

ГЛАВА 1. РАБОТА С ДАННЫМИ И ИНФОРМАЦИЕЙ

1.1 ИСПОЛЬЗУЕМЫЕ ТЕРМИНЫ

Для грамотного рассмотрения дальнейшего текста, а также для понимания хода повествования, стоит дать определения основных понятий, чтобы читатель полноценно представлял о чём идёт речь.

Нужно отметить, что часть из определений трактуется неоднозначно, указан один из возможных вариантов уместных и подходящих для данной работы.

Если какое-то определение не дано, значит автор считает, что большинство читателей будет иметь понимание о термине.

[Следующий раздел](#)

1.1.1 ОБЩИЕ ТЕРМИНЫ

Data (данные) - совокупность сведений, зафиксированных на определённом носителе в форме, пригодной для постоянного хранения, передачи и обработки.

API - программный интерфейс, который задаёт описание способов взаимодействия одного программного модуля с другим.

Неуспеваемость - это комплексная, суммарная, итоговая неподготовленность студента для дальнейшей учебы.

Information (информация) - данные имеющие ценность для субъекта.

Big Data (Большие данные) - структурированные или неструктурированные массивы данных большого объема.

Данные для обучения (data set) – обработанная и структурированная информация в виде таблицы.

1.1.2 МАТЕМАТИЧЕСКИЕ ТЕРМИНЫ

Генеральная совокупность – совокупность всех возможных объектов, которые подлежат изучению в пределах объекта исследования.

Выборочная совокупность (выборка) - множество объектов определённым способом выбранных из генеральной совокупности для использования в исследовании.

Репрезентативность выборки – свойство отражающие сопоставимость выводов по выборке выводам генеральной совокупности по свойствам.

Гистограмма частот – столбчатая диаграмма отображающая зависимость частоты встречаемости переменной от её значения.

Форма распределения – кривая, огибающая столбцы частотной гистограммы.

Мода – значение признака, которое встречается чаще всего.

Медиана (median) – значение признака, которое делит упорядоченное множество данных пополам.

1.1.3 MACHINE LEARNING ЧАСТЬ

Machine Learning (машинное обучение) - наука создания алгоритмов (моделей), которые улучшаются благодаря опыту.

Model (модель) – математическое представление системы или процесса.

Метод fit-predict – использование модели без понимания её математического аппарата.

Метрика качества – способ оценки производительности модели на основе её способности предсказывать верные ответы для тестовой выборки.

[Предыдущий раздел](#)

[Следующий документ](#)