP)
7. Сглаживание и интерполяция (Smoothing and Interpolation)
8. Карты самоорганизующихся признаков (Self-Organizing Feature Maps, SOM)
9. Деревья принятия решений (Decision Trees)
10. Сети глубокого обучения (Deep Learning)



1, 2, 3, ?

Какое число нужно поставить вместо ?