

Лабораторная работа №1 «Генерация аннотаций»

Выполнила: Буталова Юлия

Анализируемая статья: [МАШИННОЕ ОБУЧЕНИЕ](#)

Оригинальная аннотация:

В статье проведён анализ методов машинного обучения. Рассмотрены несколько методов, приведены примеры алгоритмов обучения с учителем и без учителя, а также описаны сферы применения.

Начало работы: дабы воспользоваться бесплатными возможностями и уравнять исходные данные для всех исследуемых нейросетей, анализируемая статья была разбита, на фрагменты текста и отправлена по частям. В свою очередь каждая из нейросетей получила один и тот же первый промпт:

Я отправлю тебе сейчас текст статьи, разбитой на 4 фрагмента. Тебе нужно дождаться всех фрагментов, а после выполнить задание, опираясь на этот текст.

Результат промта:

Понял, договорились. Отправляй фрагменты по очереди. Я буду сохранять их в памяти и ничего не буду делать с текстом, пока не получу все 4 части и твое задание. Жду первый фрагмент!

Модель Gemini 2.5 Flash

	Промпт 1	Промпт 2	Промпт 3	Промпт 4
Соответствие заданному объему	нет	да	да	да
Отсутствие лексических и орфографических ошибок	нет	да	да	да
Соответствие заданной структуре	нет	да	нет	да
Соответствие академическому стилю написания аннотации	нет	да	да	да

Соответствие содержанию статьи	нет	нет	нет	да
--------------------------------------	-----	-----	-----	----

Промпт 1 Составь аннотацию к этой статье.

Данная статья представляет собой комплексное введение в машинное обучение (МО), охватывая его определение, принципы работы и основные области применения. МО определяется как подход, позволяющий программным приложениям повышать точность прогнозирования без явного программирования, используя статистический анализ входных данных для прогнозирования выходных.

Статья проводит параллели между МО, интеллектуальным анализом данных и прогнозным моделированием, отмечая их общую цель — поиск закономерностей в данных. Приводятся практические примеры использования МО, такие как персонализированные рекламные рекомендации, обнаружение мошенничества, фильтрация спама, обеспечение сетевой безопасности и формирование новостных лент.

Ключевая часть статьи посвящена классификации алгоритмов МО на два основных типа:

- **Обучение с учителем:** Требуется участия специалиста для предоставления обучающих данных и обратной связи. Охватывает задачи классификации (разделение объектов на категории, например, определение языка или жанра; с примерами алгоритмов "Наивный Байес", деревья решений, метод опорных векторов) и регрессии (предсказание числовых значений, например, цен на акции; с примерами алгоритмов линейной, полиномиальной, логистической регрессии).
- **Обучение без учителя:** Направлено на выявление скрытых структур и зависимостей в данных. Включает:
 - Кластеризацию (группировку объектов по неявным признакам; методы: K-средних, DBSCAN), применяемую для задач вроде сжатия изображений.
 - Снижение размерности (упрощение многомерных данных для визуализации и анализа; методы: метод главных компонент, сингулярное разложение, T-SNE), используемое в рекомендательных системах и тематическом моделировании.
 - Нейронные сети, которые, благодаря обработке огромных объемов данных, способны решать сложные задачи, такие как распознавание изображений, преобразование речи в текст и генерация естественного языка.

Статья подчеркивает, что развитие нейронных сетей стало возможным благодаря эпохе больших данных.

Промпт 2 Прочитай эту статью и напиши краткую аннотацию. Аннотация должна содержать цель исследования, методы, основные результаты и выводы. Объем — около 50–100 слов. Используй нейтральный академический стиль на русском языке.

Настоящая статья посвящена обзору основных концепций и методов машинного обучения (МО). Целью работы является представление ключевых подходов к машинному обучению, включая обучение с учителем и без учителя. В качестве методов анализа рассматриваются классификация, регрессия, кластеризация, снижение размерности и нейронные сети, с кратким описанием их функциональности и примеров алгоритмов. Основные результаты заключаются в систематизации информации о типах задач МО, их применении в реальных сценариях (рекомендательные системы, обнаружение мошенничества и др.) и роли больших данных в развитии современных МО-алгоритмов. Выводы подчеркивают универсальность и растущее значение МО в различных областях.

Промпт 3 Проанализируй загруженную научную статью и создай аннотацию, соответствующую академическим стандартам. Требования к аннотации:

- Язык: русский, академический стиль, без субъективных оценок.
- Объем: 50–100 слов.
- Структура: необходимо кратко обозначить тему исследования и научный контекст, чётко сформулировать цель работы и ключевые исследовательские вопросы, описать, какие методы и подходы использованы, изложить основные полученные результаты, без избыточных деталей, подчеркнуть научную новизну и значимость работы. Не включай личные мнения, цитаты, таблицы или списки. Цель: получить аннотацию, которую можно использовать в журнале или репозитории научных публикаций.

Настоящая статья представляет собой обзорный анализ области машинного обучения (МО), рассматривающий его роль в создании интеллектуальных систем и прогножном моделировании. Исследование фокусируется на систематизации основных парадигм МО, включая обучение с учителем и без учителя, а также на их типовых задачах: классификации, регрессии, кластеризации и снижении размерности. Для достижения поставленной цели были проанализированы существующие алгоритмы и их практическое применение. Результаты демонстрируют эволюцию МО, выявляя ключевые алгоритмы и сценарии использования, от персонализированных рекомендаций до сложных задач обработки данных с помощью нейронных сетей. Работа подчеркивает значимость больших данных для развития современных МО-технологий и их фундаментальную роль в современных вычислительных процессах.

Промпт 4 Проанализируй эту статью и создай аннотацию. Аннотация – это один из видов сокращенной формы представления научного текста. Ее назначение – привлечь внимание читателя, пробудить читательский интерес сообщением сути исследования с помощью минимального количества языковых средств.

Требования к структуре и содержанию аннотации:

- Информативность и содержательность. Аннотация должна в обобщенном виде представлять содержание статьи. В тексте

аннотации не стоит использовать общие фразы, а также указывать несущественные детали и общеизвестные положения.

Предыстория (история вопроса) может быть приведена только в том случае, если она напрямую связана с раскрытием цели исследования.

- Оригинальность. Следует избегать прямых повторов каких-либо фрагментов работы. Название статьи не должно дублироваться в тексте аннотации.
- Четкость, логичность и связность изложения.
- Компактность. Объем аннотаций к статьям должен составлять 50-100 слов.

В аннотации должны быть представлены следующие аспекты содержания статьи:

- предмет, тема, цель работы (указываются в том случае, если они не ясны из заглавия статьи);
- метод или методология (их целесообразно описывать в том случае, если они отличаются новизной или представляют интерес с точки зрения данной работы);
- результаты работы (предпочтение отдается новым результатам и данным долгосрочного значения, важным открытиям, выводам, которые опровергают существующие теории, а также данным, которые, по мнению автора, имеют практическое значение);
- выводы (они могут сопровождаться рекомендациями, оценками, предложениями, гипотезами, описанными в статье).

Не рекомендуется:

- включать в аннотацию таблицы, рисунки, схемы, диаграммы и формулы;
- приводить ссылки на номера публикаций из пристатейного списка литературы;
- использовать сокращения и условные обозначения, кроме общеупотребительных. Если сокращения все же вводятся, то при первом употреблении необходимо дать их расшифровку

Данная работа представляет собой обзор современных методов машинного обучения (МО), исследующий их применение для создания прогнозных моделей и интеллектуальных систем. Целью исследования является систематизация и анализ

основных подходов МО, включая обучение с учителем и без учителя, с акцентом на их ключевые задачи: классификацию, регрессию, кластеризацию и снижение размерности. В качестве методологии использован анализ научной литературы и практических примеров. Полученные результаты обобщают принципы работы ведущих алгоритмов МО и их реализацию в таких областях, как персонализированный маркетинг, обнаружение мошенничества и обработка больших данных с помощью нейронных сетей. Выводы подчеркивают значимость МО как инструмента для решения сложных задач анализа и прогнозирования.

Модель DeepSeek Thinking

	Промпт 1	Промпт 2	Промпт 3	Промпт 4
Соответствие заданному объему	нет	да	да	да
Отсутствие лексических и орфографических ошибок	нет	да	да	да
Соответствие заданной структуре	нет	да	да	нет
Соответствие академическому стилю написания аннотации	нет	да	да	нет
Соответствие содержанию статьи	да	да	да	да

Промпт 1

Статья посвящена основам машинного обучения — области алгоритмов, позволяющих программам прогнозировать результаты без явного программирования. Описываются ключевые процессы, схожие с интеллектуальным анализом данных, и примеры применения: персонализация рекламы, обнаружение мошенничества, фильтрация спама и другие.

Машинное обучение делится на два основных типа:

1. Обучение с учителем: требует готовых данных и обратной связи для прогнозирования. Методы включают:
 - *Классификацию* (например, алгоритмы «Наивный Байес» и метод опорных векторов).
 - *Регрессию* (линейная, полиномиальная и другие), используемую для прогнозирования числовых значений.
2. Обучение без учителя: выявление скрытых зависимостей в данных. Основные методы:
 - *Кластеризация* (алгоритмы К-средних, DBSCAN) для группировки объектов.
 - *Снижение размерности* (метод главных компонент, T-SNE) для визуализации и анализа многомерных данных.
 - *Нейронные сети*, применяемые в сложных задачах, таких как распознавание изображений и обработка естественного языка, и требующие больших объемов данных.

Статья подчеркивает, что машинное обучение стало возможным благодаря развитию технологий и доступности больших данных, и активно используется в современных решениях для анализа и автоматизации.

Промпт 2

Целью статьи является систематизированное описание основ машинного обучения как раздела искусственного интеллекта. В работе рассмотрены ключевые методы: обучение с учителем (классификация, регрессия) и без учителя (кластеризация, снижение размерности, нейронные сети), а также приведены конкретные алгоритмы (Наивный Байес, метод опорных векторов, К-средних, метод главных компонент). Основные результаты включают классификацию задач и демонстрацию практического применения методов в различных областях, таких как анализ данных, компьютерное зрение и обработка естественного языка. Выводы подчеркивают зависимость эффективности современных алгоритмов от больших данных и их трансформирующую роль в технологических решениях.

Промпт 3

Статья представляет систематизированный обзор методов машинного обучения как раздела искусственного интеллекта. Цель работы — классифицировать и охарактеризовать ключевые подходы к прогнозированию на основе данных. В рамках исследования анализируются методы обучения с учителем (классификация, регрессия) и без учителя (кластеризация, снижение размерности, нейронные сети), а также конкретные алгоритмы их реализации. Основные результаты включают описание принципов работы и областей применения рассмотренных методов. Научная значимость работы заключается в структуризации знаний о машинном обучении и

демонстрации его возможностей для решения практических задач в условиях больших данных.

Промпт 4

Рассматриваются фундаментальные принципы и методологии машинного обучения как направления искусственного интеллекта. Цель работы — систематизация подходов к созданию алгоритмов, способных улучшать прогнозирование результатов на основе анализа данных. Детально анализируются методы обучения с учителем (классификация, регрессия) и без учителя (кластеризация, снижение размерности), а также нейронные сети для решения сложных задач. В результате определены ключевые области применения рассмотренных методов и их эволюция в контексте работы с большими данными. Подчеркивается практическая значимость машинного обучения для современных технологических решений в различных сферах деятельности.

Модель GPT-4o

Промпт 1

В данной статье обсуждается концепция и применение машинного обучения (МО), которая представляет собой категорию алгоритмов, позволяющую программным приложениям точно прогнозировать результаты, анализируя входные данные. Основное внимание уделяется различным методам МО, таким как обучение с учителем и без него. В статье описываются популярные алгоритмы, включая классификацию, регрессию и кластеризацию, а также их практическое применение в таких областях, как персонализированный маркетинг, обнаружение мошенничества и фильтрация спама. Также рассматриваются методы снижения размерности, которые помогают визуализировать многомерные данные, и нейронные сети, способные решать более сложные задачи обработки благодаря большим объемам обучающих данных. Такой подход позволяет алгоритмам выявлять тонкие корреляции и применять их для интерпретации новых данных.

Промпт 2

Статья посвящена изучению машинного обучения и его ключевых аспектов. Цель исследования — представить различные методы машинного обучения и их применение. Основные рассмотренные методы включают обучение с учителем (классификация и регрессия), обучение без учителя (кластеризация и снижение размерности), а также нейронные сети. В статье раскрываются примеры использования этих методов в реальных задачах, таких как персонализированный маркетинг и обнаружение мошенничества. Основные выводы подчеркивают важность машинного обучения в современном анализе данных и необходимость больших объемов данных для повышения эффективности алгоритмов.

Промпт 3

В статье рассматриваются методы и алгоритмы машинного обучения, направленные на повышение точности прогнозирования результатов. Основной фокус исследования сосредоточен на анализе обучающих алгоритмов с учителем и без учителя, включая классификацию, регрессию, кластеризацию, снижение размерности и нейронные сети. Применение этих методов иллюстрируется примерами из сферы персонализированного маркетинга и обнаружения мошенничества. Основные результаты подчеркивают эффективность алгоритмов в условиях больших данных. Новизна работы заключается в систематизации современных подходов и расширении области применения машинного обучения в анализе данных.

Промпт 4

В статье рассматриваются различные методы машинного обучения, включая обучение с учителем и без него, с фокусом на их применении в анализе данных. Цель работы — предоставить обзор ключевых алгоритмов, таких как классификация, регрессия, кластеризация и снижение размерности, и их использование в практических приложениях, таких как персонализированный маркетинг и обнаружение мошенничества. Особое внимание уделяется эффективности алгоритмов в условиях больших данных и потенциалу нейронных сетей. Сделан вывод о значимости машинного обучения для современного анализа данных и расширения его применимости.

	Промпт 1	Промпт 2	Промпт 3	Промпт 4
Соответствие заданному объему	нет	да	да	да
Отсутствие лексических и орфографических ошибок	да	да	да	да
Соответствие заданной структуре	нет	нет	нет	нет
Соответствие академическому стилю написания аннотации	нет	да	да	нет
Соответствие содержанию статьи	да	да	да	нет

Модель OpenAI o3

Промпт 1

Статья представляет систематизированный обзор современных методов машинного обучения. Цель работы – описать принципы, области применения и сравнительные характеристики обучения с учителем и без учителя, включая классификацию, регрессию, кластеризацию, снижение размерности и нейронные сети. Методом аналитического обобщения литературы автор выделяет ключевые алгоритмы (SVM, k-means, PCA, T-SNE) и демонстрирует их эффективность на задачах маркетинга, обнаружения мошенничества, визуализации данных и обработки речи. Показано, что масштабируемость алгоритмов обеспечивается большими данными, а интеграция нейронных сетей расширяет спектр сложных приложений.

Промпт 2

Статья предлагает системный обзор методов машинного обучения. Цель – представить принципы работы и области применения ключевых алгоритмов обучения с учителем и без учителя. Методология основана на аналитическом синтезе литературы и сравнительном описании алгоритмов классификации, регрессии, кластеризации, снижения размерности и нейронных сетей. Результатом является систематизация преимуществ, ограничений и типичных приложений – персонализированного маркетинга, обнаружения мошенничества, визуализации данных и обработки речи. Сделан вывод о ключевой роли больших данных в повышении точности и расширении спектра задач машинного обучения

Промпт 3

Работа посвящена систематическому анализу современных алгоритмов машинного обучения в контексте задач обработки больших данных. Цель исследования — определить принципы работы ключевых методов обучения с учителем и без учителя и выявить факторы, влияющие на их прикладную эффективность. Методология основана на аналитическом обзоре теоретических источников и сравнительном разборе алгоритмов классификации, регрессии, кластеризации, снижения размерности и нейронных сетей. Установлено, что масштаб данных и выбор признакового представления критически отражаются на точности моделей. Новизна состоит в интеграции разрозненных подходов в единую концептуальную схему, демонстрирующую перспективы дальнейшей оптимизации.

Промпт 4

Статья представляет обобщённый анализ алгоритмов машинного обучения, раскрывающий принципы работы методов обучения с учителем и без учителя. Цель – систематизировать ключевые подходы (классификация, регрессия, кластеризация, снижение размерности, нейронные сети) и оценить их применимость к задачам прогнозирования, обнаружения аномалий и визуализации. Использован аналитико-сравнительный метод, позволивший выделить условия, при которых точность моделей максимальна. Показано, что объём данных и выбор признаков существенно влияют на

итоговую производительность; подчеркнута роль больших данных в развитии нейросетевых решений.

	Промпт 1	Промпт 2	Промпт 3	Промпт 4
Соответствие заданному объему	да	да	да	да
Отсутствие лексических и орфографических ошибок	да	да	да	да
Соответствие заданной структуре	да	да	да	нет
Соответствие академическому стилю написания аннотации	да	да	да	да
Соответствие содержанию статьи	да	нет	нет	нет

Выводы

По результатам анализа выданных ответов выявлена модель, приведшая к более чёткому результату. Были использованы бесплатные возможности анализа текстовых документов при помощи поэтапного добавления текста статьи частями.

Для этого в самом начале генерации было написано сообщение о том, чтобы ИИ дождался полного текста статьи и только после этого приступал к выполнению запроса. Это необходимо, так как не все модели способные глубоко анализировать текст позволяют загружать текстовые документы для анализа в бесплатной версии.

	Gemini 2.5 Flash	DeepSeek Thinking	GPT-4o	OpenAI o3
Промпт 1	0	1	2	5
Промпт 2	4	5	4	4
Промпт 3	3	5	4	4
Промпт 4	5	3	2	3
Итоговая оценка модели	3	3,5	3	4

После написания промптов были получены результаты, отраженные в таблице. По ним можно определить, что наиболее эффективной является модель OpenAI o3, которая предназначена для глубоких рассуждений и мышления, что стало полезным, так как каждый из запросов привёл с первого раза к результату выше среднего.

Уточняющих промптов написано не было, анализировались «чистые» результаты работы каждой нейросети, отражающие их собственные возможности. Данный отчёт показывает, какую нейросеть можно использовать и с каким промптом, если результат нужен здесь и сейчас, когда нет времени на уточнения.

Это приводит нас к выводу, что модель, предназначенная для глубокого анализа, справляется лучше, если получает кратко и чётко сформулированное ТЗ, что не подойдёт для других моделей, для которых важно получать уточнения и пояснение. Дабы найти золотую середину, выделим в качестве идеального промпта тот, который в большинстве моделей принёс оценимые положительные результаты. Таковым является промпт №2

Прочитай эту статью и напиши краткую аннотацию. Аннотация должна содержать цель исследования, методы, основные результаты и выводы. Объем — около 50–100 слов. Используй нейтральный академический стиль на русском языке.

Наиболее подходящая аннотация, следовательно та, которая получена идеальным промптом, давшая наивысшую оценку, следовательно это аннотация модели DeepSeek, полученная вторым промптом.

Целью статьи является систематизированное описание основ машинного обучения как раздела искусственного интеллекта. В работе рассмотрены ключевые методы: обучение с учителем (классификация, регрессия) и без учителя (кластеризация, снижение размерности, нейронные сети), а также приведены конкретные алгоритмы (Наивный Байес, метод опорных векторов, K-средних, метод главных компонент). Основные результаты включают классификацию задач и демонстрацию практического применения методов в различных областях, таких как анализ данных, компьютерное зрение и обработка естественного языка.

Выводы подчеркивают зависимость эффективности современных алгоритмов от больших данных и их трансформирующую роль в технологических решениях.