

УДК 004.622

## ***КОНЦЕПЦИЯ НАУКИ О ДАННЫХ***

***Ануфриева А.С.***

*студент*

*Институт Управления и Цифровых Технологий РУТ (МИИТ),*

*Москва, Россия*

***Минаков П.А.***

*доцент, кандидат технических наук*

*Институт Управления и Цифровых Технологий РУТ (МИИТ),*

*Москва, Россия*

### **Аннотация**

В статье приведено описание науки о данных, дано определение этого термина и изложены основные понятия и принципы. Наука о данных - это концепция, объединяющая статистику, анализ данных, информатику и связанные с ними методы для понимания и анализа реальных явлений с данными, которые являются результатом реальных событий или действий. Также дано определение прогнозного моделирования и описан принцип его действия и применения.

**Ключевые слова:** Наука о данных, прогнозное моделирование, статистическое машинное обучение, анализ данных, прогностические модели, дерево решений.

### ***DATA SCIENCE CONCEPT***

***Anufrieva A.S.***

*student*

*Institute of Management and Digital Technologies RUT (MIIT),*

*Moscow, Russia*

***Minakov P.A.***

*associate professor, candidate of technical sciences*

*Institute of Management and Digital Technologies RUT (MIIT),*

*Moscow, Russia*

### **Annotation**

The article describes the science of data, defines this term and outlines the basic concepts and principles. Data science is a concept that combines statistics, data analysis, computer science and related methods to understand and analyze real phenomena with data that are the result of real events or actions. The definition of predictive modeling is also given and the principle of its operation and application is described.

**Keywords:** Data science, predictive modeling, statistical machine learning, data analysis, predictive models, decision tree.

Наука о данных — это междисциплинарная область, в которой используются научные методы, процессы, алгоритмы и системы для извлечения знаний и идей из зашумленных, структурированных и неструктурированных данных, а также применения знаний и практических

идей из данных в широком спектре областей приложений. Работой по данной тематике занимались многие ученые [2], [5].

Наука о данных включает в себя сочетание навыков и методов из разных областей, таких как статистика, инженерия, физика и информационные технологии.

Анализ данных является важным видом деятельности в науке о данных, где эксперты делят данные на различные категории для рассмотрения различной аналитики, наборов данных и статистического анализа таких данных, который включает создание данных, манипулирование данными и понимание данных посредством анализа. Тем не менее, существуют некоторые общие методы инженерии данных, которые используются при анализе данных, например структуры данных.

В большей степени, чем анализ данных, инженерия данных имеет прямое влияние на сбор данных и системы обработки данных, что включает в себя проектирование систем извлечения, обработки, хранения и доступа к данным.

Наука о данных - это концепция, объединяющая статистику, анализ данных, информатику и связанные с ними методы для понимания и анализа реальных явлений с данными, которые являются результатом реальных событий или действий. Наука о данных работает с данными, которые представляют собой коллективный набор информации, помещенной в системы и используемой организациями для развития знаний, прогнозов и оценок.

Наука о данных - это непрерывный процесс улучшения и изменения способов анализа данных. Используя новые методы и технологии, специалисты по обработке данных стремятся извлечь тенденцию или

вероятность из больших наборов данных. Наука о данных - это способ анализа любых данных.

Основной принцип науки о данных - рассматривать данные как коллективную историю о любых реальных или гипотетических явлениях. Применяя статистические методы, специалисты по данным пытаются описать сложное явление.

Приложение машинного обучения для науки о данных дает больше информации для определения прогнозной ценности фрагмента данных. Специалисты по обработке данных могут исследовать с помощью новых технологий и методов, чтобы анализировать больше информации из более крупных наборов данных. Наука о данных и машинное обучение - самые инновационные тенденции, которые оказывают огромное влияние на бизнес-процессы.

Статистическое машинное обучение - это поиск закономерности в данных, не имеющей формы или выражения. В машинном обучении алгоритмы искусственного интеллекта (ИИ) используются для наблюдения за данными и разработки закономерностей, которых не было в исходных данных. Затем алгоритм ИИ определяет закономерности или, что более вероятно, проблемы в данных.

Анализ данных также играет очень важную роль в создании прогнозных моделей. Прогнозное моделирование - это процесс создания модели для прогнозирования будущих результатов анализа или поиска закономерностей в данных. Прогнозное моделирование - это часть науки о данных, поскольку оно оказывает большое влияние на бизнес-процессы. Прогнозное моделирование позволяет создавать более точные результаты данных, чем любой анализ данных, описательная статистика или алгоритмы машинного обучения.

Прогностические модели также можно использовать для создания дерева решений для отображения наилучшего решения. Дерево решений - это простой способ принять решение. Дерево решений можно использовать для отображения различных сценариев и прогнозирования наилучшего результата на основе входных данных. Это хороший метод для планирования проектов, поскольку он использует математический алгоритм, который может обеспечить прогноз. Его можно использовать в качестве инструмента принятия решений, прогнозируя результаты на основе результатов других сценариев. Другими словами, дерево решений - это метод прогнозирования, который может предоставлять прогнозы на основе данных. Многие ученые применяли подходы такого типа в области железнодорожного транспорта, организации работы железнодорожных станций, участков и всего движения в целом [1], [3], [4], [6], [7], [8],[9].

Самым большим преимуществом анализа данных является то, что данные могут предоставлять различные сценарии для поиска наилучшего результата для каждого сценария. Прогнозное моделирование может быть очень полезным, поскольку оно может создавать отчеты и прогнозы, которые могут помочь пользователям узнать, каких результатов можно ожидать, и улучшить их проекты. Анализ данных и прогнозное моделирование позволяют создавать множество различных сценариев и упростить создание отчетов на основе различных сценариев.

Наука о данных связана с интеллектуальным анализом данных, машинным обучением и методами больших данных. Она включает в себя набор вычислительных методов для генерации или обработки данных и построения аналитических моделей. Также она включает изучение различных тем, таких как создание данных, управление данными, представление данных, принятие решений и аналитика данных. Наука о данных, как правило, является междисциплинарной областью, а работа в области науки о данных

сосредоточена в областях технологий и ИТ. Главный результат науки о данных - снижение бизнес-рисков и управление рисками. Это также включает улучшение продукта и инновацию продукта. Эти методы улучшают анализ, проектирование и реализацию бизнес-приложений. Кроме того, они делают инструменты использования и обработки данных более доступными, а также используются в приложениях реального времени. Наука о данных также позволяет более эффективно использовать различные элементы крупных организаций. Помимо этого, наука о данных также использует множество других инструментов в других областях для дальнейшего понимания и управления. Такие инструменты включают статистику, статистическое программное обеспечение, математические алгоритмы, инструменты анализа сетевых данных, программное обеспечение для управления алгоритмическими данными, инструменты научной визуализации и инструменты социальных наук. Наука о данных также направлена на улучшение качества обслуживания клиентов. В науке о данных есть различные должности, в том числе веб-аналитики, инженеры по обработке данных, архитекторы данных, инженеры данных, аналитики данных и многие другие. Инструменты обработки и анализа данных также используются для создания непрерывных конвейеров данных и операционных бизнес-приложений. Эти инструменты, особенно инструменты анализа больших данных, часто используются в организациях в качестве дополнения к существующим бизнес-инструментам или ИТ-инструментам.

Среди различных задач науки о данных задачи, связанные с бизнес-аналитикой, включают анализ и моделирование информации из разных источников. Специалисты по обработке данных обычно работают над процессами, управляемыми данными, такими как интеграция данных, извлечение данных и моделирование. Инструменты для анализа данных направлены на то, чтобы помочь пользователям данных интегрировать всю

информацию, хранящуюся в различных системах. Эти инструменты позволяют пользователям данных собирать, понимать и получать информацию из различных форм данных для дальнейшего анализа своих организационных систем. Эти инструменты включают инструменты анализа больших данных, инструменты бизнес-аналитики (BI) и инструменты визуализации данных. Специалисты по обработке данных также работают над моделированием таких вещей, как поведение потребителей.

Наука о данных - это объединение методик для обработки информации в множестве аспектов, которые могут решить множество проблем в области бизнес-аналитики и визуализации данных.

### **Библиографический список**

1. Айсина, Л. Р. Подход к автоматизации ввода исходных данных при имитационном моделировании железнодорожной станции / Л. Р. Айсина, В. И. Булгаков, В. Н. Шмаль // Реформы в России и проблемы управления - 2017 : Материалы 32-й Всероссийской научной конференции молодых ученых, Москва, 26–27 апреля 2017 года. – Москва: Государственный университет управления, 2017. – С. 130-131.
2. Воробьев, Е. А. Этический подход в науке / Е. А. Воробьев, В. Н. Шмаль // Дневник науки. – 2021. – № 3(51).
3. Инновационный центр подготовки специалистов по управлению перевозками / С. П. Вакуленко, Е. В. Копылова, Е. Б. Куликова, П. А. Егоров // Соискатель - приложение к журналу Мир транспорта. – 2010. – Т. 08. – № 2. – С. 3-6.
4. Кан, Ф. Автоматизация распознавания поездной ситуации на железнодорожном участке с использованием искусственной нейронной сети / Ф. Кан, Р. Г. Строченков, В. Н. Шмаль // Реформы в России и проблемы управления - 2017 : Материалы 32-й Всероссийской научной конференции молодых ученых, Москва, 26–27 апреля 2017 года. – Москва: Государственный университет управления, 2017. – С. 189-190.
5. Новое слово в развитии системы формирования практико-ориентированных компетенций / С. П. Вакуленко, П. А. Егоров, Р. А. Ефимов, А. Ю. Тимкова // Железнодорожный транспорт. – 2021. – № 9. – С. 14-19.

6. Тимкова, А. Ю. Модернизация учебно-тренажерного комплекса как основа подготовки высоко-квалифицированных инженеров в области управления перевозочным процессом / А. Ю. Тимкова, Р. А. Ефимов, И. С. Пашкеев // Фёдор Петрович Кочнев - выдающийся организатор транспортного образования и науки в России : Труды международной научно-практической конференции, Москва, 22–23 апреля 2021 года / Отв. редактор А.Ф. Бородин, сост. Р.А. Ефимов. – Москва: Российский университет транспорта, 2021. – С. 400-406.

7. Shmal, V. Determining the available throughput capacity of the railways using simulation for intelligent management of the transportation process / V. Shmal, E. Prokofieva, L. Aysina // MATEC Web of Conferences, Novosibirsk, 16–19 мая 2018 года. – Novosibirsk: EDP Sciences, 2018. – P. 02002. – DOI 10.1051/mateconf/201823902002.

8. Shmal, V. N. The task of organizing suburban and urban traffic on sections with branches / V. N. Shmal, P. A. Minakov, L. R. Aisina // IOP Conference Series: Materials Science and Engineering : XIII International Scientific Conference Architecture and Construction 2020, Novosibirsk, 22–24 сентября 2020 года. – BRISTOL: IOP Publishing Ltd, 2020. – P. 012085. – DOI 10.1088/1757-899X/953/1/012085.

9. Vakulenko, S. P. Technical solutions for loading and unloading operations on freight trains on an electrified track of a logistic terminal / S. P. Vakulenko, A. M. Nasybullin // T-Comm. – 2020. – Vol. 14. – No 4. – P. 66-72. – DOI 10.36724/2072-8735-2020-14-4-66-72.

*Оригинальность 76%*